PROGRAMME AND ABSTRACTS

10th International Conference on

Computational and Financial Econometrics (CFE 2016)

http://www.cfenetwork.org/CFE2016

and

9th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2016)

http://www.cmstatistics.org/CMStatistics2016

Higher Technical School of Engineering, University of Seville, Spain
9 – 11 December 2016

ISBN 978-9963-2227-1-1

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

Dear Friends and Colleagues,

We welcome you warmly to Seville, for the Tenth International Conference on *Computational and Financial Econometrics* (CFE 2016) and the Ninth International Conference of the ERCIM Working Group on *Computational and Methodological Statistics* (CMStatistics 2016). As many of you know, this annual conference has become a leading joint international meeting at the interface of statistics, econometrics, empirical finance and computing.

The conference aims at bringing together researchers and practitioners to discuss recent developments in computational methods for economics, finance, and statistics. The CFE-CMStatistics 2016 programme consists of 351 sessions, 5 plenary talks and about 1450 presentations. There are over 1550 participants. This is one to the biggest meetings 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 field.

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 Seville, Spain.

Looking forward, the CFE-CMStatistics 2017 will be held at the Senate House, University of London and Birkbeck University of London, UK, from Saturday the 16th to Monday the 18th of December 2017. Tutorials will take place on Friday the 15th of December 2017. You are invited and encouraged to actively participate in these events.

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

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

CMStatistics: ERCIM Working Group on COMPUTATIONAL AND METHODOLOGICAL STATISTICS

http://www.cmstatistics.org

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

Specialized teams

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

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

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

CFEnetwork COMPUTATIONAL AND FINANCIAL ECONOMETRICS

http://www.CFEnetwork.org

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

Specialized teams

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

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

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

SCHEDULE

2016-12-09

Opening, 08:30
A - Keynote
CMStatistics
08:40 - 09:30

B
CMStatistics
09:40 - 10:55

Opening, 09:55
C - Keynote
CFE
10:05 - 10:55

Coffee Break
10:55 - 11:25

D
CFE - CMStatistics
11:25 - 13:05

Lunch Break
13:05 - 14:35

E
CFE - CMStatistics
14:35 - 16:15

Coffee Break
16:15 - 16:45

F
CFE - CMStatistics
16:45 - 18:50

Welcome Reception
20:00 - 21:30

2016-12-10

G
CFE - CMStatistics
08:40 - 10:20

Coffee Break
10:20 - 10:50

H
CFE - CMStatistics
10:50 - 12:55

Lunch Break
12:55 - 14:25

I
CFE - CMStatistics
14:25 - 16:05

Coffee Break
16:05 - 16:35

J
CFE - CMStatistics
16:35 - 18:15

K - Keynote
CFE
18:25 - 19:15

Conference Dinner
20:30 - 23:30

2016-12-11

L
CFE - CMStatistics
08:40 - 10:20

Coffee Break
10:20 - 10:50

M
CFE - CMStatistics
10:50 - 12:05

N - Keynote
CFE - CMStatistics
12:15 - 13:05

Lunch Break
13:05 - 14:30

O
CFE - CMStatistics
14:30 - 15:50

Coffee Break
15:50 - 16:20

P
CFE - CMStatistics
16:20 - 18:00

Q - Keynote
CFE - CMStatistics
18:10 - 19:00

Closing, 19:00 - 19:15

TUTORIALS, MEETINGS AND SOCIAL EVENTS

WINTER SCHOOL AND TUTORIALS

The COST Action CRoNoS Winter Course on Econometrics and Statistics takes place on Tuesday 6th to Thursday 8th December 2016 at La Cartuja Sports Residence. The courses on Thursday are also designated as tutorials of the conference. The first tutorial is given by Prof. Elvezio Ronchetti (Higher-Order Asymptotic Methods in Statistics and Econometrics) at 9:00-13:30. The second tutorial is given by Prof. Jean-Marie Dufour (Nuisance parameters, plug-ins and C(α) tests in statistics and econometrics) at 15:00 - 19:30.

SPECIAL MEETINGS by invitation to group members

- The *CSDA Editorial Board* lunch meeting will take place on Friday 9th of December 2016, 13:00-14:15.
- The *Econometrics and Statistics (EcoSta) Editorial Board* lunch meeting will take place on Saturday 10th of December 2016, 13:00-14:15.
- The *COST Action CRONOS* meeting will take place on Saturday 10th December 2016, 18:30-19:15, Room 002.

SOCIAL EVENTS

- *The coffee breaks* will take place at the Hall of the venue. You must have your conference badge in order to attend the coffee breaks.
- *Welcome Reception, Friday 9th of December 2016, from 20:00-21:30.* The Welcome Reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the hotel TRYP Sevilla Macarena (see map at page VIII). Conference registrants must bring their conference badge and ticket and any accompanying persons should bring their reception tickets in order to attend the reception. Preregistration is required due to health and safety reasons, and limited capacity of the venue. Entrance to the reception venue will be strictly allowed only to those who have a ticket.
- *Conference Dinner, Saturday 10th of December, from 20:30 to 23:30.* The conference dinner is optional and registration is required. It will take place at the hotel TRYP Sevilla Macarena (see map at page VIII). Conference registrants and accompanying persons should bring their conference dinner tickets in order to attend the conference dinner.
- *Conference Buffet Lunches.* The conference lunches are optional and registration is required. The Buffet Lunch will be arranged at the Hall of the venue on 9th, 10th and 11th of December 2016. Conference registrants and accompanying persons should bring their conference lunch tickets in order to attend the conference lunches.

GENERAL INFORMATION

Addresses of venues

- Higher Technical School of Engineering of Seville, Camino de los Descubrimientos, s/n, 41092 Seville, Spain.
- La Cartuja Sports Residence, Glorieta de Beatriz Manchón, s/n, 41092 Seville, Spain.

Registration

The registration will be open on Thursday 8th December 2016 at the tutorials' venue, La Cartuja Sports Residence. The remaining days, that is, from Friday 9th of December to Sunday 11th of December 2016 the registration will take place at the Hall of the conference venue.

Lecture rooms

The paper presentations will take place at the Higher Technical School of Engineering of Seville, Spain (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 Auditorium.

Presentation instructions

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

Posters

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

Internet Connection

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. 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.

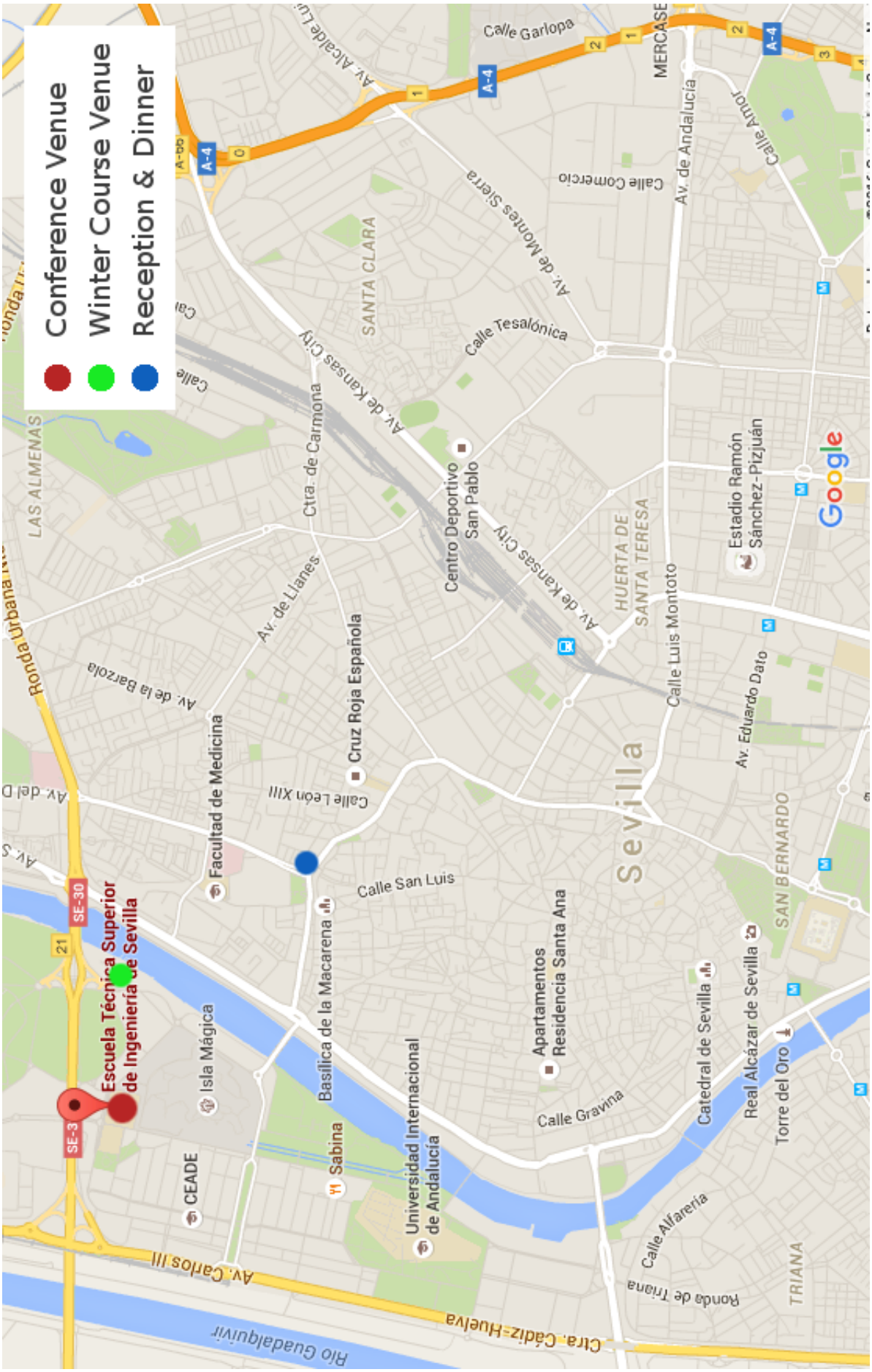
Information and messages

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.

Exhibitors

Elsevier and Springer.

Map of the venue and nearby area

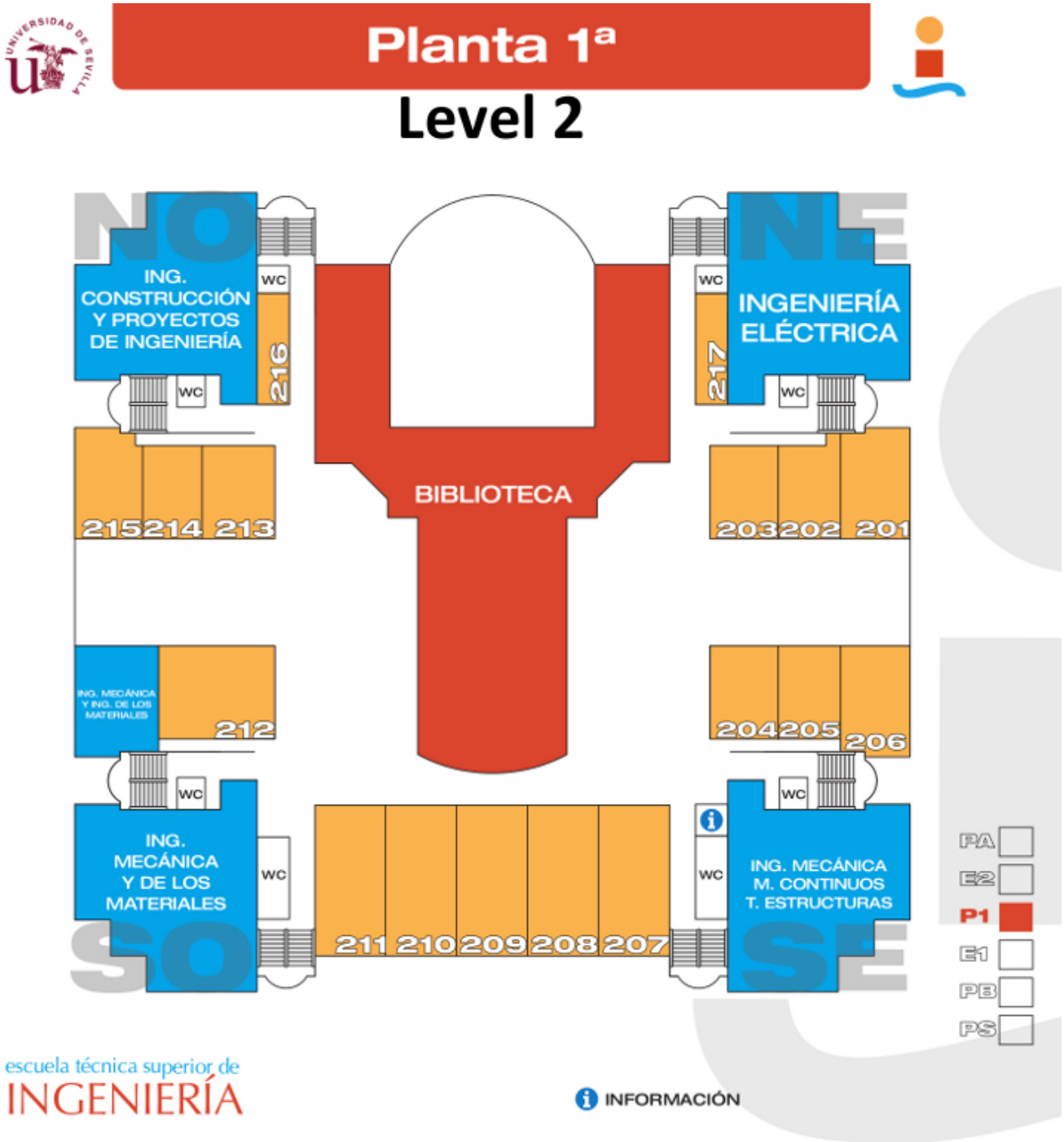


Higher Technical School of Engineering - Basement Floor

Higher Technical School of Engineering - Ground Floor

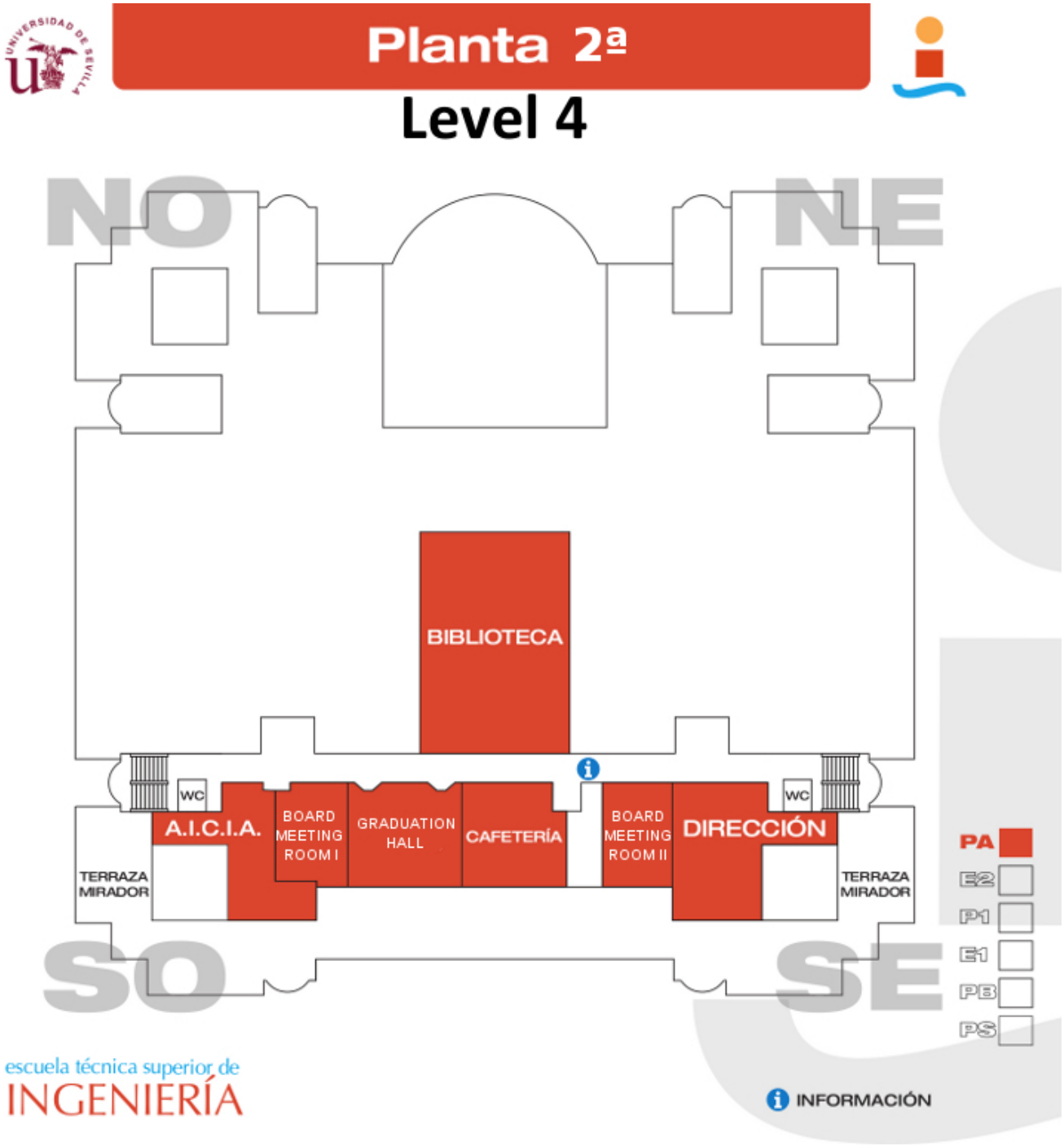
Higher Technical School of Engineering - Mezzanine 1

Higher Technical School of Engineering - First Floor



Higher Technical School of Engineering - Mezzanine 2

Higher Technical School of Engineering - Second Floor



PUBLICATION OUTLETS

Econometrics and Statistics (EcoSta)

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

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

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

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

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

Call For Papers Econometrics and Statistics (EcoSta)

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

Papers containing novel components in econometrics and statistics are encouraged to be submitted for publication in special peer-reviewed, or regular issues of the new Elsevier journal Econometrics and Statistics (EcoSta) and its supplement Annals of Computational and Financial Econometrics. The Econometrics and Statistics (EcoSta) is inviting submissions for the special issues with deadline for submissions the 30th January 2017:

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

Papers should be submitted using the Elsevier Electronic Submission tool EES: http://ees.elsevier.com/ecosta (in the EES please select the appropriate special issue). For further information please consult http://www.cfenetwork.org or http://www.cmstatistics.org.

Call For Papers Computational Statistics & Data Analysis (CSDA)

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

Papers containing strong computational statistics, or substantive data-analytic elements can also be submitted to special peer-reviewed, or regular issues of the journal Computational Statistics & Data Analysis (CSDA). The CSDA is planning for 2016 the following special issues with deadline for paper submissions the 30th June 2017:

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

Papers should be submitted using the Elsevier Electronic Submission tool EES: http://ees.elsevier.com/csda (in the EES please select the appropriate special issue). Any questions may be directed via email to: csda@dcs.bbk.ac.uk.

Contents

Friday 09.12.2016 08:40 - 09:30 Room: Auditorium Chair: Robert Serfling Keynote 1

Of quantiles and expectiles: Consistent scoring functions, mixture representations, and forecast rankings
Speaker: **Tilmann Gneiting, University of Heidelberg, Germany** Werner Ehm, Alexander Jordan, Fabian Krueger

In the practice of point prediction, it is desirable that forecasters receive a directive in the form of a statistical functional. For example, forecasters might be asked to report the mean or a quantile of their predictive distributions. When evaluating and comparing competing forecasts, it is then critical that the scoring function used for these purposes be consistent for the functional at hand, in the sense that the expected score is minimized when following the directive. We show that any scoring function that is consistent for a quantile or an expectile functional, respectively, can be represented as a mixture of elementary or extremal scoring functions that form a linearly parameterized family. Scoring functions for the mean value and probability forecasts of binary events constitute important examples. The extremal scoring functions admit economic interpretations in terms of betting and investment problems.The mixture representations allow for simple checks of whether a forecast dominates another, in the sense that it is preferable under any consistent scoring function. Plots of the average scores with respect to the extremal scoring functions, which we call Murphy diagrams, permit detailed comparisons of the relative merits of competing forecasts.

Friday 09.12.2016 10:05 - 10:55 Room: Auditorium Chair: Richard Luger Keynote 2

Exogeneity tests, weak identification, incomplete models and instrumental variables: Identification-robust inference
Speaker: **Jean-Marie Dufour, McGill University, Canada**

Recent work is reviewed on exogeneity tests in the presence of possibly weak identification, incomplete models, and non-Gaussain errors. After reviewing the finite-sample theory, we study the asymptotic distribution of Durbin-Wu-Hausman (DWH) and Revankar-Hartley (RH) tets for exogeneity, and the properties of pretest estimators where ordinary least squares (OLS) or two-stage least squares (2SLS) estimator is selected depending on the outcome of a DWH- or RH-type test. We consider linear structural models where structural parameters may not be identified and we provide a large-sample analysis of the distribution of the DWH and RH statistics under both the null (exogeneity) and the alternative (endogeneity) hypotheses. Under exogeneity, the usual asymptotic χ^2 critical values are applicable, with or without weak instruments. So, DWH and RH tests are asymptotically identification-robust. A necessary and sufficient condition is given under which all tests are consistent under endogeneity. The condition holds when the usual rank condition for identification in this type of models is satisfied. The consistency condition also holds in a wide range of cases where model identification fails. This is the case when at least one structural parameter is identified. An analysis of the bias and mean squares errors is presented for the pretest estimators. Conditions under which OLS may be preferred to an alternative 2SLS estimator are provided. Empirical results are discussed.

Saturday 10.12.2016 18:25 - 19:15 Room: Auditorium Chair: Helmut Herwartz Keynote 3

Structural vector autoregressions with heteroskedasticity: A review of different volatility models
Speaker: **Helmut Luetkepohl, DIW Berlin and Freie Universitaet Berlin, Germany** Aleksei Netsunajev

Changes in residual volatility are often used for identifying structural shocks in vector autoregressive (VAR) analysis. A number of different models for heteroskedasticity or conditional heteroskedasticity are proposed and used in applications in this context. The different volatility models are reviewed and their advantages and drawbacks are indicated. An application investigating the interaction between U.S. monetary policy and the stock market illustrates the related issues.

Sunday 11.12.2016 12:15 - 13:05 Room: Auditorium Chair: Peter Rousseeuw Keynote 4

Higher-order robust inference
Speaker: **Elvezio Ronchetti, University of Geneva, Switzerland**

Robust statistics develops statistical procedures which are still reliable in the presence of small deviations from the assumed model. Their statistical properties are typically based on approximations obtained by first-order asymptotic theory. However, when the sample size is moderate to small or even in large samples when probabilities in the tails are required, first-order asymptotic analysis is often too inaccurate. We review a class of techniques which combine robustness and higher-order accuracy. They are derived using saddlepoint methods and provide robust tests for testing hypotheses on the parameters and for overidentification which are second-order correct in terms of relative error. Their nonparametric versions are linked to empirical likelihood methods and exhibit good accuracy in finite samples even in the presence of model misspecifications. The theory is illustrated in several important domains, including generalized linear models, quantile regression, composite likelihood, measurement error models, indirect inference, and time series in the frequency domain.

Sunday 11.12.2016 18:10 - 19:00 Room: Auditorium Chair: Hedibert Lopes Keynote 5

Statistics for high-frequency observations of a stochastic process
Speaker: **Jean Jacod, Universite Paris VI, France**

It is often the case that one has to do statistical inference for a stochastic process, based on the observation of a single path of the process, at discrete times and over a finite time interval: in such a framework, estimating the law of the process is usually not feasible, but it is often the case that one can still have reasonable estimators, even consistent ones as the observation frequency increases, for some specific characteristics of the process. We will start with a quick review of those characteristics that can be consistently estimated within this framework, versus those which cannot. Then, restricting our attention to the estimation of the volatility in the case of an Ito semimartingale, we will explain some recent developments and new results, including statements about the rate-optimality and in some cases asymptotic efficiency.

Friday 09.12.2016 09:40 - 10:55 Parallel Session B – CMStatistics

EO439 Room S23 STATISTICS CHALLENGES FOR COMPLEX LARGE DATA Chair: Annie Qu

EO0153: **Method of divide-and-combine in regularized generalized linear models for big data**
Presenter: **Peter Song**, University of Michigan, United States
Co-authors: Lu Tang, Ling Zhou
When a data set is too big to be analyzed entirely once by one computer, the strategy of divide-and-combine (or division-and-conquer, MODAC) has been the method of choice to overcome the computational hurdle. Although random data partition has been widely adopted, there is lack of clear theoretical justification and practical guidelines to combine results obtained from separately analyzed sub-datasets, especially when a regularization method such as LASSO is utilized for variable selection in the generalized linear model regression. We develop a new strategy to combine separately regularized estimates of regression parameters by means of the confidence distributions of biased corrected estimators. We first establish the theory for the construction of the confidence distribution and then show that the resulting MODAC estimator enjoys the Fishers efficiency, the efficiency of the maximum likelihood estimator obtained from the analysis of entire data once. Furthermore, using the MODAC estimator we propose a variable selection procedure, which is compared analytically and numerically via extensive simulations with the existing majority-voting method and the gold standard of one-time entire data analysis.

EO0163: **Analysis of generalized semiparametric mixed varying-coefficient effects model for longitudinal data**
Presenter: **Yanqing Sun**, University of North Carolina at Charlotte, United States
Co-authors: Li Qi, Peter Gilbert
The generalized semiparametric mixed varying-coefficient effects model for longitudinal data that can flexibly model different types of covariate effects. Different link functions can be selected to provide a rich family of models for longitudinal data. The mixed varying-coefficient effects model accommodates constant effects, time-varying effects, and covariate-varying effects. The time-varying effects are unspecified functions of time and the covariate-varying effects are nonparametric functions of a possibly time-dependent exposure variable. We develop the semiparametric estimation procedure by using local linear smoothing and profile weighted least squares estimation techniques. The method requires smoothing in two different and yet connected domains for time and the time-dependent exposure variable. The estimators of the nonparametric effects are obtained through aggregations to improve efficiency. The asymptotic properties are investigated for the estimators of both nonparametric and parametric effects. Some hypothesis tests are developed to examine the covariate effects. The finite sample properties of the proposed estimators and tests are examined through simulations with satisfactory performances. The proposed methods are used to analyze the ACTG 244 clinical trial to investigate the effects of antiretroviral treatment switching in HIV infected patients before and after developing the codon 215 mutation.

EO1033: **Regularized estimation of a mix VAR model**
Presenter: **Jiahe Lin**, University of Michigan, United States
Co-authors: George Michailidis
The association between the stock market and macroeconomic variables is an important component in an economy's behavior over time. To study the nature of such association, we consider a vector-autoregressive model, where the macroeconomic variables not only depend on their past values, but are also Granger-caused by the past values of the log-returns of stocks through a low-rank matrix that encodes the basis. The latter in turn follow their own vector-autoregressive models with a sparse transition matrix. We consider a maximum-likelihood formulation of the problem with proper regularizations in the high-dimensional setting, and propose an iterative algorithm to solve the optimization problem: for each iteration, we sequentially estimate the low rank matrix and the sparse matrix, and iterate until convergence. Theoretical properties of the algorithm are also studied, taking advantage of the decomposability of the regularizers and the convexity of the objective function. The performance of the model is evaluated on synthetic and real data sets.

EO251 Room 301 THEORY AND METHODS FOR STATISTICAL MODELS WITH COMPLEX DEPENDENCE Chair: Pramita Bagchi

EO0276: **Likelihood inference for large stochastic blockmodels with covariates**
Presenter: **Sandipan Roy**, University College London, United Kingdom
Co-authors: Yves Atchade, George Michailidis
A covariate blockmodeling framework is introduced which is in the class of blockmodels that has been widely used in analysing social networks. We introduce a model that captures observations coming from a stochastic blockmodel with certain number of covariates. We devise a novel algorithm based on case-control approximation of the log-likelihood along with a subsampling approach. Our algorithm is based on dividing the subsamples in several cores and then using a single communication among the cores after every iteration. In each core we use a Monte-Carlo EM type algorithm for parameter estimation and latent node label updates. We compare our method with some other methods available for community detection in blockmodels. We also provide an application of our algorithm to a real world network comprising a collection of Facebook profiles with few specific covariates.

EO0480: **Large sample behaviour of high dimensional autocovariance matrices**
Presenter: **Monika Bhattacharjee**, University of Florida, United States
Co-authors: Arup Bose
The existence of limiting spectral distribution (LSD) of the symmetric sum of the sample autocovariance matrix is known when the observations are from an infinite dimensional vector linear process with appropriate (strong) assumptions on the coefficient matrices. Under significantly weaker conditions, we prove, in a unified way, that the LSD of any symmetric polynomial in these matrices exists. Our approach is through the more intuitive algebraic method of free probability in conjunction with the method of moments. Thus, we are able to provide a general description for the limits in terms of some freely independent variables. All the previous results follow as special cases. We suggest statistical uses of these LSD and related results in order determination and white noise testing.

EO0763: **Testing for white noise in functional time series**
Presenter: **Pramita Bagchi**, Ruhr University Bochum, Germany
Co-authors: Vaidotas Characiejus, Holger Dette
The aim is to propose a simple procedure for white noise testing of a functional time series. Our approach is based on an explicit representation of the L^2-distance between the spectral density operator and its best (L^2)-approximation by a spectral density operator corresponding to a white noise process. The estimation of this distance can be easily accomplished by sums of periodogram kernels and its shown that an appropriately standardized version of the estimator is asymptotically normal distributed under the null hypothesis (of functional white noise) and under the alternative. As a consequence we obtain a very simple test (using the quantiles of the normal distribution) for the hypothesis of a white noise functional process. In particular the test does neither require the estimation of a long run variance (including a fourth order cumulant) nor resampling procedures to calculate critical values. Moreover, in contrast to all other methods proposed in the literature our approach also allows to test for "relevant"

deviations from white noise and to construct confidence intervals for a measure which measures the deviation of the underlying process from a functional white noise process.

EO135 Room 103 LATENT VARIABLE MODELS FOR BIOSTATISTICAL APPLICATIONS Chair: Huazhen Lin

EO0340: **Bayesian regularized multivariate generalized latent variable models**
Presenter: **Xiangnan Feng**, The Chinese University of Hong Kong, China
Co-authors: Hao-Tian Wu, Xinyuan Song

A multivariate generalized latent variable model is considered to investigate the effects of observable and latent explanatory variables on multiple responses of interest. Various types of correlated responses, such as continuous, count, ordinal, and nominal variables, are considered in the regression. A generalized confirmatory factor analysis model that is capable of managing mixed-type data is proposed to characterize latent variables via correlated observed indicators. In addressing the complicated structure of the proposed model, we introduce continuous underlying measurements to provide a unified model framework for mixed-type data. We develop a multivariate version of the Bayesian adaptive least absolute shrinkage and selection operator procedure, which is implemented with a Markov chain Monte Carlo (MCMC) algorithm in a full Bayesian context, to simultaneously conduct estimation and model selection. The empirical performance of the proposed methodology is demonstrated through a simulation study. An application of the proposed method to a study of adolescent substance abuse based on the National Longitudinal Survey of Youth is presented.

EO0674: **A latent variable regression model for assessing mixed type biomarkers**
Presenter: **Zheyu Wang**, Johns Hopkins University, United States
Co-authors: Krisztian Sebestyen, Sarah Monsell

The purpose is to present a model-based clustering method via finite mixture modeling framework for mixed typed manifest variables with possible differential covariates. It is motivated by research questions in Alzheimer's disease (AD) with aims: to evaluate the accuracy of imaging biomarkers in AD prognosis, and to integrate biomarker information and standard clinical test results into the diagnoses. One challenge in such biomarker studies is that it is often desired or necessary to conduct the evaluation without relying on clinical diagnoses or some other standard references. This is because 1) biomarkers may provide prognostic information long before any standard reference can be acquired; 2) these references are often based on or provide unfair advantage to standard tests. Therefore, they can mask the prognostic value of a useful biomarker, especially when the biomarker is much more accurate than the standard tests. In addition, the biomarkers and existing tests may be of mixed type and vastly different distributions. We present a model-based clustering method to evaluate the prognostic value of biomarkers in addition to standard tests without relying on potentially inaccurate reference diagnoses. Maximum likelihood parameter estimation is carried out via the EM algorithm. Accuracy measures and the ROC curves of the biomarkers are derived subsequently. Finally, we illustrate the method with a real example in AD.

EO1041: **Measurement error correction through shrinkage estimation in multilevel imaging analysis**
Presenter: **Haochang Shou**, University of Pennsylvania, United States
Co-authors: Ani Eloyan, Amanda Mejia, Mary Beth Nebel, James Pekar, Stewart Mostofsky, Brian Caffo, Martin Lindquist, Ciprian Crainiceanu

Imaging data as high-dimensional and complex measurements are known to be observed with errors that come from multiple sources. Such errors include both random noises and systematic errors that are spatially correlated. With relatively small sample size in most of the study and limited availability of replicates for each participant, statistical inference made based on imaging data with errors might induce bias. We extend the shrinkage estimation idea in imaging data that was previously proposed to scalar-on-function regression setting and generalize the regression calibration in scalar case to functional regression. By shrinking the individual image towards population average image on the levels of individual voxels, local neighbors and the whole brain, we are able to calibrate the spatially dependent regression coefficient via the estimated attenuation ratio. Simulation studies show that the proposed approaches are able to reduce the data noise via borrowing information from the population image, and preserve the subject-specific image features. We have evaluated our methods on seed-based connectivity maps that are calculated using resting-state functional MRI from 21 healthy volunteers (publicly known as 'Kirby21' dataset). Our results have shown that we achieve substantial improvements in mean square errors for prediction, as compared to using one replicate only.

EO605 Room S22 DEALING WITH UNOBSERVED HETEROGENEITY IN COMPONENT-BASED METHODS Chair: Laura Trinchera

EO0505: **Mixture of coalesced distributions for high-dimensional data clustering**
Presenter: **Cristina Tortora**, San Jose State University, United States
Co-authors: Paul McNicholas

Finite mixture models assume that a population is a convex combination of densities; therefore, they are well suited for clustering applications. The use of flexible density distributions allows one to find clusters of different shapes. The generalized hyperbolic distribution (GHD) is a flexible distribution capable of handling skewness and heavy tails, and has many well known distributions as special or limiting cases. The multiple scaled generalized hyperbolic distribution (MSGHD) is obtained by decomposing the generalized hyperbolic distribution component scale matrix using the eigen-decomposition and adding a multidimensional weight function. The mixture of coalesced generalized hyperbolic distribution (MCGHD) arises by merging the GHD and MSGHD with common model parameters, adding more flexibility. The drawback of high flexibility is a high parametrization, the number of parameters is proportional to the number of variables. Therefore, the mentioned methods are not well suited for high dimensional data clustering. We introduce a parsimonious version of the three models based on the principal components analysis. The transformed methods have good clustering performance on high-dimensional data.

EO0650: **PATHMOX segmentation tree for discovering different models present in data: Application in alumni satisfaction**
Presenter: **Giuseppe Lamberti**, Universitat Politecnica de Catalunya UPC, Spain
Co-authors: Laura Trinchera, Tomas Aluja

Measuring satisfaction in higher education has became an essential issue for the university institutions. Competition among a growing number of institutions and learning organizations is sharpening the desire of institutions to deliver high quality service in order to attract and retain students. In this context, shaping good models adapted to students perceptions has become essential. The Partial Least Squares Path Modeling (PLS-PM) approach constitutes a reference methodology to analyze the students preferences. We present a study on the evaluation of alumni satisfaction in university ICT schools, three years after their graduation. Our aim is to ascertain which are the most important drivers of young ICT professionals regarding their perception of the school curriculum. Also we want emphasize the importance of taking into account the heterogeneity of alumni to show how can vary the satisfaction drivers when we consider different segments of them.

EO1680: **Comparing methods for discovering unobserved heterogeneity in PLS-PM: A Monte Carlo study**
Presenter: **Giorgio Russolillo**, CNAM, France
Co-authors: Laura Trinchera

PLS Path Modeling assumes homogeneity among the observed units. In particular a unique model, i.e. the global model, is estimated for the whole dataset. Real data are often affected by unobserved heterogeneity: A unique model may hide important differences. Looking for homogeneous

groups is primordial for such datasets. Response-based Unit Segmentation (REBUS-PLS) and PLS prediction-oriented segmentation (PLS-POS) are two recent approaches for dealing with unobserved heterogeneity in PLS path models. Those two methods share a common idea: both aim at identifying group-specific models with an higher predicting capability compared to the global model. We present a Monte Carlo simulation study for comparing REBUS-PLS and PLS-POS in terms of both prediction and parameter recovering. Moreover we asses their pertinence for high-dimensional data in terms of computational time and robustness of the results.

EO181 Room 308 B DECISION CRITERIA IN CLINICAL DEVELOPMENT Chair: Simon Kirby

EO0549: **Quantitative decision making: One step further**
Presenter: **Richardus Vonk**, Bayer Pharma AG, Germany
Statistical sciences are currently moving into the focus of early applied pharmaceutical research. The high costs and long duration of clinical development, paired with high levels of attrition, require the quantification of the risk when moving from early to late stage clinical development. In addition to the regulatory requirements, statistics and statistical thinking are integral parts of the internal decision making processes, particularly in early clinical development. The focus is on innovative statistical methods in different areas of early drug development that facilitate quantitative rather than qualitative decision making. We describe applications of (Bayesian) statistical techniques to improve decision making and decrease trial sizes. Furthermore, we explain how we implement this new way of thinking into our organization.

EO0863: **Decision criterion in drug development: Beyond statistical significance**
Presenter: **Andy Grieve**, Icon PLC, United Kingdom
The mandatory use of statistics in the drug approval process introduced by the FDA in the 1960s was followed by four decades of the use of statistical significance as almost a sine qua non of marketing authorisation. The impact of this requirement on the whole drug development process in my view has been negative, with clinical researchers and sponsors relying on statistical significance to support their decision making in all phases. More recently there has been an increased interest in the use of decision criteria, above and beyond pure significance, to support decision making in early phase studies. We review these developments against the background of a greater use of Bayesian and/or predictive approaches in drug development.

EO0606: **Establishing decision rules for an early signal of efficacy design strategy in clinical development**
Presenter: **Simon Kirby**, Pfizer, United Kingdom
An early signal of efficacy design strategy in clinical development is the use of a small clinical trial for the first trial of efficacy with three possible outcomes: acceleration of development of a compound; pausing of a compound before further staged investment to assess its potential; and killing a compound, i.e. stopping development of a compound. A key consideration is how decision rules should be set for individual trials to give desired portfolio level performance. We consider some portfolio level criteria and see which decision rules fulfill these criteria for conjectured prior distributions for the efficacy of a portfolio of new compounds.

EO059 Room 217 NON-REGULAR STATISTICAL MODELS Chair: Serguei Dachian

EO0565: **Multiple cusp estimation in regression models**
Presenter: **Maik Doering**, University of Hohenheim, Germany
The problem of estimating the locations of cusps in a regression model is considered. That means, we focus on regression functions, which are continuous, but not differentiable at a known number of unknown locations. We investigate the consistency with increasing sample size of the least squares estimates of the locations of the cusps. It turns out that the rates of convergence depend on the order of smoothness at the locations of the cusps and that our estimator converges to a maximizer of a Gaussian process. For a small order of smoothness at the cusps the least squares estimator for the location of the cusps shows a non-regular asymptotic. That means, we have not the asymptotic normality property, but a representation of the limit distribution as maximizer of a fractional Brownian motion with drift. In order to get confidence intervals for the least squares estimator the limit distribution will be simulated.

EO0540: **On hypotheses testing for Poisson processes: Regular and singular cases**
Presenter: **Serguei Dachian**, Universite Lille, France
The problem of hypothesis testing is considered in the situation where the first hypothesis is simple and the second one is local one-sided composite. We describe the choice of the thresholds and the power functions of different tests when the intensity function of the observed inhomogeneous Poisson process is either regular, or has a singularity of one of the two following types: cusp and discontinuity. The asymptotic results are illustrated by numerical simulations.

EC1683: **A linear mixed model approach to build a calibration function from change points**
Presenter: **Ehidy Karime Garcia Cruz**, National University of Colombia, Colombia
Co-authors: Juan Carlos Correa Morales, Juan Carlos Salazar Uribe
Linear Mixed Models have been widely studied for important authors. However, the specific approach to estimate change points subject-specific has not been worked so specifically. We present an alternative methodology to build a calibration function from change points estimated using LMMs for modeling the fixed effects and predict the random effects for each subject in a longitudinal study. The change points were estimated using evolutionary algorithms. The calibration (reverse regression) function was built under a parametric approach and it is useful to predict a change point over the time. This predicted $\hat{t}$ will allow to minimize an specific loss function, usually associated with the storage expenses for an specific product. The methodology will be illustrated using a real data set about dried cypress wood slats.

EO069 Room 008 ADVANCES IN FUZZY CLUSTERING Chair: Maria Brigida Ferraro

EO0566: **Robust fuzzy clustering based on quantile autocovariances**
Presenter: **Borja Lafuente-Rego**, Universidade da Coruna, Spain
Co-authors: Jose Vilar, Pierpaolo Durso
Three robust versions of the fuzzy C-medoids clustering algorithm for the classification of time series based on comparing estimated sequences of quantile autocovariances (QA) are introduced. Namely, (i) QA-based exponential fuzzy C-medoids clustering, (ii) QA-based fuzzy C-medoids clustering with noise cluster and (iii) QA-based trimmed fuzzy C-medoids clustering. The first one uses a robust metric to neutralize and smooth the effect of outliers, the second one is aimed at detecting outliers and classify them into a noise cluster, and with the third method the model achieves its robustness by trimming away a certain fraction of anomalous time series. The robust fuzzy methods are evaluated in different simulated scenarios, considering different structures of dependence and introducing one or more outliers. All the proposed methods take advantage of the good properties of the QA-based metric, and the results reported from the numerical study outperform the ones obtained with classical fuzzy procedures based on alternative metrics. The usefulness of the proposal is illustrated by a real data application.

EO1253: **Comparing robust fuzzy methods for clustering non-precise data**
Presenter: **Ana Belen Ramos-Guajardo**, University of Oviedo, Spain
Co-authors: Paolo Giordani
In many practical situations the data are not precise. The imprecision of the data can be managed by means of fuzzy sets. This type of data is characterized by a complex structure and, for this reason, there exist different kinds of contamination in this context. There are several proposals of robust methods for clustering fuzzy data. A type of 'robustification' is the use of medoids. Another approach consists in trimming the data. In detail, the outliers are trimmed and not used in the clustering procedure. A further proposal is to add a noise cluster, that is not a proper cluster, containing all the contaminated data. Finally, an alternative approach is the possibilistic one. In this case the membership degree is only based on the distance between the observation and the centroid. Hence, an outlier is characterized by having low membership degrees to all the clusters. We compare all the above mentioned methods by means of simulation and real-case studies in order to analyze their drawbacks and benefits.

EO0868: **Fuzzy two-mode clustering with polynomial fuzzifiers**
Presenter: **Maria Brigida Ferraro**, Sapienza University of Rome, Italy
A new two-mode clustering algorithm is proposed in a fuzzy framework. Two-mode clustering consists in simultaneously clustering modes (e.g. objects, variables) of an observed two-mode data matrix. An extension of the Fuzzy Double k-Means (FDkM) is addressed. The parameters of fuzziness in FDkM are introduced to obtain two fuzzy partitions. These parameters, analogously to the simple (one-mode) case of fuzzy k-means, are used to control the overlapping clusters by giving low membership degrees to objects/variables with unclear assignments. In this way, objects/variables are usually assigned to all clusters with non-zero membership degrees, preventing a hard classification for objects/variables that should be uniquely assigned to a single prototype. In order to overcome this kind of problem, polynomial fuzzifier functions are used. As particular cases, we obtain the FDkM and the hard Double k-Means. The adequacy of the proposal is checked by means of simulation and real-case studies.

EO115 Room 101 MIXTURE MODELS FOR COMPLEX DATA SETS Chair: Geoffrey McLachlan

EO0935: **Maximum likelihood inference for mixtures of Gaussian regression models**
Presenter: **Giuliano Galimberti**, University of Bologna, Italy
Co-authors: Gabriele Soffritti
Finite mixtures of Gaussian linear regression models have been widely studied. They represent useful tools in applied statistics, whenever the interest is in studying the effect of a set of predictors on a response, in presence of unobserved sources of heterogeneity in the data. Despite their popularity, inferential procedures for the model parameters (and in particular for the component regression coefficients) have not been deeply investigated. The score vector and the Hessian matrix of the incomplete data log-likelihood for a finite mixture of multivariate Gaussian linear regression models are derived. Approximations for the standard errors of maximum likelihood (ML) estimators are obtained from these quantities. The properties of these standard error estimates are investigated through an extensive simulation study. Particular attention is devoted to the behavior of these estimates in presence of model misspecification.

EO1206: **A fuzzy version of robust mixtures of Gaussian factor analyzers**
Presenter: **Francesca Greselin**, University of Milano Bicocca, Italy
Co-authors: Agustin Mayo-Iscar, Luis Angel Garcia-Escudero
Clustering aims at dividing a data set into groups or clusters that consist of similar data. Fuzzy clustering accepts the fact that the clusters or classes in the data are usually not completely well separated and thus assigns a membership degree between 0 and 1 for each cluster to every datum. We introduce a robust method for fuzzy clustering based on mixtures of Gaussian Factor analyzers. We illustrate our theoretical considerations by simulations and applications to real data. A comparison with probabilistic clustering is also provided.

EO1042: **Fast model-based clustering of functional data via Gaussian mixture models**
Presenter: **Hien Nguyen**, University of Queensland, Australia
Functional data analysis has become ubiquitous in recent years. It provides an effective framework for statistical modelling of infinite-dimensional functional objects, which occur frequently in practice. A popular paradigm for model-based clustering of functional data is via linear-basis filtering and mixture of mixed effects models. We demonstrate that in some situation, such a paradigm reduces to the simple Gaussian mixture model, for which there are numerous fast and free computational tools available. A demonstration of our clustering approach is performed on data from the calcium imaging of a zebrafish brain.

EO435 Room 305 A ROBUST STATISTICAL MODELLING Chair: Alfio Marazzi

EO1027: **Robust functional principal components for sparse data**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Matias Salibian-Barrera, Jane-Ling Wang
Functional principal components analysis allows us to obtain parsimonious predictions for each trajectory in the sample. The problem of robustly estimating functional principal components when there are only a few observations per curve available will be discussed. Specifically, assume that we observe $X_i(t_{ij})$, $i = 1,\ldots,n$, $j = 1,\ldots,n_i$, where the n_i's can be small for all curves. Many available methods to estimate functional principal components rely on a smoothing step of the observed trajectories, and thus require many observations per curve. A notable exception is the conditional expectation approach (PACE) which estimates the covariance function by smoothing the sparsely available cross-products being able to "combine information" from many curves. A first attempt at protecting this approach from potential outliers by using a robust smoother on the cross-products does not work because the distribution of the cross-products is generally asymmetric. We exploit the linear structure of the conditional distribution of $X_i(t)|X_i(s)$ as a function of $X_i(s)$ for elliptical processes to obtain robust estimators of the scatter function of the underlying random process. We report some numerical experiments comparing the performance of the resulting estimates and existing alternatives.

EO0849: **Robust semi-parametric estimators: Missing data and causal inference**
Presenter: **Eva Cantoni**, University of Geneva, Switzerland
Co-authors: Xavier De Luna
Situations are considered where we aim at estimating location and scale parameters from a distribution law of interest, from which a random sample has been drawn. We introduce semi-parametric estimators, which are able to deal simultaneously with two common challenges within this general context: (i) not all observations from the random sample intended are available (incomplete data due to dropout, selection, potential outcomes framework), and (ii) some of the available observations in the sample may be contaminated (generated by a nuisance distribution, outliers). Under an assumption of ignorable missingness, popular semi-parametric estimators of the parameters of interest are augmented inverse probability weighted (AIPW, doubly robust) estimators. They use two auxiliary models, one for the missingness mechanism, and another for an outcome of interest, both given observed covariates. AIPW estimators are then robust to misspecification of one of these two models (but not both simultaneously - a so-called double robustness property). We introduce versions of AIPW, which provide, moreover, robustness to contamination

of the distribution of interest. Asymptotic properties are described and finite sample results are presented. We motivate the need for robust AIPW estimators with a follow up study on BMI combining data from an intervention study and population wide record linked data.

EO1313: **Robust estimation of mixture models with skew components via trimming and constraints**
Presenter: **Agustin Mayo-Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Francesca Greselin
Impartial trimming procedures are commonly applied in many statistical settings for getting robust estimators in the presence of contamination. In order to get this robust behavior, when estimating mixture models, it is necessary to apply jointly trimming and constraints. Robust estimators based in these tools are available for estimating the model parameters in mixtures of multivariate distributions, of linear regression models, and of factor analyzers, under normal components. We attempt to extend these benefits to the case of skew-normal components. We will show robust methodology based on the joint application of trimming and constraints for different mixture models settings. A drawback of this kind of approaches is related with choosing the input parameters values that this modelling required. We have available different tools for assisting to the users in getting these values.

EO033 Room 214 OPTIMAL AND EFFICIENT DESIGNS Chair: Po Yang

EO1162: **Mismeasured response adaptive design: Theory and implementation**
Presenter: **Xuan Li**, University of Minnesota Duluth, United States
Response adaptive design represents a major advance in clinical trial methodology that helps balance the benefits of the collective and the benefits of the individual and improves efficiency without undermining the validity and integrity of the clinical research. Response adaptive designs use information so far accumulated from the trial to modify the randomization procedure and deliberately skews treatment allocation in order to assign more patients to the potentially better treatment. Little attention has been paid to incorporating the problem of errors-in-variables in adaptive clinical trials. Some important issues of response adaptive clinical design with imperfectly measured outcomes are considered. Optimal allocations under various objectives and asymptotically best response adaptive randomization procedures are investigated in the presence of measurement error. Mismeasurement effects on treatment allocation are discussed for both discrete and continuous responses. The related simulation results are also presented.

EO1507: **Minimum contamination and beta-aberration criteria for screening quantitative factors**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan
For quantitative factors, the minimum beta-aberration criterion is commonly used for examining the geometric isomorphism and searching for optimal designs. We investigate the connection between the minimum beta-aberration criterion and the minimum contamination criterion. Results reveal that ranking designs by the two criteria can be extremely inconsistent and hence the optimal designs selected by them are likely to be different. We provide statistical justifications showing that the minimum contamination criterion well controls the expected total mean square error of the estimation and demonstrate that it is more powerful than the minimum-aberration criterion on identifying geometrically non-isomophic designs.

EO089 Room 003 ANALYSIS OF MICROBIOME DATA Chair: Michele Guindani

EO1237: **A Bayesian Dirichlet-multinomial regression model for the analysis of taxonomic abundances in microbiome data**
Presenter: **Raffaele Argiento**, University of Kent, United Kingdom
Co-authors: Michele Guindani, Marina Vannucci, Duncan Wadsworth, Jessica Galloway-Pena, Samuel Shelburne
A Bayesian Dirichlet-Multinomial regression model is proposed which uses spike-and-slab priors for the selection of significant associations between a set of available covariates and taxa from a microbiome abundance table. The approach allows straightforward incorporation of the covariates through a log-linear regression parametrization of the parameters of the Dirichlet-Multinomial likelihood. Inference is conducted through a Markov Chain Monte Carlo algorithm, and selection of the significant covariates is based upon the assessment of posterior probabilities of inclusions and the thresholding of the Bayesian false discovery rate. We design a simulation study to evaluate the performance of the proposed method, and then apply our model on a publicly available dataset obtained from the Human Microbiome.

EO1239: **A Bayesian nonparametric analysis of heterogeneous data on microbial communities**
Presenter: **Sergio Bacallado**, Cambridge University, United States
Co-authors: Lorenzo Trippa, Boyu Ren, Stefano Favaro, Susan Holmes
Human microbiome studies aim to characterise the microbial communities in the body and the effect of environmental factors on them. A range of experimental techniques have been developed in recent years to catalogue the species composition of a biological sample through ribosomal DNA sequencing, to measure the transcription level of microbial genes, and the synthesis of proteins and metabolites. Modelling such heterogeneous data with a coherent assessment of uncertainty from exploratory analysis, through model selection and inference presents a significant challenge. We propose a Bayesian approach based on latent factors, which is capable of combining insights from various experiments in a parsimonious and interpretable way. We discuss how to scale up computations to massive datasets and evaluate the robustness to prior parameters.

EO0983: **Longitudinal microbiome data**
Presenter: **Snehalata Huzurbazar**, University of Wyoming, United States
Co-authors: Eugenie Jackson
As longitudinal microbiome studies become more common, it is important that we assess how to analyze such data. Methods for large, sparse multivariate count data collected over time are not common in traditional longitudinal data analysis. The first steps in most microbiome data analysis is often the use of ordination methods to explore the data and visually assess existence of patterns, especially of taxa composition with respect to covariate classes. We first consider options for such visualization for data collected over more than one time period on the same subjects. We then present a review of the literature for longitudinal inference for microbiome data, and consider other alternatives.

EO029 Room 213 RECENT ADVANCES IN SURVIVAL ANALYSIS Chair: Gongjun Xu

EO1394: **Estimation and inference of quantile regression for survival data under biased sampling**
Presenter: **Gongjun Xu**, University of Minnesota, United States

Biased sampling occurs frequently in economics, epidemiology and medical studies either by design or due to data collecting mechanism. Failing to take into account the sampling bias usually leads to incorrect inference. We propose a unified estimation procedure and a computationally fast resampling method to make statistical inference for quantile regression with survival data under general biased sampling schemes, including but not limited to the length-biased sampling, the case-cohort design and variants thereof. We establish the uniform consistency and weak convergence of the proposed estimator as a process of the quantile level. We also investigate more efficient estimation using the generalized method of moments and derive the asymptotic normality. We further propose a new resampling method forinference, which differs from alternative procedures in that it does not require to repeatedly solve estimating equations. The proposed unified framework provides researchers and practitioners a convenient tool for analyzing data collected from various designs. Simulation studies and applications to real data sets are presented for illustration.

EO1395: **Semiparametric copula quantile regression for complete or censored data**
Presenter: **Anouar El Ghouch**, The University catholique de Louvain, Belgium
Co-authors: Ingrid Van Keilegom, Mickael De Backer

When facing multivariate covariates, general semiparametric regression techniques come at hand to propose flexible models that are unexposed to the curse of dimensionality. A semiparametric copula-based estimator for conditional quantiles is investigated for complete or right-censored data. Extending recent work, the main idea consists in appropriately defining the quantile regression in terms of a multivariate copula and marginal distributions. Prior estimation of the latter and simple plug-in lead to an easily implementable estimator expressed, for both contexts with or without censoring, as a weighted quantile of the observed response variable. In addition, and contrary to the initial suggestion in the literature, a semiparametric estimation scheme for the multivariate copula density is studied, motivated by the possible shortcomings of a purely parametric approach and driven by the regression context. The resulting quantile regression estimator has the valuable property of being automatically monotonic across quantile levels, and asymptotic normality for both complete and censored data is obtained under classical regularity conditions. Finally, numerical examples as well as a real data application are used to illustrate the validity and finite sample performance of the proposed procedure.

EO1398: **The single-index/Cox mixture cure model: A new modeling approach for the mixture cure model**
Presenter: **Mailis Amico**, Universite catholique de Louvain, Belgium
Co-authors: Catherine Legrand, Ingrid Van Keilegom

Survival analysis is based on the assumption that if the follow-up period would be long enough, all observations will experience the event of interest. In some situations however, this assumption is not realistic, and survival data can contain a "cure" fraction that will never experience this particular event. In order to take into account for such a situation, classical survival models have been extended to cure models. The mixture cure model is one approach proposed in the literature. Considering that the population of interest is a mixture of cured and uncured individuals, the model is composed of the incidence part referring to probability of being uncured, and of the latency part corresponding to the survival function for uncured observations. Most often, the incidence part is modeled parametrically assuming a logistic regression model. We propose to consider a semiparametric modeling through a single-index structure, which offers more flexibility than a parametric approach but avoids the curse of dimensionality phenomenon encountered in nonparametric modeling. A kernel estimator is used for the unknown link function and an estimation method based on the EM algorithm has been developed. Based on simulations, we demonstrate the performance of the proposed method. An application on a breast cancer dataset is also presented in which we contrast our result with those obtained assuming a logistic regression model for the incidence part.

EO243 Room 306 A STATISTICS IN ASTROPHYSICS Chair: Sylvain Sardy

EO1702: **Bayesian inference with nested sampling in Astrophysics**
Presenter: **Farhan Feroz**, University of Cambridge/Winton Capital, United Kingdom

Astrophysics and cosmology have increasingly become data driven with the availability of large amount of high quality data from missions like WMAP, Planck and LHC. This has resulted in Bayesian inference methods being widely used to analyse observations, but they can be extremely computationally demanding. Over the past few years, many innovative methods for performing robust statistical analyses have been developed. We give an overview of Bayesian inference and discuss how it is done in practice. In particular, we describe the nested sampling algorithm, which is a Monte Carlo technique which has been applied successfully to numerous challenging problems in cosmology and astroparticle physics due to its capability of efficiently exploring multi-modal parameter spaces. Nested sampling can also calculate the Bayesian evidence and therefore provides means to carry out Bayesian model selection. We also discuss applications of methods based on nested sampling for gravitational wave and extrasolar planet detection.

EO1694: **GLM-lasso explores the cosmos**
Presenter: **Jairo Diaz Rodriguez**, University of Geneva, Switzerland
Co-authors: Sylvain Sardy, Dominique Eckert

An important problem in Cosmology is to reconstruct the mass distribution of a galaxy cluster based on telescope images for instance using X-ray Multi-Mirror techniques. Such images include point sources spread behind the cluster which although there are of importance, also affect the estimation of the mass distribution. Therefore it is important to robustify the reconstruction of the galaxy cluster by identifying at the same time the point sources, outliers in the response. We cast this challenging problem into a linear inverse problem involving blurring, Abel and wavelet transforms for Poisson counts. Owing to the fact that the high-dimensional (more than two millions) estimands are sparse, we employ GLM-lasso. We address the issues of selecting two regularization parameters and of solving the high-dimensional optimization problem. We show the strength of our method on simulated and real images.

EO1722: **Sparsity-based statistical methods for the estimation of the cosmological microwave background**
Presenter: **Jerome Bobin**, CEA Saclay, France

The Cosmological Microwave Background (CMB) is, by far, the oldest observable signal in the universe that has been emitted more than 13 billions years ago. This relic signal is of paramount importancefor cosmologists since it carries invaluable information about the dawn of our Universe and its evolution. However, accessing such signal is a highly challenging estimationproblem that requires the development of dedicated statistical methods. Indeed, the CMB can only be observed from multispectral data, through complex mixtures of various kinds of contaminants that include astrophysical components,noise, instrumental effect, etc. We first show how extracting one of the most important signal in cosmology can be recast as a challenging inverse problem in statistics: blind source separation. In this context, we emphasize on how modern-day sparse signal models and methods allowed for the very accurate estimation of the CMB from the latest data delivered by the Planck space mission of the European Space Agency.

EO063 Room 212 MULTIPLE TESTING Chair: Sebastian Doehler

EO0883: **Type-II generalized family-wise error rate formulas with application to sample size determination**
Presenter: **Pierre Lafaye de Micheaux**, University of New South Wales, Australia
Co-authors: Benoit Liquet
Multiple endpoints are increasingly used in clinical trials. The significance of some of these clinical trials is established if at least r null hypotheses are rejected among m that are simultaneously tested. The usual approach in multiple hypothesis testing is to control the family wise error rate, which is defined as the probability that at least one type I error is made. More recently, the q-generalized family wise error rate has been introduced to control the probability of making at least q false rejections. For procedures controlling this global type-I error rate, we define a type II r generalized familywise error rate, which is directly related to the r power defined as the probability of rejecting at least r false null hypotheses. We obtain very general power formulas that can be used to compute the sample size for single-step and step-wise procedures. These are implemented in our R package rPowerSampleSize available on the CRAN, making them directly available to end users. Complexities of the formulas are presented to gain insight into computation time issues. Comparison with Monte Carlo strategy is also presented. We compute sample sizes for two clinical trials involving multiple endpoints: one designed to investigate the effectiveness of a drug against acute heart failure and the other for the immunogenicity of a vaccine strategy against pneumococcus.

EO1336: **An extension of the Benjamini and Hochberg procedure with optimal data-driven weights**
Presenter: **Guillermo Durand**, Universite Pierre et Marie Curie, France
The BH procedure is a well-known FDR-controlling procedure which power can be improved by putting weights to the p-values. One way of doing has previously been proposed by designing optimal weights based on the distribution of the p-values under the alternative. Unfortunately this distribution is rarely known, so the weights cannot be computed. In a context of grouped p-values sharing the same distribution, we propose data-driven weights which converge to the optimal weights when the number of tests tends to infinity. The resulting step-up procedure is also shown to asymptotically control the FDR.

EO1500: **A modified Benjamini-Hochberg procedure for discrete data**
Presenter: **Sebastian Doehler**, Darmstadt University of Applied Science, Germany
The Benjamini-Hochberg procedure is a classical method for controlling the false discovery rate for multiple testing problems. This procedure was originally designed for continuous test statistics. However, in many applications, such as the analysis of next-generation sequencing data, the test statistics are discretely distributed. While it is well known that the Benjamini-Hochberg procedure still controls the false discovery rate in the discrete paradigm, it may be unnecessarily conservative. We aim to improve the Benjamini-Hochberg procedure in such settings by incorporating the discreteness of the p-value distributions. We investigate the performance of these approaches for empirical and simulated data.

EO265 Room 203 NONPARAMETRIC FUNCTIONAL DATA ANALYSIS Chair: Germain Van Bever

EO1230: **The functional outlier map for detecting and classifying functional outliers**
Presenter: **Mia Hubert**, KU Leuven, Belgium
Co-authors: Jakob Raymaekers, Peter Rousseeuw, Pieter Segaert
Functional data can be classified into regular observations, as well as shift, shape and isolated outliers. We first construct a new graphical display for the detection of outliers. This functional outlier map (FOM) roughly displays for each observation the variability of the outlyingness in each point of its domain versus its average outlyingness. This allows us to distinguish the different types of observations. Moreover we propose a rule to separate the outliers from the regular cases. This new diagnostic tool can be applied to multivariate functional data, with a univariate or multivariate domain. We give several examples and illustrate its use for detecting outlying surfaces. In particular the method is applied to fluorescence excitation-emission spectra before and after fitting a PARAFAC model.

EO1236: **Small-ball probabilities and representations for the MSE**
Presenter: **Paulo Oliveira**, University of Coimbra, Portugal
The choice of the bandwidth is well known to be crucial in kernel estimation. For finite dimensional data this problem has been extensively studied, with characterizations depending essentially on the dimension of the data. Assuming the absolute continuity means that the geometry and smoothness of the distribution is inherited from the Lebesgue measure. For infinite dimensional or functional data the geometry of the underlying space and the roughness of the distribution of the independent process play a much more relevant role. Having kernel regression estimation in mind, we look for MSE representations that can take into account this roughness through the asymptotic behavior of small-ball probabilities. Moreover, we note that some usual assumptions on the kernel function generate terms in the error representation that are not approaching zero, so we relax somewhat the assumptions of the kernel to obtain a better behavior. The error representation obtained explicitly include terms depending on the properties of the distribution of the process, its roughness, and the geometry of the functional space.

EO1382: **Smooth backfitting for additive modeling with multiple predictor functions**
Presenter: **Byeong Park**, Seoul National University, Korea, South
Smooth backfitting is studied when there are errors-in-variables, which is motivated by functional additive models for a functional regression model with a scalar response and multiple functional predictors that are additive in the functional principal components of the predictor processes. The development of a new smooth backfitting technique for the estimation of the additive component functions in functional additive models with multiple functional predictors requires to address the difficulty that the eigenfunctions and therefore the functional principal components of the predictor processes, which are the arguments of the proposed additive model, are unknown and need to be estimated from the data. The available estimated functional principal components contain an error that is small for large samples but nevertheless affects the estimation of the additive component functions. This error-in-variables situation requires to develop new asymptotic theory for smooth backfitting. Our analysis also pertains to general situations where one encounters errors in the predictors for an additive model, when the errors become smaller asymptotically. We also study the finite sample properties of the proposed method for the application in functional additive regression through a simulation study and a real data example.

EO159 Room 112 MULTIVARIATE DATA WHEN DIMENSIONS ARE LARGE, BUT SAMPLES ARE NOT Chair: Arne Bathke

EO1265: **Confidence regions for level differences in growth curve models: Low- and high-dimensional under non-normality**
Presenter: **Solomon Harrar**, University of Kentucky, United States
Co-authors: Jin Xu
In a pre-post or other kind of repeated measures study, it is sometimes clear that the mean profiles of the repeated measures are parallel across treatment groups. When for example, it can be assumed that there is no interaction between the repeated measure factor and the treatment, it would be of interest to know how much of a difference exists in the effect of the treatments. Such differences in the absence of interaction are referred to as level differences. We consider methods for constructing confidence regions for level differences in the multi-dimensional cases. We derive asymptotic expansions for some intuitively appealing pivotal quantities to construct the confidence regions corrected up to the second order. Such corrections are shown in the multivariate literature to improve the accuracy of asymptotic approximations. We evaluate the finite sample performance of the confidence regions via a simulation study. Real-data example from forestry is used to provide an empirical illustration of the features of the various confidence regions proposed.

EO1147: **Graph-based multiple testing with correlated and high-dimensional data**
Presenter: **Florian Klinglmueller**, Medical University of Vienna, Austria
Graph-based multiple testing procedures permit to tailor the testing procedure such that it reflects the importances and contextual relations of different study objectives. The procedure is specified using a directed weighted graph, where nodes correspond to hypotheses and edges determine the algorithm for reallocating the significance level. Being based on the Bonferroni test, these procedures suffer from considerable conservativeness, especially in situations where test statistics are correlated. We present extensions of graphical approaches that account for the correlation between observations even if the joint distribution is unknown. This is achieved by using multivariate permutation tests and by adapting the weighting strategy using the blinded observations. We illustrate the approach with applications from neuroscience and genetics.

EC1203: **Multiple contrast tests for high-dimensional repeated measures designs**
Presenter: **Frank Konietschke**, University of Texas at Dallas, United States
A high-dimensional setting when the number of subjects is substantially smaller than the number of conditions to be tested is widely encountered in a variety of modern longitudinal and repeated measures design studies, with applications ranging from medicine to social sciences. Recently, there have been several global testing procedures for high-dimensional repeated measures designs suggested that can be employed to assess the global null hypothesis, e.g. of no global time effect. In statistical practice, however, frequently the key question of interest is identifying the significant factor levels, along with computing simultaneous confidence intervals for treatment effects. We consider two approaches, namely, regularization and resampling, that can be employed to derive multiple contrast tests and simultaneous confidence intervals in a high dimensional setting. We discuss asymptotic properties of the proposed testing procedures and illustrate their finite-sample performance by simulations and case studies.

EC663 Room 309 B CONTRIBUTIONS IN REGRESSION ANALYSIS Chair: Natalie Neumeyer

EC0166: **Empirical likelihood based inference for fixed effects varying-coefficient panel data models**
Presenter: **Luis Antonio Arteaga Molina**, Universidad de Cantabria, Spain
Co-authors: Juan Manuel Rodriguez-Poo
Local empirical likelihood-based inference is investigated for non-parametric varying-coefficient panel data models with fixed effects. First, we show that the naive empirical likelihood ratio is asymptotically standard chi-squared when undersmoothing is employed. The ratio is self-scale invariant and the plug-in estimate of the limiting variance is not needed. Second, mean-corrected and residual-adjusted empirical likelihood ratios are proposed. The main interest of these techniques is that without undersmoothing, both also have standard chi-squared limit distributions. As a by-product, we propose also two empirical maximum likelihood estimators of the varying-coefficient models and their derivatives. We also obtain the asymptotic distribution of these estimators. Furthermore, a non-parametric version of the Wilk's theorem is derived. To show the feasibility of the technique and to analyze its small sample properties, using empirical likelihood-based inference we test for a conditional factor model in the CAPM setting and we implement a Monte Carlo simulation exercise.

EC1543: **Fitting models that will be combined**
Presenter: **Daumantas Bloznelis**, KU Leuven, Belgium
Co-authors: Gerda Claeskens
In the context of forecast combinations and model averaging, one may entertain two mutually exclusive strategies. First, one may fit the individual models as usual and then look for an optimal combination of their outcomes. Second, one may adjust model fitting aiming to improve the expected performance of their combination (even though it might compromise the expected performance of some or all individual models). We examine the two strategies in several special cases, starting with linear regression models. We evaluate the strategies from the perspective of the bias-variance trade-off and compare the forecast accuracy in terms of the mean squared forecast error.

EC1424: **Regression with compositional data and its interpretation**
Presenter: **Eva Fiserova**, Palacky University, Czech Republic
Co-authors: Ivo Muller, Karel Hron
When data are compositional, i.e. multivariate observations carrying only relative information (proportions, percentages), a special treatment for their statistical 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 statistical tools. The contribution is focused on the regression with compositional data, particularly on regression with compositional response variable, regression with compositional explanatory variables, and regression between parts of a composition. The aim is to present new orthogonal coordinates enabling analogous interpretation of coefficients in regression models, as is known in standard regression. Theoretical results will be applied to a real-world example from psychometry.

EG178 Room 216 CONTRIBUTIONS IN EXTREMES AND THEIR APPLICATIONS Chair: Yuri Goegebeur

EC1223: **A high quantile estimator based on the log-generalised Weibull tail limit**
Presenter: **Cees de Valk**, Tilburg University, Netherlands
Co-authors: Juan Juan Cai
The estimation of high quantiles for very low probabilities of exceedance $p_n \ll 1/n$ (with n the sample size) remains a major challenge. For this purpose, the log-Generalised Weibull (log-GW) tail limit was recently proposed as regularity condition as an alternative to the Generalised Pareto tail limit, in order to avoid potentially severe bias in applications of the latter. We introduce a new estimator for the log-GW tail index and a related quantile estimator, both constructed using the Hill estimator as building block. Sufficient conditions for asymptotic normality are established. For index estimation, a larger sample fraction needs to be used than for scale estimation. We show the results of simulations of finite-sample performance and of several applications, comparing the estimators to other log-GW-based estimators and for quantiles, also to an estimator based on the Generalised Pareto tail limit.

EC1544: **Maximum likelihood estimators under block maxima**
Presenter: **Ana Ferreira**, IST-CEMAT and CEAUL, UL, Portugal
Co-authors: Clement Dombry
Two fundamental methods in Extreme Value Theory are the Block Maxima (BM) and the Peaks-Over-threshold (POT) and, two widely used methods of estimation for inferences in extremes when appliyng BM or POT are the maximum likelihood (ML) and probability weighted moment (PWM). We prove asymptotic normality of ML estimators under the BM approach and maximum domain of attraction conditions. Further, we compare its asymptotic performance with PWM estimators both under BM and POT approaches.

EC1491: **On the hourly temperature data behaviour and the daily extreme temperature tail dynamics**
Presenter: **Debbie Dupuis**, HEC Montreal, Canada
It is reasonable to think that changes in hourly temperature could hold information on daily extreme temperatures, both in terms of their frequency and their size. Investigating the extent to which this holds, for both daily maximum and daily minimum temperatures, is the purpose. The seasonal nature of temperatures, along with their evolution in time due to climate change, make their analysis challenging. We take a conditional peaks over threshold approach and establish time-varying thresholds to maintain stationarity. Our analyses show that hourly changes in temperature can be better or worse predictors of extreme temperature than daily measures, depending on the location and local climate dynamics. This is illustrated by the analysis of U.S. data over the 1973 to 2015 period.

Friday 09.12.2016 11:25 - 13:05 Parallel Session D – CFE-CMStatistics

CO353 Room 003 INCOMPLETE DATA AND MEASUREMENT ERROR Chair: Laura Spierdijk

CO0165: **A semiparametric test for measurement error in panel data**
Presenter: **Laura Spierdijk**, University of Groningen, Netherlands
Co-authors: Erik Meijer, Tom Wansbeek, Roger Moon
Although measurement error is a common phenomenon, most applied regression analyses do not take it into account. One potential explanation is that there is not a standard test for the presence of measurement error. We develop such a test for linear static panel data regressions, based on the insight that under weak assumptions the bias of the OLS estimator increases by first differencing, but less so by taking differences more periods apart. The test is easy to implement and apply and has desirable statistical properties. We apply the test to models for medical expenditures and productivity and show that there is strong evidence for measurement error in some of the regressors.

CO0297: **Chain ladder with incomplete observations**
Presenter: **Ruud Koning**, University of Groningen, Netherlands
With the implementation of the Solvency II framework, best estimates for losses to calculate reserves have become more important. Besides the best estimate, an estimate of sampling variability is also needed. A framework has been previously introduced to provide an estimate of the expected loss, and to estimate its sampling variation as well. The approach is based on calendar year effects and process year effects, suitable for analysis at an aggregate level. However, in some cases information on individual files may be available, and files with certain characteristics may have different runoff characteristics from files with different characteristics. Perhaps it is possible to obtain a better estimate for the reserves required by allowing for variation between individual files. We discuss a model that allow for heterogeneity at the file level. A practical problem may be that relevant covariates may be missing in some cases. We show how it is possible to allow for incompletely observed files as well.

CO0311: **Measurement errors: Evidence from reports of program participation from multiple surveys**
Presenter: **Nikolas Mittag**, CERGE-EI, Czech Republic
Co-authors: Pablo Celhay, Bruce D Meyer
Measurement error is often the largest source of bias in survey data. Little is known about the determinants of such errors, making it difficult for data producers to reduce the extent of errors and for data users to assess the validity of analyses using the data. We study different causes of survey error using high quality validation data from three major surveys in the U.S. that are linked to administrative data on government transfers. The differences between survey and administrative records show that up to six out of ten cash welfare recipients are missed by surveys. We find that survey design and post-processing as well as misreporting by respondents affect survey errors systematically. Imputation for missing data induces substantial error. Our results on respondent behavior confirm several theories of misreporting, e.g. that errors are related to salience of receipt, respondents degree of cooperation, forward and backward telescoping, event recall, and the stigma of reporting participation in social programs. Our results provide guidance on the conditions under which survey data are likely to be accurate and suggest different ways to control for survey errors.

CO0338: **Binary response models with misclassified dependent variables applied to annuity ownership**
Presenter: **Peter van Santen**, Sveriges Riksbank, Sweden
Co-authors: Adriaan Kalwij, Rob Alessie
The problem of misclassification of the dependent variable in a binary choice setting is studied. We apply both static and dynamic models to ownership of annuity policies for a panel of Dutch households, allowing the dependent variable to be misclassified. For a subset of households, we know the duration of the policies, and therefore can construct an ownership variable which should be closer to the true value. We use this measure to benchmark our results. Our estimates of the misclassification probabilities suggest that part of the annuity puzzle can be explained by underreporting ownership of annuities.

CO311 Room S23 BAYESIAN ECONOMETRICS Chair: Deborah Gefang

CO0408: **Parsimonious inverted Wishart processes for multivariate stochastic volatility**
Presenter: **Roberto Leon-Gonzalez**, GRIPS, Japan
Co-authors: Joshua Chan, Rodney Strachan
Efficient algorithms are developed for Bayesian inference in inverted Wishart multivariate stochastic volatility models that are invariant to the ordering of variables. Furthermore, the algorithm searches for factor-type restrictions that reduce the dimension of the latent states. We combine particle Gibbs with an efficient proposal density, Hamiltonian Monte Carlo and Parameter-Expansion-Data-Augmentation, to produce an algorithm that can handle moderately large dimensions and sample sizes. We illustrate the methods with an application to several macroeconomic and financial datasets.

CO1177: **A new look at DSGE models through large Bayesian VARs**
Presenter: **Deborah Gefang**, University of Leicester, United Kingdom
Large Bayesian VARs are proposed to be used to identify the structural shocks to dynamic stochastic general (DSGE) models. We conduct extensive empirical analyses to compare the performances of our approach with that of popular existing methods such as small VARs and FAVARs.

CO0906: **An application of a Bayesian VAR copula model to the effect of macroeconomic risk appetite on the GDP growth**
Presenter: **Fuyu Yang**, University of East Anglia, United Kingdom
Co-authors: Roberto Leon-Gonzalez
GDP growth and monetary policy are closely intertwined with the size of bank balance sheets. Evidence suggests that bank balance sheets grow during booms, and shrink during recessions. We propose a flexible VAR copula model to evaluate a nonlinear association amongst the looseness of bank balance sheet constraints, GDP growth, and macroeconomic risk premium variables. Compared with the conventional VAR model, the relaxation of the distributional assumption in the error term is more realistic in terms of model fitting, which allows for an asymmetric association amongst the key variables in economic downturns. To serve the purpose of model comparisons, we evaluate the impulse response function and log likelihood using both the VAR copula and the VAR models.

CC1523: **A Bayesian estimate of the pricing kernel**
Presenter: **Chiara Legnazzi**, USI and Swiss Finance Institute, Lugano, Switzerland
Co-authors: Giovanni Barone-Adesi, Antonietta Mira
The aim is to present a Bayesian nonparametric approach to model the Pricing Kernel, defined as the present value of the ratio between the risk neutral density, q, and the modified physical density, p^*. The risk neutral density is derived by fitting an asymmetric GARCH to the cross-section of the out-of-the-money call and put options written on the S& P500 index. The modified physical density is defined as the sum with Poisson-Dirichlet

weights of the risk neutral density rescaled for the equity premium and the traditional physical density, estimated through an asymmetric GARCH fitted on the S& P500 log-returns. The nonparametric approach does not impose any a priori restriction on the shape of the Pricing Kernel and the Bayesian component allows to include into the physical density estimation some forward-looking information coming from the option market. As a result, the heterogeneity between the physical and the risk neutral measures, which is one of the main drivers of the pricing kernel puzzle, disappears and both densities are conditional on a comparable information set. The Pricing Kernel estimates gain accuracy in the tail estimation and display a monotonically decreasing shape across multiple time to maturities, consistently with the classical theory.

CO299 Room 008 QUANTILE REGRESSION METHODS IN ECONOMICS AND FINANCE Chair: Massimiliano Caporin

CO0516: **Conditional quantile-dependent autoregressive value-at-risk**
Presenter: **Giovanni Bonaccolto**, University of Padova, Italy
Co-authors: Massimiliano Caporin, Sandra Paterlini
The Conditional Value-at-Risk (CoVaR) and the Conditional Autoregressive Value-at-Risk (CAViaR) have shown to be useful tools in evaluating the connections among financial institutions in the context of extreme events. We show that it is possible to obtain relevant improvements by conditioning the estimation of CoVaR and CAViaR with respect to the support of the depedent or the explanatory variables. This corresponds to the introduction of Quantile-on-Quantile dependencies, or, equivalently, to moving toward non-parametric quantile regression estimation. Our analysis is applied on a large dataset consisting of the returns generated by the financial institutions operating in the U.S. market in the years 2000-2015 and the empirical findings support our methodological contributions.

CO0620: **Systemic risk measurement for GCC financial institutions and its role for oil**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Michele Costola, Shawkat Hammoudeh, Ahmed Khalifa
The systemic risk is analyzed across a group of financial companies traded on the financial markets of six oil rich economies, the GCC countries. The focus is first on the estimation of MES and Delta-CoVaR, and the association of these measures to oil price fluctuations. Later, the impact of oil on the CoVaR is investigated using a HAR structure. By resorting to a rolling approach, the study allows detecting the impact of a change in a fundamental factor (the oil) to the financial sector of oil-rich economies. The results show that systemic risk measures are associated with movements in the oil price and that oil has a significant impact on a large fraction of the financial companies in the database.

CC1470: **Intra-day dynamics of US-dollar exchange-rates: Evidence from quantile autoregressions**
Presenter: **Konstantin Kuck**, University of Hohenheim, Germany
Co-authors: Robert Maderitsch
The aim is to test several hypotheses regarding intra-day dynamics in exchange rates. Specifically, we investigate the presence of intra daily "over"- and "undershooting" using quantile autoregressions based on intra-day returns over time-horizons from 5-min up to several hours. The exchange rates that we consider are the USD/EUR, USD/GBP and USD/JPY. Our data-set covers an 11-year period from January 03, 2000 to September 30, 2011, which enables us to provide a long-term perspective on intra-daily dynamics and to investigate potentially changing properties in phases of increased uncertainty in financial markets. In line with previous studies, we find that the intra-daily return autocorrelation is mostly negative which is typically associated with market microstructure effects. In the tails of the return distribution, however, a tendency for positive but insignificant dependence can be observed. We moreover document that temporal dependence declines when intra-day volatility rises due to trading activity. Finally, the dynamics appear to be remarkably stable across different market conditions, that is, they seem not to be affected by increased uncertainty in the financial markets. Overall, our findings show that both over- and undershooting are absent in US-Dollar exchange rates and indicate that information are processed efficiently in the sense that future returns cannot be predicted based on historical returns.

CC1572: **Measurement of common risk factors: A panel quantile regression models for returns and volatility**
Presenter: **Frantisek Cech**, Charles University, Czech Republic
Co-authors: Jozef Barunik
The common risk factors are proposed to be measured using panel quantile regression models for returns and volatility. By exploring the fact that volatility crosses all quantiles of the return distribution, and employing panel quantile regressions, we focus on the commonalities in the quantiles of the returns in a selected portfolio. In the forecasting exercise we show that the proposed approach is able to model lower quantiles of the return distribution more precisely as widely used benchmark models. Results of our research are important for correct identification of the sources of systemic risk, and will be particularly attractive for relatively high dimensional applications for dimensionality reduction and better portfolio optimization.

CO309 Room S22 QUANTITATIVE ASSET MANAGEMENT Chair: Serge Darolles

CO0573: **Currency carry trade and the cross section of hedge fund returns**
Presenter: **Adrien Becam**, Paris-Dauphine University, France
The aim is to demonstrate that a factor mimicking the returns of a currency carry trade strategy explains strongly the cross section of hedge fund returns. Results show that funds loading on the carry trade risks outperform low loading hedge funds by 3.48% on average, despite a 1994-2015 period covering major carry trade crashes. This involvement of hedge funds in currency carry trade could explain stylized facts about foreign exchange markets.

CO0593: **Joint inference on market and estimation risks in dynamic portfolios**
Presenter: **Jean-Michel Zakoian**, CREST, France
Co-authors: Christian Francq
The purpose is to study the estimation risk induced by univariate and multivariate methods for evaluating the conditional Value-at-Risk (VaR) of a portfolio of assets. The composition of the portfolio can be time-varying and the individual returns are assumed to follow a general multivariate dynamic model. Under ellipticity of the conditional distribution, we introduce in the multivariate framework a concept of VaR parameter, and we establish the asymptotic distribution of its estimator. A multivariate Filtered Historical Simulation method, which does not rely on ellipticity, is studied. We also consider two univariate approaches based on past real or reconstituted returns. We derive asymptotic confidence intervals for the conditional VaR, which allow to quantify simultaneously the market and estimation risks. Potential usefulness, feasibility and drawbacks of the different univariate and multivariate approaches are illustrated via Monte Carlo experiments and an empirical study based on stock returns.

CO0619: **Measuring hedge fund performance: A Markov regime-switching with false discoveries approach**
Presenter: **Gulten Mero**, Universite de Cergy-Pontoise and THEMA, France
A Markov regime-switching approach is proposed which accounts for false discoveries in order to measure hedge fund performance. It enables us to extract information from both time-series and cross-sectional dimensions of panels of individual hedge fund returns in order to distinguish between skilled, unskilled and zero-alpha funds for a given state of the economy. Applying our approach to individual hedge funds belonging to

the Long/Short Equity Hedge strategy, we find that their performance cannot be explained by luck alone, and that the proportion of zero-alpha funds in the population decreases when accounting for alpha regime dependence. However, the proportion of truly skilled funds is higher during expansion periods, while unskilled funds tend to be more numerous during recession periods. Moreover, sorting on regime dependent instead of unconditional alphas improves investors' ability to select funds that outperform their benchmarks in both regimes of the economy by maximizing the performance persistence effect of top performer fund portfolios.

CO1168: **About the risks of alternative risk premia**
Presenter: **Guillaume Monarcha**, Orion Financial Partners, France
In the continuity of factor-based investing, alternative risk premia (ARP) provide investors a liquid and transparent alternative to strategies that were previously accessible only through hedge fund investments. On the model of fund of hedge funds, several asset managers have recently set up risk premia funds in order to provide investors their ability to select and allocate between the various ARP products developed by investment banks. We put a light on the fact that if ARPs are more accessible than hedge funds, their risk structures and return drivers remain complex. Thus, their low levels of correlation with traditional asset classes are an obvious source of diversification, but at the price of specific risks which need to be assessed. Using a database of investable ARPs developped by investment banks, we first perform a clustering analysis in order to identify the main ARP styles. Secondly, we test their sensitivity to traditional risk factors, through static and dynamic factor models. Finally, we test their sensitivity to various environmental factors (macroeconomic cycles, risk aversion ...), both in terms of performance and risk.

CO401 Room 103 MACRO AND FINANCIAL ECONOMETRICS Chair: Menelaos Karanasos

CO0653: **Trade credit and firm performance**
Presenter: **Yiannis Karavias**, University of Birmingham, United Kingdom
Co-authors: Stylianos Asimakopoulos, Filipa Fernandes
Theory suggests that trade credit affects firm performance non-linearly with an optimal maximum level determined through a cost-benefit trade-off. However, empirical literature fails to uncover this theoretical result. We are able to bridge this gap between theory and practice using a large panel of European firms through a novel and very general non-linear panel GMM approach. Specifically, we find a robust non-linear effect of trade credit on firm profitability. Moreover, we find a statistically significant maximum level of trade credit with a negative impact when it is exceeded and a positive when operating below that level. Disaggregating the effect of trade credit on firm performance, using account receivables and payables, we are able to uncover that account payables exhibit an inverted U-shaped effect on firm profitability, whereas account receivables exhibit a U-shaped effect. Our results also indicate that effect of account payables dominates on trade credit. Finally, the above results remain consistent irrespective of the firm performance indicator we implement.

CO1309: **Nonnegativity constraints for N-dimensional asymmetric power heavy/MEM/GARCH Models and a new mixture hormulation**
Presenter: **Yongdeng Xu**, Cardiff University, United Kingdom
Co-authors: Menelaos Karanasos
Results on the derivation of tractable nonegativity constraints for N-dimensional heavy/MEM/GARCH systems are reviewed and generalized. These nonnegativity conditions are expressed in terms of matrix inequalities which can be solved easily. Numerical examples are included to show the effectiveness of the proposed method.

CC1555: **Labor market dynamics and labor market institutions: The cointegrated VAR approach**
Presenter: **Sylwia Roszkowska**, University of Lodz, Poland
The aim is to analyze long-run relationships between main labor market variables (wages, unemployment, productivity) and labor market institutions using data for Poland. The provided analyses focus on the labor market variables, labor market institutions (such as minimum wages, union density, EPL, tax wedges) and prices, interest rate and exchange rate. The data covers period from 1995:01 to 2015:04. The cointegrated VAR (CVAR) analysis is used, which allows for identification of the long run relations, common stochastic trends and the estimation of the adjustment dynamics in the system. Having estimated parameters of aforementioned relations, the main driving forces of the labor market can be detected. The results are as follow. The increasing product market competition seems to be the main driving force during the analyzed period of convergence of Polish economy to the more advanced European economies. The Polish inflation rate has adjusted over long run to the European purchasing power parity and Balassa-Samuelson effect on consumer prices can be confirmed. Moreover, labor market tightness matters in the process of forming the real wages . The tighter labor market is, the lower the bargaining position of employees and the lower pressure on wages are. The extended Phillips curve hypothesis is also confirmed.

CC1739: **Asymptotic normality of the QML estimator of the EGARCH(1,1) model**
Presenter: **Dimitra Kyriakopoulou**, Universite Catholique de Louvain, Belgium
The aim is to investigate the asymptotic properties of the quasi-maximum likelihood estimator (QMLE) for the EGARCH model. Sufficient conditions under which the EGARCH(1,1) processes have stationary first and second order variance derivatives, and the expectation of the supremum norm of the second order log-likelihood derivative is finite are established. The existence of such moment bounds permits the establishment of the CLT of the score and the uniform SLLN of the Hessian, so that the asymptotic normality of the QML estimators is proved.

CO710 Room 216 CREDIT RISK, LIQUIDITY AND BOND PRICING Chair: Ana Escribano

CO0815: **On the effect of bond liquidity on financial contracts**
Presenter: **Yuan Wang**, Concordia University, Canada
The aim is to examine whether and how the bond market liquidity affect the contractual terms of newly issued public bonds. We find firms with better bond market liquidity issue bonds with lower cost of debt, fewer restrictive covenant and longer maturity. These results hold after we control for firm fixed effects, various bond and firm characteristics, use alternative proxies of bond liquidity, and employ an instrument variable regression to deal with the endogeneity of bond liquidity. To identify the causal effect of bond liquidity on debt contracts, we conduct a quasi-natural experiment study by using the implementation of TRACE as an exogenous shock to bond market liquidity. We find that the improvement of bond liquidity due to the introduction of TRACE indeed affects debt contracts as expected. Further investigation shows that the impact of bond liquidity is more pronounced in firms with poorer credit rating and more short-term debt. These findings are consistent with the notion that improvement of bond market liquidity reduces firms rollover and credit risk so that firms can issue bonds with more favorable contractual terms.

CO1053: **Liquidity and the size of trades around credit event news**
Presenter: **Ana Escribano**, Universidad de Castilla-La Mancha, Spain
Co-authors: M Dolores Robles, Pilar Abad, Antonio Diaz
The aim is to investigate the impact of credit rating downgrades on the liquidity and trading behavior of both segments of trading, the institutional and the retail sized ones, in the U.S. corporate bond market. Using the TRACE dataset, we analyze the information content of these events and potential information asymmetries, distinguishing between trade sizes. We propose two additional hypotheses: the regulatory constraints and the

risk tolerance limits that may force some institutional and retail bondholders to sell after certain downgrades. Our results show trading anticipation before downgrades that is consistent with the existence of both types of investors, informed and uninformed. We also observe fire sales and price concessions depending on rating specific regulatory constraints, capital requirements and risk tolerance limits.

CO1137: **Time varying illiquidity of European corporate bonds**
Presenter: **Ranko Jelic**, University of Sussex, United Kingdom
Co-authors: Dietmar Maringer, Xiaohua Chen

Excess returns and time varying illiquidity are studied for 17 European corporate bond indices during 2000-2014. It is plausible that returns and illiquidity tend to vary over time depending on the state of the economy which in turn depends on some observable external conditions which are then reflected in the state-indicator. The above scenario can be modelled through a threshold regime switching (TS) model in which the transition mechanism is driven by observable state variables, and the transition between the regimes is abrupt at a threshold identified by the optimization procedure. We therefore develop a TS model with various candidate transition variables and a genetic algorithm that enables us to identify the most suitable factors that govern regime switches. We make a further methodological contribution by addressing model over-fitting by using the Bayesian Information Criteria. The results identify changes in realized stock volatility as the best transition variable. In a regime with low volatility, illiquidity levels and shocks are mostly insignificant. In a stress regime with increasing volatility, illiquidity bears the most significant impact on bond index returns. We find significant differences in the illiquidity effects across bond indices for different maturities, ratings and industries.

CO0858: **Credit rating news and stock return synchronicity: Informational effects of regulation reforms**
Presenter: **Dolores Robles**, Universidad Complutense de Madrid, Spain
Co-authors: Pilar Abad, Rodrigo Ferreras

The informational content of changes in credit risk of corporate debt issues is analyzed. We study how credit rating variations announced by the more globally important Credit Rating Agencies convey firm-specific information affecting the synchronicity of stock returns. We analyse US companies from 1996 to 2014. We focus on the informational effects of the last regulatory reforms on the US financial market. We find informational content that depends on firm characteristics and on features of the rating announcement. The new regulations affecting the credit rating agencies has modified the informativeness level of these rating changes, sometimes on the direction expected when the rule was legislated and other to the opposite side. We also confirm that multi-agency announcements, the fallen angels and the variation on the number of notches affect directly to the information content of the rating announcement.

CO599 Room 112 MACROECONOMIC SPILLOVERS AND MONETARY POLICY Chair: Hilde Bjornland

CO1342: **Financial shocks and inflation dynamics**
Presenter: **Esteban Prieto**, Deutsche Bundesbank, Germany
Co-authors: Sandra Eickmeier, Angela Abbate

The aim is to assess the effects of financial shocks on inflation, and to what extent financial shocks can account for the "missing disinflation" during the Great Recession. We apply a vector autoregressive model to US macro and financial data over 1988-2015 and identify financial shocks through sign restrictions. Our main findings are as follows. Expansionary financial shocks temporarily lower inflation, i.e. they are disinflationary. This result withstands a large battery of robustness checks. Moreover negative financial shocks may have helped preventing a deflation during the crisis. We then explore the transmission channels of financial shocks relevant for inflation and found that effectiveness of the cost channel can explain the inflation response, a finding which could guide future theoretical work. Policy implications are twofold. First, financial shocks which raise output and, at the same time, lower inflation bring about an additional trade-off for a central bank which stabilizes output and prices. Second, a monetary policy which aims at stimulating credit supply and lowering funding costs risks pushing inflation down, something that should be avoided in a low inflation environment.

CO1334: **Economic policy uncertainty spillovers in booms and busts**
Presenter: **Giovanni Caggiano**, University of Padova, Italy

A nonlinear VAR is estimated to quantify the impact of economic policy uncertainty shocks originating in the U.S. on the Canadian business cycle in booms and busts. We find strong evidence in favor of asymmetric spillover effects. Uncertainty shocks originating in the U.S. explain about 27% of the variance of the 2-year ahead forecast error of the Canadian unemployment rate in periods of slack vs. 8% during economic booms. Counterfactual simulations lead to the identification of a novel "economic policy uncertainty spillovers channel". According to this channel, spikes in U.S. economic policy uncertainty foster economic policy uncertainty in Canada in first place and, because of the latter, lead to a temporary increase in the Canadian unemployment rate. This channel is shown to work only in periods of slack.

CO1337: **On uncertainty: A topic model approach**
Presenter: **Vegard Larsen**, Norges Bank, Norway

Newspaper content is used to create category specific measures of uncertainty. The news are classified using a topic model called Latent Dirichlet Allocation. This model allow us to label a large number of news articles by their underlying content. Uncertainty measures are calculated based on the count of uncertainty terms over time within specific types of news articles in Norways largest business newspaper. The uncertainty measures are available at a daily frequency and they capture well known events linked to both uncertainty at an aggregate level, but also category specific events. We investigate the response from changes in different types of uncertainty to aggregate macroeconomic variables in a structural VAR framework. Several of the category specific uncertainty measures foreshadow weaker macroeconomic outcomes. But, different uncertainty categories can have different effects. Grouping related categories together, we find that an increase in uncertainty related to the categories Household, Financial, and Oil lead to a sizable negative effect on the Norwegian economy. Policy uncertainty seem not to be important for the Norwegian business cycle when it is grouped together, but an increase in uncertainty related to Fiscal policy is followed by weaker economic conditions.

CO0482: **On central banks response to developments in the global economy**
Presenter: **Sepideh Zahiri**, BI Norwegian Business School, Norway
Co-authors: Hilde Bjornland, Leif Anders Thorsrud

Our result suggests that central banks are not responding timely to developments in the global economy. To arrive at this conclusion we construct a real-time data set of interest rate projections from central banks in three small open economies; New Zealand, Norway, and Sweden, and analyse if revisions to these projections can be predicted by timely information. We find a systematic role for forward looking international indicators in predicting the revisions to the interest rate projections. In contrast, using similar indexes for the domestic economy yields insignificant results. Furthermore, we find that the interest rate revisions matters! Constructing a measure of monetary policy shocks based on the forecast errors in each country, we show that financial variables react on impact to the monetary policy shocks, while there is an instant and gradual decline in output and inflation.

CO385 Room 217 FINANCIAL TIME SERIES MODELLING AND FORECASTING Chair: Alessandra Amendola

CO0424: **A simple Markovian process with hyperbolic rates of convergence.**
Presenter: **Alessandro Palandri**, Dublin City University, Ireland

A Markovian process is introduced with hyperbolic reversion rates to the fixed point. Proofs are provided regarding the uniqueness of the fixed point, its domain of attraction and the resulting uniform ergodicity. Two distinct lag-p generalizations of the baseline lag-one specification are presented alongside their properties. The proposed process is compared to standard AR(p) auto-regressions: unit-root tests, predictions, forecasts and the ability to capture various stylized facts. Furthermore, the proposed functional form is applied to discrete-time models of conditional variances resulting into a new GARCH specication as well as a new specication for the modeling of realized measures of variance.

CO0685: **Forecasting volatility with the asymmetric GARCH-MIDAS model**
Presenter: **Vincenzo Candila**, University of Salerno, Italy
Co-authors: Alessandra Amendola

Recently proposed, the GARCH-MIDAS model has gained much attention in volatility modelling literature. It allows us to forecast the volatility of a financial instrument by considering two multiplicative components, a short- and long-run one. The long-run component filters the information provided by some additional exogenous variables, observed at lower frequencies than those of the financial instrument of interest, into the short-run component. In this framework, the exogenous variables are usually some macroeconomic determinants of the variable of interest or some measures of the state of the economy. So far, the GARCH-MIDAS model only considers an average effect of these variables on the volatility of the dependent variable. However, positive and negative variations of the exogenous variables may have different effects on the volatility. And, in terms of risk management, the knowledge that only a given direction in the exogenous variable variation would produce a larger increasing of volatility could be very useful. Thus, the proposal is to forecast the volatility by considering an asymmetric GARCH-MIDAS specification, in order to investigate the impacts of exogenous variables increasing or decreasing have on the volatility of the variable of interest.

CO1305: **Combination of conditional quantile forecasts: An application to value at risk modeling**
Presenter: **Andre Portela Santos**, Universidade Federal de Santa Catarina, Brazil
Co-authors: Guilherme Valle Moura, Joao Frois Caldeira

An effective and computationally fast approach is introduced to combine conditional quantile forecasts. The approach uses the information of the relevant loss function for the quantile problem in order to define forecast combination weights in a dynamic fashion. Two important advantages of the proposed method are that i) does not require numerical optimization of the combination weights, which facilitates implementation when a large cross section of individual forecasts is considered and ii) the aggressiveness in the allocation across alternative forecasts and the trimming of worse forecasts can be easily calibrated with a single parameter. An empirical exercise based on a data set with 50 assets shows that combinations of portfolio VaR forecasts are accurate and outperform the 16 individual models in many instances. The results hold for both long and short portfolio positions as well as in high volatile subsamples.

CO0840: **Risk evaluation and pricing: Risk differential and return predictability patterns across markets and countries**
Presenter: **Francesco Violante**, CREATES Aarhus University, Denmark
Co-authors: Jeroen Rombouts, Lars Stentoft

The relationship between future aggregate market returns and variance risk premia has been recently at the center of an active debate. While some studies suggests that the variance risk premium suggest the existence of a strongly positive correlation, recent contributions show that such predictability is largely attributable to the effect of extreme variance events and more precisely to the agents perception of such events. Furthermore, mostly due to data availability, empirical evidence is generally reported for the U.S. as represented by the S& P500 portfolio. In a comprehensive empirical study, we first extend the results to different market-wide U.S. portfolios, i.e. S& P500, DJIA30, RUSSELL2000, NASDAQ100, characterized by different asset composition, degree of diversification and dimension. Second, we uncover commonalities and idiosyncrasies in return predictability patterns at international level by considering several major European and Asian market portfolio.

CO307 Room 101 BIG DATA IN ECONOMICS AND FINANCE Chair: Veronika Czellar

CO0581: **Information networks, profits, and pricing in financial markets**
Presenter: **Johan Walden**, UC Berkeley, Haas School of Business, United States

Information diffusion is studied within a network of investors in a financial market. In equilibrium, agents profits are closely related to their centrality in the network. Moreover, the network's topology affects asset pricing and trading volume dynamics. We take the model to the data, using account level datasets of all trades and traders in the stock market. Consistent with theory, we find that central investors earn higher profits and trade earlier than peripheral investors with respect to information events, and that network topology is related to asset price dynamics. Overall, our results support a very high-dimensional and decentralized view of how information is incorporated into asset prices.

CO1492: **Limited participation in the joint behavior of asset prices and individual consumptions**
Presenter: **Veronika Czellar**, EDHEC, France
Co-authors: Francois Le Grand, Rene Garcia

We propose an asset pricing model featuring both limited participation and heterogeneity, in which agents randomly participate in the bond and stock markets. Depending on their financial market participation, agents are endowed with different consumption processes, which satisfy Euler conditions for both bonds and stocks. We propose an indirect inference method to jointly estimate individual consumption, bond and stock market participation processes, and stock returns. We use individual US consumption (CEX), bond and stock return data. Our estimated model performs very well at jointly replicating the equity premium and the unequal distribution of individual consumptions. We find that our estimation is consistent with endogenous bond market participation and with a non-zero stock market participation cost. We show that the role of limited market participation is necessary to match the heterogeneity in individual consumptions, while unlimited participation leads to very inaccurate replications of consumption and return distributions.

CO1744: **News measures, volatility and jumps**
Presenter: **Francesco Poli**, University of Padova, Italy
Co-authors: Massimiliano Caporin

From two professional news providers we retrieve news stories and earnings announcements of the S&P 100 constituents and a set of 10 macroeconomic fundamentals. We thus create an extensive and innovative dataset which contains information with minute precision, useful to analyze the link between news and asset price dynamics. We develop a novel text-analysis technique to detect the sentiment of a financial text of any type, size and audience, and propose a set of more than 4K news-based variables that provide natural proxies of the information used by heterogeneous market players. We first shed light on the impact of news on daily realized volatility and select news measures by penalized regression. Then, we distinguish the relative importance between news measures and use them to forecast volatility. Finally, we investigate the relation between news and intraday jumps, within a penalized logit approach.

CO0526: **Firm characteristics and the cross section of stock returns: A portfolio perspective**
Presenter: **Alberto Martin Utrera**, Lancaster University Management School, United Kingdom
Co-authors: Victor DeMiguel, Francisco J Nogales, Raman Uppal

More than 300 characteristics have been proposed to explain the cross-section of stock returns. We study which are significant for portfolio construction and why. We highlight three findings. First, without transaction costs five characteristics are significant because they increase mean return, and also reduce the risk of the portfolio of characteristics. A sixth characteristic (beta) is significant only because it reduces risk, even though its mean return is negligible. Prominent characteristics such as momentum and book to market are not significant because their contribution to the portfolio mean is not sufficiently large to offset their contribution to portfolio risk. Second, with transaction costs the number of significant characteristics increases. This is because the rebalancing trades across characteristics offset each other, reducing marginal transaction costs by 65% relative to those from trading on the basis of single characteristics. Third, exploiting a large set of characteristics results in superior out-of-sample performance, with Sharpe ratio of returns net of transaction costs 150% larger than that of the value-weighted portfolio, and 100% larger than that of portfolios that exploit only the three traditional size, momentum, and book-to-market characteristics.

CG270 Room 109 CONTRIBUTIONS IN PORTFOLIO OPTIMIZATION AND PERFORMANCE EVALUATION Chair: Ralf Wunderlich

CC0404: **The reopening of Dubins and Savage casino in the era of diversification**
Presenter: **Isaac Meilijson**, Tel Aviv University, Israel

In Dynamic Programming, mixed strategies consist of randomizing the choice of actions. In some problems, such as portfolio management,it makes sense to diversify actions rather than choosing among them purely or randomly. Optimal betting in casinos and roulette by a gambler with fixed goal has been studied following Dubins and Savage without the element of diversification (betting simultaneously on different holes of the roulette), once it was proved (Smith's theorem) that diversification doesn't increase the probability of reaching the goal. We question the scope of this finding, that was based on the assumption that the holes on which gamblers can bet are disjoint, such as 1 and BLACK in regular roulette. A counterexample is provided in which holes are nested, such as 1 and RED. Thus, it may be rational for gamblers with fixed goal to place chips on more than one hole at the table. Unlike risk averters, these agents diversify over co-monotone assets (nested holes) but not over antithetic assets (disjoint holes).

CC1545: **Portfolio optimization based on GARCH-EVT-copula CVaR and mean-variance models**
Presenter: **Andreas Stephan**, Jonkoping University, International Business School, Sweden
Co-authors: Maziar Sahamkhadam, Ralf Ostermark

Regarding GARCH-EVT-Copula models, most studies have been performed on Student T and Gaussian copulas, with a few considering Clayton copula. By adding Frank and Gumbel, we utilize GARCH-EVT-Copula models for optimizing and comparing Min-CVaR and Mean-Variance portfolios. In order to perform the portfolio optimization, ten stock markets are considered including S& P 500, FTSE 100, DAX 30, EURO STOXX 50, MSCI World, CAC 40, OMXC 20, OMXH, OMXS 30 and TOPIX. The sample period starts at August 26, 1996 which includes 5218 total observations. By dividing the sample into three periods and applying rolling portfolio, different rebalancing strategies are considered in order to investigate the effects of Financial Crisis. To estimate the optimal weights, Linear and Quadratic Programing are implemented in Min-CVaR and Mean-Variance portfolios, respectively. The results show that Frank copula leads to the lowest CVaR. Moreover, before the crisis, Student T and Gaussian copulas have the highest Sharpe Ratios with quarterly and monthly rebalancing strategies. During the crisis, Clayton copula results to a higher Sharpe Ratio. Interestingly, after crisis, Student T and Gaussian copulas outperform other models with daily and semi-annually rebalancing. Finally, in most cases, Mean-Variance portfolio shows better results than Min-CVaR portfolio.

CC1433: **Black-Litterman model for continuous distributions and general risk measures**
Presenter: **Andrzej Palczewski**, University of Warsaw, Poland

The Black-Litterman methodology of portfolio optimization combines statistical information on asset returns with investors views within the Markowitz mean-variance framework. The main assumption underlying the Black-Litterman model is that asset returns and investors views are multivariate normally distributed. However, the empirical research demonstrates that the distribution of asset returns has fat tails and is asymmetric, which contradicts normality. Moreover, recent advances in risk measurement advocate replacing the variance by risk measures that take account of tail behavior of the portfolio return distribution. We extend Black-Litterman theory into general continuous distributions with the risk measured by general deviation risk measures. Using ideas from the Black-Litterman methodology, we design analytical and numerical methods (with variance reduction techniques) for the inverse portfolio optimization that extracts in a stable way statistical information from historical data. We introduce a quantitative model for stating investors views and blending them consistently with the market information via Bayes formula. The theory is complemented with the design of efficient numerical methods. We conclude with a number of practical examples that demonstrate significant impact of the choice of istributions on optimal portfolio weights to the extent that the classical Black-Litterman procedure cannot be viewed as an adequate approximation.

CC1372: **Testing out-of-sample portfolio performance**
Presenter: **Winfried Pohlmeier**, University of Konstanz, Germany
Co-authors: Ekaterina Kazak

While the literature on portfolio choice largely concentrates on stabilization strategies, little attention has been devoted to the quality of performance tests used to check, if a portfolio strategy can significantly outperform an alternative one in terms of a given performance measure. The quality of portfolio performance tests based on out-of-sample returns are studied. We argue that the puzzling empirical results of inferior performance of the theoretically superior strategies based on the out-of-sample comparison are partly resulting from the low power properties of these tests. We emphasize the importance of the underlying return distribution and show that the out-of-sample portfolio returns follow a mixture distribution depending on the return vector, but also the estimated portfolio weights. In our simulation study with the proposed mixture distribution design we show that in the realistic cases the test difference is overemphasized. The main issue is the low testing power, which automatically leads to a conclusion, that the benchmark strategy cannot be outperformed. For the applied researcher we provide some guidance to cope with the problem of low power.

EI673 Room Graduation hall FUNCTIONAL DATA ANALYSIS: THEORY AND APPLICATIONS Chair: Jeng-Min Chiou

EI0307: **Detection of periodicity in functional time series**
Presenter: **Siegfried Hoermann**, Univ libre de Bruxelles, Belgium
Co-authors: Gilles Nisol, Piotr Kokoszka

Periodicity is one of the most important characteristics of time series, and tests for periodicity go back to the very origins of the field. The importance of such tests has manifold reason. One of them is that most inferential procedures require that the series be stationary, but classical stationarity tests (as e.g. KPSS procedures) have little power against a periodic component in the mean. In this account we respond to the need to develop periodicity tests for functional time series (FTS). Examples of FTS's include annual temperature or smoothed precipitation curves, daily pollution level curves, various daily curves derived from high frequency asset price data, daily bond yield curves, daily vehicle traffic curves and many others. One of the important contributions of this article is the development of a fully functional ANOVA test for stationary data. If the functional time series (Y_t) satisfies a certain weak-dependence condition, then, using a frequency domain approach, we obtain the asymptotic null-distribution (for the constant mean hypothesis) of the functional ANOVA statistic. Adapting ANOVA for dependence is one way to conduct periodicity analysis. It is suitable when the periodic component has no particular form. If, however, the alternative is more specific or the period is large then we can construct simpler and more powerful tests. The power-advantage will be illustrated in simulations and by a theoretical case study.

EI0905: **Goodness-of-fit tests for models with functional data**
Presenter: **Wenceslao Gonzalez-Manteiga**, University of Santiago de Compostela, Spain
Co-authors: Juan A Cuesta-Albertos, Eduardo Garcia-Portugues, Manuel Febrero-Bande

Different goodness-of-fit tests for some null hypothesis of models with functional data are given. The simplest versions of these tests are based on the residual marked empirical process indexed by random projections for testing the hypothesis of a functional linear model with scalar response. Some asymptotic distribution of the mentioned projected empirical process is obtained under general conditions using estimation by functional principal components. Other considered models are the functional partially linear model or the functional linear model with functional response. The asymptotic distribution of the Kolmogorv-Smirnov and Cramer-von Mises tests built with the associated process is considered. The exact distribution of the tests is also calibrated by bootstrap and the finite sample properties are illustrated by simulations as well as real data applications.

EI1368: **Weighing schemes for functional data**
Presenter: **Xiaoke Zhang**, University of Delaware, United States
Co-authors: Jane-Ling Wang

Nonparametric estimation of mean and covariance functions is important in functional data analysis. We investigate the performance of local linear smoothers for both mean and covariance functions with a general weighing scheme, which includes two commonly used schemes, equal weight per observation (OBS), and equal weight per subject (SUBJ), as two special cases. We provide a comprehensive analysis of their asymptotic properties on a unified platform for all types of sampling plan, be it dense, sparse, or neither. These two weighing schemes are compared both theoretically and numerically. In addition to OBS and SUBJ schemes, we propose a new class of weighing schemes in terms of the mixture of their weights. We might also explore optimal weighing schemes.

EO009 Room 212 VINE COPULAS: THEORY AND APPLICATIONS Chair: Claudia Czado

EO0171: **Model selection for discrete regular vine copulas**
Presenter: **Anastasios Panagiotelis**, Monash University, Australia
Co-authors: Claudia Czado, Jakob Stoeber, Harry Joe

Discrete vine copulas provide a flexible modeling framework for high-dimensional data and have significant computational advantages over competing methods. A vine-based multivariate probability mass function is constructed from bivariate copula building blocks and univariate marginal distributions. However, even for a moderate number of variables, the number of alternative vine decompositions is very large and additionally there is a large set of candidate bivariate copula families that can be used as building blocks in any given decomposition. Together, these two issues ensure that it is infeasible to evaluate all possible vine copula models. Two greedy algorithms are introduced for automatically selecting vine structures and component pair-copula building blocks. The algorithms are tested in a simulation study that is itself driven by real world data from online retail. We show that both algorithms select vines that provide accurate estimates of the joint probabilities. Using three different f-divergences as criteria, our algorithms outperform a Gaussian copula benchmark, especially for data with high dependence. Finally, the selection algorithm is applied to data from the General Social Survey and outperforms a Gaussian copula benchmark using both in-sample and out-of-sample criteria.

EO0281: **Regional climate models and the spatial dependence of precipitation**
Presenter: **Ingrid Hobaek Haff**, University of Oslo, Norway
Co-authors: Arnoldo Frigessi, Douglas Maraun

How well a suite of regional climate models (RCMs) from the ENSEMBLES project represents the residual spatial dependence of daily precipitation is investigated. The study area is a 200 by 200 km region in south central Norway, with RCMs driven by ERA-40 boundary conditions at a horizontal resolution of approximately 25 by 25 km. We model the residual spatial dependence with pair-copula constructions, which allows us to assess both the overall and tail dependence in precipitation, including uncertainty estimates. The selected RCMs reproduce the overall dependence well, though the discrepancies compared to observations are substantial. All models overestimate the overall dependence in the west-east direction. They also overestimate the upper tail dependence in the north-south direction during winter, and in the west-east direction during summer, whereas they tend to underestimate this dependence in the north-south direction in summer. Moreover, many of the climate models do not simulate the small-scale dependence patterns caused by the pronounced orography well. However, the misrepresented residual spatial dependence does not seem to affect estimates of high quantiles of extreme precipitation aggregated over a few grid boxes. The underestimation of the area-aggregated extreme precipitation is due mainly to the well-known underestimation of the univariate margins for individual grid boxes, suggesting that the correction of RCM biases in precipitation may be feasible.

EO0504: **Predicting conditional quantiles using D-Vine copulas**
Presenter: **Daniel Kraus**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado

Quantile regression, that is the prediction of a random variable's quantiles conditioned on other random variables taking on certain values, has perpetually gained importance in statistical modeling and financial applications. Often used methods such as linear quantile regression fail to detect complex dependence behaviors between response and predictors. They also might result in crossing quantiles. We introduce a new semiparametric quantile regression method based on sequentially fitting a conditional likelihood optimal D-vine copula to given data resulting in highly flexible models with easily extractable conditional quantiles. As a subclass of regular vine copulas, D-vines enable the modeling of multivariate copulas in terms of bivariate building blocks, a so-called pair-copula construction (PCC). The proposed algorithm works fast and accurate even in high

dimensions, incorporates an automatic variable selection and is applicable to continuous as well as discrete data. In a simulation study the improved accuracy of the approach in comparison to established methods such as linear or nonparametric quantile regression is highlighted.

EO1181: **Precipitation modeling via spatial vine copulas**
Presenter: **Benedikt Graeler**, Ruhr University Bochum, Germany
Spatial vine copulas, where the bivariate copula building blocks are parameterized by the separating distance between locations, have proven useful to model skewed spatial random fields. Modeling and interpolating rainfall adds further challenges, as the marginal distribution is not only skewed, but also contains a majority of zeros. Furthermore, the dependence of a rainfall field may strongly differ between rainfall events ranging from extreme local rainfall to widespread continuous rainfall. These challenges are addressed with a spatial vine copula approach to model and interpolate precipitation.

EO495 Room Board meeting room I APPLYING EXTREME VALUE STATISTICS IN FINANCE AND ECONOMICS Chair: Chen Zhou

EO0402: **Tail risk in government bond markets and ECB unconventional policies**
Presenter: **Bernd Schwaab**, European Central Bank, Germany
Co-authors: Xin Zhang
A novel observation-driven model is derived to study the time variation in the tail shape for time series observations from a wide class of fat-tailed distributions. Monte Carlo experiments suggest that the model reliably captures tail shape variation in a variety of simulation settings. In an empirical study of sovereign bond yields at a high frequency, we demonstrate that unconventional monetary policies adopted by the European Central Bank between 2010-2012, specially its Securities Markets Programme and Outright Monetary Transactions, lowered the tail risk associated with holding certain sovereign bonds during the euro area sovereign debt crisis.

EO0426: **Robust bounds for multivariate extreme value distributions**
Presenter: **Sebastian Engelke**, Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: Jevgenijs Ivanovs
Univariate extreme value theory is used to estimate the value at risk of an asset in regions where few or no observations are available. It is based on the asymptotic result that the maximum of the data follows approximately a generalized extreme value distribution. The worst case bounds for high exceedance probabilities that are robust against incorrect model assumptions of the extremal types theorem has been recently studied. For two or more dependent assets, multivariate extreme value theory provides an asymptotically justified framework for the estimation of joint exceedances. Here, the strength of dependence is crucial and it is typically modeled by a parametric family of distributions. We analyze bounds that are robust against misspecification of the true dependence between assets. They arise as the explicit solution to a convex optimization problem and take a surprisingly simple form. In a financial context, these robust bounds can be interpreted as the worst-case scenarios of a systematic stress test. We show the importance of this approach in simulations and apply it to real data from finance.

EO0597: **Asymptotic analysis of portfolio diversification**
Presenter: **Fan Yang**, University of Waterloo, Canada
Co-authors: Chen Zhou
Interconnectedness and heavy-tailedness of risks may result in high systematic risk. When the number of risks are fixed, maximizing diversification benefits can lower the portfolio risk. The diversification ratio, also known as the risk concentration, is used to measure the diversification benefit. We aim to construct the most diversified portfolio when risks are heavy-tailed and dependent, modeled by the multivariate regularly variation structure. Since explicit solutions to such optimization problems are generally not available and the computation burden of such solution is very heavy and not stable, asymptotic analysis is conducted as an alternative way to study them. The asymptotic solution is showed to be a good approximation to the true solution and the approximation error is given as well. Our theoretical results are supported by extensive numerical simulations. The estimation of the asymptotic solution is provided and then our strategy is also applied to real market data.

EO1085: **Estimation of the marginal expected shortfall using extreme expectiles**
Presenter: **Abdelaati Daouia**, UMR5314 TSE-R CNRS, France
Co-authors: Stephane Girard, Gilles Stupfler
Tail expectiles are used to estimate the Marginal Expected Shortfall (MES), an important factor when measuring the systemic risk of financial institutions. Denoting by X and Y, respectively, the loss of the equity return of a financial firm and that of the entire market, the MES is equal to $E(X|Y > t)$, where t is a high threshold reflecting a substantial market decline. For a wide class of bivariate distributions of (X,Y) and for an extreme expectile t of the distribution of Y, we construct asymptotically normal estimators of the MES. The concept of expectiles is a least squares analogue of quantiles which defines itself a coherent risk measure. A rival procedure using extreme tools is based on quantiles. The expectile-based MES is definitely less pessimistic than the quantile-based MSE whenever the tail indices of X and Y are smaller than $1/2$. This is typically the case in the financial setting as illustrated via a concrete application to three large US investment banks.

EO003 Room 312 BIG DATA: CONVERGENCE OF STATISTICS AND OPTIMIZATION Chair: Stephane Chretien

EO0416: **Time-delay reservoir computers and high-speed information processing capacity**
Presenter: **Juan-Pablo Ortega**, University St. Gallen, Switzerland
Co-authors: Lyudmila Grigoryeva, Julie Henriques, Laurent Larger
The aim is to show how various ideas coming from the nonlinear stability theory of functional differential systems, stochastic modeling, and machine learning, can be put together in order to create an approximating model that explains the working mechanisms behind a certain type of reservoir computers. Reservoir computing is a recently introduced brain-inspired machine learning paradigm capable of excellent performances in the processing of empirical data. We focus on time-delay based reservoir computers that have been physically implemented using optical and electronic systems and have shown unprecedented data processing rates. Reservoir computing is well-known for the ease of the associated training scheme but also for the problematic sensitivity of its performance to architecture parameters. The reservoir design problem is addressed, which remains the biggest challenge in the applicability of this information processing scheme. Our results use the information available regarding the optimal reservoir working regimes in order to construct a functional link between the reservoir parameters and its performance. This function is used to explore various properties of the device and to choose the optimal reservoir architecture, thus replacing the tedious and time consuming parameter scannings used so far in the literature.

EO1025: **Trade-off between computational cost and estimate accuracy: Some attempts**
Presenter: **Christophe Biernacki**, Inria, France
Co-authors: Alain Celisse, Maxime Brunin
Most estimates practically arise from algorithmic processes aiming at optimizing some standard, but usually only asymptotically relevant, criteria. Thus, the quality of the resulting estimate is a function of both the iteration number and also the involved sample size. An important question is to

design accurate estimates while saving computation time, and we address it in the simplified context of linear regression here. Firstly, we fix the sample size. We focus on estimating an early stopping time of a gradient descent estimation process aiming at maximizing the likelihood. It appears that the accuracy gain of such a stopping time increases with the number of covariates, indicating potential interest of the method in real situations involving many covariates. Secondly, we authorize both the number of iterations and the (sub)sample size to be estimated for providing the optimal estimate accuracy, while respecting now a given maximum computational cost. Indeed, restricting estimation to subsamples is a standard behavior of practitioners in a "Big Data" context for computational reasons. Our aim is thus to formalize such an empirical process to provide established recommendations for selecting simultaneously the couple iteration number and subsample size.

EO1260: **A unified rule to select a threshold**
Presenter: **Sylvain Sardy**, University of Geneva, Switzerland

The last twenty years have seen the development of many thresholding estimators (best subset, group lasso, square root lasso, Subbotin lasso, total variation, elastic net, to cite a few) in many settings (GLM, low-rank matrix estimation, density estimation). They all require a threshold selection. To that aim, we propose a unified selection rule based on the concept of a zero-thresholding function and a null-thresholding statistic. We apply the methodology to a high dimensional inverse problem in Cosmology.

EO1333: **Some results in nonconvex estimation**
Presenter: **Stephane Chretien**, NPL, United Kingdom

Estimation with nonconvex criteria is crucial in low rank matrix completion, low rank matrix recovery and Semi-Definite programming problems based on the Burer-Monteiro factorization approach. It has been proved recently that local minimizers are close to the true solution in the low rank recovery problem under some restricted isometry types of conditions. We present an overview and some generalisations of these results based on random matrix theory.

EO175 Room 214 DESIGN OF EXPERIMENTS Chair: Kalliopi Mylona

EO0450: **Optimal designs for fractional polynomial models**
Presenter: **Victor Casero-Alonso**, University of Castilla-La Mancha, Spain
Co-authors: Jesus Lopez-Fidalgo, Weng Kee Wong

Fractional polynomials have been shown to be much more flexible than polynomials for fitting continuous outcomes in the biological and health sciences. Despite their increasing popularity, design issues for fractional polynomials models have never been addressed. The aim is to provide D- and I-optimal experimental designs for prediction using fractional polynomial models, evaluates their properties and provides a catalogue of design points useful for fractional polynomial models. We also construct optimal designs for selected multi-factor fractional polynomials. As applications, we re-design two studies using optimal designs developed here and show they can produce substantial gains in terms of cost and quality of the statistical inference. We also provide a user friendly applet for generating optimal designs for fractional polynomials up to degree 3.

EO0522: **Optimal design for experiments with partially observed data**
Presenter: **Stefanie Biedermann**, University of Southampton, United Kingdom
Co-authors: Kim Lee, Robin Mitra

Missing values complicate statistical analyses, but are particularly problematic when constructing optimal designs, as it is not known which values are missing at the design stage. When data are missing at random it is possible to incorporate this information into the optimality criterion that is used to find designs. However, when data are not missing at random such a framework may no longer be appropriate. We show the latest developments in designing experiments when data are missing at random, and develop a framework that addresses the specific challenges that not missing at random values present.

EO0599: **Response surface designs robust to missing observations**
Presenter: **Luzia Trinca**, Unesp, Brazil
Co-authors: Marcelo da Silva, Steven Gilmour

Design criteria which allow pure error degrees of freedom may produce designs that break down when even a single run is missing. The inclusion, in a compound criteria, of a measure of leverage uniformity is proposed in order to produce designs that are more robust to missing observations. By appropriately choosing the weights of each part of the criterion, robust designs are obtained that are also highly efficient in terms of other properties. Several examples are used for illustration.

EO0940: **Optimal design of two-stratum experiments in the presence of autocorrelation**
Presenter: **Daniel Palhazi Cuervo**, University of Antwerp, Belgium
Co-authors: Peter Goos, Kenneth Sorensen

There are several experimental scenarios in which a complete randomization is not possible. This usually happens when not all observations can be made under homogeneous conditions, or when there are factors whose levels are hard or expensive to change. For these cases, it is possible to generate experimental designs that explicitly take into account these limitations. This kind of designs, called two-stratum designs, groups the observations that are made under similar experimental conditions. The data from these experiments is usually analyzed assuming that every pair of observations within a given group has the same correlation. We study experimental scenarios in which not all pairs of observations within each group are correlated to the same extent. Instead, we consider observations that are closer to each other (either in time or space) to be more correlated than observations that are further apart. This might be expected when the observations are carried out sequentially, or when the experimental conditions are slightly different for each observation within the group. We generate optimal designs for two-stratum experiments under these conditions, and we compare their statistical efficiency to that of traditional designs (constructed assuming that all observations within a group are correlated to the same extent).

EO071 Room 305 B OUTLIERS, ROBUSTNESS AND EXTREMES IN TIME SERIES Chair: Roland Fried

EO0550: **Aspects of composite likelihood inference in time series models**
Presenter: **Xanthi Pedeli**, Ca Foscari University, Italy
Co-authors: Cristiano Varin

Statistical inference in time series models has traditionally been based on the likelihood principle. However, computation of the likelihood function at the basis of frequentist and Bayesian inferential frameworks is often intractable in high-dimensional time series models. To this end, several simulation-based methods to approximate the likelihood have been proposed in the literature. Although simulation methods may make the estimation problem feasible, they are often computationally intensive and the quality of the numerical approximations may be difficult to assess. The method of composite likelihood (CL) has been proposed as a surrogate of intractable likelihoods in highly structured models. CL is an inference function constructed from the combination of low-dimensional likelihoods that contain enough information about the dependence structure to produce meaningful results. A critical discussion about theory and application of CL in a variety of time series models is provided.

EO0378: **Outlier detection in high-dimensional time series**
Presenter: **Pedro Galeano**, Universidad Carlos III de Madrid, Spain
Co-authors: Daniel Pena
A procedure based on linear transformations is presented to clean a large collection of time series of outliers. For that, it is assumed that the series has been generated from a dynamic factor model and that the outliers can affect either the latent factors or the idiosyncratic noise. The effects of the outliers on the observed time series are analyzed. Particularly, we show that the outliers can be fairly well detected using linear transformations of the observed series constructed from eigenvectors of the sample autocovariance matrices. We propose a simple and fast procedure based on these linear transformations for outlier detection that it is illustrated with simulations and the analysis of a real data example.

EO0996: **Robust approaches for change point detection in panel data**
Presenter: **Alexander Duerre**, TU Dortmund, Germany
Co-authors: Roland Fried
In panel data we usually observe a large number of individuals over a time period. One often assumes stability of the model over time. There are different proposals to test this assumption. However, up to now little attention has been paid to robustness, although a small fraction of outliers can lead to a false rejection of the null-hypothesis or even mask a change point. We therefore propose alternative robust procedures for different situations with short-range serial and possibly cross-sectional dependence. Some simulations compare these methods with existing nonrobust tests.

EO1246: **On asymptotical methods for joint estimation of a common extreme value characteristic from multiple stations**
Presenter: **Paul Kinsvater**, TU Dortmund University, Germany
Co-authors: Roland Fried, Jona Lilienthal
In regional flood frequency analysis (RFA) a goal is to estimate high quantiles of an annual maximal river flow distribution F_1 by gathering information from many stations in the neighborhood with corresponding annual distributions F_j, $j = 2,\dots,d$. E.g. the popular Index Flood (IF) approach is based on an assumption termed regional homogeneity, which states that the quantile curves are related by $F_j^{-1} = s_j F_1^{-1}$ for some unknown $s_j > 0$, $j = 2,\dots,d$, usually with the additional assumption that $F_1 = F_\theta$ is known up to some finite-dimensional parameter $\theta \in \Theta$. While the considered time series can be regarded as independent in time, a challenge arises from the fact that river flows $\mathbf{X} = (X_1,\dots,X_d)'$ from the d stations are often highly dependent in space. To the best of our knowledge, none of the popular methods based on the IF-model is able to adequately take into account the spatial dependence. Our goal is to fill this gap. We are going to present asymptotic theory for classical parametric and also for semi-parametric RFA procedures that do take into account such dependences. We illustrate the methods on river flow series from the Mulde basin in Germany, where people have suffered several times from severe floods in the recent past, e.g. in summer of 2002 and 2013.

EO025 Room 301 NONPARAMETRIC ANALYSIS OF FINANCIAL MARKETS Chair: Stefan Sperlich

EO0552: **Recovering the yield curve evolution**
Presenter: **Davide La Vecchia**, University of Geneva, Switzerland
Co-authors: Bonsoo Koo, Oliver Linton
A novel nonparametric inferential procedure is developed for the yield curve dynamics. Our methodology makes use of cross sectional observations on bonds, either zero-coupon or coupon paying, across several time periods. The novelty of our approach is the combination of two different techniques: cross sectional nonparametric methods and kernel estimation for locally stationary time series. The resulting nonparametric time varying functional estimator is able to capture the yield curve shapes and dynamics commonly observed in the fixed income markets. Consistency, the rate of convergence and asymptotic normality of the proposed estimator are derived. A Monte Carlo exercise illustrates the good performance of the proposed method, under different scenarios.

EO0939: **Some classes of univariate and multivariate beta-generated distributions to model financial data**
Presenter: **Jose Maria Sarabia**, University of Cantabria, Spain
Co-authors: Faustino Prieto, Vanesa Jorda
Empirical features of many financial data series have motivated the study of flexible classes of distributions which can incorporate properties such as skewness and fat-tailedness. We propose the use of some models of univariate and multivariate beta-generated and generalized beta-generated distributions, for modelling financial data. We begin studying two classes of skew t distributions. The first family depends on two shape parameters which control the skewness and the tail weight, and the second family includes an extra parameter. We obtain analytical expressions for the cumulative distribution function, quantile function and moments, and some quantities useful in financial econometrics, including the value at risk. We provide several stochastic representations for these families in terms of usual distributions functions. Then, we propose some multivariate extensions and explore some of their properties. An empirical application with real data is provided, applying two alternative methods to estimate the parameters. On the one hand, we perform a conventional estimation of the parameters by maximum likelihood. We also explore the estimation of the parameters by modeling them as a function of other covariates, using non-parametric techniques.

EO0871: **Improved nonparametric estimation of the Sharpe-ratio**
Presenter: **Michael Scholz**, University of Graz, Austria
Co-authors: Stefan Sperlich, Jens Perch Nielsen, Enno Mammen
A two step approach is proposed to improve the estimation of the Sharpe-ratio, one of the most used measures of performance evaluation for investment strategies. We estimate first the conditional mean function of excess stock returns and then the corresponding conditional variance function in a predictive regression model focusing on nonlinear relationships between a set of covariates. The inclusion of prior knowledge in our nonlinear model shows notable improvement compared to a fully nonparametric model. Finally, a ratio of the estimates of both steps gives our new estimator of the Sharpe-ratio for which we also provide statistical properties. In an applied part, we show the efficiency of our estimator in the long-term view using annual data of the S& P500 in a period from 1872 to 2015.

EO1296: **An omnibus specification test of conditional asset pricing models**
Presenter: **Juan Manuel Rodriguez-Poo**, Universidad de Cantabria, Spain
Co-authors: Stefan Sperlich, Francisco Penaranda
A new omnibus specification test of conditional asset pricing models is presented. These models provide constraints that conditional moments of returns and pricing factors must satisfy, but most of them do not provide information on the functional form of those conditional moments. Hence, the main interest of this test is that it is not only robust to functional form misspecification of conditional moments, but it also detects any relationship between pricing errors and conditioning variables. This last issue is of crucial interest for power in testing conditional models. Special emphasis is given on practical issues like bias reduction, adaptive bandwidth choice, rather general but simple requirements on the estimates,and finite sample performance, including the resampling approximations.

EO691 Room 309 B ORDER RESTRICTED INFERENCE Chair: Ori Davidov

EO0555: **Detection of rhythmic signals in oscillatory systems using order restricted inference**
Presenter: **Miguel Fernandez**, Universidad de Valladolid, Spain
Co-authors: Yolanda Larriba, Cristina Rueda, Shyamal Peddada
The determination of rhythmic signals in oscillatory systems, such as cell cycle or circadian clock, is essential for biologists to know which genes are associated to the system. While there are several procedures available for this task in the literature, none of them is satisfactory enough. One of the reasons for this is the absence of a good definition of rhythmic signal. We propose a new definition of rhythmic signal using order restrictions and taking into account the needs of the biologists, and an algorithm based on order restricted inference and conditional tests to detect and classify the signals in different groups. We test the algorithm in simulations and with real databases from circadian clock, and compare it with the most usual methods available showing its good performance.

EO1215: **On comparing cumulative incidence functions in a competing risks model using empirical likelihood**
Presenter: **Hammou ElBarmi**, Baruch College, The City University of New York, United States
Co-authors: Vicente Nunez-Anton
An empirical likelihood approach is developed to testing the hypothesis that the cumulative incidence functions corresponding to k-competing risks are equal against the alternative that they are not equal or that they are linearly ordered. The proposed test statistics are formed by integrating a localized empirical likelihood statistic with respect to the empirical distribution. The asymptotic null distribution of these test statistics is found to have a simple distribution-free representation in terms of standard Brownian motion. The approach is extended to the case of right-censored survival data via multiple imputation. To illustrate the theoretical results, we discuss an example involving survival times of mice exposed to radiation.

EC1573: **Influence of microarray normalization strategies and rhythmicity detection algorithms to detect circadian rhythms**
Presenter: **Yolanda Larriba**, Universidty of Valladolid, Spain
Co-authors: Cristina Rueda, Miguel Fernandez, Shyamal Peddada
High-throughput microarray technologies are a widely used research tool in gene expression analysis. A large variety of preprocessing methods for raw intensity measures is available to establish gene expression values. Normalization is the key stage in preprocessing methods, since it removes systematic variations in microarray data. Then, the choice of the normalization strategy can make a substantial impact to the final results. Additionally, we have observed that the identification of rhythmic circadian genes depends not only on the normalization strategy but also on the rhythmicity detection algorithm employed. We analyze three different rhythmicity detection algorithms. On the one hand, JTK and RAIN which are widely extended among biologists. On the other hand, ORIOS, a novel statistical methodology which heavily relies on Order Restricted Inference and that we propose to detect rhythmic signal for Oscillatory Systems. Results on the determination of circadian rhythms are compared using artificial microarray data and publicly available circadian data bases.

EC1496: **A stochastic process approach to multilayer neutron detectors**
Presenter: **Vladimir Pastukhov**, Lund University, Sweden
Co-authors: Dragi Anevski
The feasibility of statistical determination of a neutron wavelength andenergy spectrum in the new generation of neutron detectors is discussed. The data from the multi-grid detector consists of counts of the number of absorbed particles, along the sequence of the detector cells.First, we consider the unimodal incident beam which is assumed to be a Poisson process. Using the Maximum Likelihood (ML) estimator we discuss its asymptotic properties. Next, we generalise this result for the case of the monotone spectrum of a multimodal Poisson beam. The last part is the ultimate challenge and is dedicated to the estimation the continuous spectrum of the incident beam, under the assumption of monotonicity.

EO191 Room 308 B RECENT ADVANCES ON ESTIMATING EQUATIONS Chair: Yves Berger

EO0627: **Smooth minimum distance estimation with generated regressors**
Presenter: **Daniel Becker**, University of Bonn, Germany
Co-authors: Valentin Patilea, Alois Kneip
In a recent study, a new class of smooth minimum distance (SmoothMD) estimators was proposed, that is suitable for models defined by conditional moment conditions. The SmoothMD estimator is extended to models with nuisance parameters or functions, and to generated regressors. The investigated framework includes nonlinear models with endogeneity, transformation models, as well as some semiparametric extensions. Particular attention is given to generated regressors. This topic has become quite important, especially due to the recent developments in the econometric analysis of treatment effects and in the identification and estimation of nonlinear models with endogenous covariates using control variables. The asymptotic behavior of the new SmoothMD estimator is studied under general conditions and new inference methods for complex models are proposed. A simulation experiment illustrates the performance of the methods for finite samples.

EO0638: **Bias-reduced double-robust estimation**
Presenter: **Stijn Vansteelandt**, Ghent University, Belgium
Co-authors: Karel Vermeulen
Over the past decade, double-robust (DR) estimators have been proposed for a variety of target parameters in causal inference and missing data models. These are asymptotically unbiased when at least one of two working models is correctly specified, regardless of which. While theoretically appealing, DR estimators have been the subject of recent debate. The reason is that model misspecification is likely to affect all working models in practice. Moreover, the performance of DR estimators can be sensitive to the choice of estimators used for fitting the working models, and can sometimes be worse than that of competing estimators that do not enjoy the double protection property. It will be shown that some DR estimators partially retain their robustness properties even under misspecification of both working models. In particular, we will propose a simple and generic estimation principle for the nuisance parameters indexing both working models, which is designed to improve the performance of the DR estimator, relative to the default use of MLEs for the nuisance parameters. The proposed approach locally minimises the squared first-order asymptotic bias of the DR estimator under misspecification of both working models and results in DR estimators with easy-to-calculate asymptotic variance. Results from simulation studies and data analyses confirm these theoretical properties and demonstrate major improvements relative to the default DR exposure effect estimators.

EO0710: **Estimating equations, empirical likelihood and complex sampling**
Presenter: **Yves Berger**, University of Southampton, United Kingdom
Most of mainstream statistic is based on independent and identically observations. However, this basic assumption does not hold with survey data, because units are often collected with complex sampling designs involving unequal probabilities and clustering. A novel empirical likelihood approach has been previously developed for complex sampling. We show how this approach can be used for conditional estimating equations. We

will also revisited some principles specific to survey sampling theory, such as as calibration and inclusion probabilities, and often thought not be part of mainstream statistic. We will show that these principles are just the consequence of the maximisation of an empirical likelihood function.

EO1013: **Panel-data modelling using structural mean models**
Presenter: **Paul Clarke**, University of Essex, United Kingdom
Co-authors: Yanchun Bao
Structural mean models (SMMs) are a class of semiparametric models originally designed for the estimation of causal effects of treatment regimes based on data from randomized clinical trials in which there was patient noncompliance. Until now, SMMs have not been used for modelling longitudinal panel data. We use SMMs to estimate the causal effect of employment status on mental health from the British Household Panel Study (BHPS). Estimation of these models using the generalized method of moments (GMM) is considered both with and without the assumption of no unobserved confounding. The substantive focus is on the effect of changes in employment status on mental health, which requires the use of extended SMMs which satisfy a second-order Markov assumption. Fully and locally efficient estimators are derived for this family of models. We also explore different strategies for specifying the potentially complex auxiliary for estimation. Our results are compared with those obtained using alternative types of panel-data model.

EO519 Room 305 A RECENT ADVANCES OF STATISTICAL METHODS IN SURVIVAL ANALYSIS AND MISSING DATA Chair: Lihong Qi

EO0699: **Cox regression analysis with missing covariates via nonparametric multiple imputation**
Presenter: **Chiu-Hsieh Hsu**, University of Arizona, United States
The situation of estimating Cox regression is considered when some covariates are subject to missing, and these exist additional information which may be predictive of the missing covariates. We propose to use two working models: one for predicting the missing covariates and the other for predicting the missing probabilities. For each missing covariate observation, these two working models are used to define a nearest neighbor imputing set. This set is then used to non-parametrically impute covariate values for the missing observation. Upon the completion of imputation, Cox regression is performed on the multiply imputed datasets to estimate the regression coefficients. In a simulation study, we compare the nonparametric multiple imputation approach with the augmented inverse probability weighted method. We show that the proposed nonparametric imputation method is robust to mis-specification of either one of the two working models and robust to mis-specification of the link function of the two working models. In contrast, the augmented inverse probability weighted method is not robust to mis-specification of the link function of the two working models and is sensitive to the selection probability. We apply the approaches to a breast cancer dataset from Surveillance, Epidemiology and End Results Program.

EO0790: **Re-use of nested case-control studies for the analysis of semi-competing risks data**
Presenter: **Sebastien Haneuse**, Harvard TH Chan School of Public Health, United States
Co-authors: Ina Jazic
In the semi-competing risks setting, individuals may experience some non-terminal event (e.g. disease recurrence) or a terminal event (e.g. death) or, possibly both. In instances where one or more exposures/covariates are difficult to obtain, researchers have several options at their disposal with which to sub-sample patients on whom to collect complete data. One option is the nested case-control (NCC) design, which is typically implemented by performing risk set sampling on the basis of a single outcome of interest, possibly matching on certain factors, and estimation/inference is performed for a Cox model. We propose a novel statistical framework for estimation/inference with respect to an illness-death model for semi-competing risks data by re-using data from a NCC design. In particular, we consider situations where the initial NCC study used either the non-terminal or the terminal event as the primary outcome of interest, as well as situations where one can or cannot supplement the NCC study with additional cases of the secondary outcome from the remainder of the cohort. We propose both maximum likelihood and maximum weighted partial likelihood approaches and investigate their operating characteristics as well as a series of design considerations via a simulation study. Finally, the framework is illustrated with data from an ongoing study investigating the role of circulating androgens on the risk of breast cancer risk and mortality among women.

EO1173: **MICE: Guidelines for imputing missing covariates in accelerated failure time models**
Presenter: **Lihong Qi**, University of California Davis, United States
Co-authors: Ying-Fang Wang, Rongqi Chen , Juned Siddique, John Robbins, Yulei He
Missing covariates often occur in biomedical studies with survival (or time-to-event) outcomes. Multiple imputation via chained equations (MICE) is a semi-parametric and flexible strategy that imputes multivariate data by a series of conditional models, one for each incomplete variable. MICE is a popular imputation strategy and widely implemented in statistical software. Nevertheless, the theoretical properties of MICE are difficult to establish. When applying MICE, practitioners tend to specify the conditional models in a simple fashion largely dictated by the software, which might lead to suboptimal results. Motivated by a study of time to hip fractures in the Women's Health Initiative Observational Study using accelerated failure time models, we propose and experiment with some rationales leading to appropriate MICE specifications. This strategy starts with imposing a joint model for the variables involved. We first derive the conditional distribution of each variable under the joint model, then approximate these conditional distributions to the extent which can be characterized by commonly-used regression models. Our simulation results show that some simple MICE specifications can produce suboptimal results, yet those based on the proposed strategy appear to perform well. Hence, we warn against a mechanical use of MICE and suggest a careful modeling of the conditional distributions of variables to ensure its good performance.

EO0995: **Modeling strategies to enrich a multiplexed, preemptive genomic testing program using electronic health records data**
Presenter: **Jonathan Schildcrout**, Vanderbilt University, United States
The PREDICT program (Pharmacogenomic Resource for Enhanced Decisions In Care and Treatment) is a clinical quality improvement initiative at Vanderbilt University Medical Center to prospectively identify patients for genotyping based on the likelihood of receiving medications with pharmacogenetic effects at a future time. The goal is to preemptively collect and store genetic data within patients electronic health records so that genetic information can, when necessary, be used to guide medication prescriptions. Because medications and medication related adverse events are costly, a crucial feature of an effective multiplexed preemptive testing program is the efficient identification of who will be prescribed medications with pharmacogenetic effects. We will describe and then compare several time-to-event modeling strategies for identifying such patients for pre-emptive genotyping. Once the models have been developed, weighted risk scores will be derived based on likelihood, severity and costs of each medication related adverse event. We will compare standard Cox regression methods and machine learning approaches to evaluate strategies that minimize risk under several reasonable loss functions.

EO023 Room 203 FOUNDATIONS FOR DEPTH METHODS IN MULTIVARIATE AND FUNCTIONAL DATA SETTINGS Chair: Robert Serfling

EO0726: **RKHS tools for functional depth**
Presenter: **Antonio Cuevas**, Autonomous University of Madrid, Spain
Several interesting notions of statistical depth have been so far proposed in the multivariate (finite-dimensional) case. The study of the functional (infinite-dimensional) counterparts of such depth concepts is a natural task. However, such task is hampered by some difficulties, associated with the passage to infinite dimension. Thus, unlike the multivariate situations, in the functional case there is no obvious general way to define density functions. Also, the covariance operators are typically non-invertible; in consequence, there is no immediate functional counterpart for the classical Mahalanobis distance (a relevant notion of multivariate depth). The aim is to show how the use of the Reproducing Kernel Hilbert Spaces (RKHS) associated with the underlying process, allows us to partially overcome these difficulties. On the one hand, it is known that the Radon-Nikodym derivatives of mutually absolutely continuous homoscedastic Gaussian processes can be expressed in RKHS terms. These derivatives (or suitable estimators of them) can be used as surrogates of the multivariate density functions. On the other hand, the RKHS setup suggests a suitable functional version of the Mahalanobis depth.

EO0692: **Aspect-based depth statistics for functional data**
Presenter: **Karl Mosler**, Universitaet zu Koeln, Germany
A depth statistic measures how central an observation is located in other data or, more general, in a given probability distribution. Various depth notions have been proposed for multivariate data and successfully applied to problems in data analysis and non-parametric statistics. Recently, depth statistics have been extended to the analysis of functional data. If the data are in R^d, a depth statistic is usually assumed to be affine-invariant (or invariant to translation and scale), monotonously decreasing from a point of maximum depth, vanishing at infinity, upper semicontinuous, and quasi-concave. While these postulates are well established for d-variate data, it is not obvious how to extend them to data in more general linear spaces, viz. spaces of functions. It is argued that the statistical analysis of functional data is led by certain aspects in which the analyst is interested, and discusses how these aspects enter the notion in a formalized way. A general aspect-based approach is introduced for the definition of depths for functional data and investigates sets of postulates for them. The aspects correspond to linear functionals on the function spaces. Two principal ways are considered to aggregate centrality information over the time axis: infimum depth and mean depth, where the latter includes a weight term. Known functional depths appear as special cases.

EO1191: **On the properties of statistical functional depth**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: Heather Battey
A formal definition of depth for functional data is provided, which is justified on the basis of six properties. This provides not only a sophisticated extension of those defining the multivariate depth, but also implicitly address several common or inherent difficulties associated with functional data. Particularly, we focus on the difference between the multivariate and functional notions when handling a property that regards the depth value at the center of symmetry of symmetric distributions. Furthermore, amongst the six depth defining properties is one that tackles the delicate challenge of inherent partial observability of functional data, providing a minimal guarantee on the performance of the empirical depth beyond the idealised and practically infeasible case of full observability. Robustness to the presence of outliers is often cited as one of the defining features of empirical depth and our definition of functional depth automatically yields a qualitatively robust estimator of the population depth. A further challenge, automatically addressed (if present) by our definition, pertains to functional data exhibiting little variability over a subset of the domain and significantly overlapping one another on this set. Intuitively, functional observations over such a domain ought to play a reduced role in the assignment of depth, especially in light of the partial observability and the convention to pre-process the partial observations.

EO0668: **Perspectives on multivariate depth and quantile functions**
Presenter: **Robert Serfling**, University of Texas at Dallas, United States
Various properties are desirable for multivariate depth functions, such as invariance under selected transformations, robustness, and computational ease, among others. These play a central role in choosing a particular depth function. Since such criteria trade off against each other, one must prioritize, based on the application context and the statistical inference goals. Similarly, there are desirable equivariance and other properties for the quantile and rank functions that are closely related to a given depth function. However, in the case of rank functions, for example, the relevant priorities arise from quite different conceptual orientations and may differ from those for the depth functions, even though the corresponding depth and rank functions are equivalent in some sense. It is important, therefore, to delineate precisely the correspondence between multivariate depth and rank functions and on this basis to revisit and either reaffirm or update standing priorities on desired criteria for such functions. One seeks both technical and conceptual consistency between, say, the priorities for depth function properties and those for rank function properties. Some relevant foundation and perspective on these and related issues will be provided.

EO099 Room Board meeting room II BAYESIAN METHODS IN BIOSTATISTICS Chair: Yung-Seop Lee

EO0807: **A Bayesian high-dimensional couple-based latent risk model with an application to infertility**
Presenter: **Beomseuk Hwang**, Chung-Ang University, Korea, South
Co-authors: Zhen Chen
Motivated by the Longitudinal Investigation of Fertility and the Environment (LIFE) Study that investigated the association between exposure to a large number of environmental pollutants and human reproductive outcomes, we propose a joint latent risk class modeling framework with an interaction between female and male partners of a couple. This formulation introduces a dependence structure between the chemical patterns within a couple and between the chemical patterns and the risk of infertility. The specification of an interaction enables the interplay between the female and male's chemical patterns on the risk of infertility in a parsimonious way. We took a Bayesian perspective to inference and used Markov chain Monte Carlo algorithms to obtain posterior estimates of model parameters. We conducted simulations to examine the performance of the estimation approach. Using the LIFE Study dataset, we found that in addition to the effect of PCB exposures on females, the male partners' PCB exposures play an important role in determining risk of infertility. Further, this risk is subadditive in the sense that high chemical exposure is only required on one partner of the couple to determine the highest risk.

EO0903: **Nonparametric Bayesian multivariate meta-regression: An application in the temperature-mortality association study**
Presenter: **Gyuseok Sim**, Korea Advanced Institute of Science and Technology, Korea, South
Co-authors: Yeonseung Chung, Ho Kim, Antonella Zanobetti, Joel Schwartz
In biomedical research, meta-analysis has been a popular and useful tool combining evidence from multiple studies to investigate an exposure-response association. Because of the hierarchical nature in meta-analysis, a two-stage analytical approach has been used in many studies for the computational convenience. In the first stage, study-specific exposure-response relationships are estimated as multi-parameters using individual data. These estimates are combined through meta-analysis in the second stage incorporating study-level predictors (called meta-predictors). The second stage is often called meta-regression. The currently used multivariate meta-regression assumes linearity in meta-predictors, residual normality and heteroscedasticity. However, such assumption is limited to integrate the study-specific estimates over a broad range of studies where

sub-groups may exist and the effects of meta-predictors are not linear. We propose a more flexible and generalized two-stage multivariate meta-regression by replacing the second stage by a nonparametric Bayesian multivariate nonlinear regression based on the Dirichlet process mixture. The proposed method was evaluated through a simulation study and applied to the data from 135 US cities to study the temperature-mortality association.

EO1323: **Classification of methylation status using Bayes classifier**
Presenter: **Taesung Park**, Seoul National University, Korea, South
Co-authors: Iksoo Huh, Soojin Yi
In recent methylation analysis, the whole genome sequencing of bisulfite converted DNA method provides comprehensive information of genome-wide methylation patterns. One of challenging application of these whole genome methylation maps is to classify the methylation status of each position. A most commonly used method for this purpose is a simple binomial test. Although it is intuitive and straightforward, it lacks power when sequence coverage and genome wide methylation level are low. We recently proposed a new method based on Bayes classifier to infer DNA methylation status while considering the neighborhood DNA methylation levels of a specific site. We propose a more efficient way of using the neighborhood DNA methylation levels of a specific site which results in increased power of classifying the methylation status of each position. Our new method should be widely applicable to the analyses of methylC-seq data from diverse sparsely methylated genomes.

EO1341: **Bayesian modeling of medical imaging in tumor delineation**
Presenter: **Nitai Mukhopadhyay**, Virginia Commonwealth University, United States
Medical images from dependent and independent sources are used in tumor delineation for cancer treatment. However, the analysis of these images is rather difficult, due to their massive dimensions. Therefore performing any statistical analysis is possible only after reducing the image data to a small dimensional derivative of the images. A population value decomposition (PVD) was previously proposed, which can effectively reduce the size of a population of two-dimensional images. PVD reduces the image matrix to population specific matrix and much smaller dimensional subject specific matrix. Population-specific matrices do not change across different images, and subject-specific matrix is unique to each individual image. We use PVD to reduce the dimension of two-dimensional images and make a model for the low-dimensional subject-specific matrices. Bayesian modeling of the lower dimensional matrices provides a way to denoise the images and make inference based on the actual image data and provide model based predictions.

EO481 Room 306 A STATISTICS FOR HIGH-DIMENSIONAL DATA WITH ENVIRONMENTAL APPLICATIONS Chair: Emre Barut

EO0889: **Computational challenges with big environmental data**
Presenter: **Marc Genton**, KAUST, Saudi Arabia
Two types of computational challenges arising from big environmental data are discussed. The first type occurs with multivariate or spatial extremes. Indeed, inference for max-stable processes observed at a large collection of locations 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 explore the limits of modern state-of-the-art computational facilities to perform full likelihood inference and to efficiently evaluate high-order composite likelihoods. With extensive simulations, we assess the loss of information of composite likelihood estimators with respect to a full likelihood approach for some widely-used multivariate or spatial extreme models. The second type of challenges occurs with the emulation of climate model outputs. We consider fitting a statistical model to 1 billion global 3D spatio-temporal temperature data using a distributed computing approach. The statistical model exploits the gridded geometry of the data and parallelization across processors. It is therefore computationally convenient and allows to fit a non-trivial model to a data set with a covariance matrix comprising of 10^{18} entries. We provide 3D visualization of the results.

EO0958: **Fusing multiple existing space-time land cover products**
Presenter: **Amanda Hering**, Baylor University, United States
Land cover (LC) is a critical variable driving many environmental processes, so its assessment, monitoring, and characterization are essential. However, existing LC products, derived primarily from satellite spectral imagery, each have different temporal and spatial resolutions and different LC classes. Most effort is focused on either fusing a pair of LC products over a small space-time region or on interpolating missing values in an individual LC product. We review the complexities of LC identification and propose a method for fusing multiple existing LC products to produce a single LC record for a large spatial-temporal grid. We first reconcile the LC classes of different LC products and then present a probabilistic weighted nearest neighbor estimator of LC class. This estimator depends on three unknown parameters that are estimated using numerical optimization to maximize an agreement criterion that we define. We illustrate the method using six LC products over the Rocky Mountains and show the improvement gained by supplying the optimization with data-driven information describing the spatial-temporal behavior of each LC class. Given the massive size of the LC products, we show how the optimal parameters for a given year are often optimal for other years, leading to shorter computing times.

EO0890: **Total variation depth for functional data: Properties and applications**
Presenter: **Ying Sun**, KAUST, Saudi Arabia
There has been extensive work on data depth-based methods for robust multivariate data analysis. Recent developments have moved to infinite-dimensional objects such as functional data. A new notion of depth for functional data, the total variation depth, is introduced. The proposed notion is well suited for shape outlier detection due to the fact that it considers the total variation and takes into account the necessary correlations in functional data. Effective outlier detection rules along with visualization tools are also developed, and the outlier detection performance is examined through simulation studies. The numerical results show that it outperforms many other notions under different types of outlier models. Finally, we illustrate our method using real data examples for detecting outliers in sample curves and images with environmental applications.

EO1311: **A spatio-temporal framework for modeling active layer thickness**
Presenter: **Tatiyana Apanasovich**, George Washington University, United States
The Arctic is experiencing an unprecedented rate of environmental and climate change. The active layer is sensitive to both climatic and environmental changes, and plays an important role in the functioning, planning, and economic activities of Arctic human and natural ecosystems. A methodology is developed for modeling and estimating spatial-temporal variations in active layer thickness using data from several sites of the Circumpolar Active Layer Monitoring network, and demonstrates its use in spatial-temporal interpolation. The simplest models stochastic component exhibits no spatial or spatio-temporal dependency and is referred to as the nave model, against which we evaluate the performance of the other models, which assume that the stochastic component exhibits either spatial or spatio-temporal dependency. The methods used to fit the models are then discussed, along with point forecasting. We compare the predicted fit of the various models at key study sites located in the North Slope of Alaska and demonstrate the advantages of space-time models through a series of error statistics. We find the difference in performance between the spatio-temporal and remaining models is significant for all three error statistics. The best stochastic spatio-temporal model increases predictive accuracy, compared to the naive model, of 33.3, 36.2 and 32.5% on average across the three error metrics at the key sites for a one-year hold out period.

EO477 Room 213 ABOUT 40 YEARS OF GRAPHICAL MARKOV MODELS: SOME PERSONAL PERSPECTIVES Chair: Nanny Wermuth

EO1144: **From Markov properties to causal effects: From mathematics to epidemiology**
Presenter: **Morten Frydenberg**, Aarhus University, Denmark

For many years graphs, and graphical models, were a topic for only mathematicians, statisticians and people within computer science, but during the last decade they have been adopted by epidemiologist in attempt to understand and formalize some of the central concepts in epidemiology like cause-effect, confounding, selection bias an mediation. Many researchers with no insight in nmathematics or statistics, use DAGs (Directed Acyclic Graphs) to decided how to analyse data. They do this with ease and without much concern about dynamics or complexity of the process they are studying, e.g. summarizing smoking history and habits into a simple vertex (knot) in a the DAG. A personal view on the this development will be given: Can DAGs capture the complexity of observational studies? Will epidemiology benefit by the routine use of causal DAGs?

EO1035: **Existence of the maximum likelihood estimate in discrete graphical models**
Presenter: **Helene Massam**, York University, Canada

Discrete graphical models form an important subclass of the class of discrete hierarchical loglinear models. Interest for the maximum likelihood estimate (henceforth abbreviated mle) of the parameter in contingency tables with multinomial or Poisson sampling started in the 1950 and 1960s. The first necessary and sufficient condition for the existence of the mle was given in the 70's. A hierarchical loglinear model was defined by the domain of its means which is a convex polyhedron called the marginal polytope. Those results can be interpreted in terms of the marginal polytope and are equivalent to saying that the mle exists if, and only if, the data vector belongs to the interior of the polytope. We will look at recent results and methodology developed to identify the smallest face of the polytope containing the data when the mle does not exist.

EO1154: **Conditional independence in multivariate analysis and beyond**
Presenter: **Frantisek Matus**, Institute of Information Theory and Automation, Czech Republic

The goal is to review patterns of conditional independence constraints that underpine multivariate statistical models. In the last forty years, various patterns have been applied when building e.g. graphical models and their extensions. They provide natural decompositions, make it possible to do statistical inference effectively, and sometimes they open the only way how to do it at all. Patterns of the constraints have been in use also in the network coding when studied via Shannon theory. They express also natural requirements in protocols of the information-theoretical cryptography. Philosophers study the patterns in connection to causality. The last but not least, patterns serve as appropriate tools in pure mathematics, in particular in algebraic structures, finite geometries, matroids and elsewhere. A comparison and synthesis of the above approaches might be of interest.

Friday 09.12.2016 14:35 - 16:15 Parallel Session E – CFE-CMStatistics

CI685 Room Graduation hall EFFICIENCY RESULTS IN HIGH DIMENSION Chair: Marc Hallin

CI0162: **Efficiency in the high-dimensional one-sample location problem**
Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium
Co-authors: Thomas Verdebout
The one-sample location testing problem is considered in a high-dimensional setup where the dimension p goes to infinity with the sample size n. We consider two types of problems, according to whether the direction of the possible shift is specified or not. Under a sphericity assumption, we investigate the null and non-null asymptotic distributions of two natural tests, namely the (spherical version of) the Hotelling test and the spatial sign test. Our results show in particular that these two tests share the same local asymptotic powers under power-exponential distributions, but not under t distributions. Also, the local alternatives that can be detected by both tests may depend on the underlying distribution, which is highly non-standard. We further partly read our results in terms of contiguity and local asymptotic normality. Throughout, we conduct Monte Carlo experiments to illustrate the finite-sample relevance of the results.

CI0990: **Local asymptotic normality of the spectrum of high-dimensional spiked F-ratios**
Presenter: **Alexei Onatski**, University of Cambridge, United Kingdom
Two types of spiked multivariate F distributions are considered: a scaled distribution with the scale matrix equal to a rank-k perturbation of the identity, and a distribution with trivial scale, but rank-k non-centrality. The eigenvalues of the rank-r matrix (spikes) parameterize the joint distribution of the eigenvalues of the corresponding F-matrix. We show that, for the spikes located above a phase transition threshold, the asymptotic behavior of the log ratio of the joint density of the eigenvalues of the F matrix to their joint density under a local deviation from these values depends only on the k of the largest eigenvalues $\lambda_1, ..., \lambda_k$. Furthermore, we show that $\lambda_1, ..., \lambda_k$ are asymptotically jointly normal, and the statistical experiment of observing all the eigenvalues of the F-matrix converges in the Le Cam sense to a Gaussian shift experiment that depends on the asymptotic means and variances of $\lambda_1, ..., \lambda_k$. In particular, the best statistical inference about sufficiently large spikes in the local asymptotic regime is based on the k of the largest eigenvalues only.

CI1306: **On high-dimensional robust regression and the bootstrap**
Presenter: **Noureddine El Karoui**, UC Berkeley, United States
Recent work will be discussed on high-dimensional robust regression and the bootstrap. Very interestingly, ideas connected to the analysis of robust regression estimators in high-dimension gives insight into the performance of the bootstrap. A number of surprising results will be discussed, including the fact that two equally intuitive (in low-dimension) bootstraps perform very differently in high-dimension: one leads to extremely conservative confidence intervals, the other to anti-conservative confidence intervals. Furthermore, it can be shown that maximum likelihood methods for high-dimensional regression lead to inefficient estimators. More problems with the bootstrap in moderate to high-dimension will be discussed. Generically, it seems that the bootstrap does not give statistically valid inferential statements in this context, even when it does in its low-dimensional counterparts.

CO485 Room 107 NOWCASTING AND FORECASTING MACROECONOMIC TRENDS I Chair: Gian Luigi Mazzi

CO0217: **Nowcasting real economic activity in the euro area: Assessing the impact of qualitative surveys**
Presenter: **Raisa Basselier**, National Bank of Belgium, Belgium
Co-authors: David de Antonio Liedo, Geert Langenus
The contribution of survey data is analyzed, in particular various sentiment indicators, to nowcasts of quarterly euro area GDP. We transform the real-time dataflow into an interpretable flow of news, taking into account only the first data vintage based upon the original press releases. The news is defined in our particular example as the difference between the released values and the prediction of a mixed-frequency dynamic factor model. More concretely, we specify a number of factors that is large enough to synthesize all relevant information from the data. Our purpose is twofold. First, we aim to quantify the specific value added for nowcasting of a set of heterogeneous data releases including not only sentiment indicators constructed by Eurostat, Markit, the National Bank of Belgium, IFO, ZEW, GfK or Sentix, but also hard data regarding industrial production or retail sales in the euro area and in some of the largest euro area countries. Second, we provide a ranking of these indicators, taking into account their predictive character for GDP, as well as their timeliness. We do this on the basis of the Kalman filter gains. In general, hard data contribute less to the nowcasts: their relatively late availability implies that they can to a large extent be anticipated by nowcasting models and, hence, their news component is smaller. The prevalence of survey data remains also under the counterfactual that hard data is released timely without any delay.

CO0446: **A dynamic factor model for nowcasting Canadian GDP growth**
Presenter: **Tony Chernis**, Bank of Canada, Canada
Co-authors: Rodrigo Sekkel
A Dynamic Factor Model (DFM) is estimated for nowcasting Canadian Gross Domestic Product. The model is estimated with a mix of soft and hard indicators, and it features a high share of international data. The model is then used to generate nowcasts, predictions of the recent past and current state of the economy. In a pseudo real-time setting, we show that the DFM outperforms univariate benchmarks, as well as other commonly-used nowcasting models, such as MIDAS and bridge regressions.

CO0907: **Going global: The role of international data in nowcasting German GDP**
Presenter: **Philipp Hauber**, Kiel Institute for the World Economy, Germany
Co-authors: Christian Schumacher
Factor models can summarize the co-movements of a large number of variables and have proven useful in nowcasting and short-term forecasting of GDP growth because they can tackle mixed-frequencies or missing observations at the current edge. The main aim is to assess the importance of international variables for nowcasting national developments, an issue, which - curiously - has received relatively little attention in the academic literature so far. Given the large number of variables at both the national and international level, the question arises whether all this information is useful for nowcasting or not. As such, we also contribute to the continuing debate on variable selection and the optimal size of factor models for forecasting. Rather than choosing variables ad-hoc, we employ sparse priors on the factor model's loadings for Bayesian estimation. Sparse priors can help to identify those business cycle indicators that essentially determine the factors, whereas irrelevant variables are sorted out. In an empirical exercise, we start nowcasting using a baseline factor model estimated on German GDP and a large number of monthly indicators for the German economy only. Then, we augment the national data by various indicators for the remaining G7 countries and compare the nowcast accuracy. Different model specifications are compared to identify the ones most suitable to extract useful predictive information for nowcasting in a large-data environment.

CO0845: **On the economic effect of Western sanctions on Russia**
Presenter: **Anna Pestova**, NRU HSE and CMASF, Russia
Co-authors: Mikhail Mamonov
An empirical test of Bayesian VAR model is developed and provided for the Russian economy. On the basis of Bayesian VAR forecasting model we estimate the economic effect of Western sanctions on the Russian economy controlling for the fall in oil price that occurred simultaneously. To disentangle the effects of different external shocks, we perform two counterfactual experiments with BVAR model for Russia: we make out-of-sample conditional projections of the key Russian macroeconomic variables conditioning on the two sets of realized external conditions in 2014-2015: first, only on the oil price and second, on the oil price and the Russian corporate external debt (as a proxy for the short-term effect of economic sanctions). We found that the fall in oil prices alone cannot explain the depth of the current crisis in the Russian economy, however if we add the actual path of corporate external debt deleveraging, we get more accurate results. We treat the difference between these two projections, which is significant for the most of variables, as the effect of EU and US sanctions against Russia.

CO333 Room 003 ADVANCES IN FINANCIAL FORECASTING Chair: Ekaterini Panopoulou

CO0249: **Bayesian inference and prediction for high-frequency data using particle filtering**
Presenter: **Andrea Cremaschi**, University of Kent, United Kingdom
Co-authors: Jim Griffin
Financial prices are usually modelled as continuous, often involving geometric Brownian motion with drift, leverage and possibly jump components. An alternative modelling approach allows financial observations to take discrete values when they are interpreted as integer multiples of a fixed quantity, the ticksize, the monetary value associated with a single change in the asset evolution. These samples are usually collected at very high frequency, exhibiting diverse trading operations in seconds. In this context, the observables are modelled via the Skellam process, defined as the difference between two independent Poisson processes. The intensities of the Poisson processes are modelled as functions of a stochastic volatility process, which is in turn described by a discretised Ornstein-Uhlenbeck AR(1) process. The choice of a discrete-time/discrete-space model is able to provide good fitting and out-of-sample prediction results, the latter being computed using Particle Filtering methods.

CO0321: **Dynamic quantile spacing models for forecasting asset return distributions**
Presenter: **Evangelia Mitrodima**, LSE, United Kingdom
Co-authors: Jim Griffin, Jaideep Oberoi
An approximation likelihood function is proposed for a set of quantiles for Bayesian methodology, which addresses estimation issues that arise in dynamic quantile regression. Our approximation is based on an idea from survival analysis, where the hazard rate functions represent the local scale parameters since they are the scale of the distribution in different intervals. More specifically, we jointly model the distance between quantiles via an observation-driven time series model to approximate the density of asset returns. The underlying idea of the model is that the parameters are updated by using the score function of the likelihood and so the parameters are moving in the direction of the maximum likelihood estimator (MLE) at each time point. Thus, the shape of the conditional density of the observation is associated with the dynamics of the parameters. We consider a zero median distribution function, where the distance from the median becomes the sum of autoregressive processes and generates much more persistence as we move further out in the tails. In addition, the proposed model assumes that the differences of the quantiles evolve linearly over time and therefore we avoid the crossing problem. In our empirical exercise, we find that the model fits the data well, offers robust results and acceptable forecasts for a sample of stock and index returns.

CO0351: **Evaluating the performance of risk models: A quantile score approach**
Presenter: **Christos Argyropoulos**, University of Kent, United Kingdom
Co-authors: Ekaterini Panopoulou
The evaluation of a risk model may be futile given that there is no observable alternative to verify it. Imposing specific assumptions in order to circumvent this problem reduces the power of the tests and therefore their reliability. In addition, for the cases where all models are rejected or not rejected, there is no information about which model would provide the best performing forecast. Therefore, risk model evaluation should at least be accompanied by a measurement of the performance of forecasts. The underlying idea states that even if the model is not accurate the implicit structure may approximate a part of the real, but unknown, dynamics and thus add value to the forecast. Amongst simple scoring functions, the literature proposes the usage of likelihood based scoring rules in order to evaluate risk forecasts. We argue that this is not optimal since we have to consider not just the likelihood of a density but also the tail probability. In order to do so, we focus on quantile scoring rules which target the tail of the density forecast. The proposed method evaluates the null hypothesis of equal performing models by comparing the average scores via a simple Diebold-Mariano test. Initial results for six common risk models suggest performance inequality. The use of the "least bad" model can result in significant gains compared to the alternative ones. Finally, the proposed method can serve as a performance index for the Expected Shortfall risk measure.

CO0332: **The combined effect of aggregation and the log transformation on forecasting**
Presenter: **Theologos Pantelidis**, University of Kent, United Kingdom
Co-authors: Myrsini Koutsigka
Many economic variables are often used in logarithms for forecasting and estimation analysis, as this transformation is considered to create a more homogeneous variance. The conditions under which the log transformation can help the researcher to produce more accurate forecasts are examined in previous studies in the literature. On the other hand, forecasting macroeconomic variables across a (large) number of countries/groups is also a difficult but standard task for economic analysts. A lot of work has been done about aggregated data and the way we can obtain optimal forecasts for the aggregate. However, the majority of those studies refer to linear transformations, and given that the logarithmic transformation is a nonlinear one, the effect of using logs in forecasting aggregated variables is not clear and needs further investigation. The aim is to investigate the combined effect of aggregation and the log transformation on forecasting. The study initially describes the alternatives approaches that can be followed to obtain forecasts for the variable of interest when this variable is an aggregated process. Each approach generates a different predictor. Then, we investigate the relative forecasting accuracy of each predictor by means of Monte Carlo simulations. Finally, a variety of empirical applications are carried out using aggregated economic variables.

CO447 Room 111 INFLATION EXPECTATIONS IN LOW INFLATION ENVIRONMENT Chair: Maritta Paloviita

CO0268: **Monetary shocks, central bank projections and policy and macro signals**
Presenter: **Paul Hubert**, OFCE - Sciences Po, France
The dependence of the effects of monetary shocks on the macroeconomic information disclosed by central banks is addressed. The aim is to investigate the individual and interacted effects of monetary shocks and shocks to the Bank of England's inflation and output projections on the term structure of UK private inflation expectations, to document private agents interpretation of central bank signals about policy and macroeconomic developments. First, inflation expectations respond negatively to monetary shocks, consistent with the usual transmission mechanism, and positively to projection shocks, suggesting evidence of macro signals. Second, the effects of monetary shocks are amplified by positive inflation projection shocks.

CO0323: **The long-term distribution of expected inflation in the euro area: What has changed since the great recession**
Presenter: **Geoff Kenny**, European Central Bank, Germany
Co-authors: Jonas Dovern
The distribution of long-term inflation expectations in the euro area is analyzed by using individual density forecasts from the ECB Survey of Professional Forecasters. We exploit the panel dimension in this dataset to examine whether this distribution became less stable following the Great Recession, subsequent sovereign debt crisis and period when the lower bound on nominal interest rates became binding. Our results suggest that the distribution did change along several dimensions. In particular, we document a small downward shift in mean long-run expectations toward the end of our sample although they remain aligned with the ECB definition of price stability. More notably, however, we identify a trend toward a more uncertain and negatively skewed distribution with higher tail risk. Another main finding is that key features of the distribution are influenced by macroeconomic news, including the ex post historical track record of the central bank. Overall, our evidence provides strong grounds for recent monetary policy measures aimed at bringing inflation back to levels consistent with the ECBs price stability objective.

CO0409: **ECB inflation projections and private sector forecasts**
Presenter: **Tomasz Lyziak**, National Bank of Poland, Poland
Co-authors: Maritta Paloviita
The aim is to analyse the impact of ECB inflation projections on inflation forecasts of professional economists in the euro area. We start with analysing similarity of SPF inflation forecasts and the ECB inflation projections. Then we consider potential reasons why the private sector could find it useful to rely on macroeconomic projections released by the ECB. Firstly, we examine information content of the ECB inflation projections. We follow previous work and compare relative accuracy of ECB inflation projections and SPF inflation forecasts. Secondly, using a recent test, we study whether the ECB projections contain additional information relevant for professional forecasters forecasting. Finally, we analyse to what extent inflation projections signal future monetary policy decisions. In the next step of our analysis we describe the formation of inflation expectations in the euro area, paying special attention to the role of ECB inflation projections. We consider the fact that professional forecasters, while forming expectations, can combine their own assessment, based on current inflation and forward-looking considerations, with information provided by monetary authorities, including in particular the inflation target and inflation projections. In addition to the whole EMU period, model comparisons are performed separately for the pre-crisis and post-crisis period as well as in the rolling regression manner.

CO0478: **Financial literacy and inflation expectations**
Presenter: **Fabio Rumler**, Austrian Central Bank, Austria
Co-authors: Maria Teresa Valderrama
The empirical literature on inflation expectations emphasizes financial literacy as one of the key determinants of consumers' inflation expectations. Using micro data from a survey conducted among 2,000 Austrian households, we construct a summary indicator of financial literacy from 10 questions on the economic and financial knowledge of households. We find that this indicator significantly affects both, the level and the uncertainty of inflation expectations: Households with relatively higher levels of financial/economic literacy tend to have lower and thus more realistic short-term and long-term inflation expectations. Also ex-post, we find that deviations of individual inflation expectations from the actual outturn are lower for consumers with higher levels of financial literacy. Concerning the effect on the uncertainty of inflation expectations, our results suggest that, interestingly, people with higher levels of financial literacy are less certain about their inflation expectations than people with less financial literacy. This suggests that a better knowledge of economic and financial matters obviously makes people more (not less) careful in their assessment of future macroeconomic developments.

CO545 Room Board meeting room I SPARSE BAYESIAN FACTOR ANALYSIS Chair: Sylvia Fruehwirth-Schnatter

CO0317: **Sparse Bayesian time-varying covariance estimation in many dimensions**
Presenter: **Gregor Kastner**, WU Vienna University of Economics and Business, Austria
Dynamic covariance estimation for multivariate time series suffers from the curse of dimensionality; as a consequence, parameter parsimony plays an important role in reliable statistical inference. We address this issue by modeling the underlying dynamics of a time series vector through a lower dimensional collection of latent factors that allow for time-varying stochastic volatilities. Furthermore, we apply a Normal-Gamma prior to the elements of the factor loadings matrix. This hierarchical shrinkage prior is a generalization of the Bayesian lasso and effectively pulls the factor loadings of unimportant factors towards zero, thereby increasing sparsity even more. To guarantee efficiency of the estimation procedure, we employ a fully Bayesian yet computationally feasible approach to obtain draws from the high-dimensional posterior and predictive distributions via Markov chain Monte Carlo (MCMC) samplers. We utilize several variants of an ancillarity-sufficiency interweaving strategy (ASIS) to boost efficiency when sampling the factor loadings as well as the parameters driving the time-varying volatilities. The effectiveness of the approach is demonstrated through extensive simulation studies. Furthermore, the model is applied to a 300-dimensional vector of stock returns to evaluate predictive performance for financial data. Additionally to being a stand-alone tool, the algorithm is designed to act as a "plug and play" extension for other MCMC samplers.

CO1685: **Fast Bayesian factor analysis via automatic rotations to sparsity**
Presenter: **Edward George**, University of Pennsylvania, United States
Co-authors: Veronika Rockova
Rotational transformations have traditionally played a key role in enhancing the interpretability of factor analysis via post hoc modifications of the factor model orientation. Regularization methods also serve to achieve this goal by prioritizing sparse loading matrices. We cross-fertilize these two paradigms within a unifying Bayesian framework. Our approach deploys intermediate factor rotations throughout the learning process, greatly enhancing the effectiveness of sparsity inducing priors. These automatic rotations to sparsity are embedded within a PXL-EM algorithm, a Bayesian variant of parameter-expanded EM for posterior mode detection. By iterating between soft-thresholding of small factor loadings and transformations of the factor basis, we obtain (a) dramatic accelerations, (b) robustness against poor initializations and (c) better oriented sparse solutions. For accurate recovery of factor loadings, we deploy a two-component refinement of the Laplace prior, the spike-and-slab LASSO prior. This prior is coupled with the Indian Buffet Process (IBP) prior to avoid the pre-specification of the factor cardinality. The ambient dimensionality

is learned from the posterior, which is shown to reward sparse matrices. Our deployment of PXL-EM performs a dynamic posterior exploration, outputting a solution path indexed by a sequence of spike-and-slab priors.

CO1727: **Recent advances in sparse Bayesian factor analysis**
Presenter: **Sylvia Fruehwirth-Schnatter**, WU Vienna University of Economics and Business, Austria
Factor analysis is a popular method to obtain a sparse representation of the covariance matrix of multivariate observations. We review some recent research in the area of sparse Bayesian factor analysis that tries to achieve additional sparsity in a factor model through the use of point mass mixture priors. Identifiability issues that arise from introducing zeros in the factor loading matrix are discussed in detail. Common prior choices in Bayesian factor analysis are reviewed and MCMC estimation is briefly outlined. Applications to a small-sized psychological data set as well as a financial applications to exchange rate data and stock returns serve as an illustration.

CO1747: **Unique representations of sparse factor models**
Presenter: **Sylvia Kaufmann**, Study Center Gerzensee, Switzerland
Co-authors: Markus Pape
In factor analysis, not all latent factors necessarily need to be linked to all observed data series. Hence, the loadings matrix may contain many zeros. In confirmatory factor analysis (CFA), the pattern of zero and nonzero entries is assumed to be known, and the free parameters are estimated subject to this pattern, usually by maximum likelihood. Statistical tests are available to check how well the estimated parameters from CFA actually fit the data. Recently, Bayesian exploratory approaches were suggested to estimate the sparse loading structure. In this paper, we investigate under which circumstances such sparsity patterns are unique and in which cases it may be necessary to consider the existence of multiple patterns with similar degrees of sparsity for a given data set. Moreover, we propose two MCMC approaches that can be used to discover possibly multiple sparse representations for the loadings matrix. Both approaches are further investigated in a simulation study and in an application using a Swiss inflation data set.

CO429 Room 112 NEW DEVELOPMENTS IN TIME-VARYING PARAMETER MODELS Chair: Francisco Blasques

CO0331: **Adaptive combination schemes for point and density forecasts**
Presenter: **Leopoldo Catania**, Tor Vergata University of Rome, Italy
Co-authors: Tommaso Proietti
Point and density forecast combinations based on some optimality measure are becoming increasingly popular for pooling information coming from different sources. We propose a new way of combining point and density forecasts allowing for time varying weights. Our method allows for model misspecification and does not require that the true data generating process belongs to the available set. Concerning the density combination schemes, our combined predictive density dynamically approximate the true unknown density in a Kullback–Leibler sense using a recursion based on the score of the implied mixture conditional density. Similarly, for the point forecast combination schemes we propose, the weights are updated using the gradient of the user defined loss function. The relevance of the new combination techniques is illustrated by several Monte Carlo experiments and an empirical application in time series econometrics.

CO0367: **Adaptive state space models with applications to the business cycle and financial stress**
Presenter: **Davide Delle Monache**, Bank of Italy, Italy
Co-authors: Fabrizio Venditti, Ivan Petrella
The estimation of state-space models with time-varying parameters typically implies the use of computationally intensive methods. Moreover, when volatility evolves stochastically, the model ceases to be conditionally Gaussian and requires nonlinear filtering techniques. We model parameters' variation in a Gaussian state-space model by letting their dynamics to be driven by the score of the predictive likelihood. In this setup, conditionally on past data, the model remains Gaussian and the likelihood function can be evaluated using the Kalman filter. We derive a new set of recursions running in parallel with the standard Kalman filter recursions that allows us estimate simultaneously the unobserved state vector and the time-varying parameters by maximum likelihood. Given that a variety of time series models have a state space representation, the proposed methodology is of wide interest in econometrics and applied macroeconomics. Specifically, the usefulness of the methodology is illustrated in two applications: the former aims to improve GDP measurement based on alternative noisy measures, in the letter we construct an index of financial conditions that can be used to nowcast GDP in real time.

CO0603: **Transformed polynomials for modeling clusters of conditional volatility**
Presenter: **Francisco Blasques**, VU University Amsterdam, Netherlands
Co-authors: Siem Jan Koopman, Andre Lucas, Marco Bazzi
A flexible model is proposed for filtering time-varying conditional volatilities. In particular, we make use of novel transformed polynomial functions to update the unobserved time-varying conditional volatility parameter. Our flexible class of updating equations is shown to be dense in the space of continuous function and to have known convergence rates of approximation on smooth Sobolev 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 simulation reveals good finite sample properties of the estimator. Finally, a number of applications show the good performance of the model in empirically relevant settings.

CO0758: **Dynamic models with too many zeroes**
Presenter: **Ryoko Ito**, Oxford University, United Kingdom
Co-authors: Andrew Harvey, Ryoko Ito
Situations are considered in which a significant proportion of observations are identically zero or have been set to zero because they fall below a certain threshold. When the distribution for non-zero observations is continuous, the net result is a point-mass mixture distribution. We propose an integrated approach for dynamic models with censored observations that is straightforward to apply and logically consistent. The model is fitted to daily precipitation data of selected Australian cities.

CO632 Room 002 EMPIRICAL MACROECONOMICS Chair: Kirstin Hubrich

CO0392: **Understanding the sources of macroeconomic uncertainty**
Presenter: **Barbara Rossi**, Universitat Pompeu Fabra and ICREA, Spain
A decomposition is proposed to distinguish between Knightian uncertainty (ambiguity) and risk, where the first measures the uncertainty about the probability distribution generating the data, while the second measures uncertainty about the odds of the outcomes when the probability distribution is known. We use the Survey of Professional Forecasters (SPF) density forecasts to quantify overall uncertainty as well as the evolution of the different components of uncertainty over time and investigate their importance for macroeconomic fluctuations. We also study the behavior and evolution of the various components of our decomposition in a model that features ambiguity and risk.

CO1201: **Tails of inflation forecasts and tales of monetary policy**
Presenter: **Philippe Andrade**, Banque de France, France
Co-authors: Eric Ghysels, Julien Idier
A new measure called Inflation-at-Risk (I@R) associated with (left and right) inflation tail risk is introduced. We estimate I@R using survey-based density forecasts. We show that it contains information not covered by usual inflation risk indicators which focus on inflation uncertainty and do not distinguish between the risks of low or high future inflation outcomes. Not only the extent, but also the asymmetry of inflation / deflation risks vary over time. Moreover, changes in inflation risks matter for macroeconomic outcomes: they help predict future inflation realizations and have an impact on the interest rate the central bank targets.

CO1218: **Macroeconomic implications of oil price fluctuations: A regime-switching framework for the Euro area**
Presenter: **Kirstin Hubrich**, Federal Reserve Board, United States
Co-authors: Federic Holm-Hadulla
A Markov-switching Vector Autoregressive Model is used to study the response of the Euro area economy to oil price shocks. The model identifies two regimes that are characterized by different effects of oil price fluctuations on economic activity and inflation. In the 'normal regime', oil price shocks exert only limited and short-lived effects on these variables. In the 'adverse regime', by contrast, oil price shocks trigger sizeable and sustained macroeconomic effects, with inflation and economic activity moving in the same direction as the oil price. The response of inflation expectations points to second-round effects as a potential driver of the dynamics characterising the adverse regime. By delivering (conditional) probabilities for being (staying) in either regime, the model helps interpret oil price fluctuations and assess their monetary policy implications in real-time.

CO1698: **Monetary policy, private debt and financial stability risks**
Presenter: **Eleonora Granziera**, Bank of Finland, Finland
Can monetary policy be used to promote financial stability? We answer this question by estimating the impact of a monetary policy shock on private-sector leverage and the likelihood of a financial crisis. Impulse responses obtained from a panel VAR of eighteen advanced countries suggest that the debt-to-GDP ratio rises in the short run following an unexpected tightening in monetary policy. As a consequence, the likelihood of a financial crisis increases, as estimated from a panel logit regression. However, in the long run, output recovers and higher borrowing costs discourage new lending, leading to a deleveraging of the private sector. A lower debt-to-GDP ratio in turn reduces the likelihood of a financial crisis. These results suggest that monetary policy can achieve a less risky financial system in the long run but could fuel financial instability in the short run. We also find that the ultimate effects of a monetary policy tightening on the probability of a financial crisis depend on the leverage of the private sector: the higher the initial value of the debt-to-GDP ratio, the more beneficial the monetary policy intervention in the long run, but the more destabilizing in the short run.

CO325 Room 005 RECENT ADVANCES IN HIGH DIMENSION ECONOMETRICS Chair: Weining Wang

CO0419: **Inference in high-dimensional instrumental variable models**
Presenter: **Christoph Breunig**, Humboldt-Universitat zu Berlin, Germany
A method is proposed for inference of single or low-dimensional components of a high-dimensional structural parameter in a triangular model. The method can be used for constructing confidence intervals and statistical tests for such structural components.

CO1079: **Network quantile autoregression**
Presenter: **Weining Wang**, Humboldt University at Berlin, Germany
It is a challenging task to understand the complex dependency structures in a ultra-high dimensional network, especially when the dependency between some of the nodes are highly non-linear. To tackle this problem, we consider a network quantile autoregression model (NQAR) to model the dynamic tail behavior ina complex system. In particular, we relate responses to its connected nodes and node specific characteristics in a quantile autoregression process. A minimum contrast estimation approach for the NQAR model is introduced, and its asymptotic properties are studied. Finally, we demonstrate the usage of our model by studying the financial contagions in Chinese stock market accounting for shared ownership of companies.

CO1561: **Estimation of nonlinear panel models with multiple unobserved effects**
Presenter: **Mingli Chen**, University of Warwick, United Kingdom
The aim is to propose a fixed effects expectation-maximization (EM) estimator that can be applied to a class of nonlinear panel data models with unobserved heterogeneity, which is modeled as individual effects and/or time effects. Of particular interest is the case of interactive effects, i.e. when the unobserved heterogeneity is modeled as a factor analytical structure. The estimator is obtained through a computationally simple, iterative two-step procedure, where the two steps have closed form solutions. We show that estimator is consistent in large panels and derive the asymptotic distribution for the case of the probit with interactive effects. We develop analytical bias corrections to deal with the incidental parameter problem. Monte Carlo experiments demonstrate that the proposed estimator has good finite-sample properties. We illustrate the use of the proposed model and estimator with an application to international trade networks.

CO0251: **Vine based modeling of multivariate realized volatility**
Presenter: **Yarema Okhrin**, Universitaet Augsburg, Germany
Studying realized volatility based on high-frequency data is of particular interest in asset pricing, portfolio management and evaluation of risk. We propose an approach for dynamic modeling and forecasting of realized correlation matrices that allows for model parsimony and automatically guarantees positive definiteness of the forecast. We use the one-to-one relationship between a correlation matrix and its associated set of partial correlations corresponding to any regular vine specification. Being algebraically independent the partial correlations of the vine do not need to satisfy any algebraic constraint such as positive definiteness. We present several selection methods to choose, among the large number of vines, the vine structure best capturing the characteristics of the multivariate time-series of the correlation parameters. The individual partial correlation time-series are studied using ARFIMA and HAR models to account for long-memory behavior. The dependence between assets is exibly modeled using vine copulas that allow for nonlinearity and asymmetric tail behavior. Correlation point forecasts are obtained as nonlinear transformation of the forecasts for the partial correlation vine. The usefulness of the methodology is investigated in a one-day ahead forecasting framework comparing existing methods based on a model confidence set approach.

CO539 Room Board meeting room II RECENT DEVELOPMENTS IN NONPARAMETRIC AND ROBUST INFERENCE Chair: Jana Jureckova

CO0451: **Behavior of rank tests and estimates in measurement error models**
Presenter: **Radim Navratil**, Masaryk University Brno, Czech Republic
Measurement error models (also called errors-in-variables models) are regression models that account for measurement errors in the independent variables (regressors). These models occur very commonly in practical data analysis, where some variables cannot be observed exactly, usually due to instrument or sampling error. Sometimes ignoring measurement errors may lead to correct conclusions, however in some situations it may have dramatic consequences. Behavior of standard rank procedures (both tests and estimates) in measurement error models will be investigated. The main goal is to investigate if classical rank tests and estimates stay valid and applicable when there are some measurement errors present and if not how to modify these procedures to be still able to do some statistical inference. Finally, performance of the tests and estimates will be illustrated on practical examples and simulations.

CO0532: **Generalization of L-moments in the linear regression model**
Presenter: **Jan Picek**, Technical University of Liberec, Czech Republic
Co-authors: Martin Schindler
The L-moments are analogues of the conventional moments and have similar interpretations. They are calculated using linear combinations of the expectation of ordered data. It was recently shown that L-moments are special case of L-estimators. We propose a generalization of L-moments in the linear regression model based on regression quantiles as special L-estimator. The properties of extended L-moments are illustrated on simulated data.

CO0496: **Maximum likelihood method for bandwidth detection of kernel conditional density estimation**
Presenter: **Katerina Konecna**, Masaryk University, Czech Republic
Co-authors: Ivanka Horova
Kernel smoothing technique is a suitable tool for estimation of conditional density estimation. Kernel estimations depend on a kernel which plays a role of weight function and smoothing parameters which control a smoothness of the estimation. The problem of a choice of smoothing parameters, i.e., how much to smooth, is the most significant. The aim is to compare the well-known cross-validation method and maximum likelihood method. The performance of these methods is compared by simulation studies. Application to real data is also included.

CO0498: **Regression quantile and averaged regression quantile processes**
Presenter: **Jana Jureckova**, Charles University in Prague, Czech Republic
The regression alpha-quantile and its two-step version, and weighted averages B(α) and B*(α) of both, with the regressors as the weights are considered. They are asymptotically equivalent to each other and to the location α-quantile of the model errors. As the processes in $\alpha \in (0,1)$, they differ by a drift and converge to the Brownian bridge over (0,1) after a suitable standardization.

CO411 Room 008 VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Teruo Nakatsuma

CO0473: **Multivariate stochastic volatility models with realized volatility and pairwise realized correlation**
Presenter: **Yuta Yamauchi**, University of Tokyo, Japan
Co-authors: Yasuhiro Omori
Although stochastic volatility and GARCH models have been successful to describe the volatility dynamics of univariate asset returns, their natural extension to the multivariate models with dynamic correlations has been difficult due to several major problems. We consider dynamic latent correlation variables in addition to latent volatility variables and estimate model parameters using Markov chain Monte Carlo simulations, where we sample latent correlation variables one at a time given others so that we keep the covariance matrices positive definite. Our contributions are: (1) we obtain the stable parameter estimates for dynamic correlation models using the realized measures, (2) we make full use of intraday information by using pairwise realized correlations, (3) the covariance matrices are guaranteed to be positive definite, (4) we avoid the arbitrariness of the ordering of asset returns, and (5) propose the flexible correlation structure model (e.g. such as setting some correlations to be identically zeros if necessary). Our proposed models are applied to daily returns of nine U.S. stocks with their realized volatilities and pairwise realized correlations, and are shown to outperform the existing models with regard to portfolio optimizations.

CO1016: **Stochastic volatility with regimes, skew, fat tails, and leverage using returns and realized volatility for inference**
Presenter: **Sebastian Trojan**, None, Germany
A general stochastic volatility (SV) model specification with leverage, heavy tails, skew, and switching regimes, using realized volatility (RV) as an auxiliary time series to improve inference on latent volatility is presented. The information content of the range, and of implied volatility using the VIX index, is also analyzed. Asymmetry in the observation error is modeled by the generalized hyperbolic skew Student-*t* distribution, whose heavy and light tail enable substantial skewness. Database is the S& P500 index. Number of regimes and dynamics differ dependent on the deployed auxiliary volatility proxy and are investigated for the financial crash period 2008/09 in more detail. An extended in-sample model selection is provided. An out-of-sample study comparing predictive ability of various model variants for a calm and a volatile period yields insights about the gains on forecasting performance that can be expected by incorporating different volatility proxies into the model. Findings indicate that including RV pays off mostly in more volatile market conditions, whereas in calmer environments SV specifications using no auxiliary series outperform. Results for the VIX as a measure of implied volatility point in a similar direction. The range as volatility proxy provides a superior in-sample fit, but its predictive performance is found to be weak.

CO1370: **Bayesian emulation for optimization: Multi-step portfolio decision analysis**
Presenter: **Kaoru Irie**, University of Tokyo, Japan
Co-authors: Mike West
Bayesian analysis is discussed for portfolio studies involving multi-step forecasts and decisions in financial time series. Using classes of economically and psychologically relevant multi-step ahead utility functions, we develop solutions to the resulting Bayesian expected utility optimization problem. The solution paths involve mapping the technical structure of (some) optimization problems to those of parallel, synthetic Bayesian inference problems in "emulating" statistical models. This provides access to standard Bayesian simulation and optimization methods that then yield indirect solutions of the decision problems. Study of sequential portfolio studies with multivariate currency, commodity and stock index time series illustrate the approach and show some of the practical utility and benefits of the new Bayesian decision analysis framework and methodology.

CC1520: **Hierarchical Bayes modeling of autocorrelation and intraday seasonality in financial durations**
Presenter: **Teruo Nakatsuma**, Keio University, Japan
Co-authors: Tomoki Toyabe
Intraday financial transactions are irregularly spaced, and their durations exhibit positive autocorrelation and intraday seasonality. In the literature, the former is formulated as a time-dependent duration model such as the stochastic conditional duration (SCD) model while the latter is dealt with

by filtering out any cyclical fluctuations in time series of durations with a spline smoothing method before the duration model is estimated. We propose a hierarchical Bayes approach to model both autocorrelation and intraday seasonality in durations simultaneously. In our new approach, the autocorrelation structure of durations is captured by the SCD model while the intraday seasonality is approximated with B-spline smoothing, and the parameters in both models are allowed to differ on each trading day. In B-spline smoothing, we incorporate the smoothness prior (the penalty on variations in the seasonality on the same trading day) as well as the similarity prior (the penalty on differences in the seasonality between consecutive trading days) in order to prevent overfitting. The resultant model is regarded as a non-linear non-Gaussian state space model of unbalanced panel data, for which a Bayesian approach is suitable. We developed an efficient Markov chain sampling scheme for the posterior analysis of the proposed model and applied it to high-frequency commodity futures transaction data in the Tokyo Commodity Exchange.

CO315 Room 110 ECONOMETRICS OF ART MARKETS Chair: Douglas Hodgson

CO0488: **Statistical prediction of art prices at auction**
Presenter: **Douglas Hodgson**, UQAM, Canada
Co-authors: John Galbraith

Predicting the price of a unique object, given a limited sample, is a challenging problem but of interest to market participants, including owners and insurers. The usual approach is least-squares estimation of a hedonic model for objects of a given class, such as paintings from a particular school or period. The present paper examines statistical refinements of the standard methods. First, we consider the level of aggregation that is appropriate for pooling observations into a sample, including the use of pattern recognition algorithms to identify clusters. Second, we apply model-averaging methods to estimate predictive models at the individual-artist level, in contexts where sample sizes would otherwise be insufficient to do so; averaging (ensemble prediction) is also used with regression-tree machine learning methods. Finally, we consider an additional stage in which we incorporate repeat-sale information in the subset of cases for which this information is available. The results are applied to a data set of auction prices for Canadian paintings. We compare the out-of-sample predictive accuracy of the various methods and find that those that allow us to use single-artist samples produce superior results, and that data driven averaging across predictive models produces clear gains. As well, where available,repeat-sale information appears to produce further improvements in predictive accuracy.

CO0493: **A finite mixture model to capture unobserved heterogeneity in art prices: Evidences from surrealism**
Presenter: **Juan Prieto-Rodriguez**, Universidad de Oviedo, Spain

The purpose is to analyse whether there is or not a single mechanism in price formation at the high end of the modern art market defined by auction houses at the most important art markets in the world. A serious problem in estimating functions in cultural markets is the presence of unobserved heterogeneity. For instance, the shape of the demand or value function in the modern art market relies on individual characteristics of artworks, which are may not be all directly observed. Moreover, just one value function may not be enough to capture differences in price formation if it differs across pieces of art. In order to reduce unobserved heterogeneity, we have used a sample on paintings sold at auctions between 1990 and 2007 related to different schools and movements of Surrealism. We have used a finite mixture approach to emulate the data generating process underlying the price formation. Such models have allowed us first to identify art market segments defined by a similar but unobserved price structure and, then, to investigate differences in prices determinants. Additionally, we have found some separating variables or determinants upon which art pieces are classified into a specific market segment. Thanks to this procedure we can reject the hypothesis of a unique price structure in auction markets for Surrealism. As expected, moreover, segmentation depends on local markets and on the auction houses involved.

CO0947: **Cultural and creative industries in peripheral areas: A study of the Euregion Maas Rhein area**
Presenter: **Anouk Duivenvoorden**, Maastricht University, Netherlands
Co-authors: Marina Gertsberg, Rachel Pownall

Policy on creative industries is increasingly important as an urban development tool. Current studies are primarily focused on the analysis of the cultural and creative industries (CCIs) in large urban regions; knowledge on the effects of CCIs in peripheral areas is still limited. We study the development of the CCI, and in particular the fashion industry in the Euregion Maas Rhine, which encompasses the cross-border region of Maastricht in the South of The Netherlands, Aachen in West-Germany, and Liege in Belgium. Our aim is to evaluate the potential of a change in local recent municipality policy on developing the regional fashion industry in the city of Maastricht. This policy emphasises the importance of local and sustainable production and the intention is to generate growth in employment in the fashion industry of approximately 70% within the next five years. We map the current state of CCIs in the Euregion Maas Rhine for comparison and to measure the potential spillover effects between regions. Other studies show that the city of Maastricht has a high level of sole proprietorships, in general. How this relates to the CCIs is of particular interest. We find structural differences in the developments paths and the statusquo between regions in terms of start-up activity, specifically with respect to sole proprietorships, and employment data.

CO0948: **Market evolution, bidding strategies, and survival of art dealers**
Presenter: **Marina Gertsberg**, University of Maastricht, Netherlands
Co-authors: Rachel Pownall, Daskhina De Silva

The evolution of theLondon-based fine-art dealers is studied and effects of asymmetric information is analyzed on their bidding and survival patterns during 1800-1913. While dealers captured about 17% of the total purchases between 1800 and 1850, the share increased to about 56% during 1850-1913. First, we show that dealers entered the market as the number of paintings traded and population increased. Next, we examine whether the experience of an art dealer can explain the differences in the acquisition or bidding strategies. We show that more experienced dealers payabout 23% more for an artwork of the same quality than inexperienced dealers. This price difference can be attributed to the informational advantage of experienced bidders. Additionally, our results indicate that less experienced dealers are less likely to survive in the market. Our evidence supports the conjecture that common value auctions with information asymmetries offer benefits to bidders with better information.

CO319 Room 106 COMMON FEATURES IN ECONOMICS AND FINANCE Chair: Joao Victor Issler

CO1274: **Testing for deterministic seasonality in mixed-frequency VARs**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Tomas del Barrio Castro

The mixed-frequency VAR is a multivariate modelling that stacks time series at different frequencies in order to jointly model the behavior of a low- and the high-frequency variables (respectively LF and HF hereafter). This approach complements the MIDAS regression in which a single non-linear equation from the LF to HF series is estimated. The MF-VAR modelling, although to some extent inherited from the periodic autoregressive models (i.e. PAR(p)), is often estimated on seasonally adjusted data or, at least, the consequences of such seasonality components is not really accounted for. A strategy is provided to estimate a full range of interesting hypotheses about deterministic seasonal features within raw data. We also examine the presence of common seasonal features. We apply our testing framework on the relationship between quarterly employment and monthly tourist arrivals in the Balearic Islands.

CO1291: **Inattention in individual expectations**
Presenter: **Joao Victor Issler**, Getulio Vargas Foundation, Brazil

The expectations formation process of economic agents about inflation rate is investigated. Using the Market Expectations System of Central Bank of Brazil, we perceive that agents do not update their forecasts every period and that even agents who update disagree in their predictions. We then focus on the two most popular types of inattention models that have been discussed in the recent literature: sticky-information and noisy-information models. Estimating a hybrid model we find that, although formally fitting the Brazilian data, it happens at the cost of a much higher degree of information rigidity than observed.

CO1326: **Asymptotically unbiased estimation of large dynamic panel models**
Presenter: **Jose Diogo Barbosa**, University of Michigan, United States

The estimation of a dynamic panel model with fixed effects, time-series heteroskedasticity and exogenous covariates is considered. This model allows natural rotational invariance conditions, such as re-ordering the individuals in the sample. An invariance principle determines a maximal invariant statistic. Its distribution yields an estimator for the structural parameters that is consistent and asymptotically unbiased in the large N, fixed T and in the large N, large T asymptotics. This method, therefore, solves the incidental-parameter problem created by the presence of fixed effects and time-varying heteroskedasticity even when both N and T are increasing.

CO1539: **Mixed causal-noncausal autoregressions with strictly exogenous regressors for structural expectations equations**
Presenter: **Sean Telg**, Maastricht University, Netherlands
Co-authors: Alain Hecq, Joao Victor Issler

Some authors have proposed mixed autoregressive causal-noncausal (MAR) models to estimate economic relationships involving expectations variables. Indeed, when linearly solved, those structural equations usually imply explosive roots in their autoregressive part but have stationary solutions when the future of the realized variable is considered instead. In previous work, possible exogenous variables in economic relationships are substituted into the error term and are assumed to follow an MAR process to ensure the MAR structure of the variable of interest. Following this procedure, one loses the impact of exogenous variables that are very important for understanding economic phenomena. For that reason, we instead consider a MARX representation which allows for the inclusion of strictly exogenous regressors. We develop the asymptotic distribution of the MARX parameters. We assume a Student's t-likelihood to develop closed form solutions of the corresponding standard errors. Remarkably, our method does not involve tedious numerical methods. By means of Monte Carlo simulations, we evaluate the merit of our approach as well as the accuracy of a model selection based on information criteria after estimating models by non-Gaussian MLE. Several present value specifications empirically illustrate our analysis.

CO383 Room 108 RECENT ADVANCES ON FINANCIAL TIME SERIES MODELLING Chair: Cristina Amado

CO0744: **Modelling volatility via smooth transition copulas**
Presenter: **Emilio Zanetti Chini**, University of Pavia, Italy
Co-authors: Michele Costola

A model is proposed to characterize the conditional dependence via vine-copula assuming a non-linear change in the the dependence structure through a smooth transition function. The sequential nature of the vine approach and the peculiarity of model (not identified under the null hypothesis of linearity) requires a proper statistical treatment. In particular, we show that such a form of non-linear vine copula can be efficiently estimated via peculiar version of the augmented Kalman filter. A Monte Carlo simulation study proves the efficiency of the new technique. In the empirical analysis, we apply the model to HAR-RVs of the major stock indices using the Oxford-Man Institute's realized library.

CO1450: **Forecasting volatility using long memory dynamics: How effective is the use of a realised measure**
Presenter: **Katarzyna Lasak**, VU Amsterdam, Netherlands
Co-authors: Hidde Jelsma

A closer look is taken at multiple models to estimate the unobserved volatility component of financial returns. This is done for three stocks, Nike Inc., AEGON n.v., and Idex Corporation over the period of August 3rd 2009 till August 1st 2014. We use models from the GARCH and GAS frameworks and extend these models to incorporate long memory dynamics. We look at the forecasting performance and the predictive capabilities of risk. Forecasting performance is tested by comparing the forecasts with the Realised kernel described previously in the literature, which is used as a proxy for the true volatility, and applying the Diebold-Mariano test based on squared and absolute errors. The predictive capabilities of risk are compared by looking at the Value at Risk and applying the backtesting methods also previously described. We find that models imposing long memory dynamics do not provide better one-day-ahead forecasts but become more powerful if we extend the forecasting horizon to ten days, and that the models using the Realised kernel have provided more accurate forecasts and predict risk better than models using only the returns as input, but that the Realised kernel is not always an accurate measure of the true volatility.

CC1710: **A smooth transition approach to modelling diurnal variation in models of autoregressive conditional duration**
Presenter: **Cristina Amado**, University of Minho, Portugal
Co-authors: Timo Terasvirta

The aim is to introduce a new approach for adjusting the diurnal variation in the trade durations. The model considers that durations are multiplicatively decomposed into a deterministic time-of-day and a stochastic component. The parametric structure of the diurnal component allows the duration process to change smoothly over the time-of-day. In addition, a testing framework consisting of Lagrange multiplier tests is proposed for specifying the diurnal component. Our methodology is applied to the IBM transaction durations traded at the New York Stock Exchange.

CC1577: **Quantile cross-spectral measures of dependence between economic variables**
Presenter: **Jozef Barunik**, UTIA AV CR vvi, Czech Republic
Co-authors: Tobias Kley

A quantile cross-spectral analysis of multiple time series is introduced which is designed to detect general dependence structures emerging in quantiles of the joint distribution in the frequency domain. We argue that this type of dependence is natural for economic time series but remains invisible when the traditional analysis is employed. To illustrate how such dependence structures can arise between variables in different parts of the joint distribution and across frequencies, we consider quantile vector autoregression processes. We define new estimators which capture the general dependence structure, provide a detailed analysis of their asymptotic properties and discuss how to conduct inference for a general class of possibly nonlinear processes. In an empirical illustration we examine one of the most prominent time series in economics and shed new light on the dependence of bivariate stock market returns.

EO517 Room 204 ESTIMATION OF TAIL RISK Chair: Abdelaati Daouia

EO0269: **Estimation of the marginal expected shortfall in the context of an infinite mean model**
Presenter: **Armelle Guillou**, Strasbourg, France
Co-authors: Juan Juan Cai, Valerie Chavez-Demoulin
An estimator is developed for the marginal expected shortfall in the case of an infinite mean model. We establish the asymptotic normality of our estimator under general assumptions. The finite sample performance is illustrated by a simulation study, based on which we give some recommendations about the tuning parameters of the estimation in practice. We apply our methodology to tsunami data and operational risk data.

EO1141: **Estimation of extreme expectiles from heavy tailed distributions**
Presenter: **Stephane Girard**, Inria, France
Co-authors: Abdelaati Daouia, Gilles Stupfler
The class of quantiles lies at the heart of extreme-value theory and is one of the basic tools in risk management. The alternative family of expectiles is based on squared rather than absolute error loss minimization. Both quantiles and expectiles were embedded in the more general class of M-quantiles as the minimizers of a generic asymmetric convex loss function. It has been proved very recently that the only M-quantiles that are coherent risk measures are the expectiles. Least asymmetrically weighted squares estimation of expectiles did not, however, receive yet as much attention as quantile-based risk measures from the perspective of extreme values. We develop new methods for estimating the Value at Risk and Expected Shortfall measures via high expectiles. We focus on the challenging domain of attraction of heavy-tailed distributions that better describe the tail structure and sparseness of most actuarial and financial data.

EO0920: **Local robust estimation of the Pickands dependence function**
Presenter: **Yuri Goegebeur**, University of Southern Denmark, Denmark
Co-authors: Armelle Guillou, Mikael Escobar-bach
Robust estimation is considered for the Pickands dependence function in the covariate framework. Our estimator is based on local estimation with the minimum density power divergence criterion. We provide its main asymptotic properties, in particular the convergence of the stochastic process, correctly normalized, towards a tight centered Gaussian process. The finite sample performance of our estimator is illustrated on a small simulation study involving both uncontaminated and contaminated samples.

EO1713: **Why risk is so hard to measure**
Presenter: **Chen Zhou**, Erasmus University Rotterdam, Netherlands
Co-authors: Jon Danielsson
The aim is to analyze the robustness of standard techniques for risk analysis, with a special emphasis on the Basel III risk measures. We focus on the difference between Value-at-Risk and expected shortfall, their small sample properties, the scope for manipulating risk measures and how estimation can be improved. Overall, we find that risk forecasts are extremely uncertain at low sample sizes, with Value-at-Risk more accurate than expected shortfall, while Value-at-Risk is easily manipulated without violating regulations. Finally the implications for practitioners and regulators are discussed along with best practice suggestions.

EO055 Room 309 B MACHINE LEARNING, APPROXIMATION AND ROBUSTNESS Chair: Andreas Christmann

EO0271: **Approximation analysis of distributed learning with spectral algorithms**
Presenter: **Ding-Xuan Zhou**, City University of Hong Kong, China
Analyzing and processing big data has been an important and challenging task in various fields of science and technology. Statistical learning theory has been widely used for data analysis. It aims at learning function relations or data structures from samples. Distributed learning is an important topic in statistical learning theory and is a powerful method to handle big data. Distributed learning is based on a divide-and-conquer approach and consists of three steps: first we divide over-sized data into subsets and each data subset is distributed to one individual machine, then each machine processes the distributed data subset to produce one output, finally the outputs from individual machines are combined to generate an output of the distributed learning algorithm. It is expected that the distributed learning algorithm can perform as efficiently as one big machine which could process the whole over-sized data, in addition to the advantages of reducing storage and computing costs.In this talk we discuss distributed learning with spectral algorithms, a class of kernel-based learning algorithms using spectral information of the data and the target distribution, including the least squares regularization scheme for regression. We provide some approximation analysis of these distributed learning algorithms, and demonstrate error bounds and learning rates in reproducing kernel Hilbert spaces.

EO0942: **A robust approach to summarize the location of functional data**
Presenter: **Beatriz Sinova**, University of Oviedo, Spain
Co-authors: Stefan Van Aelst
Some robust location measures to summarize the location of functional data have already been proposed in the literature, like, for instance, trimmed means (either involving depth functions or not). The aim is to extend another well-known and successful methodology used for univariate and multivariate real-valued data to the functional-valued setting, the so-called M-estimation. Such a generalization will be based on some developments from the literature for robust nonparametric density estimation. First, the theoretical results related to the existence and uniqueness of functional-valued M-estimators, as well as their consistency and uniqueness, will be established. The second part illustrates the computation of functional-valued M-estimators associated with some well-known choices for the loss function, such as Huber, Hampel and Tukey ones, and shows their finite-sample performance by means of several simulation studies.

EO0568: **Support vector machines for non-i.i.d. observations**
Presenter: **Katharina Strohriegl**, University of Bayreuth, Germany
Co-authors: Andreas Christmann
Today, support vector machines and other regularized kernel methods are an important tool to solve classification and regression problems. The goal of support vector machines is to find a function, which describes the relation between input values x and output values y, by using a given data set. So far, the overwhelming part of theoretical work is done for the assumption that data are generated by independent and identically distributed random variables. However, this assumption is not fulfilled in many practical applications, and non-i.i.d. cases increasingly attract attention. We examine important properties of statistical estimators - consistency and qualitative robustness in the case of observations which do not fulfill the i.i.d. assumption. To generalize the results of the i.i.d. case we consider stochastic processes which provide a certain convergence of the empirical measure to a limiting distribution. Consistency can be shown for assumptions on the dependence structure of the stochastic process and for the assumption on the empirical measure. Examples are α-mixing processes and weakly dependent processes. Moreover the convergence of the empirical measure together with assumptions on the estimator lead to qualitative robustness, for some processes a bootstrap approach is also qualitatively robust.

EO0406: **Robust localized learning with kernels**
Presenter: **Andreas Christmann**, University of Bayreuth, Germany
Some regularized risk methods based on kernels for the big data situation will be described. A divide-and-conquer approach is used to enable localized learning. Statistical robustness will be the main focus.

EO087 Room 308 B STATISTICAL IMAGING Chair: Michele Guindani

EO0280: **Bayesian semi-parametric modeling of near infra-red spectroscopy data**
Presenter: **Timothy Johnson**, University of Michigan, United States
Functional near-infrared spectroscopy (fNIRS) is a relatively new neuroimaging technique. It is a low cost, portable, non-invasive method to monitor brain activity. Similar to fMRI, it measures changes in the level of blood oxygen in the brain. Its time resolution is much finer than fMRI, however its spatial resolution is much courser-similar to EEG or MEG. fNIRS is finding widespread use on young children whom cannot remain still in the MRI magnet and it can be used in situations where fMRI is contraindicated-such as with patients whom have cochlear implants. Since it measures the hemodynamic response to stimulus, similar to fMRI, statistical methods that are in use simply use modifications to existing fMRI packages. We show that these methods are inadequate and we propose a fully Bayesian semi-parametric hierarchical model to analyze fNIRS data. The hemodynamic response function is modeled using a cubic B-spline basis while nuisance signals (e.g. vasomotor signal and heart beat) are modeled using a Gaussian process. We assume the residual time-series is a high-order AR process and adopt a spike-and-slab prior to shrink unnecessary AR parameters to zero. Our model is easily adapted to handle the bivariate fNIRS time-series data at a single detector (oxygenated-and deoxygenated-hemoglobin). It can also easily be adapted to handle the spatial aspects of an array of detector.

EO0927: **A variance components model for statistical inference on functional connectivity networks**
Presenter: **Mark Fiecas**, University of Minnesota, United Kingdom
Co-authors: Ivor Cribben
A variance components linear modeling framework is proposed for statistical inference on functional connectivity networks that accounts for the temporal autocorrelation inherent in functional magnetic resonance imaging (fMRI) time series data and for the heterogeneity across subjects. The novel method estimates the former in a nonparametric and subject-specific manner, and estimates the latter using iterative least squares and residual maximum likelihood. We apply the new model to a resting-state fMRI study to compare the functional connectivity networks in both typical and reading impaired young adults in order to characterize the resting state networks that are related to reading processes. We also compare the performance of the model to other methods that do not account for the temporal autocorrelation or heterogeneity across the subjects using an extensive simulation study.

EO1159: **Association of structural brain imaging markers with alcoholism using structural connectivity via a regularized approach**
Presenter: **Jaroslaw Harezlak**, Indiana University Fairbanks School of Public Health, United States
Co-authors: Mario Dzemidzic, Joaquin Goni, David Kareken, Marta Karas
Brain imaging studies collect multiple imaging data types, but most analyses are done for each modality separately. Statistical methods that simultaneously utilize and combine multiple data types can instead provide a more holistic view of brain function. We utilize cortical thickness measures obtained by FreeSurfer software to predict the alcoholism-related phenotypes while incorporating prior information from the structural connectivity between cortical regions. We develop a functional linear model with a penalty operator to quantify the relative contributions of imaging markers obtained from high resolution structural MRI (cortical thickness) as predictors of drinking frequency and risk-relevant personality traits, while co-varying for age. We estimate model parameters by a unified approach directly incorporating structural connectivity information into the estimation by exploiting the joint eigenproperties of the predictors and the penalty operator. We applied the developed methods to a sample of 148 young (21-35 years) social-to-heavy drinking male subjects from several alcoholism risk studies. Structural connectivity model was used to estimate the density of connections between 66 cortical regions based on Desikan-Killiany atlas. Using our method we found the best 11 average cortical thickness markers of the Alcohol Use Disorders Identification Test score.

EO1308: **Time varying models for brain imaging data**
Presenter: **Ivor Cribben**, Alberta School of Business, Canada
In 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 is changing over time even when the subjects are at rest. Firstly, we formulate the problem in a high-dimensional time series framework and introduce a data-driven method which detects change points in the network structure of a multivariate time series, with each component of the time series represented by a node in the network. Secondly, we introduce a new time varying approach that is model-free, data-adaptive, and is applicable in situations where the (global) stationarity of the time series from the brain regions fails, such as the cases of local stationarity and/or change points. We apply both new methods to simulated data and to a resting-state fMRI data set.

EO640 Room 206 SURVIVAL ANALYSIS WITH COPULA AND FRAILTY MODELS Chair: Takeshi Emura

EO0462: **A conditional copula model for clustered right-censored event time data**
Presenter: **Candida Geerdens**, Hasselt University, Belgium
Co-authors: Elif Acar, Paul Janssen
In many studies the response of interest is the time until a predefined event (e.g. the time to blindness). Often, the true event time is unknown and only a lower time bound can be observed (e.g. due to the end of the study period). The data are right-censored. Further, event times can be grouped into clusters and thereby be correlated (e.g. the time to blindness is registered for left and right eye, leading to a cluster of size two). Copulas provide a popular tool to describe the association in clustered time-to-event data. We propose an estimation and a testing strategy to assess the impact of a continuous cluster-level covariate on the dependence in grouped right-censored event time data, as modeled through a conditional copula. A local likelihood approach is used to estimate the functional form of the copula parameter and a generalized likelihood ratio test (with bootstrap) is described to assess its constancy. The performance of the method is evaluated in a simulation study, under different rates of right-censoring and for various parametric copula families, considering both parametrically and nonparametrically estimated margins. Data on diabetic retinopathy (blindness due to diabetic disease) are used as illustration.

EO1289: **The use of tumor dynamics and new lesions to predict survival with multivariate joint frailty models**
Presenter: **Virginie Rondeau**, University of Bordeaux INSERM, France
Co-authors: Agnieszka Krol
In oncology, the international WHO and RECIST criteria have allowed the standardization of tumor response evaluation in order to identify the time of disease progression. These semi-quantitative measurements are often used as endpoints in phase II and phase III trials to study the efficacy of new therapies. However, through categorization of the continuous tumor size, information can be lost and they can be challenged by recently developed methods of modeling biomarkers in a longitudinal way. Thus, it is of interest to compare the predictive ability of cancer progressions based on categorical criteria and quantitative measures of tumor size (left-censored due to detection limit problems) and/or appearance of new lesions on

overall survival. We propose a joint model for a simultaneous analysis of three types of data: longitudinal marker, recurrent events and a terminal event. A simulation study is performed and shows that the proposed trivariate model is appropriate for the practical use. We suggest statistical tools that evaluate predictive accuracy for joint models to compare our model to models based on categorical criteria and their components. We apply the model to a randomized phase III clinical trial of metastatic cancers.

EO1369: **Modelling unbalanced hierarchical survival data using HAC-copula functions**
Presenter: **Roel Braekers**, Hasselt University, Belgium

A copula model is introduced for hierarchically nested clustered survival times in which the different clusters and sub-clusters are possibly unbalanced. Due to the right censoring we do not fully observe each outcome variable. This, together with the hierarchical structure of the data, makes it very difficult to set-up a full likelihood function for a general copula model. To circumvent this problem, we focus to on the class of hierarchical nested Archimedean copula functions and use the properties of this copula family to simplify the full likelihood function. For the marginal survival time, we introduce both a parametric regression model and a semi-parametric Cox's regression model. Since maximizing the likelihood function for all parameters is computational difficult, we consider a two-stage estimation procedure in which we first estimate the marginal parameters and afterwards, estimate the association parameters. As a result, we obtaine the asymptotic consistency and normality of the association parameters. Next we compare the finite sample behaviour of the different estimators through a simulation study. Furthermore we illustrate the estimators on a practical example on the insemination time of dairy cows.

EO0467: **Dynamic prediction according to tumour progression and genetic factors: Meta-analysis with a joint frailty-copula model**
Presenter: **Takeshi Emura**, National Central University, Taiwan
Co-authors: Virginie Rondeau, Masahiro Nakatochi, Sigeyuki Matsui, Hirofumi Michimae

The increasing availability of genomic information and large-scale meta-analytic dataset for clinicians has motivated the extension of the traditional survival prediction based on the Cox proportional hazards model. The aim is to develop a risk prediction scheme for death according to genetic factors and dynamic tumour progression status based on meta-analytic data. To this end, we extend the existing joint frailty-copula model to a model allowing for high-dimensional genetic factors. In addition, we propose a dynamic prediction scheme to predict death given tumour progression events possibly occurring after treatment or surgery. For clinical use, we implement the computation software of the prediction scheme in R joint.Cox package. We also develop a tool to validate the performance of the prediction scheme by assessing the prediction error with the Brier score. We illustrate the method with the meta-analysis of individual patient data on ovarian cancer patients.

EO163 Room 217 STATISTICAL MODELING FOR HIGH-DIMENSIONAL AND BIOMEDICAL DATA Chair: Mauricio Castro

EO0517: **Nonparametric Bayesian regression analysis of the Youden index**
Presenter: **Vanda Inacio**, University of Edinburgh, United Kingdom
Co-authors: Miguel de Carvalho, Adam Branscum

Accurately screening diseased from nondiseased individuals and correctly diagnosing disease stage are critically important to health care on several fronts, including guiding recommendations about combinations of treatments and their intensities. The accuracy of a continuous medical test or biomarker varies by the cutoff threshold (c) used to infer disease status. Accuracy can be measured by the probability of testing positive for diseased individuals (the true positive probability or sensitivity, Se(c), of the test) and the true negative probability (specificity, Sp(c)) of the test. A commonly used summary measure of test accuracy is the Youden index,$Y_I = \max\{Se(c)+Sp(c)-1 : c \in R\}$, which is popular due in part to its ease of interpretation and relevance to population health research. In addition, clinical practitioners benefit from having an estimate of the optimal cutoff that maximizes sensitivity plus specificity available as a byproduct of estimating Y_I. We develop a highly flexible nonparametric model to estimate Y_I and its associated optimal cutoff that can respond to unanticipated skewness, multimodality and other complexities because data distributions are modeled using dependent Dirichlet process mixtures. The value of our nonparametric regression model is illustrated using multiple simulation studies and data on the age-specific accuracy of glucose as a biomarker of diabetes.

EO0524: **Analysis of incomplete high-dimensional data via mixtures of common t-factor analyzers**
Presenter: **Wan-Lun Wang**, Feng Chia University, Taiwan

Mixtures of common t-factor analyzers (MCtFA) have emerged as a sound parsimonious model-based tool for robust modeling of high-dimensional data in the presence of fat-tailed noises and atypical observations. A generalization of MCtFA to accommodate missing values is presented, as they frequently occur in many scientific researches. Under a missing at random mechanism, a computationally efficient expectation conditional maximization either (ECME) algorithm is developed for parameter estimation. The techniques for visualization of the data, classification of new individuals, and imputation of missing values under an incomplete-data structure of MCtFA are also investigated. Illustrative examples concerning the analysis of real and simulated data sets are presented to describe the usefulness of the proposed methodology and compare the finite sample performance with its normal counterparts.

EO0563: **Skew-t factor analysis models with incomplete data**
Presenter: **Tsung-I Lin**, National Chung Hsing University, Taiwan

A novel framework is presented for maximum likelihood (ML) estimation in skew-t factor analysis (STFA) models in the presence of missing values (or nonresponses). As a robust extension of the ordinary factor analysis model, the STFA model assumes a restricted version of multivariate skew-t distribution for the latent factors and unobservable errors to accommodate non-normal features such as asymmetry and heavy tails or outliers. A computationally analytical EM-type algorithm is developed to carry out ML estimation and imputation of missing values under the missing at random mechanism. The practical utility of the proposed methodology is illustrated through both real and synthetic data examples.

EO1656: **Using covariate informed partition models to identify subpopulations via curve clustering**
Presenter: **Garritt Page**, Brigham Young University, United States

Studies are considered that measure functional output on a collection of experimental units or subjects. The main objective is to flexibly model individual curves while simultaneously assigning subjects to clusters based on curve shape such that the resulting partition is scientifically meaningful. Additionally, we seamlessly incorporate covariate information to the formation of curve clusters by employing a covariate informed partition model. Once clusters are connected with covariate information, we show how this information can be used to carry out less expensive/invasive diagnosis or prediction of a subjects future behavior.

EO133 Room 305 A DECONVOLUTION AND BOUNDARY ESTIMATION Chair: Leopold Simar

EO0626: **Nonparametric boundary regression**
Presenter: **Natalie Neumeyer**, University of Hamburg, Germany
Regression models with one-sided error distribution are considered. We discuss estimation of the boundary curve by a local polynomial method and obtain uniform rates of convergence. The rate depends on the regularity of the error distribution in its end point. We further consider sequential empirical processes of residuals. We show weak convergence to a Gaussian process and apply the results for model specification tests. In particular we discuss goodness-of-fit testing for the error distribution and testing for monotonicity of the boundary curve.

EO0611: **Inference for non-parametric stochastic frontier models**
Presenter: **Valentin Zelenyuk**, University of Queensland, Australia
Co-authors: Leopold Simar, Ingrid Van Keilegom
The focus is on statistical testing in the framework of non-parametric and semi-parametric stochastic frontier via local least squares approach. In particular, we are concerned with inference about existence of production inefficiency in general and at particular points of interest as well as inference about significance of factors potentially influencing the conditional mean of efficiency. To reach our goals, we consider and adapt a previous bootstrap-based approach, as well the generalized likelihood test.

EO0766: **Estimation of the boundary of a variable observed with symmetric error**
Presenter: **Jean-Pierre Florens**, Toulouse School of Economics, France
Let X^* be a positive random variable with support $[c,+\infty[$. We observe $X = X^* + \varepsilon$, where ε is a symmetric error term. The distribution of ε is otherwise unknown. Using an iid sample the objective is to estimate c. This analysis is in particular motivated by the non-parametric estimation of stochastic cost frontiers. Our method is inspired by recent work. Let us consider the characteristic function of X (denoted by $\psi_X(t)$), which may be easily estimated non-parametrically. The phase of $\psi_X(t)$ (i.e. the ratio of $\psi_X(t)$ by its modulus) does not depend on the distribution of ε, and we use its nonparametric estimator to derive an estimator of c.

EO0798: **A mollification approach to deconvolution problems**
Presenter: **Anne Vanhems**, University of Toulouse, TBS and TSE, France
Co-authors: Pierre Marechal
The objective is to propose an alternative approach to standard regularization methods for deconvolution problems. We consider the following equation: $Y = X + \varepsilon$ and we want to recover the density function of X from the observed random sample $(Y_1,...,Y_n)$. In this setting, we will assume that ε is normally distributed with known variance. This problem is well-known to be ill-posed. Its resolution has been addressed in many publications, mostly inspired by the Tikhonov approach. The main drawback of the latter approach is that the original equation is significantly perturbed, which leads to a difficult tradeoff: a strong regularization parameter induces a strong model perturbation; a weak regularization parameter yields a unstable solution. We propose an alternative regularization scheme, in which this tradeoff will become much less crucial. The corresponding methodology appeals to the notion of mollification.

EO075 Room 216 QUANTILE REGRESSION MODELS FOR DEPENDENT DATA Chair: Ghislaine Gayraud

EO0644: **Nonparametric extreme quantile estimation for spatial data**
Presenter: **Sophie Dabo**, University-Lille, France
Co-authors: Aladji Bassene, Aliou Diop, Baba Thiam
Spatial statistics includes any techniques which study phenomenons observed on spatial sets. Such phenomenons appear in a variety of fields: epidemiology, environmental science, econometrics, image processing and many others. Complex issues arise in spatial analysis, many of which are neither clearly defined nor completely resolved, and form the basis for current researches. This is the case for instance in statistics of extremes, where data are often spatial, and so spatial location can acts as a surrogate for risk factors. More recently, there has been increased interest in non-parametric statistical models for spatial extremes. We are interested in quantile estimation for heavy tailed models when data are available in space. More precisely, we consider a non-parametric conditional quantile estimate where the explanatory and response variables are real-valued random fields. The asymptotic distribution of the proposed estimator is established under some mixing condition. The skills of the methods are illustrated on simulations.

EO0654: **Nonparametric tests for conditional independence using conditional quantiles and distributions**
Presenter: **Taoufik Bouezmarni**, Universite de Sherbrooke, Canada
The concept of causality is naturally defined in terms of conditional distribution, however almost all the empirical work focuses on causality in mean. The aim is to propose a nonparametric statistic to test the conditional independence and Granger non-causality between two variables conditionally on another one. The test statistic is based on the comparison of conditional distribution functions using an L2 metric. We use Nadaraya-Watson method to estimate the conditional distribution functions. We establish the asymptotic size and power properties of the test statistic and we motivate the validity of the local bootstrap. The power of the proposed test is better than the competitors and it has the same power compared to them. Moreover, it is simple to implement. We ran a simulation experiment to investigate the finite sample properties of the test and we illustrate its practical relevance by examining the Granger non-causality between S& P 500 Index returns and VIX volatility index. Contrary to the conventional t-test based on a linear mean-regression model, we find that VIX index predicts excess returns both at short and long horizons.

EO0756: **Local linear quantile regression for functional data**
Presenter: **Ali Laksaci**, University of sidi bel abbes, Algeria
A new nonparametric estimator is introduced for the quantile regression function when the explanatory variable is functional and when observations are i.i.d. or dependent. The proposed estimate is based on a local linear approach. Asymptotic properties of the estimate, in terms of almost-complete convergence and asymptotic normality are stated. Moreover, the usefulness of the approach is illustrated through application to some financial data.

EO1179: **Confidence bands for a recurrent quantile as a function of the covariates in recurrent event models**
Presenter: **Akim Adekpedjou**, Missouri University of Science and Technology, United States
The main driver is the need to construct confidence bands for quantile as a function of covariates with recurrent event models. Consider the quantile of the time between occurrences of a unit subject to a recurring event and let X be the possibly time varying covariates associated with the unit. Of interest is the construction of various confidence bands wherein the covariates X and interventions performed after a recurrence are incorporated via a general Cox-type model for recurrent events. We propose three type of bands: (i) the first based on the original asymptotic properties of the properly standardized quantile which are intractable; (ii) the second based on the Khmaladze transformation of the distribution in (i) leading parameter free distribution and lastly (iii) the bootstrap based bands. We present various asymptotic properties of the properly standardized quantile. Small sample properties of the results are examined via a simulation study and illustration of the methods with a real recurrent event data

is presented.

EO441 Room 306 A COMPUTATIONAL METHODS IN THE DESIGN OF EXPERIMENTS Chair: Radoslav Harman

EO0681: **A first-order algorithm for the A-optimal experimental design problem: A mathematical programming approach**
Presenter: **Selin Ahipasaoglu**, Singapore Univ. of Tech. and design, Singapore
A first-order algorithm is developed and analysed for the A-optimal experimental design problem. The problem is first presented as a special case of a parametric family of optimal design problems for which duality results and optimality conditions are given. Then, two versions of a first-order (Frank-Wolfe type-) algorithm are presented, accompanied by a detailed time-complexity analysis and computational results on various sized problems.

EO0490: **Computation of optimal experimental designs in R**
Presenter: **Lenka Filova**, Comenius University in Bratislava, Slovakia
Co-authors: Radoslav Harman
Many algorithms have been proposed for the computation of efficient designs of experiments. However, the available R packages for implementing these algorithms are limited. We introduce an R package OptimalDesign which provides a toolbox for the computation of D- , A- and IV-efficient exact and approximate designs of experiments on finite domains, for regression models with real-valued, uncorrelated observations. The package incorporates several algorithms based on significantly different principles, including search heuristics, state-of-the-art mathematical programming methods, and a combination of convex optimization methods. The algorithms are illustrated on selected computation problems in the optimal design of mixture experiments.

EO1075: **Some relations between optimum on-the-average designs and maximin designs**
Presenter: **Hans Nyquist**, Stockholm University, Sweden
Construction of optimal experimental designs often requires knowledge of unknown parameter values. Two proposed approaches out of this dilemma are using optimum on-the-average designs, which use a weighted average of criterion functions, and maximin designs, which find the best guaranteed design as the parameters vary in a specified subset of the parameter space. Some relations between these two approaches are discussed and utilized to construct an algorithm for constructing maximin designs. Specifically, examples of maximin efficient designs are discussed.

EO0630: **Computational construction of Minimax designs in binary response and heteroscedastic simple linear regression models**
Presenter: **Jesus Lopez-Fidalgo**, University of Castilla-La Mancha, Spain
Co-authors: Victor Casero-Alonso
Binary response models are used in many real applications. For these models the Fisher Information Matrix (FIM) is proportional to the FIM of a weighted simple linear regression model. The same is also true when the weight function has a finite integral. Thus, optimal designs for one binary model are also optimal for the corresponding weighted linear regression model. The main objective is to provide a tool for the construction of MV-optimal designs, minimizing the maximum of the variances of the estimates, for a general design space. MV-optimality is a potentially difficult criterion because of its non-diferentiability at equal variance designs. A methodology for obtaining MV-optimal designs where the design space is a compact interval $[a,b]$ will be given for several standard weight functions. This will allow us to build a user-friendly computer tool based on Mathematica to compute MV-optimal designs. Users can know the type of the optimal design and the exact support points and design weights. Some illustrative examples will show a representation of MV-optimal designs in the Euclidean plane taking a and b as the axes. The applet provided will be explained using two relevant models for a weighted linear regression model and for a binary response model.

EO475 Room 205 RECENT DEVELOPMENTS IN ESTIMATION AND TESTING FOR GRAPHICAL MODELS Chair: Wicher Bergsma

EO0694: **Palindromic Ising models**
Presenter: **Giovanni Marchetti**, University of Florence, Italy
Palindromic Ising models are Ising models with zero main effects. They define a subclass of multivariate Bernoulli distributions called palindromic because the joint probabilities form a sequence that reads the same forward as backward. This family has a number of interesting properties: for instance they have a graph structured concentration matrix in graphs with separators of maximum size 2. Maximum likelihood estimation and practical applications are discussed.

EO1095: **Estimation and testing in relational models**
Presenter: **Tamas Rudas**, Eotvos Lorand University, Hungary
Co-authors: Anna Klimova
Relational models generalize log-linear models to arbitrary discrete sample spaces. If an overall effect is not present in any reparameterization of the model, the Poisson and multinomial MLEs are not equivalent. In the Poisson case, the observed total is not always preserved by the MLE, and thus the likelihood ratio statistic can become negative and is not bounded asymptotically in probability. A new asymptotic framework is developed and the Bregman statistic, derived from the corresponding information divergence, is introduced for goodness-of-fit testing. This statistic is asymptotically equivalent to the Pearson statistic, and they both are asymptotically chi-squared with the degrees of freedom determined by the model.

EO1114: **A strategy for selecting mixed graphs from distributions**
Presenter: **Kayvan Sadeghi**, University of Cambridge, United Kingdom
A PC-type algorithm is developed that provides a strategy to do the following: given a probability distribution P (or an underlying distribution of data), it selects all graphs of any type to which P is faithful assuming that there exist such graphs. We exclude AMP chain graphs, and, in particular, specialize the algorithm to DAGs and LWF chain graphs. The algorithm is based on certain necessary and sufficient conditions for faithfulness of probability distributions to graphs. We provide these conditions and discuss algorithmic approaches to test them.

EC1472: **Zoom-in/out joint graphical lasso for different coarseness scales**
Presenter: **Eugen Pircalabelu**, KU Leuven, Belgium
Graphical models are estimated from data obtained at K different coarseness scales. Starting from a predefined scale $k^* \leq K$ the method zooms in or out over scales on particular edges, thus estimating graphs with similar structures, but different levels of sparsity. The graphs are jointly estimated at all coarseness scales and evaluate the evolution of the graphs from the coarsest to the finest scale or vice-versa. We select an optimal coarseness scale to be used for further analyses. The method is motivated by fMRI datasets that do not all contain measurements on the same set of brain regions. For certain datasets some of the regions have been split in smaller subregions and this gives rise to the framework of mixed scale measurements where the purpose is to estimate sparse graphical models. We accomplish this by pooling information from all subjects in order to estimate a common undirected and directed graph at each coarseness scale, accounting for time dependencies and multiple coarseness scales and by jointly estimating the graphs at all coarseness scales. The applicability of the method goes beyond fMRI data, to other areas where data on

different scales are observed and where the joint estimation of graphs is desired.

EO031 Room 203 COMPLEX AND NEXT GENERATION FUNCTIONAL DATA ANALYSIS Chair: Wenceslao Gonzalez-Manteiga

EO0795: **A functional data approach to heterogeneous aging effects on functional connectivity**
Presenter: **Jeng-Min Chiou**, Academia Sinica, Taiwan
The normal aging effects on the cerebral cortex are investigated by characterizing changes in functional connectivity using resting-state fMRI data. We proposed a functional correlation coefficient for measuring functional connectivity, which takes into account dependency between different brain regions. The distinct patterns of changes in functional connectivity within and among the cerebral lobes clarified the effects of normal aging on cortical function. The results showed heterogeneous changes in functional connectivity in normal aging and supported the frontal aging hypothesis proposed in behavioral and structural MRI studies. The functional correlation analysis enables differentiation of changes in function connectivity and characterizes the heterogeneous aging effects in different cortical regions.

EO0242: **Modeling multi-way functional data with weak separability**
Presenter: **Kehui Chen**, University of Pittsburgh, United States
Multi-way functional data refers to an extension where double or multiple indices are involved, such as a sample of brain-imaging data with spatial and temporal indices. In practice, the number of spatial grids and the number of time grids both could be very large, and a multiplication of these two dimensions easily goes beyond the capacity of most data analysis tools. To achieve efficient dimension reduction, one usually adopts the separability assumption that the covariance can be factorized as a product of a spatial covariance and a temporal covariance. We will introduce a new concept of weak separability, and discuss several open questions in factorization methods using the notion of weak separability.

EO0725: **Fréchet integration and adaptive metric selection for covariance objects in functional data**
Presenter: **Alexander Petersen**, University of California Santa Barbara, United States
Co-authors: Hans-Georg Mueller
For multivariate functional data recorded for a sample of subjects on a common domain, one is often interested in the covariance between pairs of the component functions. We generalize the straightforward approach of integrating the pointwise covariance matrices over the functional time domain by defining the Fréchet integral, which, in analogy to the Fréchet mean, depends on the metric chosen for the space of covariance matrices. This generalization is motivated by the class of power transformations on covariance matrices and the associated metrics that are more suitable for this nonlinear space. Data-adaptive metric selection with respect to a user-specified target criterion, for example fastest decline of the eigenvalues, is proposed. Asymptotic results of the functional covariance and optimal metric estimators are presented, and their practical utility is illustrated in a comparison of connectivity between brain voxels for normal subjects and Alzheimer's patients based on fMRI data.

EO1140: **Depth-based nonparametric statistics for complex data**
Presenter: **Sara Lopez Pintado**, Columbia University, United States
Research in many disciplines stands on the analysis of complex data sets of signals and images. For example, in neuroscience large collections of brain images from different subjects are obtained by either functional magnetic resonance or positron emission tomography to study variations in different neurophysiological states or modifications during psychiatric disorders. Developing new statistical tools to analyze these rich data sets is needed. In the applications mentioned above the basic unit of observation can be considered as a general function which is defined in a subset of either the real line or a higher dimensional space with dimension d, taking values in a univariate or multivariate space with dimension p. We introduce a notion of depth defined for general functions with arbitrary d and p. Based on this depth, general functions will be ranked from center-outward and robust nonparametric tests for complex functions will be developed. In particular, we propose several permutation tests for comparing two groups of images. These statistical tools will be applied to detect whether there are differences in the brain structure or function between healthy individuals and patients with specific mental disorders. Also, recent developments in the quantitative analysis of imaging data have been used to construct brain network organization. We will introduce ways of extending the idea of depth to network data, where each observation in the data sample is a graph.

EO209 Room S21 ADVANCES IN LATENT VARIABLE MODELS Chair: Giuliano Galimberti

EO0957: **Latent variable models for complex networks: Flexible modelling and scalable inference**
Presenter: **Isabella Gollini**, Birkbeck, University of London, United Kingdom
Network data arise in a wide range of applications including social and biological sciences. In many cases different relations and/or a large amount of nodal information is available. Nodal attributes and links are often in strong relation. For example, nodes having similar features are more likely to be connected to each other and vice versa. We introduce a new framework of latent variable network models which combine the information given by heterogenous relational network structures in order to analyse, identify latent traits or groups, visualise network data and predict missing links and nodes. Network data are typically of large size and all the likelihood functions of the models proposed cannot be evaluated analytically. To overcome this problem, we adopt a variational approach to estimation which turns out to improve considerably the computational efficiency without a significant loss of accuracy with respect to other existing methods. The effectiveness of this methodology is demonstrated on the analysis of a wide variety of networks, from small to large networks. The analysis is carried out with the lvm4net package for R, available on CRAN.

EO0898: **Adoption of skew distributions in mixtures of factor models**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia
Co-authors: Sharon Lee
The aim is to consider the use of mixtures of normals with latent factors introduced in their component distributions to reduce the number of free parameters in the component covariance matrices in the modelling of high-dimensional heterogeneous data. Extensions of this model to incorporate various skew versions of the multivariate normal distribution for the latent component factors are discussed. The proposed models allow one to relax the assumption of normality for the latent factors in order to accommodate skewness in the observed data. These models provide an approach to model-based density estimation and clustering of high-dimensional data exhibiting asymmetric characteristics. In the literature there has been some confusion with these models as to whether the skewness is being incorporated into the distribution of the latent factors or into the error distribution in the factor-analytic representation of the component distributions. The potential of the proposed models are demonstrated using both real and simulated datasets.

EO1249: **On the use of the contaminated Gaussian distribution in hidden Markov models for longitudinal data**
Presenter: **Antonio Punzo**, University of Catania, Italy
Co-authors: Antonello Maruotti
A joint approach is introduced to time-varying clustering and bad points detection under a longitudinal setting, extending the standard Gaussian hidden Markov model. We replace the multivariate Gaussian state-dependent distribution with a two-component Gaussian mixture where one mixture (reference) component represents the data we would expect from the given state (i.e. good points) while the other mixture component

clusters the atypical data and has a small prior probability, the same component-specific mean and an inflated covariance matrix. This change makes the model much more robust. We estimate model parameters by using an ad hoc version of the expectation-conditional maximization (ECM) algorithm, extending the Baum-Welch iterative procedure to deal with contaminated Gaussian distributions. We illustrate the proposal by analyzing an artificial longitudinal dataset.

EO1322: **A sensitivity analysis approach to model fit evaluation in latent variable models**

Presenter: **Daniel Oberski**, Utrecht University, Netherlands

Latent variable models involve restrictions on the data that can be formulated in terms of "misspecifications": restrictions with a model-based meaning. Examples include zero cross-loadings in factor analysis and local dependencies in latent class analysis, as well as measurement invariance or differential item functioning in item response theory. If incorrect, misspecifications can potentially disturb the main purpose of the latent variable analysis - seriously so in some cases. A measure is discussed to evaluate whether misspecifications seriously affect the analysis at hand. This measure, the "EPC-interest", based on the likelihood of the restricted model and therefore does not require possibly computationally intensive fitting of alternative models. Findings on the EPC-interest are discussed, and its utility is demonstrated in an application to latent class analysis.

CO395 Room 005 DEPENDENCE MODELLING IN FINANCIAL ECONOMETRICS Chair: Jean-David Fermanian

CO0160: **Some tests of the simplified assumption for conditional copulae models**
Presenter: **Jean-David Fermanian**, Ensae-Crest, France
Co-authors: Alexis Derumigny
Under the so-called "simplified assumption" (SA), conditional copulae do not depend on their conditioning subsets. We propose several nonparametric and parametric tests of this commonly used assumption. The former ones are based on some comparisons between empirical counterparts under or without (SA), or are based on a particular independence property. The latter ones are obtained though usual distances between nonparametrically and parametrically estimated distributions or densities. The consistency and the power of such tests are evaluated by simulation, through some adapted bootstrap procedures.

CO0354: **Randomization tests of copula symmetry**
Presenter: **Brendan Beare**, University of California, San Diego, United States
Co-authors: Juwon Seo
New nonparametric randomization tests of copula exchangeability and radial symmetry are proposed. The tests are simple to compute, consistently detect arbitrary forms of asymmetry, and do not require the specification of a tuning parameter. Simulations indicate excellent small sample properties.

CO0553: **Testing hypotheses for the copula of dynamic models**
Presenter: **Bruno Remillard**, HEC Montreal, Canada
The asymptotic behaviour of the empirical copula constructed from residuals of dynamic models is studied, for univariate and multivariate models. For multivariate models, it is shown that the empirical copula process sometimes behaves like if the parameters were known, a remarkable property. This is not true for the so-called auto-copula in the univariate case. Applications for goodness-of-fit, detection of structural change and tests of randomness are discussed.

CO0887: **Dynamic vine copula models for multivariate time series data**
Presenter: **Elif Acar**, University of Manitoba, Canada
Co-authors: Claudia Czado
The representation of temporal patterns is essential to time series analysis. In the case of two or more time series, one needs to account for temporal patterns not only in each univariate series but also in their joint behaviour. We propose a multivariate model which enables the specification of time-dependent dynamic patterns in multivariate time series. The model is built by first addressing the serial dependence in each series and then modelling the interdependencies among their innovations using a time-varying vine copula model. To specify the vine decomposition, we employ a heuristic model selection tool that accounts for both the magnitude and variation of the empirical Kendall's tau across different time intervals. The time variation in the strength of pairwise dependencies is inferred using nonparametric smoothing techniques, and the uncertainty in the resulting estimates is assessed using a block bootstrap. The methods are used to analyze daily exchange rate data on seven major currencies.

CO1729: **Marking to market credit derivatives on simultaneous credit events**
Presenter: **Umberto Cherubini**, University of Bologna, Italy
Credit derivative contracts on simultaneous default risk of clusters of obligors are proposed in a model that can be easily estimated and validated on CDS market data. We provide pricing formulas and hedging strategies, allowing for both systemic risk and contagion. When the systemic shock is independent (pure systemic risk), we derive static hedges for the systemic credit derivatives using a credit index. The model also provides a contagion index. We apply the model to the most representative banks of 8 European countries, during the European sovereign crisis, and we find that our model provides a good representation in 4 cases.

CO533 Room 110 FINANCIAL ECONOMETRICS UNDER UNCERTAINTY Chair: Yan Sun

CO0305: **Nonparametric estimation for optimal dividend barrier with insurance portfolio**
Presenter: **Hiroshi Shiraishi**, Keio University, Japan
Co-authors: Zudi Lu
How an insurance portfolio is used to provide dividend income for insurance company's shareholders is an important problem in application of risk theory, where the premium income as dividends is paid to the shareholders, whenever the surplus attains a level barrier, until the next claim occurs. Under the aggregate claims process taken as a compound Poisson model, we define optimal dividend barrier as the level of the barrier that maximizes the expectation of the discounted dividends until ruin. In the literature, the optimal dividend barrier was derived explicitly under some fixed models concerning claim distribution where parametric estimation is possible. In practice, it may often be hard to provide with the claim distribution parametric model either only from theoretical point of view or from a finite sample, and thus non-parametric estimation is preferred. We consider the non-parametric estimation of the optimal dividend barrier. This study would contribute in practice to decision-making on dividend barrier in the case where a new product is launched or optimality of an existing dividend barrier is tested.

CO0399: **Traders structure and the process for speculative assets price**
Presenter: **Haibin Xie**, University of International Business and Economics, China
Heterogeneity is of great importance in determining assets price. With the assumption that financial markets are structured with optimistic traders and pessimistic traders, the aim is to derive analytically from equilibrium theory that process for speculative assets price is fully determined by traders structure and could be summarized by a difference of two additive gamma processes, termed AG. The AG model not only could reproduce all these well documented facts in empirical finance but also predicts that assets of high liquidity tend to be less volatile. The AG process implies that market sentiment is of great importance in asset pricing. Anyway, the AG process is a good alternative candidate for speculative assets price.

CO0564: **An information lag component in spread decomposed model**
Presenter: **Qiang Zhang**, Beijing Universtiy of Chemical Technology, China
One order Markov property of trade indicator variables as a key assumption in MRR model contradicts with information lag as an empirical characteristic in high frequency trading process. In this paper, a nonparametrical test is employed and the Markov property of trade indicator variables is rejected in most trading days. Based on the spread decomposed structure of MRR model, a moving average structure is adopted to absorb the information lag as an extension, then a ML estimator with ARCH structure is introduced. Empirical results show that the information lag parameter is significant and the adverse selection risk parameters estimated by the original and the extended, respectively, have significant

differences.Further, the analysis suggests that the information lag parameter could measure the average speed at which the information incorporates into the price.

CO0601: **Conditional heteroskedasticity of return range processes**
Presenter: **Yan Sun**, Utah State University, United States
Co-authors: Guanghua Lian, Zudi Lu, Jennifer Loveland, Isaac Blackhurst
It is well known that price range contains important information about the asset volatility. To utilize this information towards volatility modeling, we propose to view the [low, high] daily price range as an interval-valued time series and develop an extended GARCH model, called Int-GARCH, that allows for interval-valued input, for the corresponding return range process. The aim of Int-GARCH model is to enrich the classical GARCH model that is solely based on closing price, with price fluctuation information in the surrounding time interval. Theoretical properties of the Int-GARCH model are developed under the framework of random sets, and a metric-based least squares method is presented for estimating model parameters. From the empirical analysis with stocks and indices data, the Int-GARCH model consistently outperforms GARCH for both in-sample estimation and out-of-sample prediction.

CO1689: **Fuzzy multiplication and its applications in multi-period portfolio selection**
Presenter: **Xiang Li**, Beijing University of Chemical Technology, China
A fuzzy number is a normal and convex fuzzy subset of the real line. Bsed on membership function, we give the expression of the mean value of the multiplication of two positive fuzzy numbers. Furthermore, we propose the possibilistic simulation technique and prove its effectiveness. A fuzzy multi-period mean-variance- skewness portfolio selection model is formulated, which is transformed to a single-objective model based on different risk preferences. Finally, we present some numerical examples to demonstrate the effectiveness of the proposed models.

CO375 Room 112 BAYESIAN COPULA AND TIME SERIES MODELS Chair: Rodney Strachan

CO0308: **A new approach to identify noise shocks**
Presenter: **Eric Eisenstat**, The University of Queensland, Australia
Co-authors: Luca Benati, Joshua Chan, Gary Koop
The idea that news about future productivity can play an important role in business cycle fluctuations infuses much current macroeconomic research. This paper addresses the question of whether the news is noisy, so that agents cannot quickly disentangle genuine news from a noisy signal. A logical implication of agents inability to distinguish news and noise shocks on impact is that the immediate response of the economy to the two shocks will be the same. We provide illustrations of this general property within several macroeconomic models. We then exploit this restriction, together with the fact that whereas news shocks portend future variation in the variable of interest, noise shocks do not, in order to identify news, noise and surprise TFP shocks within a structural VARMA framework. Our experiments show that sizeable systems of 8 to 15 variables are needed to identify these shocks and we develop Bayesian methods to estimate large, structural VARMAs. In an empirical application, evidence suggests that TFP noise shocks play a minor role in macroeconomic fluctuations, explaining negligible fractions of the forecast error variance of the main macroeconomic variables.

CO0357: **Large Bayesian VARs: A flexible Kronecker error covariance structure**
Presenter: **Joshua Chan**, Australian National University, Australia
A class of large Bayesian vector autoregressions (BVARs) is introduced that allows for non-Gaussian, heteroscedastic and serially dependent innovations. To make estimation computationally tractable, we exploit a certain Kronecker structure of the likelihood implied by this class of models. We propose a unified approach for estimating these models using Markov chain Monte Carlo (MCMC) methods. In an application that involves 20 macroeconomic variables, we find that these BVARs with more flexible covariance structures outperform the standard variant with independent, homoscedastic Gaussian innovations in both in-sample model-fit and out-of-sample forecast performance.

CO0366: **Computation of continuous, piecewise linear reaction functions**
Presenter: **Rodney Strachan**, The University of Queensland, Australia
Co-authors: Eric Eisenstat, Fabrizio Carmignani, Rabee Tourky
There are good economic justifications why reaction functions be linear. We extend earlier work that developed methods for efficiently estimating continuous piecewise linear functions based on a Bayesian approach to clustering. The mean of such a function will not be linear, and so not have useful economic meaning. Building upon work extending MCMC based on concepts developed in the simulated annealing literature, we obtain an estimate of the mode with measures of uncertainty based upon Frobenius norm. We apply the Reisz estimator to generate candidate paths for the CPLM. We use the approach to investigate the most likely form of the relationship between output growth and policy variables based upon fiscal and monetary interventions. We find that, in normal times, the response of output to fiscal and monetary stimuli are as expected. However, in a low growth environment, the response of growth to policy actions is very subdued.

CO1331: **Time series copulas for heteroskedastic data**
Presenter: **Ruben Loaiza Maya**, University of Melbourne, Australia
Co-authors: Michael Smith, Worapree Ole Maneesoonthorn
New parametric copulas are proposed that capture serial dependence in stationary heteroskedastic time series. We develop our copula for first order Markov series, and then extend it to higher orders and multivariate series. We derive the copula of a volatility proxy, based on which we propose new measures of volatility dependence, including co-movement and spillover in multivariate series. Using daily exchange rate returns, we show that the copula models can both capture their marginal distributions more accurately than univariate and multivariate GARCH models, as well as produce more accurate value at risk forecasts.

CO1332: **Inversion copulas from nonlinear state space models**
Presenter: **Worapree Ole Maneesoonthorn**, University of Melbourne, Australia
Co-authors: Michael Smith
While copulas constructed from inverting latent elliptical, or skew-elliptical, distributions are popular, they can be inadequate models of serial dependence in time series. As an alternative, we propose an approach to construct copulas from the inversion of latent nonlinear state space models. This allows for new time series copula models that have the same serial dependence structure as a state space model, yet have an arbitrary marginal distribution–something that is difficult to achieve using other time series models. We examine the time series properties of the copula models, outline measures of serial dependence, and show how to use likelihood-based methods to estimate the models. To illustrate the breadth of new copulas that can be constructed using our approach, we consider three example latent state space models: a stochastic volatility model with an unobserved component, a Markov switching autoregression, and a Gaussian linear unobserved component model. We use all three inversion copulas to model and forecast quarterly U.S. inflation data. We show how combining the serial dependence structure of the state space models, with flexible asymmetric and heavy-tailed margins, improves the accuracy of the fit and density forecasts in every case.

CO695 Room 003 BAYESIAN METHODS IN MACROECONOMICS AND FINANCE Chair: Nalan Basturk

CO0315: **Advances in Bayesian computations with application to stochastic volatility models**
Presenter: **Yuliya Shapovalova**, Maastricht University, Netherlands

Stochastic volatility models (SVM) belong to the class of state-space models. Volatility in this framework is an independent latent process as opposed to GARCH-type models. The presence of the latent structure makes the estimation challenging. It has been documented in the literature that standard techniques such as Quasi-Maximum Likelihood and General Method of Moments do not perform well. Particle filters have been considered as the most promising solution for quite long time, intuitively clear implementation and good performance of simulated likelihood makes this procedure appealing. However, it is computationally very demanding. For practitioners speed of estimation might be crucial and hence, methods that might work faster are of interest. We consider a few other methods in Bayesian framework, namely Approximate Bayesian Computation (ABC), Variational Bayes, Expectation Propagation algorithm and particle MCMC. We discuss advantages and disadvantages of these methods in terms of the quality of estimation and computational speed in univariate stochastic volatility models. Further, we discuss possibilities for multivariate extensions.

CO0343: **A jumping index of jumping stocks**
Presenter: **Alessandro Pollastri**, Maastricht University, Netherlands
Co-authors: Paulo Rodrigues, Norman J Seeger, Christian Schlag

Continuous-time models for the S&P 100 index and its constituents are examined. First, we find that the stylized facts found in the index literature do not carry over to single stocks. Second, parameter estimates for the stochastic processes for single stocks imply pronounced heterogeneity in the cross-section. Third, we find that a jump in the index is not necessarily accompanied by a large number of contemporaneous jumps in its constituents stocks. Consequently, fourth, index jumps can be classified as induced by either a strongly increasing correlation between the returns on individual stocks or by macroeconomic events.

CO0347: **Fundamental shock selection in DSGE models**
Presenter: **Stefano Grassi**, University of Kent, United Kingdom
Co-authors: Filippo Ferroni, Stefano Grassi, Miguel Leon Ledesma

DSGE models are typically estimated assuming the existence of certain structural or fundamental shocks that drive macroeconomic fluctuations. We analyze the consequences of introducing shocks that are non-fundamental for the estimation of DSGE model parameters. We then propose a method to select the structural shocks driving macroeconomic uncertainty. We show that forcing the existence of non-fundamental structural shocks produces a downward bias in the estimated internal persistence of the model. We then show how these distortions can be reduced by allowing the covariance matrix of the structural shocks to be rank deficient using priors for standard deviations whose support includes zero. The method allows us to accurately select fundamental shocks and estimate model parameters with precision. Finally, we revisit the empirical evidence on an industry standard medium-scale DSGE model model and find that government, price, and wage markup shocks are non-fundamental.

CO1069: **Extensions in time varying parameter shrinkage models**
Presenter: **Angela Bitto**, WU Wien, Austria

An extension will be presented of our previous work, in which we presented an efficient and novel way to induce shrinkage in state space models. We will show that forecasting results are better than usual shrinkage methods and present various univariate and multivariate applications.

CC1558: **Bayesian risk evaluation for long horizons**
Presenter: **Agnieszka Borowska**, Vrije Universiteit Amsterdam, Netherlands
Co-authors: Lennart Hoogerheide, Siem Jan Koopman

An accurate and efficient approach to Bayesian estimation of two financial risk measures, Value-at-Risk and Expected Shortfall, for a given volatility model, is presented. Precise forecasts of the tail of the distribution of returns are obtained not only for the 10-days-ahead horizon required by the Basel Committee but even for long horizons, like one-month or one-year ahead. The latter has recently attracted a considerable attention due to a different character of the short term risk and the long run one. Long horizon forecasts can also be useful, e.g. for option pricing. The key insight behind the proposed importance sampling based approach is the construction of the importance densities as mixtures of Student's t distributions sequentially. By oversampling the extremely negative scenarios and punishing them by lower importance weights, a much higher precision in characterising the properties of the left tail is achieved. Empirical studies of GARCH(1,1)-t and GAS(1,1)-t models for daily financial data show substantial accuracy gains for all the considered horizons. To illustrate the flexibility of the proposed construction method, an adjustment to the frequentist case is provided.

CO349 Room 111 TOPICS IN FINANCIAL ECONOMETRICS Chair: Leopold Soegner

CO0369: **Continuous-time regime switching models with jumps and filter-based volatility**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany
Co-authors: Elisabeth Leoff, Vikram Krishnamurthy, Elisabeth Leoff

In finance, a continuous time regime switching model, where the observation process is a diffusion process whose drift and volatility coefficients jump governed by a continuous time Markov chain, can explain some of the stylized facts of asset returns. 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 switches 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. This volatility model can be motivated by social learning arguments. 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. Further, we extend these results to a jump-diffusion model where also the intensity of the driving Poisson process is switching. This allows for jumps in the observation.

CO0904: **Exchange rate forecasting and the performance of currency portfolios**
Presenter: **Ines Fortin**, Institute for Advanced Studies, Austria
Co-authors: Jaroslava Hlouskova, Jesus Crespo Cuaresma

The potential gains of using exchange rate forecast models and forecast combination methods in the management of currency portfolios are examined for three exchange rates, the euro (EUR) versus the US dollar (USD), the British pound (GBP) and the Japanese yen (JPY). We use a battery of econometric specifications to evaluate whether optimal currency portfolios implied by trading strategies based on exchange rate forecasts out-perform single-currency and the equally weighted portfolio. We assess the differences in profitability of optimal currency portfolios for different types of investor preferences, different trading strategies, different composite forecasts and different forecast horizons. Our results indicate that the

benefits of integrating exchange rate forecasts from state-of-the-art econometric models in currency portfolios are sensitive to the trading strategy under consideration and vary strongly across prediction horizons.

CO0765: **Properties of estimators of the quantization dimension of distributions**
Presenter: **Klaus Poetzelberger**, WU Vienna, Austria
Estimators of the dimension of the support of a probability distribution are presented. These estimators are derived from the concept of quantization dimension. For the general case consistency results are discussed. Versions of the estimators may be applied for instance to estimate the dimension of the driving Brownian motion of Ito processes or the dimension of the attractor of a dynamical system. A second application of estimators of dimension is the analysis of high-dimensional data X where the stochastic properties of X are explained by a vector of factors W in the sense that for a sufficiently smooth (Lipschitz) mapping f, $X = f(W)$ and $s := \dim(W) << \dim(X) =: d$. We focus on regularity conditions that imply the consistency of the estimators, on numerical experiments to check the performance of the estimators and on the choice of the norm for the quantization.

CO1263: **Pricing Parisian option with adaptive Monte Carlo method**
Presenter: **Sercan Gur**, Vienna University of Economics and Business, Austria
Co-authors: Klaus Poetzelberger
Parisian option is a type of barrier option, which can only be exercised if the underlying value process not only reaches a barrier level but remains a certain prescribed time (so-called window period) below (or above) this level. Closed form solutions for the value of these contracts do not exist. In order to price Parisian options, we use Monte Carlo simulation instead of partial differential equations, inverse Laplace transform or lattices. We propose a new Monte Carlo method which can be used to price Parisian options not only with constant boundary but with more general boundary. The advantage of this approach is that it can easily be adapted to compute the price of an option with more complicated path-dependent pay-off. We use adaptive control variable to improve the efficiency of the Monte Carlo estimator. At last, we provide a numerical example to illustrate our method and a comparison of previous Monte Carlo methods with our technique.

CC0926: **Cross-sectional multivariate ordinal mixed regression models with an application in credit risk**
Presenter: **Rainer Hirk**, Vienna University of Economics and Business, Austria
Co-authors: Laura Vana
Two different approaches are investigated and extended for modelling multivariate categorical data. Both approaches augment cumulative link mixed models to a multivariate framework where several ordinal response variables are modelled jointly. The first approach uses the correlation structure of the error terms in order to capture the correlation among the different response variables. Composite maximum likelihood techniques in combination with a Metropolis-Hastings algorithm for optional random effects are used as estimation procedures. The second approach constitutes of a model with multivariate random effects and uncorrelated error terms where different estimation procedures using Laplace approximation, Gauss-Hermite quadrature, and EM algorithms are applied. All these procedures are implemented and compared in a simulation study. In addition, the models are fitted to real data with an application in credit risk.

CO399 Room 205 ADVANCES IN BAYESIAN REGRESSION MODELING Chair: Richard Hahn

CO0436: **Dependence priors for Bayesian regularized regression**
Presenter: **Christopher Hans**, The Ohio State University, United States
Many regularization priors for Bayesian regression assume the regression coefficients are a priori independent. In particular this is the case for Bayesian treatments of the lasso and the elastic net. While independence may be reasonable in some data-analytic settings, having the ability to incorporate dependence in these prior distributions would allow for greater modeling flexibility. The orthant normal distribution is introduced in a general form and it is shown how it can be used to structure prior dependence in Bayesian regression models that have connections to penalized optimization procedures. An L1-regularized version of Zellner's g prior is introduced as a special case, creating a new link between the literature on penalized optimization and an important class of regression priors.

CO0442: **A bivariate treed linear model for causal inference from observational studies**
Presenter: **Richard Hahn**, University of Chicago, United States
Co-authors: Carlos Carvalho
The aim is to explain how regularized regressions (linear and nonlinear) can yield poor estimators of treatment effects due to a phenomenon we call "regularization induced confounding". How to overcome this problem will be explained, starting first with the linear model and then extending the approach to the case of nonlinear, heterogeneous treatment effects using treed linear models. We show that our approach dramatically outperforms common alternatives across a range of plausible data generating processes.

CO0447: **Variable selection in non-linear regression models: A parsimony-utility approach**
Presenter: **Carlos Carvalho**, The University of Texas at Austin, United States
Co-authors: Richard Hahn, Robert McCulloch
A procedure is described for principled and pragmatic Bayesian variable selection for nonlinear mean regression models. The method is motivated from a decision theoretic perspective and uses posterior samples to gauge the adequacy of submodels relative to the unknown true model. It is shown that computationally efficient surrogate prediction models can be used to prescreen submodels, greatly easing the variable search. The procedure is various data sets from the applied regression literature.

CO0476: **Exact Bayesian variable selection and averaging for block-diagonal designs**
Presenter: **David Rossell**, University of Warwick, United Kingdom
Co-authors: Omiros Papaspiliopoulos
Variable selection is considered when the gram matrix $X'X$ is block-orthogonal, e.g. as in principal component regression, wavelet regression or certain structures with interaction terms. Conditional on the residual variance ϕ most posterior quantities of interest have closed-form, but integrating out ϕ to duly account for uncertainty has proven challenging as in principle it requires a sum over 2^p models, and led to a number of adhoc solutions in the literature. We solved this bottleneck with a fast expression to integrate phi exactly (e.g. $O(p)$ operations when $X'X$ is diagonal), avoiding MCMC or other costly iterative schemes. Coupled with an efficient model search and other tricks the framework delivers extremely exact computation for large p, as we show in our examples. It is hoped that the computational framework can serve as a basis for efficient approximations under general $X'X$.

CO0632: **Particle EM for variable selection**
Presenter: **Veronika Rockova**, University of Chicago, United States
Recent years have seen growing interest in deterministic search approaches to spike-and-slab Bayesian variable selection. Such methods have focused on the goal of finding a global mode to identify the "best model". However, the report of a single model will be a misleading reflection of the model uncertainty inherent in a highly multimodal posterior. Motivated by non-parametric variational Bayes strategies, we move beyond this

limitation by proposing an ensemble optimization approach to identify a collection of representative posterior modes. Our approach, called the Particle EM algorithm, performs deterministic posterior exploration using an ensemble of repulsive particles. These particles are geared towards uncharted areas of the posterior, providing a more comprehensive summary of posterior topography than parallel algorithms. An MCMC variant of Particle EM is also presented that explores the posterior by sampling from a set of mutually avoiding particles. Our theoretical insights indicate that the requisite number of particles need not be large in the presence of sparsity.

CO317 Room 002 THEORY OF MAJORIZATION IN STATISTICS, ECONOMICS AND ENGINEERING Chair: Artem Prokhorov

CO0439: **Schur properties of convolutions of gamma random variables with applications**
Presenter: **Fred Roosta**, UC Berkeley, United States
Co-authors: Gabor Szekely
Sufficient conditions are discussed for comparing the convolutions of heterogeneous gamma random variables in terms of the usual stochastic order. Such comparisons are characterized by the Schur convexity properties of the cumulative distribution function of the convolutions. We also present some examples of the practical applications of our results.

CO0669: **Schur-convexity of outage capacity for fading wireless channels**
Presenter: **Eduard Jorswieck**, TU Dresden, Germany
The outage or ε-capacity of fading wireless channels describes the data rate which can be reliably transmitted in a slow-fading channel with success probability of at least $1-\varepsilon$ if the receiver has perfect channel state information (CSI) while the transmitter has imperfect CSI. In fading channels with multiple degrees of freedom, such as from multiple-antenna or multi-carrier systems, the resulting diversity effect has a different impact on the outage capacity depending on the operating SNR and rate point. For small SNR and high transmission rates, the outage probability is Schur-concave with respect to the diversity weights while for high SNR and low transmission rates, the outage probability becomes Schur-convex. This enables the wireless system design by choosing a proper antenna layout and an optimal number of transmit and receive dimensions. Interestingly, the framework to show order preserving results for majorization can be extended from Rayleigh fading channels to other relevant fading distributions. Furthermore, there exists an intriguing relationship between outage capacity in wireless communications and the level-q Value-at-Risk for risk analysis in economics and finance. Adding antennas or carriers in wireless corresponds to diversification in portfolio allocation.

CO0886: **Cybersecurity risk: A majorization approach**
Presenter: **Maochao Xu**, Illinios State U, United States
Co-authors: Maochao Xu
Cybersecurity has become a problem that is threatening the economy, human privacy, and even national security. However, the contributions to study the cybersecurity risk in the literature are largely descriptive, which is mainly because the cybersecurity risk is very different from the traditional risks covered by indemnity insurance. The significant property that distinguishes Cybersecurity risk from the conventional risk is that information and communication technology resources are interconnected in a network, and therefore the analysis of risk and its related potential losses needs to take into the network topology. We discuss the applications of majorization theory in the cybersecurity risk. Particularly, we study how the majorization theory could be used to address the heavy-tail cyber risk and cyber epidemicspreading over complex networks. A Monte Carlo simulation will be presented to support the theoretical results as well.

CO0314: **Dispersion inference in small samples**
Presenter: **Paul Kattuman**, University of Cambridge, United Kingdom
Co-authors: Jun Ma
The aim is to present a family of measures of dispersion in the general sense that encompasses inequality, concentration, heterogeneity, diversity and so on for non-negative small samples. Dispersion in a sample is measured as the probability of majorization (Lorenz dominance) relations between it and a random sample drawn from a suitable symmetric multi-variate reference distribution chosen to serve as a benchmark. One example of a reference distribution is the uniform distribution on the standard n less one simplex, according to which all samples of size n are equi-probable. The probabilities of majorization so defined are the $p-$values for the hypothesis that the sample of interest and a random sample from the reference distribution are of equal dispersiveness. Unlike other summary indices of dispersion, these probability measures directly enable inference, and satisfy properties desirable in any general measure of dispersion. The choice of reference distribution in this approach permits inference on the heaviness of tails.

CO1031: **Majorization theory and optimal bundling strategies for complements and substitutes with heavy-tailed valuations**
Presenter: **Rustam Ibragimov**, Imperial College London, United Kingdom
Co-authors: Artem Prokhorov, Johan Walden
Using majorization theory and extending earlier works in the field, we develop a framework that allows one to model the optimal bundling problem of a multiproduct monopolist providing interrelated goods with an arbitrary degree of complementarity or substitutability. Characterizations of optimal bundling strategies are derived for the seller in the case of heavy-tailed valuations and tastes for the products. We show, in particular, that if goods provided in a Vickrey auction or any other revenue equivalent auction are substitutes and bidders' tastes for the objects are moderately heavy-tailed, then the monopolist prefers separate provision of the products. However, if the goods are complements and consumers' tastes are extremely thick-tailed, then the seller prefers providing the products on a single auction. We also present results on consumers' preferences over bundled auctions for complements and substituties in the case when their valuations exhibit heavy-tailedness. In addition, we obtain characterizations of optimal bundling strategies for a monopolist who provides complements or substitutes for profit-maximizing prices to buyers with heavy-tailed tastes.

CO644 Room 106 EMPIRICAL ASSET PRICING AND HIGH-FREQUENCY FINANCIAL ECONOMETRICS Chair: Roberto Reno

CO1106: **Empirical evaluation of overspecified asset pricing models**
Presenter: **Francisco Penaranda**, Queens College CUNY, United States
Co-authors: Enrique Sentana, Elena Manresa
Empirical asset pricing models with potentially too many risk factors are increasingly common. Unfortunately, they can yield misleading statistical inferences. Unlike other studies focusing on the properties of standard estimators and tests, we explicitly characterize the linear subspace of risk prices compatible with a given model pricing restrictions. We also propose tests to detect problematic cases such as economically meaningless SDFs uncorrelated to the chosen test assets. We conduct simulation exercises to assess the finite sample size and power of our tests. We confirm the empirical relevance of our methods by revisiting a linearized version of the consumption CAPM.

CO1059: **Estimating risk premia using large cross-sections**
Presenter: **Valentina Raponi**, Imperial College London, United Kingdom
A limiting theory is presented for estimating and testing linear asset-pricing models when a large number of assets, N, is available, together with a fixed, possibly small, time-series dimension, T. Since the ordinary least squares (OLS) estimator is biased and inconsistent in this case, we focus on an alternative estimator, which we show to exhibit many desirable properties. We formally prove its consistency and derive its asymptotic distribution, showing how its limiting variance can be consistently estimated. We also propose a new test of the no-arbitrage asset pricing restriction, and establish its asymptotic distribution (assuming that the restriction holds). Finally, we show how our results can be extended to deal with the more realistic case of unbalanced panels. The practical relevance of our findings is demonstrated using Monte Carlo simulations and an empirical application to asset-pricing models with traded risk factors. Our analysis suggests that the market, size, and value factors are often priced in the cross-section of NYSE-AMEX-NASDAQ individual stock returns over short time spans.

CO1282: **Systematic flatness**
Presenter: **Davide Pirino**, Universiy of Rome Tor Vergata, Italy
Co-authors: Roberto Reno, Federico Bandi
Asset prices are stale. We define a measure of systematic flatness in the price changes as a functional of the percentage of jointly and individual small price adjustments for multiple assets. Notions of idiosyncratic flatness are also established. For both systematic and idiosyncratic flatness we provide a limit theory which makes use of increasingly frequent observations over a fixed span of time. We justify structurally the proposed measures by employing a market microstructure model with asymmetries in information and various sources of liquidity, including funding liquidity.

CO1310: **Systemic cojumps**
Presenter: **Roberto Reno**, University of Verona, Italy
Co-authors: Aleksey Kolokolov, Massimiliano Caporin
The simultaneous occurrence of jumps in several stocks can be associated with major financial news, triggers short-term predictability in stock returns, is correlated with sudden spikes of the variance risk premium, and determines a persistent increase (decrease) of stock variances and correlations when they come along with bad (good) news. These systemic events and their implications can be easily overlooked by traditional univariate jump statistics applied to stock indices. They are instead revealed in a clearly cut way by using a novel test procedure applied to individual assets, which is particularly effective on high-volume stocks.

CO1344: **Jump risk and pricing implications**
Presenter: **Nancy Zambon**, University of Padova, Italy
Co-authors: Massimiliano Caporin, Walter Distaso
A new common risk factor in stock returns related to the fear of future jumps is identified. The factor can be added to standard asset-pricing models leading to a five-factor model which is directed at capturing the size, value, profitability, momentum and fear in stock returns. The model outperforms a previous four-factor model.

CO345 Room 008 REGIME CHANGE MODELING IN ECONOMICS AND FINANCE I Chair: Willi Semmler

CO1691: **Mind the output gap: The disconnect of growth and inflation during recessions & convex Phillips curves in the Euro area**
Presenter: **Marco Gross**, European Central Bank, Germany
Co-authors: Willi Semmler
A theoretical model is developed that features a business cycle-dependent relation between output, price inflation and inflation expectations, augmenting a previous model with a nonlinear Phillips curve that reflects the rationale underlying the capacity constraint theory. The theoretical model motivates our empirical assessment for the euro area, based on a regime-switching Phillips curve and a regime-switching monetary structural VAR, employing different filter-based, semi-structural model-based and Bayesian factor model-implied output gaps. The analysis confirms the presence of a pronounced convex relationship between inflation and the output gap, meaning that the coefficient in the Phillips curve on the output gap recurringly increases during times of expansion and abates during recessions. The regime switching VAR reveals the business cycle dependence of macroeconomic responses to monetary policy shocks: Expansionary policy induces less pressure on inflation at times of weak as opposed to strong growth; thereby rationalizing relatively stronger expansionary policy, including unconventional volume-based policy such as the Expanded Asset Purchase Programme (EAPP) of the ECB, during times of deep recession.

CO1679: **Inflation targeting, credit flows and financial stability in a regime change model**
Presenter: **Willi Semmler**, New School for Social Research, United States
Co-authors: Marco Gross
Several papers point to the problem that inflation targeting models do not as of yet consider financial market stability that can considerably derail inflation targeting monetary policy. Credit flows and the instability of credit appear to be at the root of the financial instability problem. It has been however recently questioned whether a too early and too strong leaning against the wind policy by central banks might have higher cost than benefits in terms of output and employment loss. We include in an inflation targeting model a financial stabilization goal. We solve the model with NMPC. We estimate a regime-switching structural VAR for the euro area. Apart from standard macroeconomic variables, such as output gap, inflation, interest rates we also include loan volume, to study credit cost as well as credit volume effects. We estimate monetary policy shocks (identifying them via sign restrictions), under different regime assumptions to reveal the state-dependent effects of both interest rate and volume-based policies. We simulate and study regime dependent VARs and IRs for conventional and unconventional policies as well as policies affecting supply and demand of loans.

CO1583: **CISS in a time-varying environment: How frequent are systemic distress?**
Presenter: **Eddie Gerba**, Bank of Spain and London School of Economics (LSE), Spain
Co-authors: Manfred Kremer
The CISS indicator is a composite indicator of systemic stress in an economy, which is highly time-varying. We incorporate this index in multiple macro-financial time-varying parameter VAR models of the Euro area to evaluate its performance and measure its predictability of systemic events. Only during times of financial distress, the indicator increases. Moreover during systemic events such as the post-2007 crises both the indicator as well as the uncertainty regarding the future path of key macro-financial variables increases in our model. This increase in the indicator precedes the increase in volatility of other variables in the model by 1-2 quarters, making it a good forecaster of crises events. Furthermore, systemic shocks (or shocks to the CISS indicator) have a stronger impact on the rest of the economy than any of the other shocks examined. Lastly we compare our model to a Markov-switching framework (ceteris paribus) and find that while the number of regime switches are fewer, increases in the mean and volatility of coefficients associated with each crisis event is higher in our current model.

CO1453: **Real-time vs ex-post monetary policy evaluation under opportunistic policy**
Presenter: **Jesus Vazquez**, The University of the Basque Country, Spain
Co-authors: Steven P Cassou
Striking differences in impulse response results are shown when using revised and real time data under opportunistic threshold structures. Notably, the impulse responses using revised data show almost no threshold behavior differences between the two sides of an opportunistic threshold structure while real time data show significant differences between the two sides. This result is important to recognize since policy makers do not have expost revised data during their decision deliberations. As a result, economists using revised data may misread policy making behavior that can only be revealed using real time data.

CO1416: **The effects of labour market reforms upon unemployment and income inequalities: an Agent Based Model**
Presenter: **Maria Enrica Virgillito**, Scuola Superiore Sant
Ánna, Italy
Co-authors: Giovanni Dosi, Andrea Roventini, Marcelo Pereira
The aim is to analyse the effects of labour market structural reforms by meansof an agent-based model. Building on a previous work we introduce a policy regimechange characterized by a set of structural reforms on the labour market, keeping constant the structure of the capital- and consumption-good markets. Confirming a recent IMF report, the model shows how labour market structural reforms reducing workers bargaining power and compressing wages tend to increase (i) unemployment, (ii) functional income inequality, and (iii) personal income inequality. We further undertake a global sensitivity analysis on key variables and parameters which confirms the robustness of our findings.

CO277 Room 107 SEASONALITY Chair: Gian Luigi Mazzi

CO0827: **Forecasting evaluation in JDemetra+**
Presenter: **David de Antonio Liedo**, National Bank of Belgium, Belgium
The new forecasting evaluation framework of the software JDemetra+ (JD+) is described, and its usefulness is proved for decision making and diagnosis. Our algorithms are inspired by recent ideas to apply small sample asymptotics to correct the well-known size distortions of popular test, e.g. Diebold-Mariano. Our example is motivated by the ESS Guidelines on Seasonal Adjustment, which suggest updating models, filters, outliers and regression parameters (henceforth, the specification) at regular time intervals. Thus, users need to fix on their own the precise updating policies because currently there are not data-driven rules in JDemetra+ to decide how to do it or how frequently. We propose calculating out-of-sample forecasts over the last two years using recursive parameter estimation. If those forecasts improve those that would have resulted using the new specification automatically proposed by JD+ we would suggest discarding the update. Using thousands of series for the US and the EU, we will provide an experimental overview of the number of times our decision rule would advise us to update the specification over the last five years, and what the gains would be. They will be given by the decrease in the root mean squared forecast error (RMSE). We will consider both the errors at forecasting unadjusted data and revision errors over a given time interval. The last concept requires approximating the true adjusted data for the whole sample averaging all accepted methods.

CO0916: **A class of periodic trend models for economic time series**
Presenter: **Gian Luigi Mazzi**, Eurostat, Luxembourg
Co-authors: Tommaso Proietti, Martyna Marczak
Trend and seasonality are the most prominent features of economic time series that are observed at the sub-annual frequency. Modelling these components serves a variety of analytical purposes, including seasonal adjustment and forecasting. We introduce unobserved components models for which both the trend and seasonal components arise from systematically sampling a multivariate transition equation, according to which each season evolves as a random walk with a drift. By modelling the disturbance covariance matrix we can encompass traditional models for seasonal time series, like the basic structural model, and can formulate more elaborate ones, dealing with time domain season specific features, such as seasonal heterogeneity and correlation, along with frequency domain features, related to the different role played like by the non-stationary cycles defined at the fundamental and the harmonic frequencies, in determining the shape of the seasonal pattern.

CO0764: **Model-based seasonal adjustment in case of seasonal heteroskedasticity**
Presenter: **Jean Palate**, National Bank of Belgium, Belgium
In many time series, the seasonal pattern presents specific characteristics following the period of the year. Typical examples are the higher variability of activity in winter or rapid changes in the production behaviour during summer. In the context of seasonal adjustment, a simple solution to deal with such seasonal heteroskedasticity consists in using the non-parametric X-11 method with specific filters for the different periods. However, decompositions based on periodic models offer a more powerful approach. Several model-based solutions are reviewed, with different strategies for identifying automatically parsimonious models. The methods were applied on large data sets and the outcomes of that large-scale study are discussed. Finally, some characteristics of the model-based approach are stressed through specific cases. All the estimations were done by means of JDemetra+, which is a free open source software.

CO0968: **A X11-based seasonal adjustment method for series with multiple periodicities**
Presenter: **Dominique Ladiray**, INSEE, France
The increasing availability of more and higher frequency data has opened new challenging ways of producing official statistics. More specifically, the question has arisen how established statistical techniques, such as seasonal adjustment, can be applied to large datasets of high frequency data and data with high periodicity (infra-monthly and infra-weekly periodicities). Even if the X-13ARIMA-SEATS method focuses on monthly and quarterly series, we show how the X11-algorithm can be generalized to multiple seasonalities. We also provide alternatives to the Reg-ARIMA feature used by X-13ARIMA-SEATS to automatically detect outliers and calendar effects and to forecast the series. Several examples are presented to illustrate the methodology using daily demography series (businesses and persons) and electricity demand.

CO0385: **Seasonal and calendar adjustment of daily time series**
Presenter: **Daniel Ollech**, Deutsche Bundesbank, Germany
The methods for seasonal adjustment of official data explicitly supported by Eurostat are X-11 and SEATS. Both methods do not allow for adjustment of data with a higher than monthly frequency, even though an increasing number of time series is available with a weekly or daily periodicity. Possible examples range from data on air pollution, web-search keywords, and traffic jam data to economic variables such as exchange rates, online prices, and the amount of Euro banknotes in circulation. The aim is the development of a procedure that makes it possible to estimate and adjust for regularly and periodically reoccurring systematic effects and the influence of moving holidays and trading days in time series with daily observations. To this end, an STL based seasonal adjustment routine is combined with a regression model with ARIMA errors for the estimation of calendar and outlier effects. The latter will also be used for forecasting of the original time series, so as to be able to compute forecasted seasonal factors. The proposed procedure successively estimates and adjusts intra-weekly, intra-monthly and intra-annual periodic movements. The prediction of the original series is based on the regARIMA model which uses trigonometric functions to incorporate monthly and annual seasonality. In addition, the intra-weekly seasonal factors are extrapolated using exponential smoothing. The procedure is evaluated

empirically using the currency in circulation in Germany.

CG390 Room 108 CONTRIBUTIONS IN RISK ANALYSIS AND BANKING Chair: Rochelle Edge

CC1600: **Term structure of variance risk premium and returns' predictability**
Presenter: **Giacomo Bormetti**, University of Bologna, Italy
Co-authors: Fulvio Corsi, Adam Aleksander Majewski
An analytic relation between equity risk premium and the term structure of variance risk premium (VRP) is derived. Motivated by this result, we estimate the VRP term structure using a general and fully analytical discrete-time option pricing framework featuring multiple volatility components and multiple risk premia. We confirm the importance of VRP in improving option pricing performances and show the ability of multi-component GARCH models to produce realistic hump-shaped VRP term structure. We finally uncover the strong predictive power of the shape of the VRP term structure, summarized by its slope, on future stock-index returns.

CC1667: **Changes to banking behaviour in response to regulatory stress testing**
Presenter: **Simon Caton**, National College of Ireland, Ireland
Co-authors: Prasad Prakash Pore, Oisin Creaner
The Global Financial Crisis of 2007-2011 resulted in the regulatory implementation of a standardized framework for determine the ability of banking sector to withstand a further crisis. This standardized framework, Third Basel Accord (Basel III), mainly concerns stress testing, capital adequacy requirement and market liquidity risk. With the help of different financial indicators like capital, risk profile, etc., regulatory stress testing assesses the ability of financial institutions to withstand further financial and economic crises. Changes in these indicators elucidate certain banking behavior in anticipation of, and in response to, regulatory stress testing. Using data from the 2014 stress tests results and 2013 transparency exercise conducted by EBA, we analyze 63 EU banks across 21 jurisdictions to unravel how the advent of institutional stress testing has resulted in a behavioral change within the European banking sector. Our results indicate that banks are instrumenting behavioral changes in response to Basel III. Whilst this is a positive result, it does, however, mean that more scrutiny may be required in future stress tests.

CC1578: **Channels of sovereign risk spillovers and investment in the manufacturing sector**
Presenter: **Sebastian H M Deininger**, University of Basel, Switzerland
Co-authors: Dietmar Maringer
The aim is to identify endogenous as well as exogenous indicators of firms' investment activity, and examine, in particular, the effect that these variables have in co-determining firms' long-term investment decisions. For this purpose, a panel vector autoregressive model extended by exogenous variables (PVARX) is applied. Two channels of spillovers from sovereign risk to the firms capital expenditures are defined. The first channel, the direct channel, describes responses in capital expenditures from an innovation in sovereign risk. The second channel, the indirect channel, is a transmission mechanism in which risk spillovers from changes in sovereign risk indirectly affect a firm's capital expenditures via its market risk and profitability. While the direct risk channel is observed to be of major importance in Emerging and Developing Economies, it is comparatively small in Advanced Economies. In the case of the latter, contagion from changes in sovereign risk on a firm's market risk plays a much more important role.

CC1483: **The effects of bank capital requirements on bank lending: What can we learn from the post-crisis regulatory reforms**
Presenter: **Rochelle Edge**, Federal Reserve Board, United States
Co-authors: Jose Berrospide
Events associated with the implementation of a number of U.S. post-crisis capital reforms are used to consider the impact of bank capital requirements on bank lending. We conduct our analysis separately for small bank holding companies (BHCs) and for large BHCs. In both cases we use the arrival of new information on capital requirements, which affected different BHCs by differential magnitudes, to estimate these impacts. The new information on capital requirements that we use varies between our small and large BHC analysis and our empirical strategies also differ. For small BHCs we rely on new information contained in the announcement of the U.S. banking agencies proposed Basel III capital rules of June 2012 and final Basel III capital rules of July 2013 and, in particular, on changes in measured regulatory capital ratios implied by differences in how these rules set various assets risk weights and defined qualifying regulatory tier 1 capital. For large BHCs we rely on new information contained in the first public release of the CCAR stress-test results in March 2012. We find negative but relatively small effects of increases in regulatory capital requirements on lending for small U.S. BHCs but larger effects for large BHCs for which, due to our use of a BHC-firm matched sample, we are able to control better for loan demand influences.

CC1541: **Towards a new taxonomy of systemic risk measures**
Presenter: **Hanane Dakhli**, Champagne School of Management PRISM Sorbonne, France
Systemic risk is one of the most studied subjects in financial literature. Since the financial crisis of 2007, a large variety of new measures has appeared constantly in scientific journals as well as in practitioners publications. The most complete and significant studies of systemic risk measures, so far, have been written by some authors. A review of the most recent literature led us to collect several measures that we classify into four families. We first present the present the systemic risk measures without any classification. Secondly, we analyze the systemic risk measures according different taxonomies and select measures with the same characteristics. Thirdly, we test empirically the different new taxonomies.

CG412 Room S21 CONTRIBUTIONS IN VOLATILITY MODELS AND THEIR APPLICATIONS Chair: Christian Conrad

CC1414: **Estimating volatility spillovers: A large t-Vector AutoRegressive approach**
Presenter: **Luca Barbaglia**, KU Leuven, Faculty of Economics and Business, Belgium
Co-authors: Christophe Croux, Ines Wilms
Commodity markets have recently experienced large price fluctuations. Volatileness and fat-tailedness have become typical features of commodity prices. We develop a general framework for the analysis of volatility spillovers accounting for fat-tailed error distribution. We propose a penalized estimator for the Vector AutoRegressive model with Student t-distributed errors that incorporates the estimation of the degrees of freedom of the t-distribution. Moreover, we provide a network analysis tool based on forecast error variance decompositions accounting for t-distributed errors to show the volatility spillovers between a large number of commodities. We study the dynamics of volatility spillovers between energy, biofuel and agricultural commodities. Our results highlight that the overall level of volatility spillovers increases in periods of low energy prices. Moreover, we find evidence of bidirectional volatility spillovers between energy and agriculture, which cannot be completely explained by presence of biofuels.

CC1681: **Modeling VIX Index dynamics with Bayesian methods and intraday data**
Presenter: **Milan Ficura**, University of Economics in Prague, Czech Republic
Co-authors: Jiri Witzany
A methodology for the Bayesian modelling of the VIX Index dynamics is presented. The proposed model captures most of the empirically observed properties of the VIX Index time series, including stochastic volatility (of the volatility), self-exciting jumps, asymmetry of the returns, as well

as the long-memory mean-reversion of the VIX index towards its long-term levels. In order to facilitate inference on the model, high-frequency power-variation estimators of the VIX volatility and the S& P500 volatility are used as additional sources of information. MCMC algorithm is proposed for the estimation of the model parameters and the latent state variables in the in-sample period, while a particle filter is used to provide out-of-sample forecasts. The ability of the model to capture VIX Index dynamics and approximate its empirical probability distribution under different market conditions is assessed on the historical VIX index time series with generally positive results.

CC1004: **Fitting a Heston's stochastic volatility model to the option quotes on the Warsaw stock exchange**
Presenter: **Katarzyna Brzozowska-Rup**, Kielce University of Technology, Poland

Estimating volatility from the underlying asset price history for discrete observations plays a key role in modelling and exploring financial data. Sequential Monte Carlo methods demonstrate considerable benefits for option pricing. Although the problem is hardly new, Monte Carlo methods have been extensively used in option pricing and it is worth it to review and expand them. The main focus is on Hestons models of stochastic volatility which uses data on an underlying market index and the prices of options written on that index. The most important property in this model is the assumption that the asset price and the volatility process are correlated. Further modifications of the model are still possible. The proposed generalisation consists in considering variables in the current and past time in the equation expressing the evolution of the asset price process. This process is not Markovian but more natural from the practical point of view. The pricing methodology is based on the maximum likelihood estimation combined with the particle filter. To facilitate the discussion simulated and real data from the Warsaw Stock Exchange are employed.

CC1542: **Asymmetric volatility connectedness on forex markets**
Presenter: **Lukas Vacha**, UTIA AV CR, v.v.i., Czech Republic
Co-authors: Jozef Barunik, Evzen Kocenda

How bad and good volatility propagate through forex markets is shown, i.e. we provide evidence for asymmetric volatility connectedness on forex markets. Using high-frequency, intra-day data of the most actively traded currencies over 2007-2015 we document the dominating asymmetries in spillovers that are due to bad rather than good volatility. We also show that negative spillovers are chiefly tied to the dragging sovereign debt crisis in Europe while positive spillovers are correlated with the subprime crisis, different monetary policies among key world central banks, and developments on commodities markets. It seems that a combination of monetary and real-economy events is behind the net positive asymmetries in volatility spillovers, while fiscal factors are linked with net negative spillovers.

EI677 Room Graduation hall RECENT ADVANCES IN ROBUST STATISTICS Chair: Juan Romo

EI1320: **Robust and sparse estimation in multivariate statistics**
Presenter: **Christophe Croux**, Leuven, Belgium

Sparse and robust estimation methods for multivariate statistics are presented. The lasso is the most popular sparse estimator for linear regression. However, it is not robust, and its breakdown point can be shown to be zero. One of the first robust alternatives for the lasso is the Sparse Least Trimmed Squares estimator. The advantage of the latter estimator is that it can be computed without the need for an initial robust sparse estimator. Moreover, it is operational and available as an R-package. We go beyond the regression model: (i) We show how a robust and sparse version of the Minimum Covariance Determinant covariance matrix estimator can be attained (ii) We give an approaches to robust and sparse principal component analysis (iii) We present a robust and sparse version of Canonical Correlation Analysis. We elaborate on the robust and sparse canonical correlation method, and show its good performance on several real data examples.

EI1701: **Wild adaptive trimming for robust estimation and cluster analysis**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Andrea Cerioli, Alessio Farcomeni

Trimming principles play an important role in robust statistics. However, their use for clustering typically requires some preliminary information about the contamination rate and the number of groups. We suggest a fresh approach to trimming that does not rely on this knowledge and that proves to be particularly suited for solving problems in robust cluster analysis. Our approach replaces the original K-population (robust) estimation problem with K distinct one-population steps, which take advantage of the good breakdown and separation properties of trimmed estimators when the trimming level exceeds the usual bound of 0.5. In this setting we prove that exact affine equivariance is lost on one hand, but on the other hand an arbitrarily high breakdown point can be achieved by "anchoring" the robust estimator. We also support the use of adaptive trimming schemes, in order to infer the population sizes and the contamination rate from the data. A further bonus of our methodology is its ability to provide a reliable choice of the usually unknown number of groups. We provide empirical evidence that our method performs well under different settings, both when the contaminant observations are well separated from the genuine populations and when there is overlap between outliers and "good" clustered data.

EO1197: **A feasible algorithm for robust fuzzy clustering throughout trimming and constraints**
Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain
Co-authors: Francesco Dotto, Alessio Farcomeni, Agustin Mayo-Iscar

A feasible algorithm for robust fuzzy clustering throughout trimming and constraints is presented. The fuzzy clustering approach adopted is aimed at achieving robustness by discarding a fixed proportion of "most-outlying" observations which are determined by the data-set itself. Constraints are useful to get mathematically well-defined problems and to avoid the detection of non-interesting "spurious" clusters. The proposed algorithm can be applied in very different statistical frameworks whenever these problems can be stated in terms of probability likelihoods. The algorithm follows from modifications of typical algorithms for fuzzy clustering but allowing for trimming and by enforcing constraints in the scatter parameters in the updating steps. The proposed algorithm is illustrated in two particular cases: heterogeneous clustering under multivariate normal assumptions and regression clustering. A brief discussion about tuning parameters is also given.

EO535 Room 306 A DEPENDENCE-BASED CLUSTERING METHODS Chair: Fabrizio Durante

EC0229: **A procedure for clustering time series**
Presenter: **Andres M Alonso**, Universidad Carlos III de Madrid, Spain
Co-authors: Daniel Pena

A new way to find clusters in large vectors of time series is presented. The procedure has two steps. Firstly, the series are split by their dependency, which is measured by the determinant of their correlation matrix including lags. Secondly, inside each groups the series are split by putting together series with a similar autocorrelation structure. The procedure is automatic and can be applied to large data sets. Several real and simulated examples are presented to illustrate the procedure.

EO1113: **Cross-dynamical analysis of the tail dependence structures of classic and fuzzy clusters of stocks**
Presenter: **Giovanni De Luca**, University of Naples Parthenope, Italy
Co-authors: Paola Zuccolotto
A dynamic clustering of financial returns time-series is performed based on the estimated bivariate lower tail dependence coefficients. The resulting groups are composed of assets characterized by high association between extremely low values. The tail dependence coefficients are estimated using the Joe-Clayton copula function. Then, the dynamics of the tail dependence structure of each group (measured e.g. by the average tail dependence) is analyzed in order to identify the group(s) able to anticipate the dynamics of the tail dependence structure of the remaining groups. The analysis is carried out considering both a deterministic and a fuzzy cluster analysis.

EO0862: **Copula-based fuzzy clustering of spatial time series**
Presenter: **Marta Disegna**, Faculty of Management, Bournemouth University, United Kingdom
Co-authors: Fabrizio Durante, Pierpaolo Durso
Model-based clustering methods allow to group different time series according to their similarity, usually detected by comparing either the estimated parameters or the residuals resulting from the fitted model performed on each time series. Recently, various methods, such as the Spearmans correlation and the copula-based dissimilarity measure, have been developed to study the association among residuals. A novel method is discussed to cluster time series data collected on different spatial locations by using the concordance (comovements) between residuals, as interpreted by the copula. In particular, the Fuzzy Partitioning Around Medoids algorithm with a suitable copula-based dissimilarity measure that include a spatial term is applied to classify time series. This clustering algorithm allows to capture both the rank-invariant dependence among time series and the uncertainty associated to the assignment of the units to each cluster taking into account possible additional features coming from the data spatial location. This method has been theoretically presented together with the illustration of the empirical case study related to the identification of touristic agglomerations of cities and towns belonging to the same macro destination and characterised by similar touristic flows over time.

EO0614: **Empirical convergence rates of dependence-based clustering methods illustrated with financial time series**
Presenter: **Gautier Marti**, Hellebore Capital Ltd / Ecole Polytechnique, United Kingdom
Co-authors: Frank Nielsen, Philippe Donnat
For clustering methods to be useful in online risk and trading systems, they have to be both robust to noise (which is partly achieved by leveraging copulas) and fast converging. Fast convergence of the clustering structures (flat or hierarchical) to the true underlying clusters is required to mitigate the non-stationarity effects of financial multivariate time series. If the methods require long time series to converge, then the underlying economic regime and its associated clustering structure may have changed several times in the meantime. In such case, the clusters dynamics are smoothed out and less useful for online risk and trading systems. At the heart of clustering algorithms is the fundamental notion of distance that can be defined based upon a proper representation of data. Copula-based dependence coefficients allow a better modelling of (non-linear) financial time series dependence than simplistic correlation measures such as the Pearson or Spearman ones. However, we may also consider what impact those novel correlation coefficients have on the convergence rate of the whole clustering methodology: does it speed it up or slow it down? We benchmark the empirical convergence rates of several state-of-the-art dependence-based clustering methods. Baseline results are obtained using a straightforward approach: Pearson's ρ, Kendall's τ, Spearman's ρ_S correlation coefficients.

EO0842: **The impact of the dependence structure in credit-risk management**
Presenter: **Silvia Romagnoli**, University of Bologna, Italy
Co-authors: Enrico Bernardi, Matteo Doti
An innovative model is proposed to value the credit risk of a portfolio. This new approach is based on the hierarchical hybrid copula-based model (HYC for short). The HYC model, involving a clusterization of the portfolio in several risks classes, is classified as hybrid because the computation of the loss cdf depends on the class cardinality: for large groups one is justified to apply a limiting approach, while for small ones one applies a more involved procedure preserving the granularity of the group itself. In order to appreciate the impact of the dependence structure and then of the systemic and contagious risks in credit-risk evaluation, a VaR analysis based on the HYC loss function (HYC-VaR for short) is compared to the Credit Metrics approach (CM-VaR for short), that is our benchmark. This comparison allows us to appreciate over/under-valuation of the capital detained from the financial institution due to the modelization used for the evaluation of risk and induced by the formalization of the dependence structure in its various connotations. Moreover the impact of an enlargement of the dependence structure is discussed with respect to systemic and contagious effects implied by a wider dependence structure in the context of a geographical area sub-portfolios analysis, now relevant to determine risk contributions of sub-groups and a portfolio optimization with constraints on a sub-portfolios risk.

EO177 Room 204 EXTREMES AND THEIR REAL-LIFE APPLICATIONS Chair: Gilles Stupfler

EO0283: **Frontier estimation based on extreme risk measures**
Presenter: **Jonathan El Methni**, Universite Paris Descartes, France
Co-authors: Laurent Gardes, Stephane Girard
Value-at-risk, Conditional Tail Expectation, Conditional Tail Variance and Conditional Tail Moment are classical risk measures. In statistical terms, the Value-at-risk is the upper α-quantile of the loss distribution where $\alpha \in (0,1)$ is the confidence level. Here, we focus on the properties of these risk measures for extreme losses (where $\alpha \downarrow 0$ is no longer fixed). To assign probabilities to extreme losses we assume that the distribution satisfies a von-Mises condition which allows us to work in the general setting, whether the extreme-value index is positive, negative or zero *i.e.* for all domains of attraction. We also consider these risk measures in the presence of a covariate. The main goal is to propose estimators of the above risk measures for all domains of attraction, for extreme losses, and to include a covariate in the estimation. The estimation method thus combines nonparametric kernel methods with extreme-value statistics. The asymptotic distribution of our estimators is established and their finite sample behavior is illustrated on simulated data and also on a motivating application in the reliability of nuclear reactors.

EO0464: **Peaks-over-thresholds modelling with multivariate generalized Pareto distributions**
Presenter: **Anna Kiriliouk**, Universite Catholique de Louvain, Belgium
Co-authors: Holger Rootzen, Johan Segers, Jenny Wadsworth
Statistical modelling using multivariate generalized Pareto distributions constitutes the multivariate analogue of peaks-over-thresholds modelling with the univariate generalized Pareto distribution. We propose a construction tool which allows to generate suitable parametric tail dependence models for extremes episodes, conditioned on having at least one threshold exceedance. Several concrete examples are proposed, and the densities necessary for censored likelihood estimation are derived. Finally, we present a new parametric model for data with structured components, and illustrate it with an application aimed at estimating the probability of a landslide in northern Sweden.

EO0625: **Extreme values, from theory to practice**
Presenter: **Leo Belzile**, EPFL, Switzerland
Co-authors: Anthony Davison
Vargas state in Venezuela was hit by abnormally high rainfalls in December 1999. A naive extreme-value model gives a return period of 17 million

years for the event. Previous statistical analysis have stressed the need for careful modelling of the series to account for non-stationarity and adequately quantify uncertainty. We revisit the dataset using models inspired by penultimate approximations and reviewing recent proposals in the literature for extremes of rainfall series.

EO0817: **On tail trend estimation and testing with application to extreme rainfall**
Presenter: **Claudia Neves**, University of Reading, United Kingdom

Extreme value theory provides a rigorous and prolific framework for analysing rare events with severe impact. The basic assumption is that the observations are independent and identically distributed. Although the celebrated extreme value theorem still holds under several forms of weak dependence, relaxing the stationarity assumption by considering a trend in extremes, for instance, leads to a changeling inference problem about the frequency of extreme events. Some studies deem that the change in climate is not so much about the mean but rather in the frequency the extreme phenomena. The average rainfall may not change much, but heavy storms may become more or less frequent, meaning that observations have different underlying distributions. We present statistical tools for dealing with changes over time and/or space by considering a trend on the frequency of high exceedances. The methodology is applied to daily rainfall data from several stations across Germany.

EO0851: **Risks to the nuclear industry from environmental extremes: Problems in return-level estimation**
Presenter: **David Walshaw**, Newcastle University, United Kingdom

In assessing the risk posed by environmental variables, a standard requirement of the nuclear industry is that systems should be able to continue operation in the face of the 1-in-100 year event, and to avoid catastrophic failure in the face of the 1-in-10000 year event. Two real-life applications are considered. In Application 1 the cooling system of a building is vulnerable to extremes of a bivariate random variable consisting of ambient air temperature and the component of wind velocity from a particular direction. In Application 2 it is desired to estimate 10000 year return levels at a particular location, of wind speed only, with as much precision as is reasonably possible. The available data are in the form of hourly wind records ranging from 10 to 50+ years, together with information on possible covariates, collected from a network of 40 to 50 stations around the location of interest. The two applications motivate a range of interesting methodological problems in bivariate/multivariate extremes. Application 1 is unusual in that there is negative dependence in the observed extremes of temperature and wind speed. The critical issue for Application 2 concerns the nature of the tail dependence in extremes across stations. Is asymptotic dependence an appropriate assumption, or asymptotic independence, or perhaps a mixture of both, depending on distance between stations? The conclusions are vital for the choice of modelling strategy.

EO129 Room 216 NEW METHODOLOGIES IN SURVIVAL AND RELIABILITY Chair: Mariangela Zenga

EO0284: **Comparison of condition-based maintenance strategies under different approaches in degradation-threshold-shock models**
Presenter: **Nuria Caballe Cervigon**, University of Extremadura, Spain
Co-authors: Inma Torres Castro

Condition-based maintenance strategies (CBM) in infinite-time and finite-time horizon are considered. This type of maintenance uses monitoring techniques for checking the state of the system. The considered systems are subject to two different causes of failure, gradual internal deterioration and sudden shocks. Internal degradation is modelled by using a time-dependent stochastic process, concretely, a gamma process and sudden shocks arrive at the system following a non-homogeneous Poisson process (NHPP). When a sudden shock takes places, the system fails. In addition, the system is regarded to fail when the deterioration level reaches a critical threshold. Under this functioning scheme, a CBM strategy is developed for controlling the reliability of the system. Traditionally, this strategy is developed under an asymptotic approach. This fact implies that the system can be replaced by a new one with the same conditions as the initial. However, considering the asymptotic approach is not always realistic. We compare a CBM strategy by asymptotic and transient approach providing numerical examples in order to illustrate the different approaches.

EO0551: **A discrete non-homogeneous Markov model in survival analysis: Application to breast cancer**
Presenter: **Juan Eloy Ruiz-Castro**, University of Granada, Spain
Co-authors: Mariangela Zenga

Homogeneous Markov processes are usual in the survival literature to analyse the behavior of an illness that evolves over time. In many cases the homogeneity is a very strong constraint and non-homogeneity arises in a natural way, since the transition probabilities between two any states are not constant by time generally. A non-homogeneous Markov model is considered to study the evolution of breast cancer. The model and the likelihood function are built to estimate the parameters. Some interesting measures, such as survival functions and mean sojourn times are worked out. The effect of treatments is incorporated as time-depending covariates. Thus, we have a multi-state model with multidimensional covariates. The results are compared with the empirical and homogeneous cases for several risk groups.

EO0991: **A note on the use of Linex loss function in parametric reliability models**
Presenter: **Inmaculada Barranco-Chamorro**, University of Sevilla, Spain
Co-authors: Pedro Luis Luque-Calvo

When using Bayesian techniques in survival analysis and reliability models, a loss function must be specified. The most widely used is squared error. However, this function has the drawback that is symmetric, and quite often, this assumption is not realistic enough in reliability models. In these cases, an asymmetric loss function may be more appropriate. In this context, we show that the Linex loss function may be a good choice, since it penalizes positive and negative errors in a different way, and it is still mathematically manageable. We review a number of results obtained from the use of Linex loss in failure models, such as Burr type XII and generalized half logistic distributions, along with methods of estimation based on records. The importance of the shape parameter in the Linex loss function is illustrated through simulations.

EO1090: **Regression approaches for competing risks models with time-dependent covariates**
Presenter: **Manuela Cattelan**, University of Padova, Italy
Co-authors: Giuliana Cortese

In the study of access to pediatric intensive care unit for children who underwent hematopoietic stem cell transplantation, the event may not be observed because of the patients death. In such instances, models for competing risks should be employed and the additional presence of internal time-dependent covariates should be considered. Indeed, children may undergo a second hematopoietic stem cell transplantation. In competing risks models there are different approaches to account for time-dependent covariates, an example is landmark analysis. The advantages and disadvantages of different models and methods for accounting for time-dependent covariates will be pointed out and a novel approach for mean residual life models in presence of competing risks will be presented. The methodology will be illustrated through an application to access to intensive care unit of children affected by oncohematologic diseases who underwent hematopoietic stem cell transplantation in the HSCT Unit of Paduas Pediatric Oncohematology Clinic in the last 15 years.

EO1290: **Coxian phase-type survival models with a hidden layer**
Presenter: **Hannah Mitchell**, Queen's University Belfast, United Kingdom
Co-authors: Adele Marshall, Mariangela Zenga

Modelling patient flow in a healthcare system has been seen as an important aspect in understanding the systems activity and how to carefully

manage its resources. The distribution of patient length of stay in hospital typically is positively skewed similar to that represented in survival analysis. The Cox proportional hazards model is are one such model used in survival analysis to investigate whether certain variables are significantly associated with a patients length of stay in hospital. An alternative to this model is the Coxian phase-type distribution which has the additional capability of being able to represent patients flow through the hospital where duration of stay is made up of phases in a Markov process. The Coxian phase-type distribution can also incorporate other variables and represent their effect on patient length of stay. Previous work is extended by incorporating a hidden layer into the Coxian phase-type distribution using the hidden Markov model. This new model highlights different pathways of patient flows within the data along with certain characteristics being of potential concern to the healthcare system. Admissions into geriatric wards within the Lombardy region of Italy will be used as a setting for these models.

EO091 Room 305 A MODEL SPECIFICATION TESTS Chair: Maria Dolores Jimenez-Gamero

EO0294: **Permutation tests in the two-sample problem for functional data**
Presenter: **Alejandra Cabana**, Universitat Autonoma de Barcelona, Spain
Co-authors: Adolfo Quiroz, Ana Maria Estrada, Jairo Ivan Pena Hidalgo

Different permutation schemes are discussed and compared in the context of the two-sample problem for functional data. The notion of functional data depth has been previously used to adapt ranking methods originally used in multivariate data. Inspired by this, we propose a Wilcoxon type test based on Fraiman and Munis' distance, and use a couple of methods inspired by meta-analysis in order to asses the significance of the depth differences. The other proposal is an adaption to the functional data setting of the k-nearest neighbours statistics of Schilling, also devised for multivariate data originally. We compare the performance of these tests with a previous test in simulated data as well as in a real data set comparing the level of air pollutants in different neighbourhoods of Barcelona during the years 2014 and 2015. After analyzing the statistical power and the computational cost of every method, we think that the best method for the two sample problem for functional data is Schillings method.

EO0372: **Testing for the generalized Poisson-inverse Gaussian distribution**
Presenter: **Apostolos Batsidis**, University of Ioannina, Greece
Co-authors: Maria Dolores Jimenez-Gamero, Virtudes Alba-Fernandez

The generalized Poisson inverse Gaussian (GPIG) family is a flexible family of distributions, useful for modeling count data with different tail heaviness. The probability generating function (PGF) of the GPIP family is the unique PGF satisfying certain differential equation. This property leads us to propose a new goodness-of-fit test for the GPIP family. It is shown that the test is consistent against fixed alternatives. The null distribution of the test statistic can be consistently approximated by means of a parametric and a weighted bootstrap. The finite sample performance of the proposed test is investigated by means of a simulation study, where the goodness of the proposed approximations is numerically studied.

EO0407: **Testing equality of a large number of densities under mixing conditions**
Presenter: **Marta Cousido Rocha**, SiDOR Research Group, University of Vigo, Spain
Co-authors: Jacobo de Una-Alvarez, Jeff Hart

In certain settings, such as microarray data, the sampling information is formed by a large number of possibly dependent small data sets. In special applications, for example in order to perform clustering, the researcher aims to verify if all data sets have a common distribution. For this reason we propose a formal test for the null hypothesis that all data sets come from a single distribution. The asymptotic null distribution of the proposed test is derived under mixing conditions. A simulation study is conducted, showing that the test respects the nominal level, and that it has a power which tends to 1 when the number of data sets tends to infinity. An illustration involving microarray data is provided.

EO0460: **Characterizations of symmetry via central order statistics and the applications to goodness-of-fit testing**
Presenter: **Bojana Milosevic**, University of Belgrade, Serbia and Montenegro
Co-authors: Marko Obradovic

In recent times, goodness-of-fit tests based on characterizations have become very popular. In case of symmetry testing, this approach is not much explored. We present some recent characterizations of symmetry around zero that use some properties of the order statistics, as well as some new classes of symmetry tests based on those characterizations. For measuring the quality of those tests we calculate their asymptotic efficiencies against some common alternatives and compare them with other symmetry tests. We also suggest some modifications of the proposed tests, so they could be applied in case of testing symmetry around unknown location. The power comparison for small sample sizes and some applications will also be presented.

EO0901: **A goodness-of-fit test for marginal distribution of linear random fields with long memory**
Presenter: **Nao Mimoto**, University of Akron, United States
Co-authors: Hira koul, Donatas Surgailis

The problem considered is the goodness-of-fit test of fitting a specified distribution function to the marginal distribution of a stationary long-memory moving average random field observed on increasing v-dimensional cubic domains when its mean and scale parameters are known or unknown. This set up includes the time series case as $v = 1$. In the case of unknown mean, if the mean is estimated by the sample mean, the first order difference between the residual empirical process and null distribution functions is known to be asymptotically degenerate at zero, and hence can not be used to fit a distribution up to an unknown mean. We show that by choosing suitable class of estimators of the mean, this first order degeneracy does not occur. Further, using a sample standard deviation as a estimate of scale, a modified Kolmogorov-Smirnov statistic based on the residual empirical process has Cauchy-type limit distribution, independent of mean and scale parameter, as well as the long-memory parameter d. Based on this result, a simple goodness-of-fit test for the marginal distribution is constructed, which does not require the estimation of d or any other underlying nuisance parameters. A simulation study investigating the finite sample behavior of size and power is presented.

EO041 Room Board meeting room II CIRCULAR MODELS AND ITS RELATED TOPICS Chair: Toshihiro Abe

EO0329: **Bayesian estimation and hypothesis tests for a circular GLM**
Presenter: **Kees Mulder**, Utrecht University, Netherlands
Co-authors: Irene Klugkist

Circular data are data measured in angles or directions. Although they occur in a wide variety of scientific fields, the number of methods for their analysis is limited. We develop a GLM-like model for circular data within the Bayesian framework, using the von Mises distribution. The model allows inclusion of covariates and categorical predictors. In a frequentist setting, this model is plagued by the likelihood surface of its regression coefficients, which is not log-concave. In a Bayesian context, a weakly informative prior solves this issue, while for other paramaters noninformative priors are available. In addition, we develop Bayesian hypothesis tests based on the Bayes factor for both equality and inequality constrained hypotheses. In a simulation study, it can be seen that our method performs well. Finally, we apply this model to a dataset from experimental psychology, and show that it provides valuable insight for applied researchers. Extensions to dependent observations are within reach by means of the multivariate von Mises distribution.

EO0334: **Interpretation and evaluation of circular effects in projected normal regression models**
Presenter: **Jolien Cremers**, Utrecht University, Netherlands
Co-authors: Kees Mulder, Irene Klugkist
Currently, coefficients for projected normal regression models come on two bivariate components. This enables predictions of a circular effect, but does not allow for the interpretation and evaluation of the size of the effect on the circle. We introduce new types of circular regression coefficients that do allow us to assess the size of this effect. These coefficients are straightforward to compute from the original bivariate output of an MCMC sampler. The bivariate coefficients allow us to distinguish between location and accuracy effects on the circle. We propose a measure to assess whether there is any effect on the circle at all. Subsequently we check whether this effect is most likely an accuracy or a location effect. The performance of these measures and of the new circular regression coefficients is investigated using a simulation study. The new measures are illustrated using example datasets for a standard multivariate regression and a mixed effects model. In addition, we use the Bayesian toolbox for model assessment and evaluation to check whether the assumption of projected normality is met, compare the fit of several models and test order constrained hypotheses for these example datasets.

EO0519: **Hidden Markov models for the analysis of environmental cylindrical data**
Presenter: **Francesco Lagona**, University Roma Tre, Italy
Correlated cylindrical data arise often in environmental research when angles and intensities are repeatedly collected across time or space. The statistical analysis of these data is complicated by several issues that include the mixed linear-circular support of the data, the skewness and the multimodality of the marginal distributions, correlations between angles and intensities and cross-correlations along time and across space. The focus is on cylindrical hidden Markov models as flexible tools that are capable to address these issues in a parsimonious and efficient way. These models are dynamic mixtures of cylindrical densities whose parameters evolve according to a latent Markov process in time or space. Cylindrical densities are exploited to approximate the joint distribution of angles and intensities. Mixtures of cylindrical densities provide an additional extension to allow for multimodal cylindrical data. Assuming that the mixture parameters vary according to the evolution of a latent process across time or space represents a further extension to capture unobserved temporal or spatial heterogeneity and to allow for temporal or spatial correlation. These models are illustrated in a study of sea wave dynamics and in a study of segmentation of sea currents.

EO0716: **Maximin tests for symmetry of circular data based on the characteristic function**
Presenter: **Thomas Verdebout**, Universite Libre de Bruxelles, Belgium
Co-authors: Simos Meintanis
Inference for circular data is considered based on the empirical characteristic function. More precisely, we provide tests for reflective symmetry on the circle based on the imaginary part of the empirical characteristic function. We show that the proposed tests enjoy many attractive features. In particular, we obtain that they are locally and asymptotically maximin in the Le Cam sense under sine-skewed alternatives in the specified mean direction case. For the unspecified mean direction case, we provide corrected versions of the original tests that keep very nice asymptotic power properties. Results are illustrated on a well-known dataset and checked via Monte-Carlo simulations.

EO1046: **Nonlinear state-space modeling for wind speed and direction**
Presenter: **Takayuki Shiohama**, Tokyo University of Science, Japan
The state space form is a useful framework for estimating unobserved state variables from some given observations. Its applications can be found in diverse areas of natural sciences and engineering such as ecology, epidemiology, meteorology, economics and finance. The wind speed and directions have complex time series probability structures involving highly non-Gaussian and nonlinear transition. We consider a simulation-based inference using the sequential Monte Carlo methods for computing the posterior distributions for the state variables given all available observations. We propose an alternative approach that allows us to extend the methods of importance sampling distributions incorporating with the class of circular Markov transition densities. The resulting methods are compared with various resampling schemes with applications real data analysis.

EO147 Room 206 MULTIVARIATE ANALYSIS AND CHANGE-POINT DETECTION Chair: Daniel Vogel

EO0395: **Change-point detection for locally dependent data**
Presenter: **Hao Chen**, University of California at Davis, United States
Local dependence is common in multivariate and object data sequences. We consider the testing and estimation of change-points in such sequences. A new way of permutation, circular block permutation with a randomized starting point, is proposed and studied for a scan statistic utilizing graphs representing the similarity between observations. The proposed permutation approach could correctly address for local dependence and make it possible the theoretical treatments for the non-parametric graph-based scan statistic for locally dependent data. We derive accurate analytic approximations to the significance of graph-based scan statistics under the circular block permutation framework, facilitating its application to locally dependent multivariate or object data sequences.

EO0430: **Multivariate changepoint detection with subsets**
Presenter: **Rebecca Killick**, Lancaster University, United Kingdom
Co-authors: Benjamin Pickering, Idris Eckley
Historically much of the research on changepoint analysis has focused on the univariate setting. However, increasingly data found in contemporary scientific fields are multivariate in nature, with each observation in a sequence containing the values of multiple variables which have been observed simultaneously. The multivariate changepoints, which may be observed within such time series can be categorized as either fully-multivariate or subset-multivariate. Fully-multivariate changepoints refer to those changes in structure which occur simultaneously in all variables. Conversely, subset-multivariate changepoints refer to those which occur in only a subset of the observed variables. Traditionally, multivariate changepoint detection methods typically assume that all changes within a series are fully-multivariate. Some recent papers have taken into account that all variables may not change but either do not explicitly output the subsets or do a fully multivariate analysis followed by subset determination. This is the first attempt to create a dynamic program specifically for detecting changes in subset-multivariate time series. We present both exact and approximate optimization methods for determining the number, location and affected subsets of changepoints. Simulation studies demonstrate the performance of the approach on fully multivariate, as well as large, medium and small proportions of affected subsets and we apply the approach to acoustic sensing data.

EO0556: **Change-point detection in series of maxima using probability weighted moments**
Presenter: **Ivan Kojadinovic**, University of Pau, France
Co-authors: Philippe Naveau
The analysis of seasonal or annual block maxima is of interest in fields such as hydrology, climatology or meteorology. In connection with the celebrated method of block maxima, we study several tests that can be used to assess whether the available series of maxima is identically distributed. It is assumed that block maxima are independent but not necessarily generalized extreme value distributed. The asymptotic null distributions of the test statistics are investigated and the practical computation of approximate p-values is addressed. Extensive Monte Carlo

simulations show the adequate finite-sample behavior of the studied tests for a large number of realistic data generating scenarios. Illustrations on several environmental datasets conclude the work.

EO0701: **Change point detection in panel data**
Presenter: **Marie Huskova**, Charles University, Czech Republic
Recent results are considered on testing and estimating for a change in panel data. Asymptotic properties of the suggested procedures will presented when the number of the panels is large and the number of the observations in each panel is either large or fixed (small w.r.t. the number of panels). Results of a simulation study and an application to a real data set will be reported.

EO1205: **Self-weighted change-point test and estimator**
Presenter: **Stefanie Schwaar**, University of Kaiserslautern, Germany
To detect a change in a time-series based on the pseudo-likelihood ratio the weighted CUSUM statistic was derived. This statistic involves a weight function and the absolute value of a partial sum. The weighted CUSUM tends to infinity unless it is transformed. But the convergence to the asymptotic distribution is quite slow. That is why modifications became of interest. Proper modifications can be derived by the use of a different weight function. Besides truncations also different powers of the weight function are analysed. For a class of deterministic functions sufficient conditions are determined such that the modified weighted CUSUM still has a known asymptotic distribtion. Taking a look at some of those weight functions by using different powers, we observe that for the test statistic other weight functions are preferable than for the change-point estimator. The preferable weight function depends on the position of the change, which is unknown. To overcome this problem a data driven change-point test and its estimator are proposed. We call these self-weighted change-point test and estimator. The asymptotics of the test statistic as well as the asymptotics of the estimator are presented.

EO013 Room 308 B MODELLING DISCRETE DATA: EXTENSIONS AND DIAGNOSTICS Chair: John Hinde

EO0776: **Joint models for repeated counts of predators and prey**
Presenter: **Rafael de Andrade Moral**, NUI Galway, Ireland
Co-authors: Clarice Garcia Borges Demetrio, John Hinde
In ecological field surveys it is often of interest to estimate the abundance of species. However detection is imperfect and hence it is important to model these data taking into account the ecological processes and sampling methodologies. In this context, N-mixture models and extensions are particularly useful, as it is possible to estimate population size and detection probabilities under different ecological assumptions. In this framework, observed animal counts are assumed to be binomial distributed with the denominator being a latent random variable with a count distribution representing the true abundance in each site, and hence only a part of this abundance, subject to a probability of detection, is observed. We extend the N-mixture modelling framework to jointly model predator and prey species abundances and obtain the correlation between them in each surveyed site. We illustrate this approach using data from the North American Breeding Bird Survey on bald eagles and mallards, obtained in the 2015 campaign for the states of Oregon, Washington and British Columbia. In each state, several 50-stop routes were surveyed and the total counts for every ten stops are available for download from their website.

EO0781: **Transition models and tests for stationarity**
Presenter: **Idemauro Antonio Rodrigues de Lara**, NUI Galway, Ireland
Co-authors: John Hinde, Cesar Augusto Taconeli
Transition models, based on Stochastic Processes and Generalized Linear Models, are an important framework that can be used to model longitudinal categorical data. A relevant issue in applying these models is the condition of stationarity, or homogeneity of transition probabilities over time. We propose two tests to assess stationarity in transition models. Also, we present three motivational studies, with ordinal and nominal data, in which the tests are applied to assess the stationarity. Their performances are assessed through simulation studies and the results show that the proposed tests have good performance. The correlation between these test statistics and the classical test available in the literature is large, which indicates concordance between the testing approaches.

EO0511: **Count data models for biological dosimetry**
Presenter: **Manuel Higueras**, Newcastle University, United Kingdom
Co-authors: Pedro Puig, Elizabeth Ainsbury
Ionising radiation overexposures are one of the major current concerns of our society. Consequently, biological retrospective dosimetry relies on quantifying the amount of damage induced by radiation at a cellular level, e.g. by counting dicentrics observed in metaphases from a sample of peripheral blood lymphocytes. This quantification is essential for predicting the derived health consequences in overexposed individuals. Moreover, it provides an accurate, personal and individual dosimeter. In biological dosimetry it is typically assumed that the number of chromosomal aberrations produced in a blood cell is Poisson distributed, whose intensity is a quadratic function of the absorbed dose. Calibration dose-response curves are calculated from cytogenetic laboratory experiments where blood samples are exposed to different doses, simulating whole body homogeneous irradiations. The dose estimation is an inverse regression problem. This classical Poisson assumption is not supported in a lot of irradiation scenarios, for instance for high linear energy transfer, partial body or gradient irradiations. These situations lead to compound Poisson, zero-inflated Poisson and Poisson finite mixture models, among others.

EO0892: **Assessing deflation or inflation of counts in count data regression**
Presenter: **Jochen Einbeck**, Durham University, United Kingdom
Co-authors: Paul Wilson
The problem of identifying inflation or deflation of individual counts in count data regression models is discussed. A particular instance of this problem is the case of testing for zero-inflation which has received considerable attention in the literature. The case of testing for inflation or deflation of an arbitrary (possibly non-zero) count has received much less attention so far. We introduce an exact test procedure based on the Poisson-Binomial distribution for this purpose, and also provide a graphical tool to which we refer to as 'Christmas tree plot' due its characteristic shape. The methods are motivated and illustrated through cytogenetic data from the field of radiation biodosimetry.

EO0583: **Poisson-Tweedie regression models for overdispersed, underdispersed, and zero-inflated data**
Presenter: **John Hinde**, NUI Galway, Ireland
Co-authors: Wagner Bonat
Models are presented based on the class of Poisson-Tweedie factorial dispersion models with variance $\mu + \phi\mu^p$, where μ is the mean, ϕ and p are the dispersion and Tweedie power parameters, respectively. This class of models provides a flexible and comprehensive family including many standard discrete models. The family provides for modelling of overdispersed count data, including Neyman Type A, Polya-Aeppli, negative binomial, Poisson-inverse Gaussian and Hermite distributions, and can also accommodate zero-inflation and underdispersion. We discuss estimation of regression, dispersion and Tweedie power (variance function) parameters and illustrate the approach with several examples.

EO201 Room 217 APPLICATIONS OF EMPIRICAL MEASURES AND EMPIRICAL PROCESSES Chair: Eric Beutner

EO0787: **Semiparametric consistent estimators for virtual age models under right censoring**
Presenter: **Laurent Bordes**, University of Pau, France
Co-authors: Eric Beutner, Laurent Doyen
A large class is considered for semi-parametric models for recurrent events based on virtual ages. The model consists of an unknown hazard rate function, the infinite-dimensional parameter of the model, and a parametrically specified effective age function. We recently derived conditions on the family of effective age functions under which the profile likelihood inference method for the finite-dimensional parameter of the model leads to inconsistent estimates. We show that we can overcome the failure of the profile likelihood method by smoothing the pseudo-estimator of the infinite-dimensional parameter of the model as it has been done previously for the accelerated failure time model. We show that the resulting estimators are asymptotically consistent and we provide a numerical study of their behavior for finite sample size.

EO1110: **The concept of local robustness with a view toward statistical estimation and risk management**
Presenter: **Volker Kraetschmer**, University Duisburg-Essen, Germany
Co-authors: Alexander Schied, Henryk Zaehle
Many standard estimators such as several maximum likelihood estimators or the empirical estimator for any law-invariant convex risk measure are not (qualitatively) robust in the classical sense. However, these estimators may nevertheless satisfy a local robustness property on relevant sets of distributions. After introducing this new concept, attention will be paid to identify sets of local robustness, and to explain the benefit of the knowledge of such sets. For instance, it will be be demonstrated that many maximum likelihood estimators are robust on their natural parametric domains. A second aim consists in extending the general theory of robust estimation to the local framework. In particular, a corresponding Hampel-type theorem is provided, linking local robustness of a plug-in estimator with a certain continuity condition.

EO1199: **Empirical measures as a tool in the analysis of Monte Carlo methods**
Presenter: **Pierre Nyquist**, KTH Royal Institute of Technology, Sweden
The aim is to discuss how empirical measure large deviations can be used to analyze the properties of various Monte Carlo methods. Covering a range of Monte Carlo methods, the focus will be on the MCMC method known as known as parallel tempering and the so-called infinite swapping limit, both which correspond to a collection of interacting particles. For infinite swapping we employ a large deviation analysis for the associated empirical measures, together with methods from stochastic optimal control, to discuss certain qualitative properties. In particular we discuss how symmetry properties of the underlying potential landscape may affect convergence properties and how the rate function identifies those parts of the state space where noise due to sampling has the greatest impact on the overall performance of the algorithms.

EO1339: **A general approach to the analysis of statistics from subsamples**
Presenter: **Stanislav Volgushev**, University of Toronto, Canada
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 us 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.

EO1176: **A new functional delta-method for the bootstrap**
Presenter: **Henryk Zaehle**, Saarland University, Germany
Co-authors: Eric Beutner
The functional delta-method (FDM) provides a convenient tool for deriving bootstrap consistency for the plug-in estimator of a "differentiable" statistical functional from bootstrap consistency of the underlying empirical process. It is commonly acknowledged that (tangential) Hadamard differentiability is a suitable notion of differentiability in this context. However it is likewise commonly acknowledged that Hadamard differentiability is a relatively strong requirement so that only "a few" plug-in estimators can indeed be treated with the help of the FDM. We will present a new version of the FDM for the bootstrap, which is based on a weaker notion of differentiability. Its benefit will be illustrated by means of several examples. It is worth mentioning that our FDM gets along without using the concepts of outer integrals and outer probabilities.

EO603 Room 309 B TIME SERIES ANALYSIS Chair: Rainer Dahlhaus

EO0813: **The predictive density of a GARCH(1,1) process**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Karim Abadir , Paolo Paruolo
The predictive probability density function of a GARCH(1,1) process is derived, under Gaussian or Student-t innovations. The analytic form is novel, and replaces current methods based on approximations and simulations.

EO1213: **Semi-parametric dynamic factor models for nonstationary time series**
Presenter: **Giovanni Motta**, Pontificia Universidad Catolica de Chile, Chile
Co-authors: Michael Eichler
A novel dynamic factor model is introduced for multivariate non-stationary time series. In a previous work, we have developed asymptotic theory for a fully non-parametric approach based on the principal components of the estimated time-varying covariance and spectral matrices. This approach allows both common and idiosyncratic components to be non-stationarity in time. However, a fully non-parametric specification of covariances and spectra requires the estimation of high-dimensional time-changing matrices. In particular when the factors are loaded dynamically, the non-parametric approach delivers time-varying filters that are two-sided and high-dimensional. Moreover, the estimation of the time-varying spectral matrix strongly depends on the chosen bandwidths for smoothing over frequency and time. As an alternative, we propose a new approach in which the non-stationarity in the model is due to the low-dimensional latent factors. We distinguish between the (double asymptotic) framework where the dimension of the panel is large, and the case where the cross-section dimension is finite. For both scenarios we provide identification conditions, estimation theory, simulation results and applications to real data.

EO0846: **Locally stationary functional time series**
Presenter: **Anne van Delft**, Maastricht University, Netherlands
Co-authors: Michael Eichler
Inference methods for functional data have received a lot of attention the last few years. So far, the literature on functional time series has focused on processes of which the probabilistic law is either constant over time or constant up to its second-order structure. Especially for long stretches of

data it is desirable to be able to weaken this assumption. We introduce a framework that allows for meaningful statistical inference of functional data of which the dynamics change over time. That is, we put forward the concept of local stationarity in the functional setting and establish a class of processes that have a functional time-varying spectral representation. Time-varying functional ARMA processes are investigated and shown to be functional locally stationary according to the proposed definition. Important in our context is the notion of a time-varying spectral density operator of which the properties are studied and uniqueness is derived. The framework is then used to construct an estimator of the spectral density operator based on a functional version of the segmented periodogram matrix. In particular, we prove it is consistent and study its asymptotic distribution.

EO0656: **Volatility decomposition and estimation in time-changed price models**
Presenter: **Sophon Tunyavetchakit**, Bank of Thailand, Thailand
Co-authors: Rainer Dahlhaus
The usage of a spot volatility estimate based on a volatility decomposition in a time-changed price-model according to the trading times is investigated. In this model clock-time volatility splits up into the product of tick-time volatility and trading intensity, which both can be estimated from data and contain valuable information. By inspecting these two curves individually we gain more insight into the cause and structure of volatility. Several examples are provided where the tick-time volatility curve is much smoother than the clock-time volatility curve meaning that the major part of fluctuations in clock-time volatility is due to fluctuations of the trading intensity. Since microstructure noise only influences the estimation of the (smooth) tick-time volatility curve, the findings lead to an improved pre-averaging estimator of spot volatility. This is reflected by a better rate of convergence of the estimator. The asymptotic properties of the estimators are derived by an infill asymptotic approach.

EO0829: **The arrow of time in multivariate time series**
Presenter: **Jonas Peters**, MPI for Intelligent Systems, Germany
Co-authors: Bernhard Scholkopf, Stefan Bauer
A time series satisfying a linear multivariate autoregressive moving average (VARMA) model is proved to satisfy the same model assumption in the reversed time direction, too, if all innovations are normally distributed. This reversibility breaks down if the innovations are non-Gaussian. Therefore, under the assumption of a VARMA process with non-Gaussian noise, the arrow of time becomes detectable. This result extends earlier work for one-dimensional time series. We present a practical algorithm that estimates the time direction for a finite sample and prove its consistency. An application to real world data from economics shows that considering multivariate processes instead of univariate processes can be beneficial. Our work relates to the concept of causal inference, where recent methods exploit non-Gaussianity of the error terms for causal structure learning.

EO039 Room Board meeting room I STATISTICAL INFERENCE FOR NETWORKS Chair: Jens-Peter Kreiss

EO1028: **Block models for clustering directed networks and nonsymmetric data matrices**
Presenter: **Cheryl Flynn**, AT&T Labs Research, United States
Much of the work on network clustering (community detection) focuses on symmetric, binary networks. In many real-world applications, including recommender systems and communications, it is more natural to treat the underlying network as directed and weighted. How to extend the block models and the popular profile likelihood-based clustering method to arbitrary nonsymmetric data matrices. This leads to a biclustering procedure with similar consistency properties to the analogous method for symmetric binary networks. The method will be demonstrated with applications to congressional voting data and microarray analysis.

EO1224: **Summaries of network characteristics**
Presenter: **Sofia Olhede**, University College London, United Kingdom
Co-authors: Patrick Wolfe
Networks are highly non-Euclidean structures. This makes it difficult to construct appropriate summaries from network data that are interpretable and strongly related to the generating mechanism of the data. We will discuss how to address the summarisation of network data, and blocks to analysis, including computational efficiency. We will also consider repeated network observations, and their appropriate analysis.

EO1363: **Convergence rate for the degree distribution in a dynamic network model**
Presenter: **Dominic Schuhmacher**, University of Goettingen, Germany
Co-authors: Fabian Kueck
In the continuous-time dynamic network model by Britton and Lindholm the number of nodes evolves according to a supercritical linear birth-and-death process. At the birth of a node a random parameter is assigned that controls the rate at which links to other nodes are established. Links break if one of their nodes dies or after an independent exponentially distributed time. In this model the degree distribution of a node picked uniformly at random at fixed time T is a mixed Poisson distribution. We use this to derive a rate of convergence in total variation distance towards a limit mixed Poisson distribution as $T \to \infty$. The main ingredient of the proof is a new result about the age distribution of a node picked uniformly at random.

EO1367: **Estimating the fibre length distribution in fibre reeinforced polymers**
Presenter: **Jan Niedermeyer**, TU Kaiserslautern, Germany
Fibre reeinforced polymers are of great import for many modern applications. The mechanical properties of fibre reinforced polymers is governed by the length and orientation of the fibres embedded in the polymer matrix. We wish to estimate the length distribution from data provided by CT images. For this we have to deal with problems like sampling bias and censoring. We propose a method based on fully segmented fibres and the EM algorithm to estimate the length distribution.

EO1404: **Spectral clustering for dynamic stochastic block models**
Presenter: **Sharmodeep Bhattacharyya**, Oregon State University, United States
Co-authors: Shirshendu Chatterjee
One of the most common and crucial aspect of many network data sets is the dependence of network link structure on time or other attributes. There is a long history of researchers proposing networks for dynamic time-evolving formation of networks. Most complex networks, starting from biological networks like genetic or neurological networks to social, co-authorship and citation networks are time-varying. This has led the researchers to study dynamic, time-evolving networks. We consider the problem of finding a common clustering structure in time-varying networks. We consider three simple extension of spectral clustering methods to dynamic settings and give theoretical justification that the spectral clustering methods produce consistent community detection for such dynamic networks. We also propose an extension of the static version of nonparametric latent variable models to the dynamic setting and use a special case of the model to justify the spectral clustering methods. We show the validity of the theoretical results via simulations too and apply the clustering methods to real-world dynamic biological networks.

EO117 Room 203 FUNCTIONAL DATA ANALYSIS Chair: Alicia Nieto-Reyes

EO1178: **Separating large and small scale variation in discretely observed functional data**
Presenter: **Victor Panaretos**, EPFL, Switzerland

Functional data analyses typically proceed by smoothing, followed by functional PCA. This paradigm implicitly assumes that any roughness is due to nuisance noise. Nevertheless, relevant functional features such as time-localised or short scale variations may indeed be rough. These will be confounded with the smooth components of variation by the smoothing/PCA steps, potentially distorting the parsimony and interpretability of the analysis. We consider the problem of recovering both smooth and rough variations on the basis of discretely observed functional data. Assuming that a functional datum arises as the sum of two uncorrelated components, one smooth and one rough, we develop identifiability conditions for the estimation of the two corresponding covariance operators. The key insight is that they should possess complementary forms of parsimony: one smooth and of finite rank (large scale), and the other banded and of arbitrary rank (small scale). Our conditions elucidate the precise interplay between rank, bandwidth, and grid resolution. We construct nonlinear estimators of the smooth and rough covariance operators and their spectra via matrix completion, without assuming knowledge of the true bandwidth or rank; we establish their consistency and rates of convergence, and use them to recover the smooth and rough components of each functional datum, effectively producing separate functional PCAs for smooth and rough variation.

EO0589: **Least trimmed squares for functional principal component analysis**
Presenter: **Stefan Van Aelst**, University of Leuven, Belgium
Co-authors: Holger Cevallos Valdiviezo, Matias Salibian-Barrera

Classical functional principal component analysis can yield erroneous approximations in presence of outliers. To reduce the influence of atypical data we propose two methods based on trimming: a multivariate least trimmed squares (LTS) estimator and a componentwise variant. The multivariate LTS minimizes the least squares criterion over subsets of curves. The componentwise version minimizes the sum of univariate LTS scale estimators in each of the components. In general the curves can be considered as realizations of a random element on a separable Hilbert space. For a fixed dimension q, we then aim to robustly estimate the q-dimensional linear subspace that gives the best approximation to the functional data. Our estimators use smoothing to first represent irregularly spaced curves in a high-dimensional space and then calculates the LTS solution on these multivariate data. The solution of the multivariate data is subsequently mapped back onto the Hilbert space. Poorly fitted observations can then be flagged as outliers. A simulation study and real data applications show that our estimators yield competitive results, both in identifying outliers and approximating regular data when compared to other existing methods.

EO0245: **Multi-dimensional functional principal component analysis**
Presenter: **Ci-Ren Jiang**, Academia Sinica, Taiwan
Co-authors: Lu-Hung Chen

Functional principal component analysis is one of the most commonly employed approaches in functional and longitudinal data analysis and we extend it to analyze functional/longitudinal data observed on a general d-dimensional domain. The computational issues emerging in the extension are fully addressed with our proposed solutions. The local linear smoothing technique is employed to perform estimation because of its capabilities of performing large-scale smoothing and of handling data with different sampling schemes (possibly on irregular domain) in addition to its nice theoretical properties. Besides taking the fast Fourier transform strategy in smoothing, the modern GPGPU (general-purpose computing on graphics processing units) architecture is applied to perform parallel computation to save computation time. To resolve the out-of-memory issue due to large-scale data, the random projection procedure is applied in the eigendecomposition step. We show that the proposed estimators can achieve the classical nonparametric rates for longitudinal data and the optimal convergence rates for functional data if the number of observations per sample is of the order $(n/\log n)^{d/4}$. Finally, the performance of our approach is demonstrated with simulation studies and the fine particulate matter (PM 2.5) data measured in Taiwan.

EO1171: **Outlier detection using random projections**
Presenter: **Paula Navarro**, Universidad de Cantabria, Spain
Co-authors: Juan A Cuesta-Albertos, Alicia Nieto-Reyes

Outlier detection is an important aspect of any statistical analysis of data. There exist multiple methods to detect outliers both for high-dimensional and functional data in the literature. An extension of a procedure to detect outliers in high-dimensional data to the funcional setting is explored. The procedure uses several one-dimensional random projections and applies an appropriate univariate outlier detection method in each projection. As a novelty, to select the number of the required directions, we use sequential analysis to ensure that the expected number of directions is relatively small in comparison with methods which fix this number beforehand to achieve the same power. Among the advantages of the proposed method, we want to emphasize the fact that some existing outlier detection methods use boostrap procedures in the calibration step which require the estimation of a high-dimensional covariance matrix. Due to the fact that our method is based on one-dimensional procedures, it does not require the estimation of these matrices. Some simulated and real datasets are studied to illustrate the method.

EO0819: **A simulation-based comparative study of functional linear regression models on remote sensing data**
Presenter: **Nihan Acar-Denizli**, Mimar Sinan Fine Arts University, Turkey
Co-authors: Pedro Delicado, Gulay Basarir, Isabel Caballero de Frutos

Functional data analysis is an important method to analyse spectral data measured on a spectrum that consists of different frequency channels. Remote Sensing (RS) data obtained from satellites are an example of spectral data. Particulary in oceanography, RS data are used to predict ocean characteristic parameters such as Sea Surface Temperature (SST), Chlorophyll-a content (Chl-a) and Total Suspended Solids (TSS). Functional Linear Regression Models (FLRM) are applied on a real data set in oceanography to predict the amount of TSS in the coastal zone adjacent to the Guadalquivir estuary. A simulation study is designed to measure the performance of FLRM in predicting a scalar response generated by using different models. The simulated models were compared in terms of Median Error of Prediction (MedEP).

EP002 Room Hall POSTER SESSION I Chair: Elena Fernandez

EP0888: **Asymptotic efficient estimation under random censorship models**
Presenter: **Mareike van Heel**, Universiteit Hasselt/ FH Aachen, Germany

The aim is to prove that under the random censorship the Kaplan-Meier integral estimators $\int \varphi(z)F^{KM}(dz)$ are asymptotically efficient with respect to all regular estimators of $\int \varphi(z)F(dz)$, where φ is an arbitrary Borel-measurable function. We calculate a lower bound for the variance of any regular, asymptotic linear estimator of $\int \varphi(z)F(dz)$ and show that this bound is equal to the asymptotic variance of the classical Kaplan-Meier integral estimators.

EP1475: **Centralized fusion estimation from multisensor observations with correlated noises and random transmission failures**
Presenter: **Raquel Caballero-Aguila**, Universidad de Jaen, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez

The study is concerned with the least-squares centralized fusion estimation problem of discrete-time stochastic signals from measured outputs obtained by multiple sensors, which are transmitted to a processing center, where all the information is combined to address the estimation problem. It is assumed that the additive noises of the measured outputs are one-step autocorrelated and cross-correlated. Moreover, due to random transmission failures, some of the data packet processed for the estimation may either contain only noise (uncertain observations), be delayed (randomly delayed observations) or even be definitely lost (random packet dropouts). These random transmission uncertainties are modelled by independent sequences of Bernoulli random variables with different probabilities for the different sensors. By an innovation approach and using the last observation that successfully arrived if a packet is lost, recursive estimation algorithms are designed for the centralized filtering and fixed-point smoothing problems. The proposed algorithms are easily implementable and the signal evolution model is not required, as only the mean and covariance functions of the processes involved are used. The accuracy of the proposed estimators, which is measured by the estimation error covariances, is examined by a numerical simulation example.

EP1487: **Centralized estimators in networked systems with bounded random delays and packet dropouts**
Presenter: **Aurora Hermoso-Carazo**, Universidad de Granada, Spain
Co-authors: Josefa Linares-Perez, Raquel Caballero-Aguila

Fusion techniques have been widely used to address signal estimation problems in networked stochastic systems. Centralized fusion estimators are obtained by processing the measurements received from all sensors in the fusion center. The unreliability and limited transmission capacity of the communication channels between the sensors and the processing center encourage the necessity of considering random delays and packet losses in data transmission. The study is concerned with the centralized estimation problem for discrete-time signals in sensor networks with bounded random delays, which can produce packet dropouts in the data transmission. The multiple-step random delays are modeled by using a collection of Bernoulli variables with known distributions and different characteristics at each sensor. Under these conditions, without requiring the knowledge of the signal evolution model, but only the first and second order moments of the processes involved, recursive algorithms for the optimal (least-squares) linear centralized estimators are derived by an innovation approach. The proposed estimators depend on the delay probabilities at each sampling time, but do not to need to know if a particular measurement is delayed or well-timed. The effects of the delays in the estimation accuracy are analyzed in a numerical example.

EP1408: **Bandwidth matrix selector for multivariate kernel regression**
Presenter: **Ivanka Horova**, Masaryk University, Czech Republic
Co-authors: Jan Kolacek

The bandwidth matrix controls both the amount and the direction of multivariate kernel smoothings and thus its choice is particularly important. Since a theoretical optimal bandwidth matrix depends on the unknown regression function to be estimated data-driven methods should be applied. A presented method is based on an optimally balanced relation between the integrated variance and integrated squared bias. The analysis of statistical properties confirms the rationale (usefulness) of this method. The performance of the method is illustrated by means of a simulation study. Application to real data is also included.

EP1476: **Distributed filtering in connected sensor networks from measurements with random matrices and transmission failures**
Presenter: **Josefa Linares-Perez**, Universidad de Granada, Spain
Co-authors: Raquel Caballero-Aguila, Aurora Hermoso-Carazo

Recently, the research on fusion estimation problems in networked systems is especially being focused on sensor networks with a given topology, to consider the possibility of coordination among the sensors to perform global tasks by the exchange of information with neighboring sensor nodes. Also, the use of random matrices is gaining a great interest as they provide a unified framework to model different random network-induced phenomena, such as stochastic sensor gain degradation, measurement multiplicative noises or missing measurements. We consider the estimation problem of discrete-time signals from measurements perturbed by random parameter matrices and white additive noises. These measurements are obtained by multiple sensors, located at the nodes of a network, which are connected according to a specified network topology. This is modeled by a directed graph and both one-step delays and packet dropouts are assumed to occur randomly during the data transmission through the network communication channels. We address the so-called distributed estimation problem, in which the signal is estimated at each node using its own measurements and those from its neighbors. Using covariance information, without requiring the evolution model generating the signal process, a recursive algorithm for the distributed least-squares linear filter is derived by an innovation approach.

EP1504: **D-optimal and quasi-optimal design for the monomer and dimer in the Adair model**
Presenter: **Maria de las Mercedes Rodriguez Hernandez**, Universidad de Castilla la Mancha (UCLM), Spain
Co-authors: Jesus Lopez-Fidalgo

The Adair equation is used to model biological macro-molecule reactions. In this equation the saturation rate is modeled through the free ligand concentration. This is not a variable under the control of the experimenter. As a matter of fact the ligand is a random variable depending on an initial ligand added by the experimenter. This is what can be designed, but the dependence of the saturation rate on the initial ligand has not been considered in the literature. We extended a model based on the Adair model of first order will be derived in order to obtain a proper model fitting an optimal designs for that. Optimal designs will be computed as well as seven point quasi-optimal designs forced to follow a harmonic, geometric or uniform progression. The parameters are estimated and compared for simulated data from these designs. A sensitivity analysis against the choice of the parameters is performed for the modfied Adairmodel. Finally and based on the good eficiencies obtained for the designs for the original first order model designs are obtained numerically for the second order model.

EP1601: **Bootstrap control charts for INR control**
Presenter: **Susana Rafaela Martins**, Universidade do Minho - Centro Algoritmi, Portugal
Co-authors: Lino Costa, Pedro Oliveira

Atrial fibrillation is the most common cardiac arrhythmia of these days particularly in elderly adults. It is characterized by an abnormal operation in the heart and it affects the patients life. Currently, the anticoagulant treatment using warfarin is the fully validated treatment to prevent the health problems. This therapy is the most used because measuring blood coagulation levels with International Normalized Ratio (INR) and through the ingestion of warfarin is possible control the treatment. Sometimes it is not easy to understand if treatment is under control, therefore, control charts are an important tool to assess whether or not the treatment is controlled. The goal is to study the application of bootstrap control charts to monitor INR levels. Bootstrap is a statistical technique that does not rely on any distributional assumptions since estimates the sampling distribution based on data. Bootstrap control charts performance was evaluated in statistical and medical terms and compared with Shewhart, EWMA, and CUSUM charts. For this purpose, data collected from a public hospital, and simulated data generated from a normal distribution assuming different parameters were considered. The advantages and drawbacks of each type of control chart are discussed.

EP1626: **Estimation of a bivariate convex function**
Presenter: **Seongoh Park**, Seoul National University, Korea, South

Three procedures are considered to estimate a bivariate convex function, and compare them numerically with synthetic and real data examples. The procedures considered are the constrained least square method, adaptive partitioning method, and two-stage procedure based on uniform approximation.

EP1650: **Optimal allocation in the item sum technique**
Presenter: **Beatriz Cobo**, University of Granada, Spain
Co-authors: Maria del Mar Rueda, Pier Francesco Perri

The item sum technique is a recent survey indirect questioning mode to elicit sensitive information on a quantitative variable and obtain estimates of parameters of interest. In its original form, the technique requires the selection of two independent samples. Units belonging to one of the two samples are presented with a Long List (LL) of items containing a sensitive question and G nonsensitive questions; units of the other sample receive only a short list (SL) of items consisting of the same nonsensitive questions in the long list. All the questions refer to quantitative variables possibly measured on the same scale of the sensitive one. The respondents are asked to report the total score of the answers to all the questions in their list without revealing the individual score of each question. The mean difference of answers between the LL-sample and the SL-sample is then used as an unbiased estimator of the population mean of the sensitive variable. We consider the problem of the optimal allocation under cost constraints in the two samples in order to minimize the variance of the estimator of the mean of the sensitive variable. Theoretical results are investigated on the basis of a simulation study based on real data from a survey on the use of cannabis in the University of Granada. Optimal allocation is also theoretically and computationally discussed in the case of multiple sensitive variables where the selection of more than two samples is required.

EP1765: **Statistical analysis of the relationship between fluvial sediment and coastal dunes in the Cantabrian coast**
Presenter: **Elena Fernandez**, University of Oviedo, Spain
Co-authors: Jorge Marquinez, Gil Gonzalez-Rodriguez

During the last decades, fluvial systems of the Cantabrian region (Northwest of Spain) have experienced important geomorphological changes. These changes are characterized by main channel incision and narrowing, vegetal colonization of fluvial bars and channel morphology simplification. This trend has been described in other spanish and Atlantic European rivers. Some authors suggest that these changes are caused by a general decrease in sediment supply rates. Current data available for Cantabrian basins are consistent with this interpretation, but no systematic analysis has been carried out yet to support this hypothesis. In this context, evidences of this decrease in the sediment supply could find in the extent variations of coastal beaches and dunes, which are built mainly with the sand that comes from the rivers. Databases about waves height, fluvial flows and changes in the extension of dunes in the West coast of Asturias have been analyzed by fitting different regression models using R to identify trends in the sediment supply from the rivers and to discuss the influence of other marine variables, taking in account the potential impact of the dams and sea level changes related with the climatic change.

Saturday 10.12.2016 08:40 - 10:20 Parallel Session G – CFE-CMStatistics

CI681 Room Graduation hall NONCAUSAL TIME SERIES MODELS Chair: Alain Hecq

CI1280: **Detecting co-movements in asymmetric cycles: A noncausal time series approach**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Alain Hecq, Lenard Lieb, Sean Telg
The notion of common noncausal feature is introduced and tools are proposed for detecting the presence of co-movements in stationary economic and financial time series such as variables with asymmetric cycles or bubbles. For purely noncausal models, i.e. forward looking VARs, we estimate reduced rank regressions in reverse time in order to highlight the potential presence of such noncausal co-movements. For more than one lead or lag, we are able to determine whether the VAR is better represented by purely causal or purely noncausal reduced rank models. Using both sets of lag and lead instruments within a canonical correlation or a GMM framework, additional relationships are discovered between series, both in the Monte Carlo simulations and in empirical illustrations. For mixed causal-noncausal models though, an approximate maximum likelihood estimator assuming non Gaussian disturbances is needed.

CI1287: **A new time-varying parameter autoregressive model for U.S. inflation expectations**
Presenter: **Markku Lanne**, University of Helsinki, Finland
Co-authors: Jani Luoto
The evolution of U.S. inflation is studied by means of a new noncausal autoregressive model with time-varying parameters that outperforms the corresponding constantant-parameter causal and noncausal models in terms of fit and forecast accuracy. We find the coefficients of both lagged and expected inflation in the new Keynesian Phillips curve based on our model constant in time and important although the latter dominates in determining current inflation. The implied trend inflation estimate evolves smoothly and is well aligned with survey expectations, and its variation is found to follow from the underlying marginal cost rather than time-varying steady state inflation.

CI1403: **Semi-parametric estimation of noncausal vector autoregression**
Presenter: **Christian Gourieroux**, University of Toronto and CREST, Canada
Co-authors: Joann Jasiak
Consistent semi-parametric estimation methods are introduced for mixed causal/noncausal multivariate non-Gaussian processes. We show that in the VAR model, the second-order identification is feasible to some extent, and it is possible to distinguish the mixed processes with different numbers of causal and noncausal components. For detecting the causal and noncausal components, we introduce a semi-parametric estimation method that does not require any distributional assumptions on the errors. This method is based on the generalized covariance estimator of the autoregressive coefficients. Although this estimator is not fully efficient, it provides the estimates in one single optimization, while the MLE would assume a parametric speciffication and require a number of optimizations equal to the number of all possible causal dimensions. The method is illustrated by a simulation study and applied to commodity prices.

CO327 Room 112 RECENT ADVANCES IN NONLINEAR AND NONSTATIONARY TIME SERIES Chair: Weining Wang

CO0181: **Functional coefficient time series models with trending regressors**
Presenter: **Tingting Cheng**, Nankai University, China
A functional coefficient time series model with trending regressors is studied, where the coefficients are unknown functions of time and random variables. We propose a local linear estimation method to estimate the unknown coefficient functions. An asymptotic distribution of the proposed local linear estimator is established under mild conditions. For practical use, we further propose a Bayesian approach to select bandwidths involved in this local linear estimator. Several numerical examples are provided to examine the finite sample performance of the proposed local linear estimator. The results show that the local linear estimator works well and the Bayesian bandwidth selection method is better than cross-validation method. Furthermore, we employ the functional coefficient model to study the relationship between consumption per capita and income per capita in U.S. and the results show that functional coefficient model with our proposed local linear estimator and Bayesian bandwidth selection method performs best in both in sample fitting and out of sample forecasting.

CO0188: **Monitoring multivariate variance changes**
Presenter: **Dominik Wied**, University of Cologne, Germany
Co-authors: Katharina Pape, Pedro Galeano
A model-independent multivariate sequential procedure is proposed to monitor changes in the vector of componentwise unconditional variances in a sequence of $p-$variate random vectors. The asymptotic behavior of the detector is derived and consistency of the procedure stated. A detailed simulation study illustrates the performance of the procedure confronted with different types of data generating processes. We conclude with an application to the log returns of a group of DAX listed assets.

CO0802: **RKHS-based approach to SCAD-penalized regression in high-dimensional partially linear models**
Presenter: **Jiajing Sun**, University of Chinese Academy of Sciences, China
Co-authors: Wenquan Cui, Haoyang Cheng
The high-dimensional partially linear models relevant for the problem of simultaneous variable selection and estimation, are studied under the assumptions that the function of the nonparametric part is derived from a reproducing kernel Hilbert space (RKHS) and the vector of regression coefficients for parametric component is sparse. A double penalties is provided to deal with the problem, with the roughness penalty of squared semi-norm on RKHS deployed to estimate the nonparametric component and the SCAD penalty used to achieve sparsity in the parametric part. Under some regular conditions, we establish the rate of convergence and consistency of the parametric estimation together with the consistency of variable selection. Furthermore, the presented estimators of the non-zero coefficients are shown to have the asymptotic oracle property. They are asymptotically normal with the same means and covariances that they would have if the zero coefficients were known in advance. Simulations are conducted to demonstrate the performance of the proposed method. Empirical analysis is also included.

CO0830: **Testing rational bubbles using high-frequency financial data**
Presenter: **Yang Zu**, University of Nottingham, United Kingdom
Within a model of continuous-time Ornstein-Uhlenbeck process with stochastic volatility, we develop a novel test for explosive behavior in daily level financial asset prices. The test exploits the information in intraday high-frequency data,it consists of devolatizing daily price increments with the daily realized volatility measures and performing the classical supreme-type recursive Dickey-Fuller test on the reconstructed devolatized sample. The asymptotic distribution is derived. The test is easy to implement. In finite sample simulations, the proposed test shows good size and remarkable improvement in the power as compared to existing tests under time-varying volatility. The test is applied to a up-to-date OMI Realized Library dataset. We find some results for the explosive behavior in the major equity markets and pose warnings for bubbles.

CO331 Room 103 REGIONAL MACROECONOMICS Chair: Michael Owyang

CO0202: **Local and aggregate fiscal policy multipliers**
Presenter: **William Dupor**, Federal Reserve Bank of St. Louis, United States
Co-authors: Bill Dupor
The effect of defense spending on the U.S. macroeconomy since World War II is estimated. First, a new panel dataset of state-level federal defense contracts is constructed. Second, observations across states are summed and, using the resulting time series, the aggregate effect of defense spending on national income and employment is estimated via instrumental variables. Third, local multipliers are estimated using the state-level data, which measures the relative effect on economic activity due to relative differences in defense spending across states. Comparing the aggregate and local multiplier estimates, it is found that the two differ dramatically. It is inferred that the local multiplier estimates alone do not provide useful information about the aggregate effects of policy. Finally, the panel aspect of the data are used to dramatically increase the precision of estimates of the aggregate multiplier (relative to using the aggregate data alone) by including a spillover term in the panel regressions. The baseline aggregate findings are a long-run multiplier on income equal to 1.6, a moderate long-run effect on employment, and no effect on income or employment effect in the short run. The results suggest that lags in the effects of defense spending are so long that they render countercyclical spending policies ineffective. In addition, negative short-run spillovers on employment of spending across state borders are found.

CO0203: **A state-level analysis of Okun's law**
Presenter: **Amy Guisinger**, Lafayette College, United States
Co-authors: Michael Owyang
Okun's law is an empirical relationship that measures the correlation between the deviation of the unemployment rate from its natural rate and the deviation of output growth from its potential. We estimate Okun's coefficients for each U.S. state and examine the potential factors that explain the heterogeneity of the estimated Okun relationships. We find that indicators of more flexible labor markets (higher levels of education achievement in the population, lower rate of unionization, and a higher share of non-manufacturing employment) are important determinants of the differences in Okun's coefficient across states. Finally, we show that Okun's relationship is not stable across specifications, which can lead to inaccurate estimates of the potential determinants of Okun's coefficient.

CO0209: **Assessing state and national labor market conditions**
Presenter: **Laura Jackson Young**, Bentley University, United States
Co-authors: Kevin Kliesen, Michael Owyang
Assessing the condition of the labor market is an important task in evaluating the overall state of the economy for both private decision-making and for policymaking. The labor market itself is a complex entity with workers flowing in and out of employment or participation, jobs being created and destroyed, and vacancies being posted and filled. A factor model was previously used to develop the Labor Market Conditions Index to summarize the common fluctuations in a vector of 19 national labor market indicators. However, labor markets tend to be regional entities with concentrated flows. This phenomenon was highlighted during the Great Recession when the economic downturn was accompanied by a housing downturn, making it difficult for workers to relocate to find new employment opportunities. Thus, it may also be important to analyze the fluctuations of local labor market variables and differentiate between large isolated fluctuations and large pervasive fluctuations affecting many different areas. We develop both national and state-level labor market indexes. The national index isolates and removes state-level fluctuations that are not pervasive—i.e., not occurring across many states simultaneously. We are also able to obtain state indexes that reflect the common movements of the panel of state-level labor market variables.

CO1001: **Regional beveridge curves**
Presenter: **Michael Owyang**, Federal Reserve Bank of St Louis, United States
Co-authors: Hannah Shell
The Beveridge curve - the relationship connecting the unemployment rate with job vacancies - is typically estimated at the national level. Labor markets, however, are largely local level phenomena, with usually only a small fraction of workers migrating from place to place to seek employment. Moreover, local labor markets have been found to exhibit some heterogeneity, depending in part on things like industrial composition and the local minimum wage. We construct MSA-level unemployment and vacancy data using techniques similar to those used in previous papers for the national level. We then estimate MSA-level Beveridge curves for 51 U.S. cities. Previous estimates of local Beveridge curves used time fixed effects as identifying restrictions, imposing that the slope of the Beveridge curve was invariant across locations. Our method identifies the slope of the Beveridge curve by restricting it to be constant for locations with similar labor market characteristics but allows it to vary across labor markets with, say, different industrial compositions. In this sense, we form regions with similarly sloped Beveridge curves. We then study whether the advent of the internet and subsequent adoption of online job postings has reduced the heterogeneity in local labor markets, reducing the number of regions.

CO341 Room 105 FORECASTING I Chair: Francesco Ravazzolo

CO0237: **A new monthly indicator of global real economic activity**
Presenter: **Joaquin Vespignani**, University of Tasmania, Australia
In modelling macroeconomic time series, often a monthly indicator of global real economic activity is used. We propose a new indicator, named World steel production, and compare it to other existing indicators, precisely the Kilians index of global real economic activity and the index of OECD World industrial production. We develop an econometric approach based on desirable econometric properties in relation to the quarterly measure of World or global gross domestic product to evaluate and to choose across different alternatives. The method is designed to evaluate short-term, long-term and predictability properties of the indicators. World steel production is proven to be the best monthly indicator of global economic activity in terms of our econometric properties. Kilians index of global real economic activity also accurately predicts World GDP growth rates. When extending the analysis to an out of sample exercise, both Kilians index of global real economic activity and the World steel production produce accurate forecasts for World GDP, confirming evidence provided by the econometric properties. Specifically, a forecast combination of the three indices produces statistically significant gains up to 40% at nowcast and more than 10% at longer horizons relative to an autoregressive benchmark.

CO1100: **Using analysts' forecasts for stock predictions: An entropic tilting approach**
Presenter: **Christoph Frey**, University of Konstanz, Germany
Predictive density forecasts for US stock returns from Bayesian vector autoregressions are combined with financial analysts' forecasts via entropic tilting. The predictive density of the asset returns is modified to match certain moment conditions that are formed based on average analysts' forecasts, for example, sell and buy recommendations or target prices. The advantage of this approach is that we can combine model-based time-series information with external, forward-looking information in a parsimonious way using closed-form solutions. We show that the tilting approach based on the (possibly biased) professional forecasts leads to an increase in prediction accuracy for both point and density forecasts. This also translates into portfolio gains for an investor who maximizes her expected utility.

CO1490: **Spatio-temporal diffusion of US house prices with unknown spatial weights**
Presenter: **Achim Ahrens**, Heriot-Watt University, United Kingdom

The aim is to examine the relationship between house prices and their fundamental drivers, such as real per capita income, interest rates and inflation, using a panel data set of US states over the sample period 1976Q1-2014Q4. The issue of spatial, or cross-sectional, dependence is addressed using a novel estimation method based on the Lasso estimator, a well-established regularisation technique. The vast majority of the spatial econometric literature assumes that the spatial weights matrix, which determines interaction effects between units, is known and, in practice, spatial weights are typically specified on an ad hoc basis using observable distance measures. The proposed method allows for unknown interaction effects and employs the Lasso estimator to appropriately control for spatial effects under the identification assumption of sparsity. The methodology is of more general interests for large T panels where spatial dependence is an issue or where the interest lies in modelling spatio-temporal diffusion processes. Estimation results show a long-run relationship between house prices and real per capita income. The most striking result, however, are strong and complex spatial effects that would not have been captured with standard specifications of the spatial weights matrix such as the binary contiguity matrix.

CC0259: **Forecasting UK house prices during turbulent periods**
Presenter: **Efthymios Pavlidis**, Lancaster University Management School, United Kingdom
Co-authors: Alisa Yusupova

The latest boom and bust in housing markets and its role in the Great Recession has generated a vast interest in the dynamics of house prices. A substantial empirical literature has developed that deals with predicting future house price movements. This literature concentrates almost entirely on the US, leaving national and regional markets of other countries, where housing has also played a central role, mostly unexplored. We contribute to this literature by conducting an extensive investigation of the ability of a battery of econometric models to forecast UK national and regional housing prices over the last two decades. The econometric models considered include ARDL, BVAR, Factor Augmented BVAR, TVP-VAR, DMA, DMS, and a previous DSGE model. In summary, our findings suggest that models that allow both the underlying specification and the parameter estimates to vary over time produce more (and, in some cases, dramatically more) accurate forecasts than methods where the number of predictors is kept fixed.

CO567 Room Board meeting room II PERSISTENCE CHANGE AND STRUCTURAL BREAKS Chair: Paulo Rodrigues

CO0393: **Testing shock induced asymmetries under unknown form of conditional heteroskedastisity**
Presenter: **Nazarii Salish**, BGSE and University of Cologne, Germany

A Lagrange multiplier test statistic is developed, and its variants to test for the null hypothesis of no asymmetric effects of shocks on time series are analyzed. In asymmetric time series models that allow for different responses to positive and negative past shocks the likelihood functions are, in general, non-differentiable. By making use of the theory of generalized functions Lagrange multiplier type tests and the resulting asymptotics are derived. The test statistics possess standard asymptotic limiting behavior under the null hypothesis. Monte Carlo experiments illustrate the accuracy of the asymptotic approximation and show that conventional model selection criteria can be used to estimate the required lag length. We provide an empirical application to the U.S. unemployment rate.

CO0612: **A structural change test in duration of bull and bear markets**
Presenter: **Joao Nicolau**, ISEG and CEMAPRE, Portugal

A recursive test is proposed, derived from the fluctuations test of Ploberger-Kramer-Kontrus, with a finite sample adjustment, to test possible structural changes in duration of bull and bear markets. Using the Dow Jones Industrial Average index, we detected a single structural break in the bull market duration in April, 1942, but none in bear market duration. Possible explanations for this occurrence are discussed.

CO0697: **Tests for segmented cointegration**
Presenter: **Luis Filipe Martins**, ISCTE-IUL, Portugal
Co-authors: Paulo Rodrigues

There is a growing literature that documents that the persistence of economic time series may change over time. Hence, it is natural to expect that these changes in persistence may also lead to changes in the long-run equilibrium of economic time series. According to previous research, failure to find a unique cointegration relationship between economic time series may be due to, for instance, the testing procedures used, the span of the data set analysed, the choice of the lag length in generating the test statistics, the presence of structural breaks and the presence of cointegration only beyond some threshold. This led to introduce the concept of regime-sensitive cointegration, according to which the underlying series need not be cointegrated at all times. The purpose is to evaluate existing cointegration test procedures and to propose new ones that can be used in the context of segmented cointegration.

CO1189: **CUSUM based ratio tests for parameter constancy: With application to variance stability**
Presenter: **Paulo Rodrigues**, Universidade Nova de Lisboa, Portugal
Co-authors: Uwe Hassler, Mehdi Hosseinkouchack, Antonio Rubia

New CUSUM based ratio tests for parameter constancy are introduced following a class of ratio tests recently proposed, which are based on a Karhunen-Loeve expansion. Analytical results as well as an in-depth Monte Carlo analysis is performed to evaluate the finite sample performance of the procedure and a comparison with available procedures is also provided. The new procedure presents interesting finite sample size and power behaviour. An empirical application is also included.

CO287 Room 102 BAYESIAN METHODS IN ECONOMETRICS Chair: Enrique ter Horst

CO0557: **Bayesian nonparametric estimation of ex-post variance**
Presenter: **John Maheu**, McMaster University, Canada
Co-authors: Jim Griffin, Jia Liu

Variance estimation is central to many questions in finance and economics. Until now ex-post variance estimation has been based on infill asymptotic assumptions that exploit high-frequency data. A new exact finite sample approach is offered to estimating ex-post variance using Bayesian nonparametric methods. In contrast to the classical counterpart, the proposed method exploits pooling over high-frequency observations with similar variances. Bayesian nonparametric variance estimators under no noise, heteroskedastic and serially correlated microstructure noise are introduced and discussed. Monte Carlo simulation results show that the proposed approach can increase the accuracy of variance estimation. Applications to equity data and comparison with realized variance and realized kernel estimators are included.

CO0753: **Bayesian nonparametric time varying vector autoregressive models**
Presenter: **Maria Kalli**, Canterbury Christ Church University, United Kingdom
Co-authors: Jim Griffin

Although stationary time series models are theoretically appealing, macroeconomists consider them to be too restrictive. A popular alternative

framework is time varying vector autoregression, with or without stochastic volatility (TVP-VAR or TVP-SV-VAR). Under this framework the parameters of the stationary vector autoregressive (VAR) model are allowed to change over time. This accounts for nonlinearity in the conditional mean, and heteroscedasticity in the conditional variance. We considered a Bayesian nonparametric stationary VAR (BayesNP-VAR) model and found that it outperformed the TVP-SV-VAR in terms of out-of-sample prediction for monthly macroeconomic series from the USA and Eurozone. Our aim is to extend the Bayes NP-VAR model to a time varying parameter specification, creating a nonparametric, non-stationary.

CO1468: **A Bayesian time-varying approach to risk neutral density estimation**
Presenter: **Enrique ter Horst**, CESA, Colombia
Co-authors: Roberto Casarin, German Molina
The aim is to expand the literature of risk neutral density estimation across maturities from implied volatility curves, usually estimated and interpolated through cubic smoothing splines. The risk neutral densities are computed through the second derivative, which we extend through a Bayesian approach to the problem, featuring: (1) an extension to a multivariate setting across maturities and over time; (2) a flexible estimation approach for the smoothing parameter, traditionally assumed common to all assets, known and fixed across maturities and time, but now potentially different between assets and maturities, and over time; and (3) information borrowing about the implied curves and risk neutral densities not only across different option maturities, but also dynamically.

CO1673: **Modeling seasonality in high-frequency data**
Presenter: **Hedibert Lopes**, INSPER, Brazil
Co-authors: Ruey Tsay, Audrone Virbickaite
High-frequency time series data often exhibit seasonality. The transaction-by-transaction data of financial market show not only diurnal patterns but also weekly and annual seasonality. Similar features are observed in hourly measurements of particulate matter, e.g. PM2.5, in many locations. We investigate statistical models (both structural and reduced-form) that are flexible and can adequately describe the stochastic seasonality of such time series. Special attention is paid to the relatively weak low-frequency seasonality in the high-frequency data. Practical implications of the proposed models are discussed. We also consider the effects of outliers and missing values.

CO565 Room Board meeting room I JUMPS AND VOLATILITY Chair: Eduardo Rossi

CO0572: **Models for jumps in trading volume**
Presenter: **Eduardo Rossi**, EC-JRC, Italy
Co-authors: Paolo Santucci de Magistris
The specification of continuous-time models for high-frequency trading volume series is considered. Analogously to stock prices, the stochastic process for trading volume might be characterized by diffusive and jump components. First, we study the activity level of volume jumps using both parametric and nonparametric techniques. We also analyze if the activity of positive and negative moves in the log-trading volume may differ. Second, we consider alternative specifications of the trading volume dynamics, with finite and infinite variation processes. We also include in the analysis self exciting jump processes, also known as Hawkes processes. Indirect inference is used to estimate the models.

CO0578: **A model for jumps in volatility and volume**
Presenter: **Paolo Santucci de Magistris**, Aarhus University, Denmark
Co-authors: Eduardo Rossi
A multivariate model is proposed for stock returns, volatility and trading volume that allows for idiosyncratic and common jump terms, where the latter is typically associated with important news arrivals, such as significant announcements and/or extreme economic events. In particular, the jump terms are modeled a the sum of random variables where the number elements in the sum is Poisson distributed with time-varying intensity. The interaction between the jump terms makes the model able to provide high flexibility in explaining the contemporaneous presence of jumps and the dynamic dependence between return, volatility and volume. The model is estimated by maximum likelihood and it exploits the information coming from realized volatility measures based on high frequency data. The empirical results testify the ability of the proposed model to fit most of the features that characterize the volatility and the volume relationship.

CO0980: **A model-free option-implied volatility for forecasting returns and realized volatility**
Presenter: **Xingzhi Yao**, Lancaster University, United Kingdom
Co-authors: Marwan Izzeldin
The VIX has long been identified to suffer from non-trivial measurement errors due to flaws in implementation of the index by the Chicago Board Options Exchange (CBOE). This has stimulated the development of alternative measures such as the model-free implied volatility (MFIV) and corridor implied volatility (CX). The aforementioned estimators (MFIV and CX) are enhanced by using different interpolation-extrapolation techniques and highlight the importance of the risk-neutral measure in recovering extreme option prices. The refined measures are found superior in terms of forecasting realised volatility and predicting future returns as evident in both the Monte Carlo simulations and an empirical application. Our findings indicate that the tail risk captured by the measures of implied volatility is critical in predicting returns but hampers the forecasting power for future volatility. Our conclusions are robust across different forecasting models, market conditions, and the various procedures of extrapolating option prices.

CO1335: **Regimes in stochastic volatility**
Presenter: **Alessandro Rossi**, European Commission, Joint Research Centre, Italy
Co-authors: Christophe Planas, Eduardo Rossi
The regime-switching behavior in the stock market volatility is analyzed via a Markov switching specification of the log-stochastic volatility. The model is based both on realized measures and daily returns. The formers can biased due to the presence of microstructure noise but are more informative regarding the true latent integrated volatility whereas the latter even if less subject to microstructure noise have less information on the true volatility. The model is parameterized such that it can be considered as a dynamic mixture model, i.e. a conditionally Gaussian state space model. Dynamic mixture models typically include a continuous unobserved state vector, some discrete latent variables that control discontinuities or change-points, plus the model parameters. The results show that allowing for Markov-switching specification in the stochastic volatility process improves the fitting properties as long as the model's forecasting performance.

CO369 Room 104 INFERENCE IN TIME SERIES VOLATILITY MODELS Chair: Jean-Michel Zakoian

CO0587: **Inferring volatility dynamics and risk premia from the S& P 500 and VIX markets**
Presenter: **Elise Gourier**, Queen Mary University of London, United Kingdom
Co-authors: Markus Leippold, Chris Bardgett
The information content of the S& P 500 and VIX markets on the volatility of the S& P 500 returns is studied. We estimate a flexible affine model based on a joint time series of underlying indexes and option prices on both markets. An extensive model specification analysis reveals that jumps and a stochastic level of reversion for the variance help reproduce risk-neutral distributions as well as the term structure of volatility smiles and of variance risk premia. We find that the S& P 500 and VIX derivatives prices are consistent in times of market calm but contain conflicting information on the variance during market distress.

CO1210: **Cholesky-GARCH, theory and application to conditional beta**
Presenter: **Serge Darolles**, Paris Dauphine, France
Co-authors: Sebastien Laurent, Christian Francq
The class of Cholesky-GARCH models, based on the Cholesky decomposition conditional variance matrix, are studied. We first consider the one-step and multi-step QML estimators. We prove the consistency and the asymptotic normality of the two estimators and derive the corresponding stationarity conditions. We then show that this class of models is useful to estimate conditional betas and compare it to other approaches proposed in the financial literature. Finally, we use real data to show that our model performs very well compared to other multivariate GARCH models.

CO1175: **Goodness-of-fit tests for log and exponential GARCH models**
Presenter: **Christian Francq**, CREST and University Lille III, France
Co-authors: Jean-Michel Zakoian, Olivier Wintenberger
Goodness-of-fit tests and specification tests are studied for an extension of the log-GARCH model which is both asymmetric and stable by scaling. A Lagrange-Multiplier test is derived for testing the extended log-GARCH against more general formulations including the Exponential GARCH (EGARCH). The null assumption of an EGARCH is also tested. Portmanteau goodness-of-fit tests are developed for the extended log-GARCH. An application to real financial data is proposed.

CO1447: **Noncausal heavy-tailed AR(p) processes**
Presenter: **Sebastien Fries**, Crest, Paris-Saclay University, France
Co-authors: Jean-Michel Zakoian
The adjunction of a noncausal component to standard causal linear autoregressive processes often yields a better fit to economic and financial time series. The general framework of noncausal AR(p) with possibly asymmetric α-stable errors is investigated. The nonlinear causal dynamics is derived and shown to display quadratic GARCH effects in direct time. The existence of a unit root in this causal dynamics is of particular interest as it allows to exhibit linear noncausal processes which are stationary, and even positive, martingales. Finally, under the broader assumption that the errors belong to the domain of attraction of a stable distribution, it is shown that contrary to the OLS estimator, the LAD estimator of the autoregressive parameters is able to identify causal and noncausal structures.

CO636 Room 106 CONTINUOUS-TIME AND HIGH FREQUENCY ECONOMETRICS Chair: Ruijun Bu

CO0673: **Term structure of forward moments and equity premium predictability**
Presenter: **Abderrahim Taamouti**, Durham University Business School, United Kingdom
A recently established aggregation property of the second and third moments of returns is considered to construct forward moments extracted from option prices. We show that according to standard affine no-arbitrage models, the forward moments should exhibit a factor structure, while the equity risk premium should also be an affine function of the same state variables. In light of this, we use dimensionality reduction techniques to extract, from the forward moments, the common factors that maximize the covariance between these factors and the equity premium. We show empirically that a small number of factors can explain the equity premium, both in-sample and out-of-sample, better than most traditional predictors. Moreover, we document that the inclusion of forward skewness into the analysis improves the asset return predictability and thus show that forward moments encapsulate important information about future market returns.

CO0346: **Modelling the relationship between future energy intraday volatility and trading volume with wavelet**
Presenter: **Fredj Jawadi**, University of Evry, France
Co-authors: Louhichi Wael, Zied Ftiti
Although there has been substantial research to explore the relationship between volatility and trading volume in stock markets, few researchers have investigated this relationship in energy markets. Moreover, previous studies did not describe its nature or impetus. We investigate this relationship using intraday data from the oil and gas markets and we extend the previous studies through the use of a frequency. More specifically, we employ a continuous wavelet transform to identify the lead-lag phase between volatility and volume. This framework supplants usual time series modelling, as it uses a measure of coherence for different frequencies and time-scales to capture further changes and time variation in the volume-volatility relationship. Our results provide supportive evidence for the well-known positive relationship between realized volatility and trading volume, thereby supporting the Mixture Distribution Hypothesis (MDH). In particular, our results show that volume causes volatility only during turbulent times, while volatility causes volume during good times. Furthermore, there is no relationship between volume and volatility in the long term, due to the absence of noise traders and liquidity traders in the long run. These findings are helpful for investors and policymakers as they contribute to better forecast the trading volume and price volatility during turbulent and calm periods and over several investment horizons.

CO0735: **Forecasting low-frequency return density using high-frequency information: MC simulation and FAR**
Presenter: **Minjoo Kim**, University of Glasgow, United Kingdom
Motivated by a stylized fact that intraday returns can provide additional information on the distribution of daily returns, we propose a novel forecasting framework combining MC simulation and Function Autoregressive Model (FAR) of nonparametric density functions. First, we simulate daily returns by a continuous-time data generating process (DGP). We estimate the DGP using intraday returns and it simulates daily returns. We then approximate the empirical density function of simulated daily returns relying on the nonparametric density estimator. We replicate this procedure for each day and the empirical densities are used as a functional sample. The advantage of this approach is that the continuous-time filter is able to control serial dependence often observed in the intraday data. Next, we simulate daily returns without requiring specific distribution family on the intraday returns. This MC stimulation is the most popular and standard in the financial engineering literature. Second, we apply FAR to the empirical densities. Based on the in-sample estimation, we forecast a daily return density function. The forecast can provide various information on the financial risk such as extreme event, event probability and interval. Finally, we evaluate our approach by both simulation study and empirical evaluations. Furthermore, we introduce its usage in financial applications.

CO0326: **Analysing market volatility dynamics: Evidence from a latent factor-based regime-switching continuous-time model**
Presenter: **Jie Cheng**, Xi-an Jiaotong-Liverpool University, China
Co-authors: Ruijun Bu
Understanding the volatility dynamics is crucial for derivative pricing, financial investment, and risk management. We analyze the behavior of the S&P 500 Volatility Index(VIX) by a novel latent-factor based regime-switching continuous-time approach. We assume that in each regime the VIX follows a continuous-time, nonlinear and non-Gaussian diffusion process and that the switching between regimes is driven by a latent time series potentially correlated with volatility shocks. Our focus is on the nonlinearity, long-run and short-run behaviors in the regime-dependent dynamics of the VIX, the endogeneity in regime changes, and finally the potential determinants of regime changes. Evidence from the VIX data at various frequencies all confirmed clear presence of endogenous regime-switching effects predominantly aecting the short-run behavior (i.e. the volatility) of VIX and also strong nonlinearity in its regime-dependent dynamics. To understand what is driving the regime changes and to what extent, we investigate the potential composition of the extracted regime-driving latent process by considering several regressions on a large dimension of market variables. Relying on several dimension reduction techniques, we find that as much as 40-65 percent of the variations in the latent process can be explained by observable economic and financial variables.

CO387 Room 111 PERSISTENCE AND ASYMMETRIES IN FINANCIAL TIME SERIES Chair: Josu Arteche

CO0786: **Outliers and misleading leverage effects in asymmetric GARCH-type models**
Presenter: **Ana Perez Espartero**, University of Valladolid, Spain
Co-authors: Angeles Carnero
It is known that outliers could mislead the estimation of both parameters and volatilities in symmetric GARCH models. We face the problem of how outliers affect the parameter estimation and test of significance in asymmetric GARCH models where the volatility response is different depending on the sign of past returns. We analyze the impact of positive and negative outliers on the leverage effect estimates in the context of the Threshold GARCH model. Different estimation methods are compared: Gaussian quasi-maximum likelihood (QML), quasi-maximum likelihood based on maximizing the Student likelihood and Least Absolute Deviations Estimation. We also look at the impact of outliers on testing the significance of the leverage effect parameter, when the usual t-test and the Likelihood Ratio test are used. As expected, inference based on Gaussian QML can be severely distorted by even a single outlier. Moreover, the presence of two consecutive outliers could lead to estimating either spurious asymmetries or asymmetries of the wrong sign and could make the tests fail to reject the null of no leverage effect even for moderately large samples. However, inferences based on robust estimators are more reliable. We illustrate these results with a series of daily DJIA returns.

CO0610: **A robust LM test for long memory**
Presenter: **Michael Will**, Leibniz University Hannover, Germany
Co-authors: Matei Demetrescu, Philipp Sibbertsen
Existing statistical tests for long memory in general require the assumption of at least conditional homoskedasticity in the error terms of the data generating process. Furthermore many parametric or frequency-domain based procedures display heavy size distortions if the analyzed model is misspecified. Therefore, we propose a nonparametric time-domain based test for long memory. Our test is of the Lagrange Multiplier type and therefore does not require any estimation of the memory parameter under the null. Due to its nonparametric nature, it does not require any specification of a model, making it robust to the influence of short-run dynamics. Moreover it displays robustness to conditional and unconditional heteroskedasticity (nonstationary volatility) in the error terms of the data generating process of quite general form. We illustrate the performance of our test by comparing it to other existing testing procedures in the literature via Monte Carlo simulation.

CO0615: **Assessment of Value-at-Risk estimation of long and short-memory GARCH-class using filtered historical simulation methods**
Presenter: **Pilar Grau**, Universidad Rey Juan Carlos, Spain
The use of both filtered historical simulation and bootstrap filtered historical simulation of Value-at-Risk (VaR) is proposed under different realistic generating processes that include short and long-memory. Various non-linear short and long-memory GARCH-class under three different density function, Gaussian, Student and skewed Student are used to evaluate the predictive performance of the method. Additionally, daily data on three well known active stock indices are used to empirically evaluate the VaR estimates. The predictive performance is evaluated in terms of different criteria, such as the tests of unconditional and conditional coverage and the independence test. Results from different models specifications and methods are ranked showing that VaR using bootstrap filtered historical simulation under long-memory GARCH-class models outperforms over short memory ones.

CO0456: **Estimation of the volatility in SV models using singular spectrum analysis**
Presenter: **Josu Arteche**, University of the Basque Country, Spain
Co-authors: Javier Garcia
One of the main difficulties that Stochastic Volatility models have to face when applied to financial time series is estimating the in-sample volatility. To solve this problem,a non-parametric strategy based on Singular Spectrum Analysis is proposed. Its main advantage is its generality because it does not impose any parametric restriction on the volatility component: only some spectral structure is needed to identify it separately from noisy components. Its convincing performance is shown in an extensive MonteCarlo analysis that includes stationary and nonstationary long memory, short memory and level shifts in the volatility component, which are models often used for financial time series. Its applicability is finally illustrated in a daily Dow Jones Industrial index series and an intraday series from the Spanish Ibex35 stock index.

CO487 Room 107 BUSINESS CYCLE ANALYSIS AND FORECASTING Chair: Gian Luigi Mazzi

CO0422: **The evolution of regional economic interlinkages in Europe**
Presenter: **Ana Gomez-Loscos**, Bank of Spain, Spain
Co-authors: Lola Gadea, Danilo Leiva-Leon
Changes, across time and space, in the synchronization of European regional business cycles are analyzed, and the role that the sectoral composition has played in explaining such changes is investigated. Motivated by the lack of high frequency data at the regional level, we propose a new method to measure time-varying synchronization in small samples that combines regime-switching models and dynamic model averaging. Our main findings can be summarized as follows: (i) in only two years the Great Recession synchronized Europe twice more that what the European Union process did in decades, (ii) Ile de France is the region acting as main channel in the transmission of business cycles shocks in Europe, (iii) increases in regional sectoral composition similarity have a positive eect on business cycles synchronization, only for regions that already experience high levels of similarity in their productive structure, and in particular, after the Great Recession.

CO0469: **Stochastic processes in discrete and continuous time**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom
The relationship is considered between discrete-time ARMA models and the corresponding stochastic differential equations in continuous time. The precise relationship depends on the assumptions that are made regarding the forcing function. In the classical theory of stochastic differential

equations, the forcing function is a continuous-time white-noise process, which is derived from the increments of a Weiner process that is unbounded in frequency. We develop a theory that accommodates forcing functions that are limited in frequency. The forcing function of a discrete-time ARMA model is limited in frequency to the Nyquist interval $[-p_i, p_i]$, where the frequency is measured in radians per unit interval. A one-to-one correspondence with a continuous model can be established. A continuous trajectory is created by replacing the sequence discrete-time ordinates by a superposition of suitably scaled sinc functions at unit displacements. The stochastic differential equation for such a frequency-limited continuous-time process can be derived readily. An application to business-cycle analysis is presented.

CO0931: **Structural analysis using factor augmented VARs and three-pass regression filters**
Presenter: **Anindya Banerjee**, University of Birmingham, United Kingdom
Co-authors: Massimiliano Marcellino, Igor Masten

A new approach is proposed to sign-restriction-based identification of structural shocks in FAVAR models. Using a new method to estimate factors called the three-pass regression filter (3PRF), sign restrictions are imposed on the responses of industrial production (IP) and consumer price index (CPI) in the first twelve months following a monetary policy (MP) shock. Each of the N variables in our U.S. dataset is then used individually as a proxy for the 3PRF factors and for each of these proxies up to seven factors are extracted and used in a BBE-type FAVAR model. From the models satisfying the sign restrictions we attempt to obtain the one that is most representative. We observe that the impulse responses of non-sign-restricted variables are in accord with economic intuition. Moreover, the adjustment of prices seems to be much faster (and without permanent effect on the price level) than in BBE or similar studies. A further step of our analysis compares the 3PRF responses with impulse responses based upon one to seven factors estimated from standard principal component analysis (PCA). The responses of IP are quite similar under both approaches but price puzzles emerge when looking at the responses of CPI using PCA. Looking through 3PRF lenses we argue that adding more information which may be irrelevant leads to biased responses and erroneous identification of shocks.

CO1105: **A multivariate system of turning points detection for all Euro area countries**
Presenter: **Rosa Ruggeri Cannata**, European Commission, Luxembourg
Co-authors: Gian Luigi Mazzi, Monica Billio

The recent 2008-2009 financial and economic crises, as well as the 2012-2013 recessions, have shown the importance of having an accurate system able to detect, almost in real-time, the occurrence of turning points. At the Euro area level, such a system has been developed by Eurostat in the last years and then extended to the largest Euro area economies. Nevertheless, the increasing idiosyncratic behaviour among the Euro area economies, started around 2008, does not guarantee that concentrating our attention to the euro area as a whole, or to its largest economies, provides a reliable picture of the cyclical situation. For this reason, we have decided to extend our turning points detection system to all Euro area countries starting from the same multivariate Markov-Switching based model. The main problems and challenges encountered when extending our system to all Euro area countries are discussed in details. In particular, we are focusing on data availability problems, on the complexity of properly identifying cyclical movements for small and very open economies, as well as on the short span of available data, at least for some countries.

CC735 Room 109 CONTRIBUTIONS IN MACROECONOMETRICS AND TIME SERIES I Chair: Thanasis Stengos

CC1457: **Evolving macroeconomic dynamics: A time-varying structural approach**
Presenter: **Michael Ellington**, University of Liverpool, United Kingdom
Co-authors: Costas Milas

A comprehensive reduced–form and structural analysis of evolving macroeconomic dynamics is provided using theoretically founded Divisia money aggregates and a time-series spanning the Great Recession. We fit a Bayesian time-varying parameter VAR model with stochastic volatility to US and UK data from 1979 to 2015. Models using Divisia money growth rates pseudo-forecast GDP growth and inflation with a higher precision than simple-sum aggregates up to a 2-year horizon. Structural variance decompositions reveal that monetary policy shocks during the Great Recession contribute the lion's share of variation in real GDP growth and inflation volatility.

CC1642: **Frequency response analysis of monetary policy transmission**
Presenter: **Lubos Hanus**, Charles University, Czech Republic
Co-authors: Lukas Vacha

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

CC1517: **Measuring interconnectedness between financial institutions with Bayesian time-varying vector autoregressions**
Presenter: **Marco Valerio Geraci**, Universite libre de Bruxelles, Belgium
Co-authors: Jean-Yves Gnabo

A market-based framework is proposed that exploits time-varying parameter vector autoregressions to estimate the dynamic network of financial spillovers. We apply it to financials listed in the Standard & Poor's 500 index and uncover interesting features at the individual, sectorial and system wide level. This includes a gradual decrease in interconnectedness after the crisis, not observable using the classical rolling window approach, and more stable interconnectedness-based rankings that can be used for monitoring purposes.

CC0401: **Endogenous monetary-fiscal regime change in the United States**
Presenter: **Boreum Kwak**, Martin Luther University Halle-Wittenberg and Halle Institute for Economic Research, Germany
Co-authors: Yoosoon Chang

U.S. monetary and fiscal policy regime interactions in a regime switching model are estimated where regimes are represented by endogenous latent policy factors. Policy regimes interact strongly: shocks that switch one policy from active to passive tend to induce the other policy to switch from passive to active, consistent with existence of a unique equilibrium. In some periods, though, both policies are active and government debt grows rapidly. We also observe relatively strong interactions between monetary and fiscal policy regimes after the recent financial crisis. Latent policy regime factors exhibit patterns of correlation with macroeconomic time series, suggesting that policy regime change is endogenous.

EO139 Room 207 RECENT DEVELOPMENTS OF STATISTICAL METHODS FOR COMPLEX LONGITUDINAL DATA Chair: Yanqing Sun

EO0164: **Regression analysis of mixed and incomplete recurrent event data**
Presenter: **Jianguo Sun**, University of Missouri, United States
Event history studies occur in many fields including economics, medical studies, and social science. In such studies concerning some recurrent events, two types of data have been extensively discussed in the literature. One is recurrent event data that arise if study subjects are monitored or observed continuously. In this case, the observed information provides the times of all occurrences of the recurrent events of interest. The other is panel count data, which occur if the subjects are monitored or observed only periodically. This can happen if the continuous observation is too expensive or not practical, and in this case, only the numbers of occurrences of the events between subsequent observation times are available. We discuss a third type of data, a mixture of recurrent event and panel count data and present an estimating equation-based approach for regression analysis of the data.

EO0339: **Fused lasso with the adaptation of parameter ordering in combining multiple studies with repeated measurements**
Presenter: **Lu Wang**, University of Michigan, United States
Co-authors: Fei Wang, Peter Song
Combining multiple studies is frequently undertaken to increase sample sizes for statistical power improvement. We consider repeated measurements collected in several similar studies with potentially different variances and correlation structures. It is of great importance to examine whether there exist common parameters across study-specific marginal models so that simpler models, sensible interpretations, and meaningful efficiency gain can be obtained. We develop a new method of fused lasso with the adaptation of parameter ordering (FLAPO) to scrutinize only adjacent-pair parameter differences, leading to a substantial reduction for the number of involved constraints. Our method enjoys the oracle properties as does the full fused lasso based on all pairwise parameter differences. We show that FLAPO gives estimators with smaller error bounds and better finite sample performance than the full fused lasso. We also establish a regularized inference procedure based on bias-corrected FLAPO. We illustrate our method through both simulation studies and an analysis of HIV surveillance data collected over five geographic regions in China, in which the presence or absence of common covariate effects is reflective to relative effectiveness of regional policies on HIV control and prevention.

EO0652: **Dynamic predictions of disease comorbidity using longitudinal Big Data from EHR systems**
Presenter: **Hulin Wu**, University of Texas Health Science Center at Houston, United States
An EHR database with big clinical record data from millons of people is used to develop dynamic prediction models for disease comorbidity. This Big Data modeling process involves several steps, which includes: 1) identifying and extracting the right data from the large database; 2) establishing a directed-graph causal network for disease comorbidity; 3) dealing with incomplete data; 4) using variable selection approaches to identify significant predictors for high-dimensional longitudinal data with high rate of missing; 5) model comparisons; and 6) result evaluation and validation. Each of these steps requires sophisticated or novel statisitcal methods. We will introduce this multi-disciplinary research project and discuss novel statistical methodologies, in particular, some new thoughts on how to use the novel Big Data concept in practice to produce big impact and big number of scientific findings.

EO0806: **Individualized subgroup variable selection**
Presenter: **Annie Qu**, University of Illinois at Urbana-Champaign, United States
A novel individualized variable selection method is proposed which performs coefficient estimation, subgroup identification and variable selection simultaneously. In contrast to traditional model selection approaches, an individualized regression model allows different individuals to have different relevant variables. The key component of the new approach is to construct a separation penalty which utilizes cross-subject information and assumes that within-group subjects share the same homogeneous effect. This allows us to borrow information from subjects within the same subgroup, and therefore improve the estimation efficiency and variable selection accuracy for each individual. Another advantage of the proposed approach is that it combines the strengths of homogeneity and heterogeneity in modeling, and therefore enhances prediction power. We provide theoretical justification for the proposed approach, and propose an effective algorithm to achieve an individualized variable selection. Simulation studies and an application to HIV longitudinal data illustrate comparisons of the new approach to existing penalization methods.

EO157 Room S23 RISK MEASURES: THEORETICAL AND PRACTICAL ASPECTS Chair: Elena Di Bernardino

EO0207: **Extreme estimation of multivariate return levels based on maximization problems**
Presenter: **Fatima Palacios Rodriguez**, University of Seville, Spain
Co-authors: Elena Di Bernardino
The classic univariate risk measure in environmental sciences is *Return Level*. The return level is the quantile x_p which expresses the magnitude of the event that is exceeded with a probability equal to p, with $p = 1/T$ (T is called the *return period*). Since an individual risk may strongly be affected by the degree of dependence amongst all risks, it is gaining ground the need of providing multivariate return levels. In contrast to the univariate case, the return level definition in the multivariate setting presents some troubles. Particularly, the multivariate return level was previously defined as the vector that maximizes a weight function given that the risk vector belongs to a given critical layer of its joint multivariate distribution function. We provide the explicit expression of the aforementioned multivariate return level and we estimate the measure by using extreme value theory techniques. The consistency of the proposed estimators is studied. In addition, the performance of the proposed estimators is evaluated on simulated data and on hydrological real dataset.

EO0757: **Estimation of directional multivariate extremes at high level**
Presenter: **Raul Andres Torres Diaz**, Universidad Carlos III de Madrid, Spain
Co-authors: Elena Di Bernardino, Henry Laniado Rodas, Rosa Lillo
In extreme value analysis, the research is focused on the quantification of the multivariate risk outside of the observable sampling zone; that is the meaning of a region of interest located at high levels. An out-sample estimation method is provided for the recently introduced Directional Multivariate Quantiles. The asymptotic normality of the proposed estimator is derived. Finally, the methodology is illustrated with simulated examples for which the theoretical directional multivariate quantiles are known and also an application to real data is shown.

EO1186: **Expectiles, omega ratios and stochastic ordering**
Presenter: **Fabio Bellini**, University of Milano-Bicocca, Italy
Co-authors: Bernhard Klar, Alfred Mueller
The expectile order is introduced, defined by $X \leq_e Y$ if $e_\alpha(X) \leq e_\alpha(Y)$ for each $\alpha \in (0,1)$, where e_α denotes the α-expectile. We show that the expectile order is equivalent to the pointwise ordering of the Omega ratios, and we derive several necessary and sufficient conditions. In the case of equal means, the expectile order can be easily characterized by means of the stop-loss transform; in the more general case of different means we provide some sufficient conditions. In contrast with the more common stochastic orders such $\leq_{st}$ and $\leq_{cx}$, the expectile order is not generated by a class of utility functions and is not closed with respect to convolutions. As an illustration, we compare the $\leq_{st}$, $\leq_{icx}$ and $\leq_e$ orders in the family of Lomax distributions.

EO1297: **Estimating covariate functions associated to multivariate risks: A level set approach**
Presenter: **Thomas Laloe**, LJAD - Universite de Nice, France
Co-authors: Elena Di Bernardino, Remi Servien
The aim is to study the behavior of a covariate function in a multivariate risks scenario. The first part deals with the problem of estimating the c-upper level sets $L(c) = \{F(x) \geq c\}$, with $c \in (0,1)$, of an unknown distribution function F on $\mathbb{R}_+^d$. A plug-in approach is followed. We state consistency results with respect to the volume of the symmetric difference. In the second part, we obtain the L_p-consistency, with a convergence rate,for the regression function estimate on these level sets $L(c)$. We also consider a new multivariate risk measure: the Covariate-Conditional-Tail-Expectation. We provide a consistent estimator for this measure with a convergence rate. We propose a consistent estimate when the regression cannot be estimated on the whole data set. Then, we investigate the effects of scaling data on our consistency results. All these results are proven in a non-compact setting. A complete simulation study is detailed and a comparison with parametric and semi-parametric approaches is provided.

EO173 Room 217 DETERIORATION MODELS FOR RELIABILITY Chair: Sophie Mercier

EO0223: **Log-concavity for shock models with applications in inventory models**
Presenter: **Carmen Sanguesa**, University of Zaragoza, Spain
Results concerning log-concavity of the distribution function in compound distributions are presented. Compound distributions refer to the distribution of sums of independent random variables, in which the number of summands is random. These results will be applied to shock models, in which the number of summands represent the number of shocks to a device in a given interval. For the proof of the main results, reliability properties of the individual shocks, as well as reliability properties of the the counting process describing the arrivals of shocks will be taken into account. These results can be also applied to insurance models (in which the compound distribution represents the losses of an insurance company), as well as to inventory model (cumulative demand). In particular, it is shown how the results can be applied to generalize results concerning the quasi-convexity of the cost function in certain inventory models, by considering a random lead-time (delay between orders and supplies).

EO0494: **Where statistical tools are unable to choose between two degradation models based on different physical assumptions**
Presenter: **Massimiliano Giorgio**, Second University of Naples, Italy
Co-authors: Gianpaolo Pulcini
The gamma process is often used to describe degradation processes where the degradation growth does not depend on the current degradation level but at most on the current age. In particular, the gamma process with random scale parameter is often used to describe degradation phenomena in presence of random effects. We introduce a new degradation process, the Transformed Beta (TB) process, where the degradation growth depends on both the current age and degradation level. Then, we show that the likelihood function relative to degradation data under the assumption that the degradation process is gamma with gamma-distributed (random) scale parameter, is the same as the likelihood function under the assumption that the process is TB without random effects. In particular, we show that the distribution of the degradation increment during a future time interval under the above two alternatives is the same, and hence the residual reliability and the mean residual lifetime are the same. However, if the unit is subject to a so-called imperfect maintenance, where the current degradation level is assumed to be lowered through an age reduction model, then the distribution of the degradation growth under the two alternatives is no longer the same, and hence an inappropriate degradation model produces wrong estimates of residual reliability and mean lifetime. Finally, a numerical application shows the errors produced by a wrong model when the unit is imperfectly maintained.

EO0643: **Computational tools for the calculus of some performance measures in a degradation system**
Presenter: **Inma Torres Castro**, University of Extremadura, Spain
Performance measures are used in various stages of the system life cycle to measure the system behaviour. From a mathematical point of view, Markov models with finite state spaces are commonly used tools for the evaluation of these transient performance measures. For Markov models,the holding times are exponentially distributed which eases the calculus of the performance measures. However, the assumption of exponential holding times is often too restrictive. Considering more general holding-times leads to semi-Markov processes, which are less amenable to analyze but provide more flexible models. The functioning of a degradation process subject to a maintenance program is evaluated. The degradation of this system is modelled using a gamma process. A gamma process is a stochastic process with independent, non negative gamma distributed increments. The functioning of the system is described as a semi-regenerative process with continuous state space. To evaluate the performance of this system, some transient performance measures are calculated. These performance measures fulfill Markov renewal equations. To solve these Markov renewal equations, a numerical method that combines numerical integration, interpolation techniques and Monte Carlo simulation is used. The accuracy of this numerical method is analyzed by making comparisons with the results obtained by using strict Monte Carlo simulation.

EO0923: **Some advancements on degradation empirical models for reliability evaluation of solid oxide fuel cells**
Presenter: **Fabio Postiglione**, University of Salerno, Italy
Solid oxide fuel cells (SOFCs) are electrochemical devices working at high temperature and producing electricity with high efficiency. An important open issue of this promising technology is the characterization of the degradation process. Being a direct observation of degradation phenomena difficult to implement, indirect performance indicators based on voltage measurements are frequently adopted, that are influenced also by the temperature of the furnace containing the cells. The voltage measurements collected during degradation tests on SOFC stacks can be effectively modelled by empirical random-effects regression models. They allow us to describe the variability components present in the measurements, such as the slow decay of voltage over time of each single cell of the stack, the variability of voltage decay among cells, and the fluctuations of voltage due to experimental noise and lack of fit, depending also on the temperature fluctuations. Some advancements are introduced on the available degradation empirical models in order to cope with correlation structure induced by temperature. Point and interval estimates are also derived for some performance measures of interest, for instance the prediction of cell voltage and the reliability function. Finally, the proposed methodology is applied to a real degradation test of a SOFC prototype.

EO215 Room 006 SOFTWARE FOR HIGH-PERFORMANCE AND BIG-DATA STATISTICS AND MACHINE LEARNING Chair: Alireza Mahani

EO0240: **Combining data and task parallelism in big-data analytics platforms**
Presenter: **Asad Hasan**, Sentrana Inc, United States
Co-authors: Syeed Mansur, Alireza Mahani
Open-source software platforms such as Apache Spark are motivated by a desire to produce scalable performance for big-data analytics using commodity servers connected via an unreliable network. As such, these platforms focus on data parallelism to achieve scalability. Two factors, however, necessitate a revision of this framework: 1) Rising core count and available RAM per compute node means larger and larger data sets can be efficiently parallelized within each node, 2) In most machine learning applications such as cross-validated tuning and model diagnostics there is ample opportunity for task parallelism, with longer execution times and less need for synchronization. Both of these properties lead to better performance scaling compared to a pure data parallelism approach. We present evidence supporting our claim that, generally speaking, task parallelism - where available - is the preferred mode compared to data parallelism, unless data sizes exceed RAM per compute node. We also

discuss ways to combine task and data parallelism in Apache Spark, and highlight the resulting performance improvement as well as limitations of current solutions.

EO0634: **HiCMA: Hierarchical computations on manycore architectures**
Presenter: **Hatem Ltaief**, KAUST, Saudi Arabia

The Hierarchical Computations on Manycore Architectures library (HiCMA) aims to tackle the challenge that is facing the linear algebra community due to an unprecedented level of on-chip concurrency, introduced by the manycore era. HiCMA is a high performance numerical library, which implements hierarchical numerical algorithms (e.g. matrix computations, eigenvalue decomposition, H-matrix, FMM, etc.) on emerging architectures. The hierarchy expression of the algorithms allows the enhancement of data locality (communication-reducing), while still ensuring embarrassingly parallel workloads (synchronization-reducing). The core idea is to redesign the numerical algorithms and to formulate them as successive calls to hierarchical computational tasks, which are then scheduled on the underlying system using a dynamic runtime system to ensure load balancing. The algorithm is then represented as a Directed Acyclic Graph (DAG), where nodes represent hierarchical tasks and edges show the data dependencies between them. Performance assessment on spatial statistics applications will be highlighted.

EO1182: **High-performance maximum likelihood estimation of multinomial logit regression for big-data applications**
Presenter: **Alireza Mahani**, Sentrana Inc, United States
Co-authors: Asad Hasan

Multinomial logit regression (MLR) is a popular framework for discrete choice analysis in statistical and econometric settings, as well as a multi-label classifier in machine learning applications. While smoothness and concavity of the log-likelihood functions of MLR models allows Newton optimization to be applied for Maximum-Likelihood (ML) estimation of coefficients, naive implementations suffer from long computation times and high memory consumption. We discuss several strategies for high-performance ML estimation of MLR models that rely on the unique structure of the Hessian matrix of the log-likelihood function. These strategies fall under two broad categories. First in the context of Newton optimization, we show a strategy of exploiting the Hessian structure that improves computation efficiency by over an order of magnitude in terms of running time and memory. Second, we propose an algorithm for efficiently computing the product of an implicit representation of the Hessian matrix with an arbitrary vector. We show that when used in an inexact Newton procedure this method gives a further speedup that is proportional to the number of model covariates. Finally we also present efficient parallelized versions of both our algorithms.

EO0930: **The Yin-Yang method for merging parallel MCMC output, with application to Austrian stroke data**
Presenter: **Alexandra Posekany**, Danube University Krems, Austria
Co-authors: Sylvia Fruehwirth-Schnatter

In many applied fields like economics and medical statistics large amounts of data are available, leading to a growing interest of applied Bayesian researchers and machine learners in approaches which split big data into subsets, performing inference independently in parallel and then merge these outputs. Often these data are too large for a single analysis due to the computational burden. This led to the development of approaches for combining parallelly obtained inference results, e.g. samples from posterior distributions, and subsequently obtaining a joint result which recovers the full posterior distribution and resulting posterior estimators or decisions. We propose the Yin-Yang sampler, a mathematically well-founded approach merging two samples from posterior distributions inferring different data partitions. Correcting for reusing the same prior for each subset instead of only once for the full data set is the key notion of our method. Sequential usage of Yin-Yang sampling steps retains the joint posterior from separate subsamples' posteriors for any given number of reasonably large subsets which have to contain enough information for sound inference results. To demonstrate our approach, we provide several simulation studies. In addition, we add a medical statistics application using the Austrian stroke unit data set which contains observations of more than 550 variables for over 130000 patients.

EO167 Room S24 BIG DATA ANALYTICS AND ITS APPLICATIONS IN BIOMEDICINE Chair: Yi Li

EO0274: **Classification with ultrahigh-dimensional features**
Presenter: **Yi Li**, University of Michigan, United States

Although much progress has been made in classification with high-dimensional features, wherein the features much outnumber the sample size, defies most existing work. A novel and computationally feasible multivariate screening and classification method is introduced for ultrahigh-dimensional data. Leveraging inter-feature correlations, the proposed method enables detection of marginally weak and sparse signals and recovery of the true informative feature set, and achieves asymptotic optimal misclassification rates. We also show that the proposed procedure provides more powerful discovery boundaries compared to previous ones. The performance of the proposed procedure is evaluated using simulation studies and demonstrated via classification of patients with different post-transplantation renal functional types.

EO0286: **Smoothing spline mixed-effects density models for clustered data**
Presenter: **Yuedong Wang**, University of California - Santa Barbara, United States

Smoothing spline mixed-effects density models are proposed for the nonparametric estimation of conditional density functions with clustered data. The random effects in a density model introduce within-cluster correlation and allow us to borrow strength across clusters by shrinking cluster specific density function to the population average, where the amount of shrinkage is decided automatically by data. Estimation is carried out using the penalized likelihood and computed using a Markov chain Monte Carlo stochastic approximation algorithm. We apply our methods to investigate evolution of hemoglobin density functions over time in response to guideline changes on anemia management for dialysis patients.

EO1380: **Interaction selection and screening for high dimensional data**
Presenter: **Hao Zhang**, University of Arizona, United States

The topic of identifying important interaction effects for high dimensional data is investigated. New and scalable methods are developed for both interaction selection and screening, with or without the hierarchy-preserving constraint. Both theoretical and numerical results are presented.

EO1381: **Measurement error case series model of infection-cardiovascular risk: Application to database of patients on dialysis**
Presenter: **Danh Nguyen**, University of California, Irvine, United States

The accumulation of patient data across practices over time provides a rich source of patient-level data to design and conduct population-based studies. Examples include national registries, electronic medical data and medical claims linked across health systems, and adverse events reporting systems of adverse drug reactions. We introduce the case series method and discuss its use for big data analytics in biomedicine. As an example, an application of the measurement error case series method (MECS) using the national registry, United States Dates Renal Data System (USRDS), which collects data on >99% of all patients in the U.S. is presented. Infection and cardiovascular (CV) disease are leading causes of hospitalization and death in patients on dialysis. A challenge in modeling the infection-CV risk is that the exact time of infection onsets cannot be ascertained based on hospitalization data. Only imprecise markers of the timing of infection onsets are available. The new MECS model is used to account for measurement error in time-varying exposure onsets. We describe the general nature of bias resulting from estimation that ignores measurement error and proposed a bias-correction method. Hospitalization data from the USRDS is used to illustrate the new method. The results suggest that the estimate of the CV incidence following the 30 days after infections is substantially attenuated in the presence of infection onset measurement error.

EO541 Room 203 HIGH-DIMENSIONAL OR MULTIVARIATE FUNCTIONAL DATA ANALYSIS Chair: Juhyun Park

EO0333: **Outlier detection of high dimensional data via dual rotations**
Presenter: **Jeongyoun Ahn**, University of Georgia, United States
Co-authors: Myung Hee Lee, Jung Ae Lee, Hee Cheol Chung
Despite the popularity of high dimension, low sample size data analysis, little attention has been paid to the outlier detection problem. We propose a two-stage procedure to detect outliers for high dimensional data. The first step screens out pre-determined most outlying points one by one, based on the distance between each data vector and the affine space generated by the remaining data. At the second step we test whether each of the screened observations is significantly outlying or not. The reference values for the significance test are based on random rotations of the data in the "dual" space. We show that the rotation procedure generates null data sets with the same volume as the original data, but without any outliers. High dimensional asymptotics is used to justify the proposed remoteness measure. The proposed method shows superior performance with various simulation settings compared to alternative approaches.

EO0690: **Estimation of ecological networks from multidimensional functional data with optimal control**
Presenter: **Serge Barbosa**, Telecom SudParis, France
Co-authors: Nicolas Brunel
Ecological interactions between species are classically inferred through the dynamics of the corresponding populations. Several mathematical models can be used for identifying the structure of the global network, and characterizing the nature of the relationships (cooperation, competition), from multidimensional time series (noisy and discretely) of potentially high dimension. We consider the generalized Lotka Volterra model that remains a standard although simple ordinary differential equations model. Our aim is then to identify the interaction matrix in this ODE model from a population of multidimensional time series, corresponding to various initial conditions and perturbations of the same network of species. The main challenge is that the estimation of differential equations is often ill posed and the estimation is unstable and is notably affected by noise and model misspecification. We propose a perturbation approach for designing a robust estimator to misspecification and we solve a trade off between fidelity to data and to model by using optimal control, with an appropriate discretization scheme.We show on data coming from intestinal microbiota that we can deal with the variability among multiple experiments and the potential modeling error of the Lotka Volterra model.

EO0736: **Bayesian region selection in functional regression**
Presenter: **hongxiao zhu**, Virginia Tech, United States
Co-authors: Yizhi Sun, Fengrong Wei
Detecting important regions in functional data has great significance in guiding decision-making. Although variable selection among multiple functional predictors is by no means new, the selection of unknown regions within a functional object has been less studied. We propose a novel Bayesian method for region selection in the framework of functional data regression. The selection of regions is achieved through encouraging sparse estimation of the regression coefficient. Nonzero regions of the estimated coefficient function correspond to the selected regions. In particular, we adopt compactly supported and potentially over-complete basis to capture local features of the regression coefficient function, and assume spike-slab priors to coefficients of the bases functions. To encourage continuous shrinkage of nearby regions, we adopt an Ising hyper-prior to take into account the neighboring structure of the bases functions, represented by an undirected graph. Posterior sampling is performed through Markov chain Monte Carlo algorithms. We finally verify the practical performance by applying the proposed approach to the near-infrared and sonar data.

EO1345: **Smoothing of surfaces and spatial fields with anisotropy using penalties involving partial differential operators**
Presenter: **Laura Sangalli**, Politecnico di Milano, Italy
Co-authors: Mara Sabina Bernardi, Michelle Carey, James Ramsay
Regression with partial differential regularization allows for the accurate estimation of surfaces or spatial fields starting from discrete and noisy observations. We show that, in presence of spatial anisotropy, we can use the parameters of the partial differential operator in the roughness penalty to estimate the direction and intensity of anisotropy.

EO017 Room 211 PREDICTIVE INFERENCE Chair: Fumiyasu Komaki

EO0337: **Predictive density estimation: Recent results**
Presenter: **Eric Marchand**, Universite de Sherbrooke, Canada
The estimation of predictive densities and their efficiency as measured by frequentist risk is addressed. For Kullback-Leibler, alpha-divergence, L_1 and L_2 losses, we review several recent findings that bring into play improvements by scale expansion, as well as duality relationships which lead to the study of plug-in estimators. A range of models are studied and include multivariate normal, scale mixtures of normals, Gamma, as well as models with restrictions on the parameter space.

EO0816: **Relationship between minimax predictive densities under Kullback-Leibler risk and conditional regret**
Presenter: **Mutsuki Kojima**, Mitsui Sumitomo Insurance Co., Ltd., Japan
The aim is to discuss the relationship between two minimax predictive densities. One is the Bayesian predictive density based on the latent information prior (LIP) that is minimax under Kullback-Leibler risk, and the other is CNML3, one type of conditional normalized maximum likelihood distributions (CNMLs) that is minimax under conditional regret. The Bayes projection of a predictive density, which is an information projection of the predictive density on a set of Bayesian predictive densities, is considered. We prove that the Bayes projection of CNML3 (BPCNML3) is asymptotically identical to the Bayesian predictive density based on LIP. In addition, under several regularity conditions, we show that BPCNML3 exactly coincides with the Bayesian predictive density based on LIP. Finally, we touch on the difference between CNML3 and the Bayesian predictive density based on LIP.

EO0823: **Minimax optimality of sparse Bayes predictive density estimates**
Presenter: **Gourab Mukherjee**, University of Southern California, United States
The problem of predictive density estimation is studied under Kullback-Leibler loss in ℓ_0 sparse Gaussian sequence models. We propose a proper Bayes predictive density estimate and establish its asymptotic minimaxity in sparse models. The minimax risks of predictive density estimates based on popular spike and slab approaches are also studied. Comparing our proposed Bayes predictive density estimate with thresholding based minimax optimal rules, we explain new similarities and contrasts with the parallel theory of point estimation of a multivariate normal mean under quadratic loss.

EO0836: **Improved simultaneous prediction intervals for autoregressive models**
Presenter: **Paolo Vidoni**, University of Udine, Italy
The specification of well-calibrated multivariate prediction regions may be useful in time series applications, whenever the aim is to consider not just one single forecast but a group of consecutive forecasts. However, the definition of multivariate prediction regions, having coverage probability

closed to the target nominal value, is still a challenging problem both from the theoretical and the practical point of view. An important results on improved multivariate prediction, based on higher-order asymptotic calculations, is reviewed. Although this solution is asymptotically superior to the estimative one, which is simpler but it may lead to unreliable predictive conclusions, it is usually hard to apply, since it requires complicated asymptotic expansions. A new asymptotically equivalent solution, giving improved simultaneous prediction intervals, is presented. It has a simple and intuitive form and, when computations are hard to perform, it is readily available an approximation based on bootstrap simulation methods. An application of this simpler bootstrap-based procedure to general autoregressive models is presented, focusing in particular on AR and ARCH models.

EO035 Room 201 RECENT DEVELOPMENT IN NONPARAMETRIC STATISTICS AND ITS APPLICATIONS Chair: Xin Zhang

EO0398: **On the estimation of ultra-high dimensional semiparametric Gaussian copula models**
Presenter: **Qing Mai**, Florida State University, United States
The semiparametric Gaussian copula model has wide applications in econometrics, finance and statistics. Recently, many have considered applications of semiparametric Gaussian copula model in several high-dimensional learning problems. We propose a slightly modified normal score estimator and a new Winsorized estimator for estimating both nonparametric transformation functions and the correlation matrix of the semiparametric Gaussian copula model. Two new concentration inequalities are derived, based on which we show that the normal score estimator and the new Winsorized estimator are consistent when the dimension grows at an exponential rate of the sample size. As demonstration, we apply our theory to two high-dimensional learning problems: semiparametric Gaussian graphical model and semiparametric discriminant analysis.

EO1195: **Nonlinear sufficient reductions**
Presenter: **Efstathia Bura**, The George Washington University, United States
Co-authors: Liliana Forzani
A methodology is developed for identifying and estimating sufficient reductions in regressions with predictors that, given the response, follow a multivariate exponential family distribution. This set-up includes regressions where predictors are all continuous, all categorical or mixtures of categorical and continuous. We derive and show that the minimal sufficient reduction comprises of linear and nonlinear functions of the predictors. In particular, we consider the case of binary predictors, a common occurrence in fields like gene association studies, image processing, natural language processing, social networks, and spatial statistics. We use the Ising model, an undirected graphical model, to model the binary predictors and compute the minimal sufficient reduction. Our approach is compared with Reproducing kernel Hilbert spaces (RKHS)-based methods, the only other methodology that computes nonlinear reductions.

EO1038: **Sufficient directional forecasting using factor models**
Presenter: **Wei Luo**, Baruch College, United States
Co-authors: Lingzhou Xue, Jiawei Yao
Factor models have been widely used for forecasting in economical research. Instead of the traditional regression of the target response variable on the factors, sliced inverse regression has been applied to further reduce the dimension of the factors, which enhances the accuracy of forecasting while avoids the risk of model mis-specification. We apply directional regression for dimension reduction on the factors, which is more comprehensive than sliced inverse regression. In addition, we allow the dimension of the factors to increase with the sample size. Consequently, compared to the previous approach, ours allows a larger number of factors in finite samples, and is applicable for a more general relationship between the factors and the target response variable. We develop relative asymptotic theory, and conduct simulation studies and real data analysis to illustrate the effectiveness of the proposed method.

EO0979: **Subject-wise empirical likelihood inference in partial linear models for longitudinal data**
Presenter: **Lianfen Qian**, Florida Atlantic University, United States
Co-authors: Suojin Wang
In analyzing longitudinal data, within-subject correlations are a major factor that affects statistical efficiency. Working with a partially linear model for longitudinal data, we consider a subject-wise empirical likelihood based method that takes into consideration the within-subject correlations to estimate the model parameters efficiently. A nonparametric version of the Wilks' theorem for the limiting distribution of the empirical likelihood ratio, which relies on a kernel regression smoothing method to properly centered data, is derived. We also consider the estimation of the nonparametric baseline function. A simulation study and an application are reported to investigate the finite sample properties of the proposed method and compare it with the block empirical likelihood method. These numerical results demonstrate the usefulness of the proposed method.

EO195 Room 208 BAYESIAN MODELING AND APPLICATIONS Chair: Christopher Hans

EO0438: **Quantifying discovery in astrophysics and particle physics**
Presenter: **David van Dyk**, Imperial College London, United Kingdom
Co-authors: Sara Algeri, Jan Conrad, David Jones
The question of how best to compare models and select among them has dogged statisticians for decades. The difficulties with interpreting p-values and the strong dependence of Bayes Factors on the choice of prior distribution are well known. We explore a class of non-standard model comparison problems that are important in astrophysics and in high-energy physics. The search for the Higgs boson, for example, involved quantifying evidence for a narrow component added to a diffuse background distribution. The added component corresponds to the Higgs mass distribution and cannot be negative. Thus, not only is the null distribution on the boundary of the parameter space, but the added components location is unidentifiable under the null. Because many researchers have a strong preference for frequency-based statistical methods, they employ a sequence of likelihood ratio tests on a grid of possible null values of the unidentifiable location parameter. We compare this with a Bayesian strategy that employs a prior distribution on the location parameter and show how this prior automatically corrects for the multiple testing inherent in the standard method. The Bayesian procedure is significantly more conservative in that it avoids the well-known tilt of p-values toward the alternative when testing a precise null hypothesis. Finally, we discuss the circumstance under which the dependence of the Bayes Factor can be interpreted as a natural correction for multiple testing.

EO0465: **Bayesian clustering of player styles for multiplayer games**
Presenter: **Shane Jensen**, The Wharton School of the University of Pennsylvania, United States
Co-authors: Aline Normoyle
Empirical approaches to video game data are typically based solely on game outcomes, e.g. kills, deaths, and score for each player. We investigate a method for clustering players based on how a players choices relate to outcomes, i.e. the latent player styles exhibited by players. Our approach is based on a Bayesian semi-parametric clustering model with several advantages: the number of clusters is not specified a priori; the technique can work with a very compact representation of each match; a player can belong to multiple clusters and hence can have a hybrid play style; and the resulting clusters often have a straight-forward interpretation. We apply our method to multiplayer match logs from Battlefield 3 consisting of over 1200 players and 500,000 matches.

EO1165: **Black box versus do-it-yourself methods for causal inference: Lessons learned from a data analysis competition**
Presenter: **Jennifer Hill**, New York University, United States
Co-authors: Vincent Dorie, Uri Shalit, Dan Cervone, Marc Scott
Statisticians have made great strides towards assumption-free estimation of causal estimands in the past few decades. However this explosion in research has resulted in a breadth of inferential strategies that both create opportunities for more reliable inference as well as complicate the choices that an applied researcher has to make and defend. Relatedly, researchers advocating for new methods typically compare their method to (at best) 2 or 3 other causal inference strategies and test using simulations that may or may not be designed to equally tease out flaws in all the competing methods. The causal inference data analysis challenge, Is Your SATT Where Its At?, launched as part of the 2016 Atlantic Causal Inference Conference, sought to make progress with respect to both of these issues. The researchers creating the data testing grounds were distinct from the researchers submitting methods whose efficacy would be evaluated. Results from over 30 competitors in the two parallel versions of the competition (Black Box Algorithms and Do It Yourself Analyses) are presented along with post-hoc analyses that reveal information about the characteristics of causal inference strategies that yielded stronger inferences under a variety of circumstances. The most consistent conclusion was that the black box methods performed better overall than the user-controlled methods in all scenarios tested.

EO0914: **Analysis of ordinal non-numeric data: networks and causal inference**
Presenter: **Alexander Volfovsky**, Duke University, United States
Many outcomes of interest in the social and health sciences are ordinal and do not have a meaningful scale. Frequently, analysis of these data relies on one of two extremes: dichotomization to avoid scale issues or making assumptions that define a numeric scale. We demonstrate the pitfalls of these approaches and develop Bayesian estimation procedures based on variants of the rank likelihood, a type of marginal likelihood that does not depend on the unknown scale parameters of the data. We illustrate the wide applicability of rank likelihood methods on applications in network analysis and causal inference.

EO101 Room S22 THE STEIN METHOD IN STATISTICS Chair: Christophe Ley

EO0441: **Stein's method for noncentral chi-squared approximation**
Presenter: **Yvik Swan**, Universite de Liege, Belgium
Co-authors: Benjamin Arras, Gesine Reinert
A method is provided (including operators, characterization and bounds on the constants) for the noncentral chi-squared distribution. We apply our theory to problems of noncentral chi-square asymptotic convergence.

EO0510: **Monte Carlo integration using Stein's method**
Presenter: **Francois-Xavier Briol**, University of Warwick, United Kingdom
Co-authors: Chris Oates, Jon Cockayne, Mark Girolami
The recent surge in data available to scientists has led to an increase in the complexity of mathematical models, rendering them much more computationally expensive to evaluate. This is a particular challenge for many tasks of interest such as making predictions or inferring parameter values, since these will require the extensive use of numerical integration methods, which will tend to be slow due to the high computational costs. A novel Monte Carlo integration scheme is introduced which makes use of properties of the integrand (e.g. smoothness or periodicity) in order to obtain fast convergence rates in the number of integrand evaluations. Furthermore, we will demonstrate how Steins method can enhance this scheme by also making use of properties and allow it to deal with un-normalised densities, which is a common problem in Bayesian computation. The advantages of the proposed methodology will be illustrated on a Bayesian inverse problem for a PDE model of subsurface flow.

EO0520: **Bounds for the normal approximation of the maximum likelihood estimator**
Presenter: **Andreas Anastasiou**, The London School of Economics and Political Science, United Kingdom
The asymptotic normality of the maximum likelihood estimator (MLE) under regularity conditions is a long established and famous result. This is a qualitative result and the assessment of such a normal approximation is the main interest. Stein's method is partly used, which is a probabilistic technique that can be used to explicitly measure the distributional distance between two distributions. Explicit upper bounds on the distributional distance between the distribution of the MLE and the normal distribution are derived. First, the focus is on independent and identically distributed random variables from both discrete and continuous single parameter distributions. The bounds are of order $n^{(}-1/2)$. Furthermore, often the MLE can not be obtained analytically. Even in such cases, optimal order bounds are given. The case of independent but not necessarily identically distributed random vectors from multi-parameter distributions is also covered and specific examples are given. Going back to the single-parameter setting a different approach to get an upper bound on the distributional distance of interest, based on the Delta method, is also developed. Finally, the independence assumption can be relaxed and results for the case of locally dependent random variables are obtained.

EO0536: **The geometry of spherical random fields: Theory and applications**
Presenter: **Maurizia Rossi**, Universite du Luxembourg, Luxembourg
In modern Cosmology, Cosmic Microwave Background radiation is viewed as a single realization of a random field on the sphere; the investigation of its statistical properties for instance, Gaussianity and isotropy - has attracted an enormous amount of attention in the last decade. In this framework, asymptotic theory must be developed in the high-frequency (fixed-domain) sense; one is therefore interested in the geometrical properties of random spherical eigenfunctions for large eigenvalues, such as the area of their level sets, as well as their EP characteristic and their boundary length. Our results concern the limiting distribution of such geometric functionals; the proofs rely on a pervasive use of chaotic expansions as well as Fourth Moment Theorems, which combine Stein's method and Malliavin calculus on Wiener chaoses.

EO523 Room 101 INFERENCE AND APPLICATIONS OF DYNAMIC AND TIME-DEPENDENT NETWORKS Chair: Yulia Gel

EO0454: **High-dimensional multivariate time series with additional structure**
Presenter: **Michael Schweinberger**, Department of Statistics, Rice University, United States
High-dimensional multivariate time series are challenging due to the dependent and high-dimensional nature of the data, but in many applications there is additional structure that can be exploited to reduce computing time along with statistical error. We consider high-dimensional vector autoregressive processes with spatial structure, a simple and common form of additional structure. We propose novel high-dimensional methods that take advantage of such structure without making model assumptions about how distance affects dependence. We provide non-asymptotic bounds on the statistical error of parameter estimators in high-dimensional settings and show that the proposed approach reduces the statistical error. An application to air pollution in the US demonstrates that the estimation approach reduces both computing time and prediction error and gives rise to results that are meaningful from a scientific point of view, in contrast to high-dimensional methods that ignore spatial structure. In practice, these high-dimensional methods can be used to decompose high-dimensional multivariate time series into lower-dimensional multivariate time series that can be studied by other methods in more depth.

EO0805: **Latent space models for complex networks**
Presenter: **Aram Galstyan**, USC Information Sciences Institute, United States
Studies of social systems have traditionally focused on analyzing networks induced by social interactions, while discarding rich contextual information on nodes and their properties. At the same time, empirical evidence points to strong correlations between node attributes and their interactions. We suggest a viable framework for analyzing attribute-rich and multi-modal social data based on latent space models. In this approach, each node is assigned an unobserved (latent) position in some space, so that both the nodes attributes and their interactions depend on their coordinates in this space. This shared latent space allows to capture observed correlations between the attributes and network structure. We perform extensive experiments where the goal is predict missing links in a network using attributes, or predict user attributes based on network information, and observe that the proposed method outperforms other baselines in both prediction tasks.

EO1350: **Community change-point detection in time-dependent networks**
Presenter: **Leto Peel**, UCLouvain, Belgium
Time-dependent networks are frequently used to represent complex systems in which dynamic interactions occur over time. Understanding if, when and how such networks change over time can provide valuable insights into how complex systems evolve. However, the variation in individual pairwise interactions can be hard to interpret and is often not particularly meaningful on its own. Instead, community detection can be used to identify a more readily interpretable, coarse-grained representation of the system. We use a hierarchical model to capture stochastically equivalent nodes across multiple resolutions and introduce an efficient spectral method for inferring the community assignment. We then apply methods from statistical change-point detection to detect significant changes in the large-scale structure of the network. Finally, we demonstrate the real-world changes identified in networks of resource transportation and social interactions.

EO1366: **Anomaly detection in time-evolving networks using tensor spectrum**
Presenter: **Yulia Gel**, University of Texas at Dallas, United States
In analysis of dynamic networks, one of the key tasks are anomaly detection. Its applications range from new gang formation to brain damages to money laundering. Most of the currently available methods for anomaly detection have two disadvantages: either they focus only on two-dimensional structures, that is, edges connecting pairs of nodes; or they neglect the important temporal dependence structure of change point statistics in networks, which in turn leads to distorted false positive and false negative rates. We circumvent these problems by introducing a new anomaly detection method based on tensor spectral characteristics. The new data-driven approach is distribution-free and allows to detect change points in higher-order network motifs. We evaluate our new anomaly detection procedure on synthetic networks and benchmark case studies.

EO513 Room 204 MULTIVARIATE EXTREMES Chair: Miguel de Carvalho

EO0471: **Angular volatility for multivariate extremes**
Presenter: **Miguel de Carvalho**, The University of Edinburgh, United Kingdom
Co-authors: Simone Padoan
The angular volatility is proposed as a measure of the dynamics of extremal dependence over time. The building block for the notion of angular volatility is that of a family of time-changing spectral densities of a multivariate extreme value distribution. For modeling, our preferred implementation is through a stochastic volatility method which we fit through Bayesian methods.

EO0746: **Bayesian inference for the extremal dependence**
Presenter: **Isadora Antoniano-Villalobos**, Bocconi University, Italy
Co-authors: Giulia Marcon, Simone Padoan
A simple approach for modeling multivariate extremes is to consider the vector of component-wise maxima and their max-stable distributions. The extremal dependence can be inferred by estimating the angular measure or, alternatively, the Pickands dependence function. A nonparametric Bayesian model is proposed that allows, in the bivariate case, the simultaneous estimation of both functional representations through the use of polynomials in the Bernstein form. The constraints required to provide a valid extremal dependence are addressed in a straightforward manner, by placing a prior on the coefficients of the Bernstein polynomials which gives probability one to the set of valid functions. The prior is extended to the polynomial degree, ensuring great modelling flexibility demonstrated through an analysis of the prior support. Although the analytical expression of the posterior is unknown, inference is possible via a trans-dimensional MCMC scheme.

EO0821: **Flexible modeling of non-stationary spatial extremes over large heterogeneous regions using factor copula models**
Presenter: **Daniela Castro**, King Abdullah University of Science and Technology, Saudi Arabia
Co-authors: Raphael Huser
In order to model complex dependence structures in spatial extremes, an approach is proposed based on factor copula models. The latter, which can be seen as Gaussian location mixture processes, assume the presence of a common factor affecting the joint dependence of all measurements. When the common factor is exponentially distributed, the resulting copula is asymptotically equivalent to the Husler-Reiss copula; therefore, the so-called exponential factor model is suitable to capture tail dependence. Under the assumption of local stationarity, the exponential factor model is used to model non-stationary extreme measurements over high thresholds. Inference is performed using a censored local likelihood. Performance is assessed using simulation experiments, and illustrated using a daily rainfall dataset.

EO0956: **Non linear models for extremal dependence**
Presenter: **Linda Mhalla**, University of Geneva, Switzerland
Co-authors: Valerie Chavez-Demoulin, Philippe Naveau
The dependence structure of max stable random vectors can be characterized by its Pickands dependence function. In many applications such as climatology, the extremal dependence measure varies with covariates (time, altitude, etc). A new flexible semi parametric method for the estimation of non stationary multivariate Pickands dependence functions is developed. The proposed construction is based on an accurate max projection allowing to pass from the multivariate to the univariate setting and to rely on the generalized additive modelling framework. The resulting estimator of the Pickands function is regularized, in the bivariate case, using constrained median smoothing B splines and bootstrap confidence intervals are constructed. Finally, the results from a simulation study are presented and the new methodology is applied to the U.S temperature dataset.

EO213 Room 214 BUILDING OPTIMAL EXPERIMENTAL DESIGNS Chair: Victor Casero-Alonso

EO0485: **Methods of computing efficient exact designs based on optimal approximate designs**
Presenter: **Radoslav Harman**, Comenius University in Bratislava, Slovakia
Co-authors: Lenka Filova
Computing optimal exact designs of experiments is a difficult problem of discrete optimization. A common solution is to use a "rounding" method, which first applies a rapid convex optimization procedure to obtain the so-called optimal approximate design, and then finds a "similar" exact design. However, the existing rounding methods have significant limitations; for instance, the size of the required exact design cannot be smaller than the number of support points of the optimal approximate design. Moreover, the existing rounding methods can only be applied under the simplest constraints. We will propose two main ideas of how to overcome these limitations by using: 1) the integer linear programming applied to an extension of the standard efficient rounding procedure, and 2) an integer quadratic programming to solve a problem resulting from a quadratic approximation of the target optimality criterion in the neighborhood of the optimal approximate information matrix.

EO0621: **Computing c-optimal designs for non-independent observations**
Presenter: **Juan M Rodriguez-Diaz**, University of Salamanca, Spain
c-optimality is one of the most employed optimality criteria. For a given model it looks for the design that minimizes the variance of the linear combinationof the parameters' estimators given by vector c. c-optimal designs are needed when dealing with standardized criteria, and are frequently used to check how good a design is for the estimation of each of the parameters of the model. The Elfving's procedure for independent observations gives the idea of the procedure for correlated observations. Some results can be obtained for simple covariance structures; however when moving from these situations the computations become harder. An analysis of covariance structures suitable for the application of the procedure is performed, and its behavior will be illustrated with convenient examples.

EO1078: **Some tools to measure the performance of adaptive dose finding experiments**
Presenter: **Jose Moler**, Universidad Publica de Navarra, Spain
Tools are presented to study the performance of adaptive designs in phases I or II of Clinical Trials. In the early phases of a clinical trial, the main statistical target is to estimate a dose with a targeted percentage of toxicity or efficiency in the response variables. We analyze the way to combine inferential and ethical criteria to evaluate the performance of some relevant adaptive procedures presented in the specialized literature. Finding appropriate measures for these criteria is a challenging point which we deal with.

EO0876: **Building optimal designs of experiments in multi-factor settings from their univariate counter-parts**
Presenter: **Rainer Schwabe**, Otto-von-Guericke University Magdeburg, Germany
To use an optimal or, at least, efficient design is an essential precondition to perform a good experiment. Theoretical and numerical approaches to generate optimal designs are well developed for regression models with a single explanatory variable. These can be used as bricks to build optimal or efficient designs for more complicated situations. We start with a survey on the well-known result that product-type designs are optimal in multi-factorial models with a complete interaction structure and are, at least, highly efficient for additive models when their marginals are optimal in the corresponding univariate models with a single explanatory variable. We extend these findings to recent results for multivariate models of the seemingly unrelated regression type and present an alternative approach for polypod designs in a multiple generalized linear model setup.

EO107 Room 209 RECENT ADVANCES IN COMPLEX DATA MODELING AND COMPUTATIONAL METHODS Chair: Tsung-I Lin

EO0539: **Hypothesis test of a block compound symmetric covariance matrix for two-level multivariate data**
Presenter: **Anuradha Roy**, The University of Texas at San Antonio, United States
Co-authors: Carlos Coelho
The purpose is to study the problem of testing the hypothesis of a block compound symmetry covariance matrix with two-level multivariate observations, taken for m variables over u sites or time points. Through the use of a suitable block-diagonalization of the hypothesis matrix, it is possible to obtain a decomposition of the main hypothesis into two sub-hypotheses. By using this decomposition it is then possible to obtain the likelihood ratio test statistic as well as its exact moments in a much simpler way. The exact distribution of the likelihood ratio test statistic is then analyzed. Because this distribution is quite elaborate, yielding a non-manageable distribution function, a manageable but very precise near-exact distribution is developed. Numerical studies conducted to evaluate the closeness between this near-exact distribution and the exact distribution show the very good performance of this approximation even for very small sample sizes. A real data example is presented and a simulation study is also conducted.

EO0680: **Likelihood-based inference for Tobit confirmatory factor analysis using the multivariate Student-t distribution**
Presenter: **Mauricio Castro**, Universidad de Concepcion, Chile
Factor analysis models have been one of the most popular multivariate methods for data analysis among psychometricians, behavioral and educational researchers. But these models, originally developed for normally distributed observed variables, can be seriously affected by the presence of influential observations and censored data. Motivated by this situation, we propose a likelihood-based estimation for a multivariate Tobit confirmatory factor analysis model using the Student-t distribution (t-TCFA model). An EM-type algorithm is developed for computing the maximum likelihood estimates, obtaining as a byproduct the standard errors of the fixed effects and the exact likelihood value. Unlike other approaches proposed in the literature, our exact EM-type algorithm uses closed form expressions at the E-step based on the first two moments of a truncated multivariate Student-t distribution with the advantage that these expressions can be computed using standard statistical software. The performance of the proposed methods is illustrated through a simulation study and the analysis of a real dataset of early grade reading assessment test scores.

EO0797: **Reduced rank analysis in GMANOVA-MANOVA models**
Presenter: **Tatjana von Rosen**, Stockholm University, Sweden
Co-authors: Dietrich von Rosen
The maximum likelihood estimation is discussed in general multivariate linear models where the rank restrictions are imposed on the matrix of regression coefficients in order to enable parsimonious modeling. In particular, the models which mean is the sum of two components, a growth curve and a multivariate analysis of variance models are of interest. Maximum likelihood estimators are derived for the parameters in this type of models for the mean and covariance matrix. Additionally, a growth curve model with a reduced rank structure on the matrix of regression coefficients as well as on the covariance matrix is studied. A numerical example is provided to illustrate the obtained results.

EO0875: **Spatial Bayesian hierarchical model for small area estimation of proportions with constraint**
Presenter: **Zhengyuan Zhu**, Iowa State University, United States
Motivated by the need to produce small area estimates for the National Resources Inventory survey, we develop a spatial hierarchical Bayesian model based on a generalized Dirichlet distribution to construct small area predictors of proportions in several mutually exclusive and exhaustive land cover classes. The standard survey estimators are judged unreliable at the county level due to small sample sizes, and the hierarchical model

is used to obtain more efficient predictors. At the observation level, the standard survey estimators of the proportions are assumed to follow the generalized Dirichlet (GD) distribution. After proper transformation of the survey based estimators, beta regression is applicable. We consider a logit mixed model for the expectation of the beta distribution, which incorporates covariates through fixed effects and random effects with spatial structure through a conditionally autoregressive (CAR) process. Three special cases of the Bayesian hierarchical model, with different random effects structures, are compared using Bayesian model comparisons tools. In the application, the survey data are from the National Resources Inventory survey, and the covariate is derived from the Cropland Data Layer (CDL), a land cover map based on satellite data. In a design-based evaluation study, the Bayesian estimators are shown to have smaller relative root mean squared error than direct estimators.

EO011 Room 007 BAYESIAN BIOSTATISTICS Chair: Timothy Johnson

EO0541: **Flexible Bayesian models for causal inference and missing data**
Presenter: **Jason Roy**, University of Pennsylvania, United States

Bayesian methods have not been widely used for causal inference in observational studies. A possible reason for this is that causal inference in a likeihood-based framework often requires modeling the joint distirbution of all of the observed data, including covariates. However, recent developments in Bayesian nonparametric (BNP) modeling, along with increasing computing capacity, have opened the door to a new, potentially powerful approach to causal inference in a variety of settings. We develop a general joint Dirchlet process mixture model and show how it can be used to obtain posterior inference for any causal effect of interest. The extra effort needed to model a full observed data distribution has many potential benefits, including efficiency gains, full posterior inference rather than just point estimates and confidence intervals, automatic imputation of missing data, and a general way to account for uncertainty about a variety of assumptions. We compare our method with inverse probability of treatment-weighted estimators in simulation studies.

EO1156: **A Bayesian non-parametric causal inference model for comparative effectiveness research**
Presenter: **Chenguang Wang**, Johns Hopkins University, United States

Comparative effectiveness research (CER) is designed to synthesize evidence of the benefits and harms of a treatment option from disparate sources including randomized clinical trials, observational studies and registry studies. The task of addressing study-specific heterogeneities is one of the most difficult challenges in CER. Bayesian hierarchical model with non-parametric extension provide a powerful and convenient platform that formalizes the information borrowing strength across the studies. We propose a propensity score-based Bayesian non-parametric Dirichlet process mixture model that summarizes information from multiple observational and randomized studies to draw inference on the causal treatment effect. Simulation studies are conducted to evaluate the model performance under different scenarios.

EO0684: **Bayesian nonparametric analysis of longitudinal studies in the presence of informative missingness**
Presenter: **Antonio Linero**, Florida State University, United States

How to address non-ignorable missing data when estimating causal effects in longitudinal clinical trails remains a challenging problem, particularly when missingness is non-monotone. Any successful method must, either implicitly or explicitly, confront two problems: (1) the non-identifiability of most interesting causal effect and (2) the curse of dimensionality. We propose a Bayesian nonparametric approach to this problem which addresses (2) in a flexible manner by placing a nonparametric prior on the space of data generating mechanisms such that inferences are shrunk towards a structured parametric model. To address (1), we introduce the notion of a "working prior" which allows our Bayesian nonparametric prior to be used in conjunction with identifying restrictions. In simulations based on a real data set, our approach is shown to provide uncertainty quantification which is more honest than several model-based and hot-deck imputation techniques based on chained equations (MICE). In addition to competitive performance, we provide a theoretical foundation for our approach and present posterior consistency results. Additionally, we introduce a new family of identifying restrictions for non-monotone missingness which may be of independent interest.

EO0727: **Bayesian approaches for integrative genomics**
Presenter: **Francesco Stingo**, University of Florence, Italy

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. We propose a novel Bayesian model to identify genomics markers that are associated with survival time by incorporating the regulatory networks through prior distributions. We assume that biomarkers involved in regulatory networks are likely 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. 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.

EO153 Room 213 RECENT ADVANCES IN QUANTITLE REGRESSION AND SURVIVAL ANALYSIS Chair: Roel Braekers

EO0543: **Conditional quantile inference for generalized autoregressive conditional heteroscedastic models: A hybrid approach**
Presenter: **Guodong Li**, University of Hong Kong, China

It is an essential task in modern risk management to estimate the conditional quantiles for conditional heteroscedastic models, among which, the generalized autoregressive conditional heteroscedastic (GARCH) model has the greatest popularity. However, so far feasible quantile regression methods in this area have been limited to the linear GARCH models, owing to its tractable conditional quantile structure. A surprising equivalence of the GARCH model and the linear GARCH model is established through a simple yet nontrivial transformation. This gives us a chance to develop a hybrid, yet easy-to-implement two-stage conditional quantile estimation (CQE) procedure for the GARCH model. It merges the convenience of the GARCH form in obtaining the Gaussian quasi-maximum likelihood estimator at the first stage and that of the linear GARCH form in calculating the conditional quantiles at the second stage. To approximate the asymptotic distribution of the CQE, we employ a bootstrap method, where the equivalence of the two models further allows us to significantly reduce the computation time through a hybrid approach. Moreover, a goodness-of-fit test based on the residual quantile autocorrelation function is constructed to check the adequacy of the fitted conditional quantiles. Simulation results indicate the usefulness of the proposed inference procedure. A real example further corroborate the merits of our approach.

EO1387: **Two layer EM algorithm for ALD mixture regression models: A new solution to composite quantile regression**
Presenter: **Liming Xiang**, Nanyang Technological University, Singapore
Co-authors: Shanshan Wang

Motivated by the link between quantile regression and a likelihood-based approach under assumption of asymmetric Laplace distributed errors, we introduce linear regression by modeling the error term through a nite mixture of asymmetric Laplace distributions. The model expands the flexibility of the linear regression by accounting for heterogeneity among data, and allows us to establish the equivalence between maximum likelihood estimation of the model parameters and the composite quantile regression (CQR) estimation, providing a new likelihood-based solution to CQR. We propose a two-layer EM-based algorithm for implementing the estimation procedure. An appealing feature of the proposed algorithm is that the closed form updates for the parameters in each iteration are obtained explicitly, instead of resorting to linear programming optimization

methods, as in the existing work. The computational complexity can be reduced significantly. We evaluate the performance through simulation studies and illustrate its usefulness by analyzing a gene expression dataset.

EO0592: **Informative censoring models**
Presenter: **Gerhard Dikta**, FH-Aachen, Germany
Based on an identifying Volterra-type integral equation for a lifetime distribution F and randomly right censored observations, we solve the corresponding estimating equation by an explicit and implicit Euler scheme. Depending on the assumptions we make about the conditional expectation of the censoring indicator given the observation time, we derive the well-known Kaplan-Meier and other established estimators of F under the explicit Euler scheme. Moreover, under the implicit Euler scheme, we obtain new pre-smoothed and semi-parametric estimators of F. Some properties of the new semi-parametric estimator and a real data application are discussed.

EO0635: **The focused information criterion for a mixture cure model**
Presenter: **Ingrid Van Keilegom**, Universite catholique de Louvain, Belgium
Co-authors: Gerda Claeskens
In many situations in survival analysis, it may happen that a fraction of the subjects under study will never experience the event of interest: they are considered to be cured. The mixture cure model is a common regression model in survival analysis that takes this feature into account. It supposes that the population consists of a mixture of two sub-populations, the cured ones and the non-cured ones, and it supposes a logistic model for the probability of being cured. For the non-cured sub-population we suppose a Cox proportional hazards regression model. We are interested in doing variable selection in this model using the focused information criterion (FIC). Of interest is therefore the asymptotic distribution of the estimators of the parameters and of the baseline hazard in a mixture cure model under local misspecification. Once this asymptotic distribution is obtained, the MSE can be used to guide selection of variables to be included in the logistic and Cox proportional hazards parts of the model. The method is illustrated by means of simulations and data regarding a UK financial institution.

EO073 Room 202 STATISTICAL APPLICATIONS IN GENETICS Chair: Florian Frommlet

EO0608: **Hidden Markov models for QTL-mapping in haploid and diploid yeast**
Presenter: **Jurgen Claesen**, Hasselt University, Belgium
Co-authors: Tomasz Burzykowski
The analysis of polygenic, phenotypic characteristics such as quantitative traits or inheritable diseases requires reliable scoring of many genetic markers covering the entire genome. The advent of high-throughput sequencing technologies provides a new way to evaluate large numbers of single nucleotide polymorphisms as genetic markers. Combining the technologies with pooling of segregants, as performed in bulk segregant analysis, should, in principle, allow the simultaneous mapping of multiple genetic loci present throughout the genome. We propose homogeneous and non-homogeneous hidden Markov-models to analyze the marker data obtained by bulk segregant next generation sequencing. The model includes several states, each associated with a different probability of observing the same/different nucleotide in an offspring as compared to the parent. The transitions between the molecular markers imply transitions between the states of the model. After estimating the transition probabilities and state-related probabilities of nucleotide (dis)similarity, the most probable state for each SNP is selected. The most probable states can then be used to indicate which genomic regions may be likely to contain trait-related genes.

EO0609: **Model selection for network inference from gene expression data with small sample size**
Presenter: **Melina Gallopin**, Universite Paris Sud, France
Co-authors: Emilie Devijver
Gaussian graphical models are widely utilized to infer networks from gene expression data. However, inferring the graph is difficult when the sample size is small compared to the number of genes. To reduce the number of parameters to estimate in the model, we propose a non-asymptotic model selection procedure supported by strong theoretical guarantees based on an oracle type inequality and a minimax lower bound. The covariance matrix of the model is approximated by a block-diagonal matrix. The structure of this matrix is detected by thresholding the sample covariance matrix, where the threshold is selected using the slope heuristic. Based on the block-diagonal structure of the covariance matrix, the estimation problem is divided into several independent problems: subsequently, the network of dependencies between variables is inferred using the graphical lasso algorithm in each block. The performance of the procedure is illustrated on simulated data. An application to a RNA-seq gene expression dataset with a limited sample size is presented: the dimension reduction allows attention to be objectively focused on interactions among smaller subsets of genes, leading to a more parsimonious and interpretable modular network.

EO0770: **Stochastic modelling of PCR to estimate and correct for unobserved molecules in quantitative NGS experiments**
Presenter: **Florian Pflug**, University of Vienna, Austria
Co-authors: Arndt von Haeseler
Many protocols in modern-day biology use next-generation sequencing (NGS) as a quantitative method, i.e. to measure the abundance of particular DNA molecules. Then, any molecule that remains unsequenced causes a measurement error, and if this affects molecules non-uniformly, results are systematically biased. A major source of such biases is the Polymerase Chain Reaction (PCR), used to amplify DNA prior to sequencing. If it can be adequately modelled, its biases can be predicted and corrected for. Different models of PCR haven been proposed, but none have yet found their way into standard analysis pipelines, owing to a lack of parameter estimates for specific conditions. We thus focus on describing a model whose parameters can be estimated from actual experimental data, while still capturing the main source of biases. We show that this is achieved by viewing PCR as a branching process which, during each cycle, duplicates each DNA molecule with a certain probability, called the reactions efficiency. We combine this model with a simple model of the sampling behaviour of NGS and apply it to published RNA-Seq data. We demonstrate that the reaction efficiency can be estimated from the data, and that the data matches the models predictions well. In particular, we find that the model explains the main observed stochastic effects. Finally, we explore how well we can correct for unobserved molecules, and how much this improves the accuracy of the measured gene transcript abundances.

EO1192: **Controlling the rate of GWAS false discoveries**
Presenter: **Malgorzata Bogdan**, University of Wroclaw, Poland
Co-authors: Damian Brzyski, Christine Peterson, Piotr Sobczyk, Emmanuel Candes, Chiara Sabatti
With the rise of both the number and the complexity of traits of interest, control of the false discovery rate (FDR) in genetic association studies has become an increasingly appealing and accepted target for multiple comparison adjustment. The nature of this error rate is intimately tied to the precise way in which discoveries are counted, and the performance of FDR controlling procedures is satisfactory only if there is a one-to-one correspondence between what scientists describe as unique discoveries and the number of rejected hypotheses. The presence of linkage disequilibrium between markers in genome-wide association studies (GWAS) often leads researchers to consider the signal associated to multiple neighboring SNPs as indicating the existence of a single genomic locus with possible influence on the phenotype. This a posteriori aggregation of rejected hypotheses results in inflation of the relevant FDR. We propose a novel approach to FDR control that is based on pre-screening to identify the level of resolution of distinct hypotheses. We show how FDR controlling strategies can be adapted to account for this initial selection both with

theoretical results and simulations that mimic the dependence structure to be expected in GWAS.

EO569 Room 210 HIGH AND INFINITE DIMENSIONAL TIME SERIES ANALYSIS Chair: Pramita Bagchi

EO0617: **A flexible regression model for functional time series**
Presenter: **HyeYoung Maeng**, London School of Economics, United Kingdom
Co-authors: Piotr Fryzlewicz
A prediction model is introduced for functional time series. Based on the classical scalar-on-function regression, the idea is to split the observed daily curves into several pieces to apply different smoothness. The proposed model allows more smoothing on observations located far from the prediction point compared to the closely located ones. In contrast to the typical time series prediction, our proposal gives flexibility in modelling in the sense that it offers less weight on the interval which is considered less important than others by fitting functional variables. The model in its simplest form includes one functional and one or more scalar covariates which is classified as semi-functional regression. In our approach, the change point which divides scalar and functional variable can be estimated from data. Illustrations on real data sets are given to show that the new model outperforms existing competitors. The asymptotic properties will also be presented.

EO0993: **An open source architecture for online monitoring, statistical analysis and forensics of multi-gigabit streams**
Presenter: **Shrijita Bhattacharya**, University of Michigan, United States
Co-authors: Stilian Stoev
The Internet, as a global system of interconnected networks, carries an extensive array of information resources and services. Key requirements include good quality-of-service and protection of the infrastructure from nefarious activity (e.g. cyber-attacks, password or port scans, etc). We develop an open source and readily deployable architecture, AMON (All-packet MONitor), for online monitoring and sequential analysis of multi-gigabit data streams under relatively stringent time and space constraints. AMON examines all packets passing through an interface, partitions traffic into sub-streams by using rapid hashing and computes certain real-time statistical summaries. The resulting data structures provide views of the intensity and connectivity structure of network traffic at the time-scale of routing. Internet data being heavy tailed in nature requires the estimation of its tail index to allow for the detection online fluctuations in traffic. However the tail index is susceptible to anomalies in the data. We thereby propose a robust version of the Hill estimator, which is optimal among the class of all estimators immune to extreme outliers in the data. We demonstrate our framework in the context of detection/identification of heavy-hitters as well as the visualization and statistical detection at the time-of-onset of high-connectivity events such as DDoS (distributed denial of service).

EO1172: **Parameter estimation and inference in a continuous piecewise linear regression model**
Presenter: **Georg Hahn**, Imperial College London, United Kingdom
The estimation of regression parameters in one-dimensional broken stick models is a research area of statistics with an extensive literature. We are interested in extending such models to two dimensions and hence in recovering two intersecting planes in a two-dimensional broken plane model. In contrast to approaches which either aim to recover two lines (two planes) using a grid search for the change point (intersection line) as well as approaches using local smoothing around the change point, we show how to use Nesterov smoothing to obtain a smooth and everywhere differentiable approximation to the broken plane model with a uniform error bound. The parameters of the smoothed approximation are then efficiently found by minimizing a least squares objective function using the Newton algorithm. The main contribution is threefold: we show that the estimates of the Nesterov smoothed approximation of the broken plane model are also $\sqrt{n}$-consistent and asymptotically normal, where n is the number of data points on the two planes. Moreover, we show that as the degree of smoothing goes to zero, the smoothed estimates converge to the unsmoothed estimates and present an algorithm to perform parameter estimation. We conclude by presenting simulation results on simulated data as well as guidance for practical applications.

EO1169: **Inference for the cross covariance of stationary functional time series**
Presenter: **Gregory Rice**, University of Waterloo, Canada
When considering two or more time series of functions, for example those derived from densely observed intraday stock price data of several companies, one commonly wishes to understand the dependence structure across series. This can be measured by means of the empirical cross covariance operator, but statistical procedures for measuring the significance of such estimators are limited. We present methodology for conducting statistical inference on the cross covariance operator estimated between two stationary functional time series. Two separate problems are considered: testing for a specified cross covariance structure, and testing for a single change point in the cross covariance within the sample.

EO727 Room 212 DEPENDENCE MODELS AND COPULAS I Chair: Fabrizio Durante

EO0618: **Weak convergence of the empirical copula process with respect to weighted metrics**
Presenter: **Betina Berghaus**, Ruhr University Bochum, Germany
Co-authors: Axel Buecher, Stanislav Volgushev
The empirical copula process plays a central role in the asymptotic analysis of many statistical procedures which are based on copulas or ranks. Among other applications, results regarding its weak convergence can be used to develop asymptotic theory for estimators of dependence measures or copula densities, they allow to derive tests for stochastic independence or specific copula structures, or they may serve as a fundamental tool for the analysis of multivariate rank statistics. Weak convergence is addressed for the empirical copula process (for observations that are allowed to be serially dependent) with respect to weighted supremum distances. The usefulness of this result is illustrated by an application to estimation procedures for the Pickands dependence function arising in multivariate extreme-value theory.

EO1050: **The empirical beta copula and its applications**
Presenter: **Hideatsu Tsukahara**, Seijo University, Japan
Co-authors: Johan Segers, Masaaki Sibuya
The empirical beta copula is introduced by a simple idea of rearranging uniform random variates in the order specified by the componentwise ranks of the original sample. It turns out to be a special case of the empirical Bernstein copula, the degrees of all Bernstein polynomials being equal to the sample size. Necessary and sufficient conditions are given for a Bernstein polynomial to be a copula, and they imply that the empirical beta copula is a genuine copula. Furthermore, the empirical process based on the empirical Bernstein copula is shown to be asymptotically the same as the ordinary empirical copula process under fairly weak assumptions. A Monte Carlo simulation study shows that the empirical beta copula outperforms the empirical copula and the empirical checkerboard copula in terms of both bias and variance. Compared with the empirical Bernstein copula with the smoothing rate suggested in the literature, its finite-sample performance is still significantly better in several cases, especially in terms of bias. Some resampling schemes using the empirical beta copula are explored to see if any beneficial effect on the accuracy of resampling schemes for the empirical copula process.

EO0172: **Modeling the serial dependence of financial returns with copulas**
Presenter: **Fabian Spanhel**, LMU Munich, Germany
The aim is to investigate to what extent stylized facts of financial returns, such as the martingale difference property, volatility clustering and the leverage effect, can be characterized in terms of copulas. We derive sufficient and necessary conditions for a conditionally symmetric martingale difference sequence in terms of copulas. These conditions reveal that the only possible martingale difference sequence that can be generated by commonly used copula families is a sequence of independent random variables. We then focus on the construction of copula-based first-order Markov processes that exhibit volatility clustering. For this purpose, we introduce two dependence properties that are sufficient for a positive correlation between squared (or absolute) symmetric random variables or an increasing transition variance in the absolute value of one conditioning variable. Moreover, several construction methods of copulas with the desired dependence properties are presented and compared. Finally, we explore the construction of higher-order Markov processes that exhibit volatility clustering. An application of the copula-based time series models to the returns of three major stock indices and one currency exchange rate documents the competitiveness with established GARCH models.

EO1746: **Using the Rosenblatt transformation to compute joint probabilities for random vectors**
Presenter: **Klaus Herrmann**, Concordia University, Canada
The problem of numerically computing the measure of a hyper-rectangle under a given multivariate distribution function is revisited. While most approaches are based on integrating the corresponding density function we take clues from recent approaches that focus on geometrical aspects. Instead of working in the original coordinate system we are using the Rosenblatt transformation to standardize the problem. This consequently leads to a transformation of the hyper-rectangle into a corresponding set within the unit hypercube. We then discuss numerical algorithms to efficiently approximate the measure of this corresponding set under the given transformation. Finally we discuss applications to certain copula families and multivariate distributions.

EO561 Room 216 MODERN STATISTICAL METHODS FOR ANALYSIS OF COMPLEX DATA Chair: Shaojun Guo

EO0663: **Spectral analysis of high-dimensional time series with applications to the mean-variance frontier**
Presenter: **Alexander Aue**, UC Davis, United States
Co-authors: Haoyang Liu, Debashis Paul
The limiting spectral behavior of the covariance and symmetrized autocovariance matrices of a class of high-dimensional linear time series is discussed, where the asymptotic regime is such that dimensionality and sample size grow proportionally. The results extend the classical Marcenko-Pastur law to the time series case. The form of the limiting spectral distribution is exploited to estimate the mean-variance frontier, an important measure of the minimum risk required for a fixed expected return in a portfolio of financial assets. The results may help alleviate the risk underestimation of the mean-variance frontier well documented in the finance and econometrics literature.

EO1294: **Combined hypothesis testing on graphs with applications to gene set enrichment analysis**
Presenter: **Ming Yuan**, University of Wisconsin-Madison, United States
Motivated by gene set enrichment analysis, we investigate the problem of combined hypothesis testing on a graph. We introduce a general framework to effectively use the structural information of the underlying graph. A new test is proposed within this framework. We show that the test is optimal in that it can consistently detect departure from the collective null at a magnitude that no other test could improve, for almost all graphs. We also provide general performance bounds for the proposed test test under any specific graph, and illustrate their utility through several common types of graphs. Numerical experiments are presented to further demonstrate the merits of our approach.

EO1303: **On the maximum cross correlations under high dimension**
Presenter: **Han Xiao**, Rutgers University, United States
Multiple time series often exhibits cross lead-lag relationship among its component series. It is very challenging to identify this type of relationship when the number of series is large. We study the lead-lag relationship in the high dimensional context, using the maximum cross correlations and some other variants. Asymptotic distributions are obtained. We also use moving blocks bootstrap to improve the finite sample performance.

EO1321: **Bayesian multiple changepoint detection**
Presenter: **Robert Lund**, Clemson University, United States
A method is presented to estimate the number of changepoint times and their locations in time-ordered data sequences. Our methods are driven by the need to identify mean shifts in temperature time series; in this case, a record exists of some (not necessarily all) of the times where the station moves locations or instrumentation is changed. The mixing of multiple undocumented and documented changepoint times is the key nuance. A prior distribution of changepoint times is constructed for this situation, and a penalized likelihood objective function is developed by modifying minimum description length information theory principles. Optimizing the objective function, which is done by Markov chain Monte Carlo simulations, yields estimate(s) of the changepoint numbers and their location time(s). Our techniques permit autocorrelation in the observations and seasonal means. Asymptotic consistency of the estimated configuration is established under infill asymptotics.

EO467 Room 002 DIMENSION REDUCTION AND ROBUSTNESS IN CAUSAL INFERENCE Chair: Xavier de Luna

EO0706: **Doubly-robust estimation of semiparametric additive hazards models**
Presenter: **Oliver Dukes**, Ghent University, Belgium
Co-authors: Stijn Vansteelandt, Torben Martinussen
Additive hazards models are becoming increasingly popular in survival analysis. Their parameters are easily interpretable in terms of relative survival risks, and are moreover collapsible, which makes the development of mediation and instrumental variable approaches more manageable. Estimation of the effect of an exposure in such models typically demands adjustment for a high-dimensional covariate. Standard estimation is then less desirable, as misspecification of the effect of these covariates may induce large bias in the exposure effect estimate. To overcome this, we consider a novel class of semiparametric additive hazards models which leave the effects of baseline covariates unspecified. We derive the efficient score for the exposure effect in these models, which we assume to be constant over time. We argue that using this score may not be advisable in practice, as it requires a model for the conditional distribution of the exposure, given covariates. We therefore derive the efficient score in a subclass of estimators, which require this conditional distribution merely to be correctly specified up to the mean. We show that the resulting estimator enjoys a double-robustness property. Finally, we give guidance on efficiency, and discuss how the proposed estimators lay the foundations for G-estimation of structural nested cumulative survival models for the effect of time-varying exposures.

EO0999: **Data-driven confounder selection via Markov and Bayesian networks**
Presenter: **Jenny Haggstrom**, Umea University, Sweden
To unbiasedly estimate a causal effect on an outcome unconfoundedness is often assumed. If there is sufficient knowledge on the underlying causal structure then existing confounder selection criteria can be used to select subsets of the observed pretreatment covariates, X, sufficient for unconfoundedness, if such subsets exist. The estimation of these target subsets is considered when the underlying causal structure is unknown. The

proposed method is to model the causal structure by a probabilistic graphical model, e.g. a Markov or Bayesian network, estimate this graph from observed data and select the target subsets given the estimated graph. The approach is evaluated by simulation both in a high-dimensional setting where unconfoundedness holds given X and in a setting where unconfoundedness only holds given subsets of X. Several common target subsets are investigated and the selected subsets are compared with respect to accuracy in estimating the average causal effect. The proposed method is implemented with existing software that can easily handle high-dimensional data, in terms of large samples and large number of covariates. The results from the simulation study show that, if unconfoundedness holds given X, this approach outperforms alternative approaches based on random forests and LASSO.

EO1148: **Collaborative targeted learning using regression shrinkage**
Presenter: **Mireille Schnitzer**, Universite de Montreal, Canada
Co-authors: Matthew Cefalu
Causal inference practitioners are routinely presented with the challenge of wanting to adjust for large numbers of covariates despite limited sample sizes. Collaborative Targeted Maximum Likelihood Estimation (CTMLE) is a general framework for constructing doubly robust semiparametric causal estimators that data-adaptively reduce model complexity in the propensity score in order to optimise a preferred loss function. This stepwise complexity reduction is based on a loss function placed on a strategically updated model for the outcome variable, assessed through cross-validation. New work involves integrating penalized regression methods into a stepwise CTMLE procedure that may allow for a more flexible type of model selection than existing variable selection techniques. Two new algorithms are presented and methods to reduce computational complexity, based on previous work, are assessed.

EO0679: **Matching using sufficient dimension reduction for causal inference**
Presenter: **Yeying Zhu**, University of Waterloo, Canada
Co-authors: Wei Luo
Sufficient dimension reduction is employed and the resulting low dimensional covariates in subsequent matching is used to estimate the causal effects. Under the ignorability assumption, the reduced covariates are balancing scores, and the proposed causal effect estimator is consistent. Compared to the original covariates and the propensity scores, which are commonly used for matching in the literature, the reduced covariates are estimable nonparametrically under a mild assumption and are sufficient in imputing the missing potential outcomes. In addition, it requires a weaker common support condition, which makes the resulting matching procedure more applicable.

EO125 Room 215 DETECTING STRUCTURAL CHANGES IN MULTIDIMENSIONAL DATA Chair: Zuzana Praskova

EO0732: **Uniform change point tests in high dimension**
Presenter: **Moritz Jirak**, TU Braunschweig, Germany
Consider d dependent change point tests, each based on a CUSUM-statistic. We provide an asymptotic theory that allows us to deal with the maximum over all test statistics as both the sample size n and d tend to infinity. We achieve this either by a consistent bootstrap or an appropriate limit distribution. This allows for the construction of simultaneous confidence bands for dependent change point tests, and explicitly allows us to determine the location of the change both in time and coordinates in high-dimensional time series. If the underlying data has sample size greater or equal n for each test, our conditions explicitly allow for the large d small n situation, that is, where $n/d \to 0$. The setup for the high-dimensional time series is based on a general weak dependence concept. The conditions are very flexible and include many popular multivariate linear and nonlinear models from the literature, such as ARMA, GARCH and related models. The construction of the tests is completely nonparametric, difficulties associated with parametric model selection,model fitting and parameter estimation are avoided. As an application, we analyze all S& P 500 companies over a period of one year.

EO1044: **Detecting changes in panel data**
Presenter: **Jaromir Antoch**, Charles University in Prague, Czech Republic
The analysis of panel data will be considered when the design vectors are assumed nonrandom and the panels can be dependent via common factor. We will present results concerning the wild bootstrap and show how to get simulated critical values for statistics under consideration. Application for real data about firms' financial status will be presented too.

EO1018: **Change point detection in a dynamic panel data model**
Presenter: **Zuzana Praskova**, Charles University in Prague, Czech Republic
A panel data model with lagged dependent variables and unobserved individual effects is considered and a procedure to detect change in coefficients of lagged variables is proposed. The individual effects are eliminated by a proper transformation that enables to represent the data by a vector autoregressive model. Test statistic to detect change is based on quasi likelihood scores and quasi maximum likelihood estimators. Asymptotic properties of the test statistic are studied in case that both the number of panels and the number of observations is sufficiently large.

EO1301: **Detection of changes and inference for high-dimensional covariance matrices**
Presenter: **Ansgar Steland**, University Aachen, Germany
New results about inference and change point analysis of high dimensional vector time series are discussed. The results address bilinear forms of sample variance-covariance as arising when studying projection statistics. The results hold true without any constraint on the dimension, the sample or their ratio, contrary to many known results, provided the weighting vectors are uniformly l1-bounded. Those results are in terms of (strong resp. weak) approximations by Gaussian processes, which imply (functional) central limit theorems under certain conditions. More generally, one can also study an increasing number of such bilinear forms. One may obtain approximations by Gaussian processes without any constraint on the number of forms, the dimension, the sample size or their ratios. We also discuss applications and related extensions of the theoretical results, focusing on shrinkage covariance matrix estimation. Other fields of applications are sparse principal component analysis and dictionary learning.

EO161 Room 206 ROBUST METHODS FOR RELIABLE FINANCIAL DECISION MAKING Chair: Marjan Wauters

EO0880: **Nowcasting economic sentiment for forecasting economic growth**
Presenter: **Keven Bluteau**, Institute of Financial Analysis, Neuchatel University, Switzerland
Co-authors: Kris Boudt, David Ardia
Economic sentiment surveys are carried out by all European union member states and often considered as early indicators for future economic development. By construction, there is however a publication lag between the time at which the survey is carried out and the time of publication. We show how textual analysis can be used to nowcast economic sentiment and that these nowcasts are useful for forecasting economic growth.

EO0852: **Robust nearest comoment estimation using unobserved factors**
Presenter: **Dries Cornilly**, Vrije Universiteit Brussel, Belgium
Co-authors: Tim Verdonck, Kris Boudt
A flexible estimator is proposed for the higher order comoments under a linear factor model with unobserved factors. Robustness of the estimator

is obtained by cell-wise winsorization of the sample comoments, while the factor structure is imposed by searching for the nearest comoment satisfying the factor model assumptions. By construction, the resulting covariance estimate is positive semi-definite, all estimated comoments are robust to outliers and the estimation remains feasible in moderately large dimensions. We show the usefulness of the method in finance, where the robust nearest comoment estimates lead to more accurate estimates of the Cornish-Fisher value-at-risk than the sample based approach.

EO0824: **Beyond risk-based portfolios: Balancing performance and risk contributions in asset allocation**
Presenter: **Giang Nguyen**, Vrije Universiteit Brussel, Belgium
Co-authors: Kris Boudt, David Ardia
In a risk-based portfolio, there is no explicit control for the performance per unit of risk taken. We propose a flexible framework to evaluate and optimize the balance between risk and performance at both the portfolio and component level. The first objective is achieved by imposing a target relative performance constraint at the aggregate portfolio level. The aggregate alignment may mask large discrepancies at the component level. We thus complement the constraint with the objective of minimizing the so-called Performance/Risk Contribution Concentration (PRCC) measure. This measure is designed to be minimal when, for all portfolio components, the performance and risk contributions are perfectly aligned. We investigate the theoretical properties of this measure and show its usefulness in asset allocation. Compared to the traditional risk-based portfolios (equally weighted, minimum variance, equal risk contribution and maximum diversification), minimizing the PRCC leads to higher total and relative performance.

EO0843: **Block bootstrap impact evaluation of the equity portfolio choice on CPPI performance**
Presenter: **Marjan Wauters**, VUB, Belgium
Co-authors: Kris Boudt, David Ardia
Determining whether combining smart beta and portfolio insurance is mutually beneficial or not is complex, because of the non-linear structure of portfolio insurance products and the market regime dependence due to the long investment horizons. We propose to use a block bootstrap evaluation framework to simulate historical performance of CPPIs under several scenarios. We find that portfolio insurance preserves the improved risk-adjusted performance of the smart beta strategies in normal market regimes and offers protection against the non-diversifiable systematic risk of sudden market downturns.

EO111 Room 205 GRAPHICAL MODELS: DEPENDENCE IN NETWORKS Chair: Giovanni Marchetti

EO1089: **Testing conditional independence**
Presenter: **Wicher Bergsma**, London School of Economics, United Kingdom
In a multivariate normal setting, conditional independence is equivalent to certain partial correlations being zero. In more general settings, however, it is more involved. We show some ways in which conditional independence can be understood, and how this can lead to appropriate conditional independence tests without having to assume multivariate normality.

EO0867: **The networked partial correlation and its application to the analysis of genetic interactions**
Presenter: **Robert Castelo**, Universitat Pompeu Fabra, Spain
Co-authors: Alberto Roverato
Genetic interactions confer robustness on cells in response to genetic perturbations. This often occurs through molecular buffering mechanisms that can be predicted using, among other features, the degree of coexpression between genes, commonly estimated through marginal measures of association such as Pearson or Spearman correlation coefficients. However, marginal correlations are sensitive to indirect effects and often partial correlations are used instead. Yet, partial correlations convey no information about the (linear) influence of the coexpressed genes on the entire multivariate system, which may be crucial to discriminate functional associations from genetic interactions. To address these two shortcomings, and in a similar vein to Sewall Wright's path coefficients, we propose to use the edge weight derived from the covariance decomposition over the paths of the associated gene network. We call this new quantity the networked partial correlation and use it to analyze genetic interactions in yeast.

EO1049: **Credit risk model with network effects for a large panel of companies**
Presenter: **Veronica Vinciotti**, Brunel University London, United Kingdom
Co-authors: Elisa Tosetti, Francesco Moscone
A credit risk model is developed with network effects for a very large panel of companies. We assume a probit specification with group random effects having a non-diagonal, sparse covariance matrix. We propose a penalised maximum likelihood estimation approach and develop an Expectation-Maximization algorithm where we exploit the properties of truncated Normals to efficiently approximate conditional expectations. A simulation study shows good properties of our approach at a significantly reduction in computational cost. We fit this model on a dataset of over one million accounts for small and medium-sized enterprises in the United Kingdom over the period from 2009 to 2015. We group companies according to their sector and geographical location. We find that accounting for network effects makes a significant contribution to increasing the default prediction power of risk models built specifically for SMEs, compared to a more conventional probit prediction model. Our results may help bankers to improve their credit scoring of SMEs ultimately reducing their propensity to apply excessive lending restrictions.

EO1352: **Sparsity in Gaussian totally positive distributions**
Presenter: **Piotr Zwiernik**, Universitat Pompeu Fabra, Spain
Probability distributions that are multivariate totally positive of order 2 (MTP2) appeared in the theory of positive dependence and in statistical physics through the celebrated FKG inequality. The MTP2 property is stable under marginalization, conditioning and it appears naturally in various probabilistic graphical models with hidden variables. Models of exponential families with the MTP2 property admit a unique maximum likelihood estimator. In the Gaussian case, the MLE exists also in high-dimensional settings, when $p >> n$, and it leads to sparse solutions, which provides an interesting alternative to the graphical lasso method.

CO525 Room 111 TIME-VARYING DEPENDENCIES Chair: Dominik Wied

CO0158: **Sequential detection of parameter changes in dynamic conditional correlation models**
Presenter: **Katharina Pape**, TU Dortmund, Germany
Co-authors: Pedro Galeano, Dominik Wied
A multivariate monitoring procedure is presented to detect changes in the parameter vector of the dynamic conditional correlation model proposed by Robert Engle in 2002. The benefit of the proposed procedure is that it can be used to detect changes in both the conditional and unconditional variance as well as in the correlation structure of the model. The detector is based on the contributions of individual observations to the gradient of the quasi log likelihood function. More precisely, standardized derivations of quasi log likelihood contributions of points in the monitoring period are evaluated at parameter estimates calculated from a historical period. The null hypothesis of a constant parameter vector is rejected if these standardized terms differ too much from those that were expected under the assumption of a constant parameter vector. Under appropriate assumptions on moments and the structure of the parameter space, limit results are derived both under null hypothesis and alternatives. In a simulation study, size and power properties of the procedure are examined in various scenarios.

CO0192: **A mixed frequency stochastic volatility model for intraday stock market returns**
Presenter: **Bastian Gribisch**, University of Cologne, Germany
Co-authors: Jeremias Bekierman
A mixed frequency stochastic volatility (MFSV) model is proposed for the dynamics of intraday asset return volatility. In order to account for long-memory we separate stochastic daily and intraday volatility patterns by introducing a long-run component that changes at daily frequency and a short-run component that captures the remaining intraday volatility dynamics. An additional component captures deterministic intraday patterns. The resulting non-linear state-space model is estimated in a single step using simulated maximum likelihood based on Efficient Importance Sampling (EIS). In addition to intraday volatility estimation and forecasting the model can be applied in order to forecast volatility at daily frequency incorporating intraday information on stock price fluctuations and daily realized volatility measures. We apply the model to intraday returns of five New York Stock Exchange traded stocks. The estimation results indicate distinct dynamic patterns for daily and intradaily volatility, where most of the volatility dynamics are explained by the daily volatility component. In-sample diagnostic tests and out-of-sample Value-at-Risk (VaR) forecasts show that already the very basic model specification successfully accounts for the strong persistence of intraday asset return volatility.

CO0200: **Add the beef and combine: Dynamic density combinations from point forecasts**
Presenter: **Laura Hersing**, Karlsruhe Institute of Technology, Germany
Co-authors: Oliver Grothe
A real-time forecasting problem is faced, where one-period-ahead point forecasts for a certain variable of interest are available at each point in time. Such forecasting problems are relevant in many areas of the economy, e.g. in energy and financial markets. Typically, different commercial providers offer point forecasts for the corresponding variables of interest. We aim at constructing density forecasts out of these point forecasts and at combining these forecasts dynamically to a joint forecasting density. Transforming the point forecasts to density forecasts allows for considering more advanced risk metrics than in the point forecasting case. Forecasting the density allows the user of the forecast to dynamically calibrate its risk over time, particularly in a high-frequency setup. Our approach consists of two main steps. First, we transform the available point forecasts to univariate, conditionally optimal density forecasts (adding the beef to the bone). Second, we infer the time dynamic copula between the forecast densities and use it to combine the density forecasts (combination part). We show that our method outperforms other techniques that are using or combining the different point forecasts with respect to a variety of loss measures.

CO0302: **Econometric methods for detecting imbalances in house prices**
Presenter: **Andre Anundsen**, Norges Bank, Norway
The question whether house price bubbles can be detected in real time is addressed. The answer seems to be yes. Four different econometric methods are considered to construct indicators of housing market imbalances for the US, Finland and Norway. We investigate if the house price development in these countries in the 2000's can be explained by underlying economic fundamentals, or whether it is best characterized by bubble-dynamics. For the US, all measures suggest a bubble in the early to mid 2000's, while only one measure indicates imbalances in the Finnish housing market. For the case of Norway, none of the measures provide evidence of systematic overheating. The global savings glut and the relaxation of credit standards have been put forth as alternative explanations of the US housing bubble. The results suggest that the bubble cannot be explained by capital inflows, but that it was caused by the increased exposure to the subprime mortgage market.

CO0503: **New backtests for unconditional coverage of the expected shortfall**
Presenter: **Robert Loeser**, TU Dortmund, Germany
Co-authors: Dominik Wied, Daniel Ziggel
While the Value-at-Risk (VaR) has been the standard risk measure for a long time, the Expected Shortfall (ES) has become more and more popular in recent times, as it provides important information about the tail risk. We present a new backtest for the unconditional coverage property of the ES. The test is based on the so called cumulative violation process. The main advantage is that the distribution of the test statistic is available for finite out-of-sample size which leads to better size and power properties compared to existing tests. Moreover, it can be easily extended to a multivariate test.

CO628 Room S22 ADVANCES IN MACROECONOMETRICS Chair: Luca Fanelli

CO0168: **DSGE Pileups**
Presenter: **Stephen Morris**, Bowdoin College, United States
Restrictions from economic theory are usually placed on the parameter space in DSGE models. Such restrictions are also utilized to ensure identifiability. But even when the population parameter value is consistent with theory, the likelihood maximizer need not be. The aim is to document the resulting "pileup" phenomenon in the maximum likelihood estimator and to illustrate its consequences for Bayesian inference. The results help to clarify why one frequently observes that the small sample distribution of the MLE and the posterior probability are multimodal, even when regularity conditions including global identifiability are apparently satisfied.

CO0360: **Improving inference for dynamic economies with frictions: The role of qualitative survey data**
Presenter: **Andreas Tryphonides**, European University Institute, Italy
A new inferential methodology is proposed that is robust to misspecification of the mechanism generating frictions in a dynamic stochastic economy. The frictions can be real, nominal or informational. The approach treats economies with frictions as perturbations of a frictionless economy that are consistent with a variety of mechanisms. Models and their parameters are therefore set identified. Three contributions are made. First, we derive a characterization of the model economy that provides identifying restrictions on the solution of the model. Second, we show how qualitative

survey data can be linked to the expectations of agents and how this link generates an additional set of identifying restrictions on the probability of observing a distortion in a macroeconomic variable. We provide conditions under which the additional restrictions lead to a smaller set of admissible models. Third, we show how the framework can be used to validate mechanisms that generate frictions. We propose a test statistic, derive its large sample properties and provide a bootstrap procedure to compute the critical values. Finally, we apply the methodology to estimate the distortions in the Spanish economy due to financial frictions using qualitative survey data collected by the European Commission on the financial constraints of agents, and compute a robust optimal Taylor rule.

CO0833: **Expectations correction and DSGE model selection**
Presenter: **Giovanni Angelini**, University of Bologna, Italy
The focus is on a bootstrap refinement of the Expectations Correction idea already proposed in the literature. Expectations Correction is a feasible remedy to the dynamic misspecification that characterizes the small-scale New-Keynesian monetary policy models. These models are generally not able to take into account all the dynamic correlation structure present in quarterly data hence the empirical performance of these models could be problematic. To solve this problem a pseudo-structural model is built from the baseline of a DSGE model by adding a number of lags present in the statistical model for the data. We refine this idea using bootstrap techniques for both the lag structure selection and both in the empirical evaluation. The selection of the true number of lags is a crucial point in the definition of the pseudo-structural model and it is particularly important in the forecasting performance. In this sense bootstrap techniques are particularly useful to identify the true lag structure. Moreover, the addition of new lags in the structural model produces a higher number of non-linear cross equation restrictions and this is the cause of a high rejection rate of the theoretical model and a bootstrap approach can reduce this overejection. To evaluate the performances of our method we provide a Monte Carlo simulation study and an empirical illustration based on U.S. quarterly data.

CO0676: **Measuring nonfundamentalness for structural VARs**
Presenter: **Stefano Soccorsi**, ECARES, Universite Libre de Bruxelles, Belgium
As nonfundamental vector moving averages do not have causal VAR representations, standard structural VAR methods are deemed inappropriate for recovering the economic shocks of general equilibrium models with nonfundamental reduced forms. In previous literature it has been pointed out that, despite nonfundamentalness, structural VARs may still be good approximating models. I characterize nonfundamentalness as bias depending on the zeros of moving average filters. However, measuring the nonfundamental bias is not trivial because of the simultaneous occurrence of lag truncation bias. I propose a method to disentangle the bias based on population spectral density and derive a measure for the nonfundamental bias in population. In the application, I find that the SVAR exercises of Sims (2012) are accurate because the nonfundamental bias is mild.

CO0502: **On the fundamentalness of nonfundamentalness in DSGE models**
Presenter: **Marco Sorge**, University of Goettingen, Germany
The role of indeterminacy (e.g. equilibrium non-uniqueness) is explored for the VAR-based analysis of dynamic stochastic general equilibrium (DSGE) models. By means of simple examples, we first show that determinate models may well entail nonfundamental (i.e. noninvertible) equilibrium reduced forms - which only (and uniquely) depend on the structural (fundamental) shocks - whereas indeterminate ones may actually deliver fundamental representations in the observables. We then investigate conditions for the emergence of nonfundamentalness in a general class of (linearized) DGSE models, which nests the prototypical New Keynesian framework as a special case. Monte Carlo simulations show that an indeterminate equilibrium model need not engender nonfundamentalness, even when its determinate counterpart always proves nonfundamental. As a main implication, detecting a causal representation of the data cannot be interpreted as evidence of determinacy.

CO489 Room 107 INDICATORS FOR RISK MONITORING AND IMBALANCING DETECTION Chair: Gian Luigi Mazzi

CO0241: **How to predict financial stress: An assessment of Markov switching versus logit models**
Presenter: **Thibaut Duprey**, Bank of Canada, Canada
Co-authors: Benjamin Klaus
The gap between the business cycle literature and the literature on currency, banking and financial crises is bridged by (i) comparing a Markov switching (MS) and a binary logit model to assess their ability of predicting the occurrence of high financial stress episodes, and (ii) applying a particular form of MS model to predict the transition to a high financial stress regime and back to a tranquil regime. The dependent variable used by the MS model to capture the financial cycle in the EU countries is a financial stress index. The results indicate that both models have a relatively similar ability to predict high financial stress episodes, with the MS model outperforming the logit model between six to one quarters prior to the onset of financial stress episodes. Based on cross-country estimations, debt service ratios and housing variables indicate a transition to a high financial stress regime, while equity price growth and economic sentiment indicators provide signals for a transition to a tranquil state. The MS model results based on country-specific data suggest that country-specific factors can be identified and might be valuable for the design of early warning models.

CO0435: **Financial crisis indicators based on implied correlations computed from option prices**
Presenter: **Antoine Kornprobst**, University Paris Sorbonne, France
The aim is to build financial crisis and systemic risk indicators based on the study of the differences between implied volatilities and correlations, which are computed using available option (and option baskets) prices, and realized volatilities and correlations, which are computed from historical spot prices. This forward-looking approach using computed implied quantities is popular at the moment and many market tools, like the CBOE Volatility Index (VIX) and several CBOE Implied Correlation Indices, are based on similar ideas. Our approach is novel in the sense that we intend to build financial crises indicators and market forecasting tools based on implied correlations between the components or sector components of several equity indices. The first part of the study builds practical tools based on the daily comparison between implied and realized correlations, explore the link between our indicators and the CBOE indicators and also attempt to replicate the results of the CBOE Implied Correlation Indices. Then, in a second part, optimal active trading strategies based on those financial crisis indicators are constructed and tested against typical passive investment strategies.

CO1204: **Risk measure estimates in quiet and turbulent times: An empirical study**
Presenter: **Marie Kratz**, ESSEC Business School, CREAR, France
The aim is to test the relevance in practice, especially in time of financial crisis, of the Sample Quantile Process as a risk measure compared with the use of the historical VaR, or when using a GARCH model. We also specifically analyse the VaR following a period of crisis characterized by a high volatity. To do so, we develop an empirical study inspired by a previous one performed on S& P 500 data and show that in time of high volatility the risk measure calculated on the historical data overestimate the risk measure computed on the future data. This expected feature is evidenced by our empirical results. Moreover, this fact introduces a pro-cyclical behavior of the risk measure estimate, tending to overestimate the risk in the future, whereas, in quiet time (low volatility), the risk measure estimate tends to underestimate the future risk by a large amount.

CO0913: **Dissecting the financial cycle with dynamic factor models**
Presenter: **Christian Menden**, University of Bamberg, Germany
Co-authors: Christian Proano
The analysis of the financial cycle has become a central topic since the 2007-08 financial crisis. So far, the great majority of studies has analyzed the cyclical properties of the financial cycle by means of a small number of financial risk indicators at a rather aggregated level. An empirical approach is proposed to extract information about the financial cycle from a large data set of macroeconomic and financial variables for the US. Using a dynamic factor model we estimate three synthetic financial cycle components that account for the majority of the variation in the complete data set. Peaks in our synthetic financial cycle components coincide with economic recessions giving rise to potential properties as early warning indicators for financial and economic distress. Thus, we investigate whether our financial cycle components have significant predictive power for GDP growth, output gap, inflation and short-term interest rates by means of Granger causality tests. Further, we analyze the forecasting power of recessions by estimating the recession probabilities based on a probit approach.

CO1068: **Use of unit root methods in early warning of financial crises**
Presenter: **Eero Tolo**, Bank of Finland, Finland
Co-authors: Matti Viren, Timo Virtanen, Katja Taipalus
Unit-root based methods have for long been used to study the existence of financial bubbles in asset prices. The basic idea is that certain fundamental changes in the autocorrelation structure of relevant time series imply the presence of a rational price bubble. We provide cross-country evidence for performance of unit-root based early warning systems in ex-ante prediction of financial crises in 15 EU countries in the past 3 decades. The early warning signals have quite high relative usefulness. To further improve the crisis prediction, the signals from multiple time series can be combined into a composite indicator. It is also possible to use a mix of data with different frequencies, which can be useful for providing more timely warning signals. The results suggest that this approach has a place in the toolkit of financial stability supervision.

CO415 Room 105 TOPICS IN MODELLING DEPENDENCE AND SAMPLE SELECTION Chair: Artem Prokhorov

CO0262: **Efficient GMM estimation with incomplete observations**
Presenter: **Chris Muris**, Simon Fraser University, Canada
The standard MAR model has 2 strata of missingness, complete observations and completely missing observations. A GMM missing data model is proposed for settings where observations may be incomplete, where some observations are missing some instruments and other observations are missing different instruments. This model is identified under MAR. The implied estimators attain the semiparametric efficiency bound and are easy to implement.

CO1104: **Conditional moment restriction models with missing data**
Presenter: **Valentin Patilea**, CREST-Ensai, France
A general statistical model is considered which is defined by moment restrictions when a subvector of data are missing. The main incomplete data situations we have in mind are missing at random and endogenous selection. Using the inverse probability weighting, we show that such a model is equivalent to a model for the observed variables only, augmented by a moment condition defined by the missingness mechanism. In particular, our framework covers parametric and semiparametric mean regressions and quantile regressions. We allow for missing responses, missing covariates and any combination of them. We present a general equivalence result, obtained under minimal technical conditions, that sheds new light on various aspects of interest in the missing data literature. It also provides guidelines for building (efficient) estimators. Moreover, as an immediate consequence of our general theory, we derive the efficiency of the complete cases analysis in a general semiparametric regression model with responses missing at random.

CO0186: **Consistent estimation of linear regression models using matched data**
Presenter: **Masayuki Hirukawa**, Setsunan University, Japan
Co-authors: Artem Prokhorov
Economists often use matched samples, especially when dealing with earnings data where a number of missing observations need to be imputed. We demonstrate that the ordinary least squares estimator of the linear regression model using matched samples is inconsistent and has a nonstandard convergence rate to its probability limit. If only a few variables are used to impute the missing data then it is possible to correct for the bias. We propose two semiparametric bias-corrected estimators and explore their asymptotic properties. The estimators have an indirect-inference interpretation and they attain the parametric convergence rate if the number of matching variables is no greater than three. Monte Carlo simulations confirm that the bias correction works very well in such cases.

CO0637: **Bayesian adaptive sparse copula**
Presenter: **Martin Burda**, University of Toronto, Canada
Co-authors: Artem Prokhorov
A new method is proposed for Bayesian multivariate nonparametric density estimation that is both adaptive and sparse. Our approach extends recent work on univariate multiscale nonparametric densities by requiring sparsity via an alternative functional approximation implemented with a spike-and-slab prior structure. Implementation of the resulting sparse multiscale density model has a flavor of multiscale importance sampling whereby major functional approximation components are preserved while minor components known to fall below a given threshold are not evaluated. As a result, the nonparametric density approximation requires only a fraction of the implementation time and memory size relative to its non-sparse counterpart at the expense of a negligible loss of precision controlled by the user. This makes our approach suitable for multivariate scenarios and usage in wider structural models; indeed so far the multiscale density methods proposed in the literature have been univariate. We further embed the sparse multiscale estimator within a multivariate nonparametric copula density model with countably infinite mixtures of location-scale marginals.

CO0986: **Copula by triangulation**
Presenter: **Yajing Zhu**, Concordia University, Canada
Co-authors: Artem Prokhorov, Edward Anderson
The focus is on simple arrangements for approximating copula densities with spline type surfaces while guaranteeing that our estimator is indeed a copula density. We start by showing the difficulty of approximating copula densities with piecewise linear surface while guaranteeing the uniform margin property because this implies that the estimation procedure would involve mixed integer optimization problems. We then turn to a straightforward method of applying the spline as basis functions to approximating copula densities. We propose a semi-parametric copula density estimation procedure that guarantees that the estimator is indeed a copula density. The estimation procedure involves a maximum likelihood estimation of the coefficients of the splines. With simple linear constraints included in the maximization problem, we are solving a convex optimization problem which is easy to solve numerically. Our estimation procedure can be easily generalized to irregular grid on the unit square instead of regular grid with equidistant knots, which implies good localization property. Our estimator also can be easily generated to higher dimensions. We construct a simulation-based study to examine the effect of sample sizes and extent of dependence on the performance of our copula density estimation method and compare with the leading copula density estimators.

CO389 Room 104 MEASURING SYSTEMIC RISK Chair: Monica Billio

CO0295: **Systemic risk contribution and bank business models**
Presenter: **Federico Nucera**, Luiss Guido Carli, Italy
Co-authors: Andre Lucas, Julia Schaumburg, Bernd Schwaab
A novel observation-driven dynamic finite mixture model is proposed for the study of high-dimensional banking data. A factor structure allows us to incorporate many bank-level measurements. The model accommodates time-varying component means and covariance matrices, as well as normal and Student-t distributed mixture densities. In an empirical study of 208 European banks between 2008q1-2015q4, we identify six different business models and demonstrate that they differ in their contribution to systemic risk

CO0750: **Time to consensus in financial causality networks: A Von Neumman entropy approach to contagion**
Presenter: **Lorenzo Frattarolo**, Ca Foscari Venezia, Italy
Co-authors: Roberto Casarin, Monica Billio, Michele Costola
Time to consensus is the time a networked system needs to reach a condition in which all of its constituents have the same state and can be considered a measure of information and shock diffusion in the system. Extending results in literature, to directed networks, we show that in the classical consensus dynamics, Shannon entropy decreases, while for quantum case, the von Neumann entropy (VNE) is non-decreasing and that they are equivalent frameworks, in terms of dynamics. We propose to measure classical time to consensus by measuring the time an equivalent quantum system takes to maximize its VNE. The relevance of consensus dynamics of stock returns for financial contagion could be understood by considering the convergence of returns to a common negative-valued state and the implied dynamical increase of dependence, a standard signature of financial contagion. We use rolling Granger Causality to approximate the networked dynamics of returns, obtain, by the Sinkorn-Knopp decomposition, a doubly stochastic matrix (the adjacency matrix of the corresponding balanced network) and compute a parsimonious Birkoff Von Neumman decomposition of the latter. Permutation matrices and convex weights from the decomposition are then used to compute, the quantum evolution operator, maximal VNE and time to consensus. We investigate the relevance of those measures also as an early warning for financial contagion episodes.

CO1073: **Hidden leaders: Identifying latent lead-lag structures in multivariate ultra-high-frequency returns**
Presenter: **Fulvio Corsi**, Ca Foscari University Venice and City University London, Italy
Co-authors: Giuseppe Buccheri, Stefano Peluso
A test is proposed for the presence of latent lead-lag structures in the dynamics of multivariate ultra-high-frequency returns. To account for the non-synchronous trading and microstructure noise in the observed prices, the model is formulated in a state space representation with missing data, where the state vector of latent returns follows a VAR process. The likelihood is computed using the Kalman filter and optimized through an EM algorithm. In addition, by disentangling contemporaneous covariances from autocovariances in the latent VAR process, the proposed method is the first providing unbiased and consistent covariances estimates in presence of microstructure noise, asyncronicity and lead-lag dependencies. Extensive simulation analysis shows the accuracy of the estimator in recovery the true lead-lag structure in the latent return process and how neglecting the lagged dependencies causes severe distortions in the estimation of contemporaneous covariances. Finally, the empirical application to equity data provides useful information on latent lead-lag relationships among high-frequency returns.

CO0946: **Systemic risk measures relevance: A permutation test approach to European banks**
Presenter: **Francesca Parpinel**, University CaFoscari of Venice, Italy
Co-authors: Claudio Pizzi, Lorenzo Frattarolo
The financial stability board defines the systemically important financial institutions as characterized by an important size, complexity and systemic interconnectedness, such that a possible their failure would produce a breakdown to the whole financial system and economic activity. The current determination of these institutions is based on yearly balance-sheet variables and expert judgment. A statistical procedure based on a permutation test is proposed in order to cluster systemically important financial institutions splitting the financial system. In particular, the weights of a combination of partial permutation tests are optimized, basing on several well-known systemic risk measures, available at daily frequency. The procedure will be applied to the European banking institutions, for which the European Banking Union fully discloses information used in the choice of systemically important financial institutions. The sample of banks in European Banking Authority is chosen, for which a stock quote is available. The results are tested to reproduce as close as possible the selection of systemically important financial institutions made by the Basel committee and to investigate in this way their relevance in the choice.

CO1553: **Capturing systemic risk by robust and sparse network estimation**
Presenter: **Gabriele Torri**, University of Bergamo, Italy
Co-authors: Rosella Giacometti, Sandra Paterlini
Network analysis is becoming a fundamental tool in the study of systemic risk and financial contagion. Still, the network structure has to be typically estimated from noisy data, as the true network structure is usually unobservable, and standard statistical methods return dense network structures, which are hard to be interpreted. We introduce an approach that allows to estimate sparse networks, capturing only the relevant links, and better deal with estimation error due to outliers. Empirical analysis on CDS spreads and equity returns highlights the ability of our approach to capture/infer the most relevant European bank system interconnectedness and contagion dynamics.

CO403 Room 106 ADVANCES IN COMPLEX SPATIAL AND SPATIO-TEMPORAL ECONOMETRICS Chair: Maria Kyriacou

CO0365: **Improving the UK index flood estimation by catchment characteristics with additive and spatial regression analyses**
Presenter: **Marinah Muhammad**, University of Southampton, United Kingdom
Co-authors: Zudi Lu
Flood modelling at ungauged catchment has always been a challenging problem. Regionalization is an important procedure widely used with the assumption flood peak can be explained by catchment characteristics. The Flood Estimation Handbook (FEH) model for index flood is a well-established model of this kind in the UK. There are however unexpected or unknown features in flood dataset that need to be taken account of. Two research questions are to be investigated: (1) Is the FEH model reliable in characterising the nonlinear effects of the catchment characteristics for the UK flooding? (2) Could we improve the FEH model with a better accuracy in index flood estimation?Potential nonlinear effects of the covariates are incorporated into the FEH model by an additive regression analysis. Moreover, spatial autocorrelation is examined on the regression residuals, and potential spatial neighbouring effect is incorporated into the FEH regression by spatial econometric models. The results show that: (i) our additive model analysis confirms the nonlinear impact of the catchment characteristics identified by the FEH model is reliable; (ii) the identified statistical significance of spatial autocorrelation indicates the spatial neighbouring effect that is not taken account of in the FEH regression should be well considered; (iii) it has been detected,the FEH regression with spatial error model is a most appropriate alternative to the FEH model, helping to improve index flood estimation.

CO1003: **Variable selection with spatially autoregressive errors: A generalized moments LASSO estimator**
Presenter: **Arnab Bhattacharjee**, Heriot-Watt University, United Kingdom
Co-authors: Liqian Cai, Roger Calantone, Taps Maiti
A new penalized variable selection and estimation method under a spatial regression model - a generalized moments LASSO estimator - is proposed as a combination of the LASSO with GMM estimation. We consider the spatial error model where the error term is spatially autoregressive across cross-section units. Statistical properties of penalized estimation methods like LASSO have not been studied under spatial regression models where the errors are not IID. We establish parameter consistency and selection sign consistency of the proposed estimator in the low dimensional setting when the parameter dimension p is fixed and smaller than the sample size n, as well as the high dimensional setting when p is greater than and growing with n. In both cases, we assume sparsity, that is, the number of non-zero components of the parameter are assumed to be small relative to the number of observations. Finite sample performance of the proposed method is examined by simulation studies, and compared with the traditional LASSO for independent data. The methods are illustrated with an application to hedonic house price model for the Aveiro-lhavo urban housing market in Portugal.

CO1187: **Continuously updated indirect inference in SAR models with unobserved heterogeneity**
Presenter: **Maria Kyriacou**, University of Southampton, United Kingdom
Co-authors: Peter CB Phillips, Francesca Rossi
Spatial units are often heterogeneous as they vary in many of their observed characteristics such as income and so the assumption homoskedasticity may not hold. In the presence of unobserved heterogeneity the (Q)LE of both the spatial parameter and the exogenous regressors coefficients become, in general, inconsistent. There is an evident lack of estimation methods that account for the presence of hetersokedasticity while allowing for a wider class of heteroskedastic designs and also more realistic weight matrix designs. A Robust Generalized Methods of Moments (RGMM) estimator has been previously proposed which is consistent in heteroskeskadic situations. Also a GMM method robust to heteroskedasticity has been considered. Modifying the QMLE/MLE estimator has been proposed to restore consistency under mild forms of heteroskedasticity. There is yet a method that provides finite sample refinements for moderate sample sizes for general forms of hetersoskedasticy without being restrictive in the design of the exogenously given weights matrix. We propose an indirect inference based method robust to unobserved heterogeneity, the Continuously Updated Indirect Inference (CUII) which is derived using a binding function with continuously updated diagonal variance/covariance matrix. Simulation results reveal that our proposed CUII estimator is effective in reducing both bias and MSE compared to QML/ML and the RGMM estimator.

CO0959: **Testing misspecification in spatial autoregressive models**
Presenter: **Jungyoon Lee**, Royal Holloway, University of London, United Kingdom
Spatial autoregressive (SAR) and related models offer flexible, yet parsimonious, ways to model spatial or network interaction. However, SAR specifications rely on a particular parametric functional form and an exogenous choice of the so-called spatial weight matrix. Moreover, the choice of SAR model over other alternatives, such as Spatial Durbin (SD) or Spatial Lagged X (SLX) models, is often arbitrary and unwarranted. The aim is to develops a specification test in SAR model that can detect general forms of misspecification including that of the spatial weight matrix, functional form and the model itself. We extend the framework of conditional moment testing to a general spatial setting. We derive the asymptotic distribution of our test statistic under the null hypothesis of correct SAR specification and show consistency of the test. We also carry out a small Monte Carlo study of finite sample performance.

CO1021: **Some recent progress on nonlinear spatial modelling: A review**
Presenter: **Zudi Lu**, University of Southampton, United Kingdom
Larger amounts of spatial or spatio-temporal data with more complex structures collected at irregularly spaced sampling locations are prevalent in a wide range of disciplines. With few exceptions, however, practical statistical methods for nonlinear modeling and analysis of such data remain elusive. A review is provided on some developments and progress of our research. In particular, we will look at some nonparametric methods for probability, including joint, density estimation, and semiparametric models for a class of spatio-temporal autoregressive partially nonlinear regression models permitting possibly nonlinear relationships between responses and covariates. In the setting of semiparametric spatio-temporal modelling, we will also show a computationally feasible data-driven method for spatial weight matrix estimation. For illustration, our methodology is applied to investigate some land and housing prices data sets.

CO279 Room 101 EMPIRICAL PROCESSES Chair: Bent Nielsen

CO0421: **Tightness of M-estimators for multiple linear regression in time series**
Presenter: **Bent Nielsen**, University of Oxford, United Kingdom
Tightness of a general M-estimator for multiple linear regression in time series is shown. The positive criterion function for the M-estimator is assumed lower semi-continuous and sufficiently large for large argument. Particular cases are the Huber-skip and quantile regression. Tightness requires an assumption on the frequency of small regressors. We show that this is satisfied for a variety of deterministic and stochastic regressors, including stationary an random walks regressors. The results are obtained using a detailed analysis of the condition on the regressors combined with some recent martingale results.

CO0427: **The bivariate Kakutani division process**
Presenter: **Winfried Stute**, University of Giessen, Germany
The univariate Kakutani division process is a sequential procedure which generates data in the unit interval in such a way, that given data $U_1, \ldots, U_n$, the next U is uniform over the largest interval between two successive (ordered) U's. The empirical process for such data is non-standard. We present the bivariate extension of the Kakutani division process and study the associated empirical process.

CO0444: **Residual empirical processes**
Presenter: **Hira koul**, Michigan State University, United States
Residual empirical processes are known to play a central role in the development of statistical inference in numerous additive models. The aim is to discuss some history and some recent advances in the asymptotic uniform linearity of residual empirical processes and their usefulness in developing ADF GOF tests for error distribution functions in nonparametric ARCH(1) models.

CO0500: **Statistical functionals of residuals**
Presenter: **Vanessa Berenguer Rico**, University of Oxford, United Kingdom
Co-authors: Bent Nielsen
The analysis of residuals is central in statistical modelling. This is often conducted using specification tests that involve statistical functionals of residuals. We provide a general theoretical framework to systematically analyse the asymptotic properties of a wide class of residual-based specification tests. The framework is built using empirical processes techniques and it is general enough to handle standard least squares regressions as well as robust estimation procedures in models that include stationary and non-stationary regressors. The theoretical framework is then applied

to study the asymptotic properties of some well-known specification tests, i.e., normality test based on cumulants, Kolmogorov-Smirnov statistics, or White heteroskedasticity test. These are analysed in least squares and robust regression contexts.

CO0838: **Testing constancy of structural parameters in the direction of random alternatives**
Presenter: **Miguel A Delgado**, Universidad Carlos III de Madrid, Spain
The problem of testing the hypothesis of constancy of structural parameters versus the alternative that they are an unknown function of an observed variable is considered. This includes testing that marginal effects are constant in regression models versus non-parametric interactive effects, and specification testing of panel data models. Test statistics depend on partial sums of concomitants involving residuals, which resemble classical parameter stability tests, where parameters vary with time under the alternative hypothesis. We also consider the problem of testing that a subset of coefficients is constant, which includes testing the specification of partially linear regressions versus random slope coefficients. The test is implemented using bootstrap approximations of the distribution of the statistic under the null hypothesis. We study the finite performance of the test by means of Monte Carlo experiments.

CO705 Room S24 INVESTMENT STRATEGIES Chair: Fotis Papailias

CO0579: **Returns signal momentum**
Presenter: **Jiadong Liu**, Queens University Belfast, United Kingdom
Co-authors: Fotis Papailias, Dimitrios Thomakos
A new type of momentum based on the probability of past returns signs is introduced. Position signals are generated when the equally weighted average of past returns signs exceed a certain threshold. We consider various fixed and time-varying values for this threshold. We form a portfolio which consists of 55 of the world's most liquid commodity and financial futures. Investment strategies using returns signal momentum result in higher returns, Sharpe ratio and lower maximum drawdown when compared to the $1/N$, moving average strategies and the time series momentum. Additional statistical tests illustrate the robustness of the method. Returns signal momentum can, hence, be considered as an effective strategy for speculation and hedging by market participants.

CO1128: **Investment strategies for energy assets**
Presenter: **Thomas Alexopoulos**, University of Peloponnese, Greece, Greece
Co-authors: Dimitrios Thomakos, Rafael Yahlomi
A new investment strategy for Energy trading products is utilized, that can be used for hedging and risk management while accounting for the high volatility of this asset class. The strategy consists of exploiting three stylized empirical facts of asset returns, momentum, mean reversion and bubbles, by taking sequentially different segments of the data that are used in a functional-type of analysis. We illustrate the mechanics of the proposed method with real data either on Energy related ETFs or stocks, but also in Energy futures as their volume appears an increasing trend in many energy exchanges. Our results show that the proposed method can outperform other strategies, like a simple rebalancing strategy or the buy and hold benchmark, as it exhibits better risk return characteristics. More importantly, it appears that it can identify turning points relatively fast and is thus suitable for being used as a hedging and risk management tool in the highly unstable energy markets.

CO1115: **A top-down approach to identifying bull and bear market states**
Presenter: **Alan Hanna**, Queens University Belfast, United Kingdom
Bull and bear markets receive considerable media and academic attention. It is widely believed that such states are important determinants of wider market dynamics, yet no agreed definition exists. Frameworks are investigated for ex-post classification of asset prices in two-state (bull and bear) markets. A number of potential difficulties with existing methodologies are highlighted. A new, flexible, hierarchical methodology is proposed that addresses these issues and permits varying degrees of resolution, allowing secondary trends such as bear rallies to be incorporated. The methodology is shown to be optimal under one measure of performance.

CO1312: **Testing common time-varying coefficients in semiparametric panel data models with fixed effects**
Presenter: **Alev Atak**, City University London, United Kingdom
Co-authors: Yonghui Zhang
A nonparametric Wald-type test is proposed for common time-varying coefficients functions in semiparametric panel data models with fixed effects. The idea is to compare the fitted homogeneous time-varying coefficients functions by a local linear dummy variable regression with the estimated heterogeneous time-varying coefficient functions panel data model by running nonparametric regression for each individuals time series observations. We show that after appropriate standardization the test statistic is asymptotically normally distributed under both the null hypothesis and a sequence of Pitman local alternatives. We prove test consistency and propose a bootstrap procedure to obtain p-values. Monte Carlo simulations indicate that the test performs well in finite samples.

CO0955: **Forecasting and profiting from rare events**
Presenter: **Fotis Papailias**, Queens University Belfast, United Kingdom
A simple, yet intuitive, algorithm for outliers forecasting is presented. Given a suitable time series where outliers can be spotted in past observations, we perform the following operations: (i) detect past outliers using a rolling window scheme, (ii) create the events time series with 1 indicating an outlier and 0 otherwise, (iii) calculate the one-step ahead event probability using a recursive forecasting exercise. Then, based on these forecasts, an investment strategy could be created. An example using real time data illustrates the applicability of the method.

CO329 Room 103 UNCERTAINTY: MEASUREMENT AND INFERENCE Chair: Wojciech Charemza

CO0837: **Understanding the role of uncertainty in the Euro area business cycle**
Presenter: **Cecilia Melo Fernandes**, Goethe-University Frankfurt am Main, Germany
Co-authors: Geoff Kenny
Parametric and non-parametric measures of economic uncertainty extracted from the density forecasts of professional forecasters are proposed as a statistically and economically well-grounded concept of uncertainty that is linked to the ex ante predictability of economic outcomes. It includes a block of three uncertainty indicators consisting of GDP uncertainty, unemployment uncertainty and inflation uncertainty. The properties of different aggregate uncertainty measures are compared also with reference to other proxies for uncertainty commonly used in the literature. We exploit the timing of the survey in order to identify an exogenous uncertainty shock in a Bayesian Vector AutoRegression (BVAR). The relevance of uncertainty shocks during the Great Recession and subsequent sovereign debt crisis in the Euro area is then assessed both in sample and also in terms of out-of-sample predictive content.

CO0616: **On the effect of uncertainty on non-response to the European central bank survey of professional forecasters**
Presenter: **Victor Lopez-Perez**, Universidad Politecnica de Cartagena, Spain
The purpose is to explore how changes in macroeconomic uncertainty have affected the decision to reply to the European Central Bank Survey of Professional Forecasters (ECB SPF). The results suggest that higher (lower) aggregate uncertainty increases (reduces) non-response to the survey.

This effect is statistically and economically significant. Therefore, the assumption that individual ECB SPF data are missing at random may not be appropriate. Moreover, the forecasters that perceive more individual uncertainty seem to have a lower likelihood of replying to the survey. Consequently, measures of uncertainty computed from individual ECB SPF data could be biased downwards.

CO0853: **Quasi ex-ante inflation forecast uncertainty**
Presenter: **Svetlana Makarova**, University College London, United Kingdom
Co-authors: Wojciech Charemza, Carlos Diaz
A new measure of inflation uncertainty is proposed, which is, to an extent, net of the effects of the monetary policy decisions. The measure is computed using the parameters of the weighted skew-normal distribution fitted to inflation forecast errors, which can be interpreted as representing monetary policy outcomes. These parameters are then used for recovering a hypothetical distribution free of such outcomes, called the distribution of the pseudo-ex-ante uncertainty. Under some general assumptions, it can be interpreted analogously to that of the distribution of the ex-ante uncertainty in a survey of professional forecasters. The measure we propose, named the uncertainty ratio, is the ratio of variance of this hypothetical distribution to the mean squared error of the original weighted skew-normal distribution fitted to forecast errors. This measure is computed for annual inflation measured monthly for 38 countries. With the use of this ratio, we provide some evidence that the central banks independence contributes positively towards a reduction in forecast uncertainty. We also evaluate ex-post and pseudo-ex-ante forecast uncertainty term structure for U.K. and U.S. using new indicators and compare with earlier results.

CO0855: **Uncertainty, parameters' heterogeneity and cross-sectional dependence in fiscal reaction functions estimation**
Presenter: **Roberto Golinelli**, University of Bologna, Italy
Co-authors: Irene Mammi
The estimation of fiscal reaction functions parameters, in order to assess the cyclicality of discretionary fiscal policies, has to tackle uncertainty over several dimensions. Uncertainty may stem from different sources including the timing and composition of fiscal measures implemented by policymakers; the difficulties in predicting variables such as potential GDP and output gap in real time, and data revisions; fiscal/volatility shocks and the increase in volatility of macroeconomic indicators. Additional empirical challenges are represented by cross-sectional dependence in fiscal shocks due to common latent factors or spatial spillovers and by the non-constancy of parameters across countries. An empirical estimation of fiscal reaction functions for the Euro area countries between 1996 and 2016 is provided. A dynamic fiscal rule is estimated for the cyclically-adjusted primary balance using dynamic-panel-data estimators, allowing for both parameters heterogeneity and common correlated random effects in the errors. The sensitivity of the estimates to the use of ex-post, real-time or forecast data is investigated and the impact of macroeconomic indicators revisions/forecast errors is evaluated. The role of interest-rate spreads is also analysed as determinants of both discretionary actions by policymakers and of fiscal policy volatility, measured from the residuals of estimated fiscal rules on real-time data.

CO0755: **A measure of ex-ante inflation uncertainty based on density forecast revisions**
Presenter: **Carlos Diaz**, University of Leicester, United Kingdom
A new measure of ex-ante inflation uncertainty is derived from UK inflation density forecasts. Based on the revision of these forecasts the density of the shocks of information perceived by the Bank between two consecutive releases is obtained. The variance of these shocks can be interpreted as a measure of the uncertainty that the Bank expects to affect inflation in the following quarter due to information obtained in the previous one. The results show that inflation uncertainty was decreasing during the period leading to the financial crisis, increasing sharply during the period 2007-2011, and remaining stable after that but at levels higher than its pre-crisis values. A similar exercise is done with the GDP growth density forecasts and the uncertainty spillovers between both variables are investigated.

CO419 Room Graduation hall RECENT ADVANCES IN LARGE PANEL DATA ANALYSIS Chair: Herman van Dijk

CO0929: **Local-to-unity asymptotics for the factor analytical approach**
Presenter: **Milda Norkute**, Lund University, Sweden
Co-authors: Joakim Westerlund
A recent study proposes a new factor analytical (FA) method for the estimation of stationary dynamic panel data models with fixed effects. Our interest in this method originates with the fact it does not require explicit detrending of the data, a practice that is known to cause problems of bias and low power in (near) unit root panels, especially in models that include a linear trend. Our purpose is to study the properties of FA when applied to such (near) unit root panels with a possible trend.

CO0759: **Point optimal panel unit root testing with covariates**
Presenter: **Arturas Juodis**, University of Groningen, Netherlands
Co-authors: Joakim Westerlund
Asymptotic optimality results are provided for panel unit root tests with covariates by deriving the Gaussian power envelope. The results extend previous ones which consider the special case when no information on covariates is available. The main conclusion is that the use of covariates holds considerable promise in the panel data context, much more so than in the time series context. The biggest difference occurs in the empirically relevant case with incidental trends. Specifically, while the power envelope without covariates is defined within $N^{-1/4}T^{-1}$-neighborhoods of unity, the envelope with covariates is in general defined within $N^{-1/2}T^{-1}$-neighborhoods. Therefore, the use of the covariates not only leads to increased power, but it can actually have an order effect on the shrinking neighborhoods around unity for which power is non-negligible. As a way to explore this increase in power in practice, a feasible point optimal unit root test is proposed, whose small-sample properties are investigated by means of Monte Carlo simulation.

CO1696: **Cross-section average-based confidence intervals for diffusion index forecasts**
Presenter: **Hande Karabiyik**, VU University Amsterdam, Netherlands
Co-authors: Jean-Pierre Urbain, Stephan Smeekes, Joakim Westerlund
A situation is considered where is a single time series that can be forecasted by using a set of observed regressors and a set of latent factors. We assume there exists a set of panel data variables with a large number of cross-section (N) and time series observations (T) that are affected by the same latent factors. We propose to estimate the factors by using the cross-sectional averages of the panel data variables and then use these estimates in the forecasting equation to obtain the forecasts by using the least squares estimation method. We show that when the number of panel data variables (m) is equal to the number of unobserved factors (r), predicted conditional mean of the variable to be forecasted is consistent and asymptotically normal as $\sqrt{(T)}/N$ goes to zero. We provide an estimation method for the variance as well that lead to a successful way of constructing confidence intervals. Additionally we show how this result fails to hold when (m) is not equal to (r).

CO1317: **A Hausman test for cross-section dependence in linear panel regression models**
Presenter: **Simon Reese**, Lund University, Sweden
Testing for the presence of cross-section dependence in panel data is becoming an ever more relevant model specification issue given the increasing use of interactive fixed effects models. We contribute to this strand of the literature by proposing a Hausman test for cross-section dependence. By

focusing on differences in the parameter vectors, our focus is thus different from that of dominating existent tests which are based on correlations in the model residuals. The distribution of the interactive fixed effects estimator under the null hypothesis is derived using a normalisation techniques for the factor estimates, allowing to circumvent the problem of asymptotic singularity. Simulations reveal good performance results relative to other tests.

CO591 Room S23 ADVANCES IN OPTION PRICING AND HEDGING USING STOCHASTIC VOLATILITY MODELS Chair: Juan-Pablo Ortega

CO1036: **Non-affine GARCH option pricing models, variance dependent kernels, and diffusion limits**
Presenter: **Alex Badescu**, University of Calgary, Canada
Co-authors: Zhenyu Cui, Juan-Pablo Ortega
The pricing and weak convergence of an asymmetric non-affine, non-Gaussian GARCH model is investigated when the risk-neutralization is based on a variance dependent exponential linear pricing kernel with stochastic risk aversion parameters. The risk-neutral dynamics are obtained for a general setting and its weak limit is derived. We show how several GARCH diffusions, martingalized via well-known pricing kernels, are obtained as special cases and we derive necessary and sufficient conditions for the presence of financial bubbles. An extensive empirical analysis using both historical returns and options data illustrates the advantage of coupling this pricing kernel with non-Gaussian innovations.

CO1235: **Option valuation with IG-GARCH model and an U-shaped pricing kernel**
Presenter: **Christophe Chorro**, University, France
Co-authors: Fanirisoa Rahantamialisoa
Empirical and theoretical studies have attempted to establish the U-shape of the log-ratio of conditional risk-neutral and physical probability density functions. The main subject is to question the use of such a U-shaped pricing kernel to improve option pricing performances. Starting from the so-called Inverse Gaussian GARCH model (IG-GARCH), known to provide semi-closed form formulas for classical European derivatives when an exponential affine pricing kernel is used, we build a new pricing kernel that is non-monotonic and that still have this remarkable property. Using a daily dataset of call options written on the S&P500 index, we compare the pricing performances of these two IG-GARCH models proving, in this framework, that the new exponential U-shaped stochastic discount factor clearly outperforms the classical exponential affine one. What is more, several estimation strategies including options or VIX information are tested taking advantage of the analytical tractability of these models.

CO0569: **GARCH option pricing models with Meixner innovations**
Presenter: **Alexander Melnikov**, University of Sankt Gallen, Switzerland
Co-authors: Matthias Fengler
GARCH option pricing models are presented with Meixner distributed innovations. The risk-neutral dynamics are derived by means of the conditional Esscher transform. The models are estimated from time series data and price options by simulating the transition density of stock prices and the Radon-Nikodym process under the historical measure jointly. Assessing the option pricing performance both in-sample and out-of-sample, one finds that the model compares favorably against the benchmark models. Simulations suggest that the driver of these results is the impact of conditional skewness and conditional excess kurtosis on option prices.

CO0324: **On different pricing approaches for options under GARCH with non-normal innovations**
Presenter: **Lars Stentoft**, University of Western Ontario, Canada
Two different pricing frameworks are typically used in the literature when pricing options under GARCH with non-normal innovations: the equilibrium approach and the no-arbitrage approach. Each framework can accommodate various forms of GARCH and innovation distributions, but empirical implementation and tests are typically done in one framework or the other because of the computational challenges that are involved in obtaining the relevant pricing parameters. We contribute to the literature by comparing and documenting the empirical performance of a GARCH specification which can be readily implemented in both pricing frameworks. The model uses a parsimonious GARCH specification with skewed and leptokurtic Johnson s_u innovations together with either the equilibrium based framework or the no-arbitrage based framework. Using a large sample of options on the S&P 500 index, we find that the two approaches give rise to very similar pricing errors when implemented with time-varying pricing parameters. However, when implemented with constant pricing parameters, the performance of the no-arbitrage approach deteriorates in periods of high volatility relative to the equilibrium approach whose performance remains stable and at par with the models with time-varying pricing parameters.

CO1155: **Pricing and hedging of non-affine ARSV options using volatility dependent kernels**
Presenter: **Lyudmila Grigoryeva**, University of Konstanz, Germany
Co-authors: Alex Badescu, Juan-Pablo Ortega
New pricing and hedging strategies are proposed for a non-affine auto-regressive stochastic volatility (ARSV) models with non-predictable drift which allows to account for leverage effects. We consider a volatility dependent exponential linear pricing kernel with stochastic risk aversion parameters and implement both pricing and hedging for ARSV models estimated via the hierarchical-likelihood method. This technique proves to outperform standard GARCH and Heston-Nandi based strategies in terms of a variety of considered criteria in an empirical exercise using historical returns and options data.

CO363 Room 102 RECENT DEVELOPMENTS IN COINTEGRATION ANALYSIS Chair: Martin Wagner

CO1043: **Residual based consistent bubble detection**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Martin Wagner
Consistent testing and monitoring procedures are presented for detecting bubbles, defined as episodes of mildly explosive behavior, based on residuals from estimated equilibrium relationships. Under the null these relationships correspond to cointegrating relationships, whereas the errors are mildly explosive under the bubble alternative. The goal of the procedures is to detect transitions in both directions, based on backward sup ADF statistics. This monitoring procedure is used to investigate arbitrage parities, such as the Covered Interest Rate Parity and the Triangular Arbitrage Parity.

CO0856: **Linear fully modified OLS estimation of cointegrating polynomial regressions**
Presenter: **Oliver Stypka**, TU Dortmund, Germany
Co-authors: Martin Wagner
In the empirical environmental Kuznets curve (EKC) literature, which investigates a potentially inverted U-shaped relationship between measures of economic development and pollution respectively emissions, it is common practice to use cointegration methods developed for the linear setting in nonlinear contexts. This means that integrated processes and its powers are both considered to be $I(1)$ and estimators designed for asymptotically valid inference in standard cointegration settings like the FM-OLS estimator are commonly applied as if there were multiple integrated regressors included in the regression. We show that this "linear" estimator has, surprisingly, the same limiting distribution as the FM-CPR estimator developed

specifically for cointegrating polynomial regressions, which takes into account that nonlinear functions of integrated processes are not I(1). We also show, by means of simulations, that inference based on the FM-CPR estimator outperforms inference based on the "linear" FM-OLS estimator in terms of both lower size distortions and higher (size-corrected) power.

CO1385: **Some results on the structure theory of cointegrated state space systems**
Presenter: **Massimo Franchi**, University of Rome La Sapienza, Italy
Minimality of the state space representation of a stochastic process places restrictions on the rank of certain matrices that show up in the leading coecient of the principal part of its implied transfer functions; when unit roots are allowed for, those restrictions and the reduced rank structure of the state matrix polynomial at one shape the integration and cointegration properties of the state and the observed processes. A characterization of cointegration and polynomial cointegration in integrated state space systems is presented in the $I(d)$ case for generic integer d and the leading $I(1)$ and $I(2)$ cases are discussed in detail. Extensions to multiple frequency unit roots and connections with the results in the literature are also covered.

CO0779: **Cointegrating multivariate polynomial regressions: Fully modified OLS estimation and inference**
Presenter: **Martin Wagner**, Technical University Dortmund, Germany
Co-authors: Oliver Stypka
A fully modified OLS (FM-OLS) estimator is developed for cointegrating multivariate polynomial regressions, i.e. regressions that include as explanatory variables deterministic variables, integrated processes and products of non-negative integer powers of the integrated processes. The stationary errors are allowed to be serially correlated and the regressors are allowed to be endogenous. We extend existing variants of the FM-OLS estimator to cointegrating multivariate polynomial regressions, which overcomes the additive separability restriction typically used in nonlinear cointegration analysis for the polynomial case. The FM-OLS estimator has a zero-mean Gaussian mixture limiting distribution that allows for standard asymptotic inference. In addition to hypothesis testing on the parameters also Wald and LM specification tests are derived, as well as a KPSS-type test for cointegration. The theoretical analysis is complemented by a simulation study. Since the developed estimator immediately leads to a RESET-type specification test, we also compare in the simulation section the performance of our FM-OLS RESET with other (partly more restrictive) cointegration specification tests.

CO0531: **A topological view on the identification of structural vector autoregressions**
Presenter: **Klaus Neusser**, University of Bern, Switzerland
The notion of the group of orthogonal matrices acting on the set of all feasible identification schemes is used to characterize the identification problem arising in structural vector autoregressions. This approach presents several conceptual advantages. First, it provides a justification for the use of the normalized Haar measure as the natural uninformative prior. This specification of priors corresponds to the invariance principle. Second, using the multivariate beta distribution we derive the joint distribution of blocks of parameters defining an identification scheme. Finally, it allows the coherent introduction of perturbed identification schemes which is of relevance for the specification of time-varying vector autoregressions.

CO373 Room 112 COMPUTATIONAL APPROACHES IN ECONOMETRICS Chair: Sandra Paterlini

CO1065: **Default contagion & systemic risk in the intercorporate loan guarantee network**
Presenter: **Wenwei Li**, EBS Universitaet fuer Wirtschaft und Recht, Germany
Co-authors: Shenglin Ben, Ulrich Hommel, Sandra Paterlini, Jiefang Yu
A one-mode network is proposed to model the intercorporate loan guarantee activities in China and test it with the field research data from Zhejiang Province. By stress testing the network, we also reveal the shock transmission mechanism and loss contagion process within the loan guarantee market. The empirical results show that the topology of the presented network is close to a scale-free structure, which is robust against accidental failures but vulnerable to coordinated attacks. The simulation analysis further highlights some important findings. First, the rising level of corporate debt (including private lending) will weaken sample companies capabilities to absorb losses from the contingent guarantee liabilities, in turn leading to a larger scale of domino effect. Second, the continuous credit support from financial institutions during crisis will significantly increase the survival rate of the sample companies, and reduce the losses from default contagion.

CC0477: **Statistical and computational tradeoff in econometric models building by genetic algorithms**
Presenter: **Manuel Rizzo**, Sapienza University of Rome, Italy
Co-authors: Francesco Battaglia
When a Genetic Algorithm (GA), or in general a stochastic algorithm, is employed in a statistical problem, the result is affected by both the variability due to sampling error, due to the fact that only a sample is observed, and the variability due to the stochastic nature of the algorithm. Such issues can be analyzed by understanding the trade-off between statistical accuracy and computational efforts. We focus on statistical estimation problems for which the variability of the GA estimates can be decomposed in the two sources of variability by means of cost functions, related to both data acquisition and runtime of the algorithm. Simulation studies will be presented to discuss the statistical and computational tradeoff question.

CO0960: **Sorting out your investments: Sparse portfolio vonstruction via the ordered l_1 norm**
Presenter: **Philipp Johannes Kremer**, EBS Universitaet fuer Wirtschaft und Recht, Germany
Co-authors: Sandra Paterlini, Malgorzata Bogdan
Since its introduction to the statistics literature, the desiring features of simultaneous model selection and estimation have gained Lasso a wide recognition in statistics and also recently in financial portfolio optimization. Still, the Lasso has well-known shortcomings when applied to the setting of highly dependent financial data. We move away from the standard framework of orthogonal design and apply the recently developed sorted l_1 penalized estimation, called SLOPE, to the framework of correlated data. SLOPE relies on the idea of penalizing coefficients with a stronger signal more heavily and clumping equally correlated assets together. In fact, in a simulated factor model, SLOPE is able to identify and to cluster assets with the same underlying risk factor exposures into one group. This enables the investor to improve his ex-ante portfolio risk management. Furthermore, our empirical analysis on the SP100 and SP500 from 2004-2016 confirms the validity of SLOPE in developing effective investment strategies.

CO0752: **Effects of diversification and capital buffers on the EU sovereign network**
Presenter: **Margherita Giuzio**, EBS Universitaet fur Wirtschaft und Recht, Germany
Co-authors: Sandra Paterlini
After the European sovereign debt crisis, it has become clear that government bonds are not risk-free investments and are heterogeneous in their risk profiles. Therefore, recent policy discussions focused on the introduction of stricter capital and diversification requirements for EU government bonds holdings. We investigate the effect of these requirements on the European banking system, relying on the EBA stress-test dataset and on the Securities Holdings Statistics including sovereign exposures of the entire German banking system. Having access to such data, we point out some biases of the EBA sample that may undermine its representativeness. Furthermore, we analyse the sovereign portfolios of both EBA

and German banks, by calculating their risk and diversification at the level of single monetary financial institutions and of the entire system, and discuss the implications of home-bias, i.e very concentrated positions of banks toward their home country. By requiring higher portfolio diversification, regulators may force connections and dependence between banks through common asset holdings, which represent an important channel of contagion in presence of financial distress. By using the German banking portfolios at an ISIN-level, we can then estimate the spillover risk and calculate the response of German banks to potential fire sales.

CC0170: **Modelling multidimensional extremal dependance for operational risk**
Presenter: **Sandra Paterlini**, European Business School Germany, Germany
Co-authors: Claudia Klueppelberg, Oliver Key
A statistical model of operational losses is introduced based on extreme value distributions and bipartite graphs, which perfectly capture the event type and business line structure of operational risk data. The model explicitly takes into account the Pareto tails of losses as well as heterogeneous dependence structures between them. We then derive estimators for individual as well as aggregated tail risk, measured in terms of Value-at-Risk and Conditional Tail Expectations for very high confidence levels. Asymptotic properties as well as upper and lower bounds are discussed. We introduce then two estimation methods for such risk-measures and test their validity on simulated data. Finally, by having access to real-world operational risk losses from the Italian banking system, we show that the estimated quantities converge to the asymptotic values and show that quantifying dependence by means of the empirical severity and frequency distribution can have quite an impact on estimates at both individual and aggregate level.

CO365 Room 109 MACROECONOMETRICS Chair: Toshiaki Watanabe

CO1077: **Large time varying parameter VARs for macroeconomic forecasting**
Presenter: **Gianni Amisano**, Federal Reserve Board, United States
Co-authors: Domenico Giannone, Michele Lenza
Bayesian time varying parameter models have been widely used for forecasting purposes and as exploratory devices. In spite of their obvious overparameterization, these models are capable of reproducing salient features of the data. The main challenge is to use them with many endogenous variables. The problem is that VAR models are over-parameterized and each coefficient is endowed with its own drift term. Therefore the state equation shock covariance matrix becomes huge. In addition, in order to contain the effects of random walk dynamics on coefficients it is important to specify a prior that greatly limits the amount of time variation and to provide a sensible initialization. We propose to use an informative Bayesian VAR on the pre-sample to obtain a sensible initialization and calibrate the prior in a conservative way. In addition, we use the VAR Kronecker structure to reduce the number of free parameters in the state equation. We also augment the model with a suitable stochastic volatility specification. We conduct two experiments, the first one in VARs with 7 and 20 variables. We show that our specification works very well in prediction, both in terms of point and density forecasts, and yields substantial gains especially in periods of recessions and economic turbulence.

CO0850: **Macroeconomic forecasting in times of crises**
Presenter: **Pablo Guerron**, Boston College, United States
A parsimonious semiparametric method is proposed for macroeconomic forecasting during episodes of sudden changes. Based on the notion of clustering and similarity, we divide the series into blocks, search for the closest blocks to the most recent block of observations, and with the matched blocks we proceed to forecast. One possibility is to compare local means across blocks, which captures the idea of matching directional movements of a series. We show that our approach does particularly well during the Great Recession and for variables such as inflation, unemployment, and real personal income. When supplemented with information from housing prices, our method consistently outperforms parametric alternatives for the period 1990 - 2015.

CO1032: **Trend inflation and exchange rate dynamics: A new Keynesian approach**
Presenter: **Takashi Kano**, Hitotsubashi University, Japan
Exchange rate implications of trend inflation are studied within a two-country new Keynesian model under incomplete international financial markets. A new Keynesian Phillips curve generalized by persistent variations in trend inflation with a positive long-run mean implies a second-order expectational difference equation of inflation rate in each country. A smoother and persistent process of the cross-country inflation differential as well as a slower moving dynamics of within-sector price dispersions leads to (i) a persistent real exchange rate with an autoregressive root close to one, (ii) a volatile real exchange rate relative to cross-country inflation differential, and (iii) an almost one-to-one comovement between real and nominal exchange rates within an otherwise standard two-country new Keynesian model. Trend inflation, therefore, approaches the persistence and volatility puzzles of exchange rates dynamics in actual data.

CO0966: **Money market interest rate uncertainty and macroeconomic implications**
Presenter: **Michele Lenza**, European Central Bank, Germany
Co-authors: Carlo Altavilla, Giacomo Carboni
The aim is to evaluate the macroeconomic impact of uncertainty about future interest rates in the Euro area interbank market for unsecured loans. The results suggest that the volatility of money market rates proxied by either realised or implied volatility of future interest rates contains information on the macroeconomic environment over and above standard measures of economic uncertainty. The effects of this uncertainty shock on the economy resemble those associated to a demand shock. Remarkably, these macroeconomic effects remain robust after controlling for a number of uncertainty indicators, stock market volatility, financial stress measures, and volatility of macroeconomic news.

CO1574: **Cross-country evidence on the effectiveness of monetary policy: A time-varying parameter VAR approach**
Presenter: **Jouchi Nakajima**, Bank for International Settlements, Switzerland
Co-authors: Andrew Filardo
Base on a time-varying parameter VAR approach, the aim is to estimate the time-varying effectiveness of monetary policy, identifying unconventional monetary policy shocks from market surprises and highlighting a lending channel. US, the euro area and Japan experiences are analysed to highlight the similarities and differences in the experiences of quantitative easing.

CO721 Room Board meeting room I REGIME CHANGE MODELING IN ECONOMICS AND FINANCE II Chair: Willi Semmler

CO1478: **A three-phase model of climate change mitigation**
Presenter: **Sergey Orlov**, Inernational Institute for Applied System Analysis, Austria and Lomonosov Moscow State University, Russia
Co-authors: Elena Rovenskaya, Julia Puaschunder, Willi Semmler
A three-phase optimal control model is considered for economic growth and climate change mitigation through transiting to low-carbon technologies. In the first phase, companies invest in adopting low-carbon technologies and a part of the required investment is reimbursed through 'climate' bonds issued by the government. The second phase starts when the greenhouse gases' concentration is reduced to a pre-industrial level and firms have fully adopted low-carbon technologies. In this phase, bonds are being repaid through taxation. In the last phase, the economy is accumulating capital and consume enjoying the stabilized climate. The standard consumption based utility function in the logarithmic form is applied over all three phases. We use GPOPS-II software to find a locally optimal solution in the model, including the optimal switching times between phases. We compare the obtained optimal solution with a business-as-usual version of this model, in which no mitigation action is undertaken. The social welfare function in the mitigation policy model turns out to be greater than the one in the business-as-usual model over almost the entire time horizon of consideration with the exception of a short period in the beginning of the first phase.

CO1704: **Application of receding horizon optimal control to DICE integrated assessment models**
Presenter: **Timm Faulwasser**, Karlsruhe Institute of Technology, Germany
Co-authors: Lars Gruene, Christopher M Kellett, Steven Weller
To quantify the damages from anthropogenic emissions of heat-trapping greenhouse gases, specifically carbon dioxide (CO2), integrated assessment models are used to describe the dynamics of climate-economy interactions. We consider the computation of the Social Cost of Carbon (SCC) via the Dynamic Integrated model of Climate and the Economy (DICE). Typically, any SCC computation is based on the solution of a single (long-horizon) optimal control problem. We show that receding-horizon strategies can also be used to compute the SCC. In receding-horizon optimal control, also known as model predictive control in the field of systems and control, one solves a sequence of short-horizon optimal control problems, of which only the first part of the optimal solution is used, instead of tackling the long-horizon optimal control problem directly. We demonstrate that tools developed in a systems and control context can be used to analyze the receding-horizon approximation of the SCC. Furthermore, we show that the receding-horizon strategy facilitates feedback, introducing an element of robustness. Additionally, we comment on different strategies of computing the SCC and a recently published open-source MATLAB implementation of DICE.

CO0219: **The nonlinear nature of country risk**
Presenter: **Jacek Kotlowski**, Narodowy Bank Polski, Poland
Co-authors: Michal Brzoza-Brzezina
Country risk premia can substantially affect macroeconomic dynamics. We concentrate on one of their most important determinants - a country's net foreign asset position and - in contrast to the existing research - investigate its nonlinear link to risk premia. The importance of this particular nonlinearity is twofold. First, it bears a close relationship to debt crises. Second, such a nonlinear relationship is a standard ingredient of DSGE models, but its proper calibration/estimation is missing. Our estimation shows that indeed the link is highly nonlinear and helps to identify the NFA position where serious and posibly dangerous nonlinearities kick in at -70% to -80% of GDP. We also provide a proper calibration of the risk premium - foreign debt relationship applicable to DSGE models and demonstrate that the steady state NFA position matters for economic dynamics in such a model.

CO1647: **On the way out of a recession**
Presenter: **Stephen Nicar**, Franklin and Marshall College, United States
The recession that occurred in the United States from 2007 through 2009, along with the sluggish recovery since, has generated a spirited debate among policy makers and academics on the desirability and effects of fiscal policy. A central question that has emerged from this debate is whether the size of the multiplier on fiscal policy varies across the business cycle. Recent theoretical and empirical work provides evidence that it does. Almost all of the theoretical work has focused on how the zero bound on nominal interest rates may result in non-linear effects of fiscal policy, while much of the empirical work supports the idea that non-linearities may arise under less restrictive circumstances. One drawback of many of these empirical studies is that the results are not motivated by an alternative theory of the mechanism through which fiscal policy effects vary across the cycle. A notable exception within the theoretical literature is a model previously developed by Michaillat that provides a labor-market mechanism for different government spending multipliers across the business cycle. While Michaillat's model is consistent with the existing empirical evidence, the mechanism it proposes has not been tested explicitly. We propose fill this gap in the literature, evaluating the model with respect to the U.S. using a Bayesian mixed-frequency VAR.

CO1664: **Stochastic games with endogenous transitions**
Presenter: **Reinoud Joosten**, Twente University, Netherlands
A stochastic game is introduced in which transition probabilites between states depend on the history of the play. We analyze this game under the limiting average reward criterion. We determine the set of jointly-convergent pure-strategy rewards which can be supported by equilibria involving threats. For expository purposes we analyze a stylized fishery game. Each period, two agents choose between catching with restraint or without. The resource is in either of two states, High or Low. Restraint is harmless to the fish, but it is a dominated action at each stage. The less restraint shown during the play, the higher the probablities that the system moves to Low. The latter may even become 'absorbing temporarily', i.e. transition probabilities to High become zero temporarily.

CG282 Room 108 CONTRIBUTIONS IN BOOTSTRAP IN ECONOMETRICS Chair: Peter Boswijk

CC0491: **Bootstrap inference for impulse response functions in factor-augmented vector autoregressions**
Presenter: **Yohei Yamamoto**, Hitotsubashi University, Japan
Residual-based bootstrap methods are considered to construct the confidence interval for structural impulse response functions in factor-augmented vector autoregressions. In particular, we compare the bootstrap with factor estimation (Procedure A) with the bootstrap without factor estimation (Procedure B). In theory, both procedures are asymptotically valid under a condition T/N0, where N and T are the cross-sectional dimension and the time dimension, respectively. Even when T/N0 is irrelevant, Procedure A still accounts for the effect of the factor estimation errors on the impulse response function estimate and it achieves good coverage rates in most cases. On the contrary, Procedure B is invalid in such cases and tends to undercover if N is much smaller than T. However, Procedure B is implemented more straightforwardly from the standard structural VARs and the length of the confidence interval is shorter than that of Procedure A in finite samples. Given that Procedure B still gives a satisfactory coverage rate unless N is very small, it remains in consideration of empirical use, although using Procedure A is safer as it correctly accounts for the effect of the factor estimation errors.

CC1591: **Testing for heteroscedasticity in jumpy and noisy high-frequency data: A resampling approach**
Presenter: **Ulrich Hounyo**, Aarhus University, Denmark
Co-authors: Kim Christensen, Mark Podolskij
A new way is proposed to measure and test the presence of time-varying volatility in a discretely sampled jump-diffusion process that is contaminated by microstructure noise. We use the concept of pre-averaged truncated bipower variation to construct our t-statistic, which diverges in the presence of a heteroscedastic volatility term (and has a standard normal distribution otherwise). The test is inspected in a general Monte Carlo simulation setting, where we note that in finite samples the asymptotic theory is severely distorted by infinite-activity price jumps. To improve inference, we suggest a bootstrap approach to test the null of homoscedasticity. We prove the first-order validity of this procedure, while in simulations the bootstrap leads to almost correctly sized tests. As an illustration, we apply the bootstrapped version of our t-statistic to a large cross-section of equity high-frequency data. We document the importance of jump-robustness, when measuring heteroscedasticity in practice. We also find that a large fraction of variation in intraday volatility is accounted for by seasonality. This suggests that, once we control for jumps and deflate asset returns by a non-parametric estimate of the conventional U-shaped diurnality profile, the variance of the rescaled return series is often close to constant within the day.

CC1615: **Stationary bootstrapping for realized covolatility with high frequency noisy and asynchronous observations.**
Presenter: **Eunju Hwang**, Gachon University, Korea, South
Co-authors: Dong Wan Shin
Under two important modern financial market features such as noise and non-synchronicity, the average realized volatility matrix is adopted for consistent estimators of the integrated covariations, and their asymptotic normal theories are established. Furthermore, stationary bootstrapping is applied to propose the bootstrap realized volatility matrix and its validity for the high frequency heterogeneous returns is proved by showing that there exist parameters of the stationary bootstrap blocks so that the bootstrap consistencies hold. Confidence intervals and hypothesis tests for the integrated covariance, regression coefficient and correlation coefficient are constructed. The proposed bootstrap methods extend the existing i.i.d. bootstrapping methods for realized covariations that are confined to synchronous noise-free sampling. A Monte Carlo experiment shows good finite sample performances of the proposed bootstrap methods for high frequency noisy asynchronous samples.

CC1477: **Estimated Wold representation and spectral density driven bootstrap for time series**
Presenter: **Jonas Krampe**, TU Braunschweig, Germany
Co-authors: Jens-Peter Kreiss, Efstathios Paparoditis
For purely nondeterministic stationary processes, the spectral density is factorized to get coefficients of a moving average representation of the process which, appropriately normalized, are identical to those of the Wold representation. This relation together with a spectral density estimator is used to obtain estimators of these coefficients. A moving average bootstrap for time series is then developed which uses the entire sequence of estimated moving average coefficients together with appropriately generated pseudo-innovations to obtain new pseudo-time series. It is shown that if the underlying process is linear and the pseudo-innovations are generated by means of an i.i.d. wild bootstrap which mimics, to the necessary extend, the moment structure of the true innovations, then this bootstrap asymptotically works for a wide range of statistics. The relations of the new bootstrap procedure to the linear process bootstrap and to the autoregressive sieve bootstrap are discussed, with the latter being a special case of the moving average bootstrap, when an autoregressive spectral density estimator is used. Simulations investigate the performance of the new bootstrap procedure in finite samples and a real-life data example is presented.

CC0178: **Wild bootstrap seasonal unit root tests for time series with periodic non-stationary volatility**
Presenter: **Anton Skrobotov**, Russian Presidential Academy of National Economy and Public Administration, Russia
Co-authors: Giuseppe Cavaliere, Robert Taylor
The behaviour of the well-known HEGY regression-based seasonal unit root tests is investigated in cases where the driving shocks are allowed to display periodic non-stationary volatility and conditional heteroskedasticity. Our set up allows for periodic heteroskedasticity, non-stationary volatility and (seasonal) GARCH as special cases. We show that the limiting null distributions of the HEGY tests depend, in general, on nuisance parameters which derive from the underlying volatility process. Monte Carlo simulations show that the standard HEGY tests can be substantially over-sized in the presence of such effects. As a consequence, we propose bootstrap implementations of the HEGY tests, based around a seasonal block wild bootstrap principle. This is shown to deliver asymptotically pivotal inference under our general conditions on the shocks. Simulation evidence is presented which suggests that our proposed bootstrap tests perform well in practice, largely correcting the size problems seen with the standard HEGY tests even under extreme patterns of heteroskedasticity, yet not losing finite sample relative to the standard HEGY tests.

CG324 Room Board meeting room II CONTRIBUTIONS IN TIME-SERIES ECONOMETRICS Chair: Stanislav Anatolyev

CC0452: **Fractal methods for fractional cointegration**
Presenter: **Ladislav Kristoufek**, Institute of Information Theory and Automation, Czech Academy of Sciences, Czech Republic
Detrended fluctuation analysis (DFA) and detrending moving average (DMA) methods are standardly used for fractional differencing parameter d estimation. Recently, the DFA and DMA based estimators of standard regression parameters have been proposed. The estimators possess some desirable properties with regards to long-range dependence, trends, seasonalities and heavy tails. We study properties of both estimators beyond the general fractional cointegration framework, i.e. we examine a simple model $y_t = \alpha + \beta x_t + u_t$, where $x_t \sim I(d)$ and $u_t \sim I(d-b)$ which implies $y_t \sim I(\max[d, d-b])$. The fractional cointegration requires $b > 0$ while the standard cointegration $CI(1,1)$ assumes $x_t, y_t \sim I(1)$ and $u_t \sim I(0)$. We are interested in various combinations of d and b parameters ($0 \leq d, b \leq 1$, i.e. we cover not only the fractional cointegration framework). We provide a broad Monte Carlo simulation study focusing on different time series lengths, combination of d and b parameters, and on possible spurious relationships. Specifically, we compare the estimators based on DFA and DMA with the standard OLS procedure under true and spurious relationships ($\beta = 0$ and $\beta \neq 0$). Based on the bias, standard error and mean squared error of the estimators, the new procedures outperform OLS for various settings (e.g. with $d = 1$ and $b < 0.5$).

CC0629: **Asymptotic behavior of temporal aggregation in mixed-frequency datasets**
Presenter: **Cleiton Guollo Taufemback**, Universidad Carlos III de Madrid, Spain
It is well known that discarding intermediate data in a time series mixed-frequency dataset, also called temporal aggregation, imply a loss of information. We demonstrate that temporal aggregation results in inconsistent estimates by showing how aliasing affects the linear regression model. We also propose a new method to circumvent this inconsistency. We analyze stationary and non-stationary linear semiparametric/nonparametric models. Monte Carlo simulations illustrate that the proposed method has good finite sample properties. Finally, an empirical application to quarterly GDP and monthly US indicators data highlights the applicability of our approach.

CC0981: **Specifying autoregressive processes: A horse race of frequentist model selection methods**
Presenter: **Niels Aka**, DIW Berlin, FU Berlin, Germany
Co-authors: Rolf Tschernig
In a Monte Carlo simulation study, autoregressive processes are selected using three model selection methods: (i) standard information criteria,

(ii) the model confidence set (MCS), (iii) jackknife model averaging (JMA). The autoregressive processes that create synthetic data have zero restrictions for some lags, thus instead of just testing for the maximum lag order, a full specification search is performed. Using standard criteria in this case poses a multiple testing problem and could therefore impair subsequent analyses. To assess the severity of this problem and the ability of more elaborate methods to address it, implications of each method's model choice are judged against the true model. In particular, multi-period forecasts and impulse response functions serve as testing grounds. The results indicate that in small samples JMA consistently produces better forecasts than standard criteria, while the MCS approach improves forecasts only at longer forecast horizons. For larger samples and fixed DGPs, model uncertainty becomes negligible and relying on standard criteria is a dominant strategy. Further extensions are in progress, namely the addition of penalized methods and the use of vector autoregressive moving average processes to create synthetic data.

CC1494: **Volatility filtering in estimation of kurtosis (and variance)**
Presenter: **Stanislav Anatolyev**, CERGE-EI and New Economic School, Czech Republic
The kurtosis of the distribution of financial returns characterized by high volatility persistence and thick tails is notoriously difficult to estimate precisely. We propose a simple but effective procedure of estimating the kurtosis coefficient (and variance) based on volatility filtering that uses a simple GARCH model. In addition to an estimate, the proposed algorithm issues a signal of whether the kurtosis (or variance) is finite or infinite. We also show how to construct confidence intervals around the proposed estimates. Simulations indicate that the proposed estimates are much less median biased than the usual method-of-moments estimates, their confidence intervals having much more precise coverage probabilities. The procedure also works well when the underlying volatility process is not the one the filtering technique is based on. We illustrate how the algorithm works using several actual series of returns.

CC1690: **A data-cleaning augmented Kalman filter for robust estimation of state space models**
Presenter: **Martyna Marczak**, University of Hohenheim, Germany
Co-authors: Tommaso Proietti, Stefano Grassi
The aim is to present a robust augmented Kalman filter that extends the data cleaning filter to the general state space model featuring nonstationary and regression effects. The robust filter shrinks the observations towards their onestepahead prediction based on the past, by bounding the effect of the information carried by a new observation according to an influence function. When maximum likelihood estimation is carried out on the replacement data, an Mtype estimator is obtained. We investigate the performance of the robust AKF in two applications using as a modeling framework the basic structural time series model, a popular unobserved components model in the analysis of seasonal time series. First, a Monte Carlo experiment is conducted in order to evaluate the comparative accuracy of the proposed method for estimating the variance parameters. Second, the method is applied in a forecasting context to a large set of European trade statistics series.

EO217 Room 006 GRAPHICAL TOOLS FOR STATISTICAL DATA ANALYSIS Chair: Lola Martinez-Miranda

EO0159: **SiZer for untraditional data**
Presenter: **Cheolwoo Park**, University of Georgia, United States
Co-authors: Yongho Jeon, Kee-Hoon Kang
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.

EO0282: **Choosing graphics for exploring data: Scagnostics and interaction**
Presenter: **Katrin Grimm**, University of Augsburg, Germany
Choosing the right graphics for understanding and analyzing new data sets with a huge number of variables can be time intensive and challenging. To get first insights into important variables and interesting structures a small number of well chosen criteria and the right interactive tools can provide really helpful support. This talk concentrates on scagnostics, which are measures for characterizing scatterplots. Scagnostics (from scatterplot and diagnostics) were first mentioned by Paul and John Tukey in the 1980s. Concrete measures were previously proposed and implemented in the R-package scagnostics. Specifically, clustering plots using nine scagnostics were suggested to identify both unusual plots (outliers) and frequently occurring plots (summarised by exemplars). Providing flexible interactive interfaces for the criteria gives the user more control of the analysis. The aim is to introduce an interactive apps programmed with the R-package shiny and shows the functionality using a dataset of the 2013 German election.

EO0336: **Graphical tools for modal clustering diagnostics**
Presenter: **Jose E Chacon**, Universidad de Extremadura, Spain
Modal clustering is a nonparametric methodology that uses density estimation to find clusters. The basic idea is to estimate the modes of the density, and then assign points to the modes by finding the basins of attraction of the modes. We present diagnostics for modal clustering. Particularly, we provide graphical tools to assess the uncertainty in the cluster assignment of a data point. As expected, the higher uncertainty zones are located around the cluster boundaries.

EO0415: **Statistical significance of features using CircSiZer**
Presenter: **Maria Oliveira**, University of Santiago de Compostela, Spain
Co-authors: Rosa Crujeiras, Alberto Rodriguez-Casal
Nonparametric smoothing methods, such as kernel methods, are useful tools for exploring underlying structures in data samples. However, both for nonparametric density and regression estimation, the smoothing parameter controls the global appearance of the estimator and its dependence on the sample, in such a way that an unsuitable choice of this value may provide a misleading estimate of the density or the regression curve. Hence, the assessment of the statistical significance of the observed features through the smoothed curve should be required for not compromising the extracted conclusions. SiZer methods provide a graphical tool to assess significance features of the underlying curve without relying on a smoothing parameter. Several adaptations of SiZer have been proposed in the statistical literature, making it possible to extend this graphical tool to a variety of contexts, such as circular data, yielding the CircSiZer. The CircSiZer method, both for circular density and circular-linear regression, will be described and its performance will be checked with some simulated data examples and illustrated with real datasets from different scientific fields.

EO1131: **Visual inference: Statistics at street corners**
Presenter: **Heike Hofmann**, Statistics and Statistical Laboratory, Iowa State University, United States
Co-authors: Di Cook
Visual inference allows us to determine whether something seen in a data plot is really there similarly to classical statistical hypothesis testing. Visual inference is based on non-parametric inferential methods using human observers to establish the relevance of graphical findings, thereby creating a bridge between classical statistical inference and exploratory data analysis. Based on a previous work, we will start with the lineup protocol, give examples where visual inference has allowed us to gain insight beyond classical modeling situations, and go into a discussion of some of the still open questions on visual inference.

EO057 Room 213 BAYESIAN METHODS FOR DEPENDENT DATA Chair: Antonio Linero

EO0182: **Dependent Bayesian density estimation with species sampling models**
Presenter: **Jaeyong Lee**, Seoul National University, Korea, South
A novel Bayesian nonparametric model is considered for density estimation with an underlying spatial structure. The model is built on a class of species sampling models, which are discrete random probability measures that can be represented as a mixture of random support points and random weights. Specifically, we construct a collection of spatially dependent species sampling models and propose a mixture model based on this collection. The key idea is the introduction of spatial dependence by modeling the weights through a conditional autoregressive model. We will also give examples in time series and climate data. The computation methods with Markov chain Monte Carlo and variational method will be discussed.

EO0234: **Treed covariate segmentation models for soil carbon and other nonstationary spatial processes**
Presenter: **Veronica Berrocal**, University of Michigan, United States
A key component of statistical models for environmental applications is the spatial covariance function, which is traditionally assumed to belong to a parametric class of stationary models whose parameters are estimated using observed data. While convenient, the assumption of stationarity is rarely realistic; as a result, there is a rich literature on alternative strategies that allow the covariance function to vary across space. We consider a Bayesian nonstationary spatial model constructed through the convolution of locally stationary spatial processes. Regions of local stationarity underlying our model are determined via segmentation of the geographic space into homogeneous regions based on the within-region distribution of spatially-varying covariate information. Via Bayesian model averaging, we account for uncertainty in the segmentation process, as well as the situation where multiple spatially-varying covariates are relevant to describing the nonstationary behavior of the process of interest. In this way, spatially-referenced covariates in our model are allowed to drive both the first and second moments of the spatial process. One advantage of our modeling framework over other covariate-driven nonstationary spatial models is that our approach does not require the covariates to be observed at the same location as either the observations or prediction locations. We illustrate our methodology through an application in soil science.

EO0717: **Process-based hierarchical models for coupling high-dimensional LiDAR and forest variables over large geographic domains**
Presenter: **Yuzhen Zhou**, University of Nebraska-Lincoln, United States
Co-authors: Andrew Finley, Sudipto Banerjee, Bruce Cook
Recent advancements in remote sensing technology, specifically Light Detection and Ranging (LiDAR) sensors, provide the data needed to quantify forest characteristics at a fine spatial resolution over large geographic domains. From an inferential standpoint, there is interest in prediction and interpolation of the often spatially misaligned LiDAR signals and forest variables. We propose a fully process-based Bayesian hierarchical model for above ground biomass (AGB) and LiDAR signals. The process-based framework offers richness in inferential capabilities (e.g. inference on the entire underlying processes instead of their values only at pre-specified points) and their easier interpretability. Key challenges we obviate include misalignment between the AGB observations and LiDAR signals and the high-dimensionality in the model emerging from LiDAR signals in conjunction with the large number of spatial locations. We offer simulation experiments to evaluate our proposed models and also apply them to a challenging dataset comprising LiDAR and spatially coinciding forest inventory variables.

EO0896: **Models for high-dimensional non-Gaussian dependent data**
Presenter: **Jonathan Bradley**, Florida State University, United States
Co-authors: Scott Holan, Christopher Wikle
A Bayesian approach is introduced for analyzing high-dimensional dependent data that are distributed according to a member from the exponential family of distributions. This problem requires extensive methodological advancements, as jointly modeling high-dimensional dependent data leads to the so-called 'big n problem'. The computational complexity of this problem is further exacerbated by allowing for non-Gaussian data models. Thus, we develop new computationally efficient distribution theory for this setting. In particular, we introduce a class of conjugate multivariate distributions for the exponential family. We discuss several theoretical results regarding conditional distributions, an asymptotic relationship with the multivariate normal distribution, parameter models, and full-conditional distributions for a Gibbs sampler. We demonstrate the modeling framework through several examples, including an analysis of a large dataset consisting of American Community Survey (ACS) period estimates.

EO103 Room 007 BAYESIAN SEMI-AND NONPARAMETRIC MODELLING I Chair: Antonio Lijoi

EO0190: **Frequentist analysis of the posterior for high-dimensional models**
Presenter: **Johannes Schmidt-Hieber**, Leiden University, Netherlands
Recently, various methods have been proposed for estimation and model selection in high-dimensional statistical settings. The most widely known procedure is the LASSO which can be interpreted as a maximum a posteriori probability estimate. Generalizing this, it seems natural to study high-dimensional statistics using the Bayesian method. Firstly, we will summarise recent results concerning posterior shrinkage and model selection for spike-and-slab type priors. These methods are known to perform well theoretically but are hard to compute. Secondly, priors which can be represented as scale mixtures of normals are discussed. The well-known horseshoe prior is for instance of this form. We derive sharp conditions under which such priors are sparsity inducing and show some simulations.

EO0682: **Posterior contraction of the population polytope in finite admixture models**
Presenter: **Long Nguyen**, University of Michigan, United States
The posterior contraction behavior is studies for the latent population structure that arises in admixture models as the amount of data increases. We adopt the geometric view of admixture models - alternatively known as topic models - as a data generating mechanism for points randomly sampled from the interior of a (convex) population polytope, whose extreme points correspond to the population structure variables of interest. Rates of posterior contraction are established with respect to Hausdorff metric and a minimum matching Euclidean metric defined on polytopes. Tools developed include posterior asymptotics of hierarchical models and arguments from convex geometry.

EO0590: **Bayesian nonparametric calibration and combination of predictive distributions**
Presenter: **Federico Bassetti**, Univeristy of Pavia, Italy
Co-authors: Roberto Casarin, Francesco Ravazzolo
A Bayesian approach is introduced to predictive density calibration and combination that accounts for parameter uncertainty and model set incompleteness through the use of random calibration functionals and random combination weights. Building on previous work, we use infinite beta mixtures for the calibration. The proposed Bayesian nonparametric approach takes advantage of the flexibility of Dirichlet process mixtures to achieve any continuous deformation of linearly combined predictive distributions. The inference procedure is based on Gibbs sampling and allows accounting for uncertainty in the number of mixture components, mixture weights, and calibration parameters. The weak posterior consistency of the Bayesian nonparametric calibration is provided under suitable conditions for unknown true density. We study the methodology in simulation examples with fat tails and multimodal densities and apply it to density forecasts of daily S&P returns and daily maximum wind speed at the Frankfurt airport.

EO1255: **Bayesian detection of image boundaries**
Presenter: **Meng Li**, Duke University, United States
Co-authors: Subhashis Ghosal
Detecting boundary of an image based on noisy observations is a fundamental problem of image processing and image segmentation. For a d-dimensional image ($d = 2,3,\ldots$), the boundary can often be described by a closed smooth $(d-1)$-dimensional manifold. We propose a nonparametric Bayesian approach based on priors indexed by $\mathbb{S}^{d-1}$, the unit sphere in $\mathbb{R}^d$. We derive optimal posterior contraction rates for Gaussian processes or finite random series priors using basis functions such as trigonometric polynomials for 2-dimensional images and spherical harmonics for 3-dimensional images. For 2-dimensional images, we show a rescaled squared exponential Gaussian process on $\mathbb{S}^1$ achieves four goals of guaranteed geometric restriction, (nearly) minimax optimal rate adapting to the smoothness level, convenience for joint inference and computational efficiency. We conduct an extensive study of its reproducing kernel Hilbert space, which may be of interest by its own and can also be used in other contexts. Several new estimates on the modified Bessel functions of the first kind are given. An R package is available online to implement the developed fast posterior sampling algorithm. Simulations confirm excellent performance and robustness of the proposed method.

EO1225: **Bayesian estimation of the threshold of a generalised Pareto distribution for heavy-tailed observations**
Presenter: **Cristiano Villa**, University of Kent, United Kingdom
A method is discussed to define prior distributions for the threshold of a generalised Pareto distribution, in particular when its applications are directed to heavy-tailed data, and considering a semi-parmatric model set up. We propose to assign prior probabilities to the order statistics of a given set of observations. In other words, we assume that the threshold coincides to one of the data points. We show two ways of defining a prior: by assigning equal mass to each order statistic, that is a uniform prior, and by considering the worth that every order statistic has in representing the true threshold. Both proposed priors represent a scenario of minimal information, and we study their adequacy through simulation exercises and by analysing two applications from insurance and finance.

EO185 Room 202 DEALING WITH FINANCIAL TIME SERIES Chair: Pedro Galeano

EO0221: **Modelling high dimensional stock dependence using factor copulas**
Presenter: **Hoang Nguyen**, Universidad Carlos III de Madrid, Spain
Co-authors: Concepcion Ausin, Pedro Galeano
Multivariate GARCH models are widely used to derive the dynamic dependence structure of financial time series. However, the number of parameters becomes explosive in high dimensions which results in most of the models in the literature being static. Alternatively, factor copulas as the truncated C-vines rooted at the latent variables are proposed for tackling the problem. As in any factor model setting, it is assumed that each financial time series is affected by some common latent factors. Factor loadings are modelled as generalized autoregressive score (GAS) processes imposing a dynamic dependence structure in their densities. We employ the Bayesian approach to estimate the different specifications of the factor copula models. Condition on the latent factors, time series become independent which allows the algorithm to run in a parallel setting. Besides, the independence assumption of the latent models also reduces the computational burden for the conditional posterior distribution. A simulation study shows the performance for each models. We suggest DIC criteria for model selection. Finally, we provide an illustration on the stock price of companies listed in S&P100 and compare the prediction power of models using Value-at-Risk.

EO0368: **A robust theoretical regression tree to detect structural breaks in financial time series**
Presenter: **Carmela Cappelli**, University of Naples Federico II, Italy
Co-authors: Francesca Di Iorio
In several real life and research situations data are naturally bounded by intervals, but usually they are analyzed summarizing the original data into single values. By doing so, some relevant information in the original data is lost. In the last years, efforts have been done either to extend classical methods or to develop new approaches to deal with interval-valued data. We address the problem of locating multiple structural breaks in financial Interval-Valued Time Series (IVTS), which represent a special case of IVTS because, for a given time unit, it is available, besides the extremes values (lowest and highest) of the corresponding interval, also the last value (closing) that is the most widely used in standard analysis. Moreover, financial time series are characterized by high variability and the presence of outliers that can affect the correct identification of break dates. In order to cope with the above issues, we define the lower and upper bound of the intervals as a function of the closing value, and then we employ, in the framework of a regression tree based approach, a robust distance that is able to neutralize the impact of outliers. We present the results of an empirical application to the prices of the American International Group (AIG) that shows the usefulness of the proposed procedure.

EO0672: **Bayesian semiparametric multivariate stochastic volatility**
Presenter: **Martina Danielova Zaharieva**, University of Muenster, Germany
The proposed framework is a Cholesky-type multivariate stochastic volatility model (MSV), in which the multivariate distribution of the error term is modeled as univariate infinite scale mixture of Gaussian distributions. A Bayesian non-parametric approach, in particular a Dirichlet process mixture, is adopted. This allows for high flexibility of the return distribution with respect to the kurtosis. Furthermore, the Cholesky decomposition allows for parallel univariate process modeling, creating potential for higher dimension MSV. Markov Chain Monte Carlo methods are applied for the posterior simulation and the computation of the predictive density. Finally, the predictive abilities of the model are studied by means of the predictive density evaluation.

EO1010: **Non-parametric mixture copula models**
Presenter: **Audrone Virbickaite**, University of Konstanz, Germany
Co-authors: Hedibert Lopes
A Bayesian non-parametric Gaussian mixture copula model is proposed. This allows us to model asymmetric and fat-tailed dependence structures between multiple variables and also permits parsimonious models in high dimensions. While the copula is obtained from an infinite location-scale mixture of multivariate Gaussian distributions, inference and prediction is carried out using MCMC methods, in particular slice sampling

techniques. The proposed model is investigated using simulated data and applied to model dependence between financial returns. Finally, extensions are considered to time-varying infinite Gaussian mixture copula models.

EC1153: **Out-of-sample forecasting performance of predictive regression models**
Presenter: **Anibal Emiliano da Silva Neto**, Universidad Carlos III de Madrid, Spain
Co-authors: Jesus Gonzalo, Jean-Yves Pitarakis
Several studies have shown that most of the linear predictive regression models could not forecast better than the simple historical average for the US equity premium; besides, techniques such as economically motivated constraints as well as bootstrap aggregation (henceforth, bagging) were shown to decrease the mean squared forecast errors of some linear models and outperform the historical average. However, there are no studies that investigate the performance of such techniques in the context of nonlinear models, and also whether these improvements are robust to the window size choice or not. Our aim is to verify whether applying these constraints as well as bagging indeed provide more accurate forecasts than the historical average, by means of both Clark-West and Rossi-Inoue's forecast comparison tests. By employing linear (bivariate and a factor model) and nonlinear (Markov-Switching and Threshold) predictive models, an empirical application to the S& P500 and Goyal and Welch financial data shows that the gains in terms of out-of-sample forecasting accuracy provided by applying only the constraints and constraints with bagging outperform the historical average. However, for most of the variables, these gains are only robust to the window size choice for the case of constraints with bagging.

EO065 Room 212 DEPENDENCE MODELS AND COPULAS II Chair: Fabrizio Durante

EO0231: **Dependence models and copulas in coherent systems**
Presenter: **Jorge Navarro**, Universidad de Murcia, Spain
Representations for the reliability function of coherent systems with possibly dependent components are obtained. The representations are similar to copula representations and are based on distortion (aggregation) functions of the component reliability functions. These representations can be used to compare different systems in some stochastic orders. Similar representations can be obtained for the residual lifetimes of the system at a time t under different assumptions. We consider the following cases: (1) At time t all the components are working. (2) At time t we just know that the system is working. (3) At time time the system is working but some components do not work and we know the failure times of the broken components.

EO0289: **Stochastic properties of conditionally dependent frailty models**
Presenter: **Rosario Rodriguez-Grinolo**, Universidad Pablo de Olavide, Spain
Co-authors: Jose-Maria Fernandez-Ponce, Pellerey Franco
The frailty approach provides a convenient tool in survival analysis and reliability to model dependence between lifetimes of individuals, or between components, due to common environmental conditions. The frailty (an unobservable random vector that describes environmental conditions) acts simultaneously on the hazard functions of the lifetimes. Comparisons between frailty models have been studied by assuming independence for the baseline survival functions and the corresponding environmental parameters. A generalization of these models is developed, which assumes conditional dependence between the components of the random vector. Some conditions for stochastic comparisons are provided and several examples of frailty models satisfying these conditions are also described.

EO1048: **Distributions with fixed marginals maximizing the mass of the endograph and the graph of a function**
Presenter: **Wolfgang Trutschnig**, University of Salzburg, Austria
Motivated by the question of maximizing the probability that X does not default before Y within the class of all random variables X,Y with given (continuous) distribution functions F and G respectively, we solve the more general problem of maximizing the mass of the endograph $\Gamma^{\leq}(T)$ of a not necessarily monotonic transformation $T:[0,1]\to[0,1]$ within the class of all copulas. In particular, we show that the maximum is attained for a completely dependent copula and derive a simple and easily calculable formula for the maximum. Based on an analogous formula for the maximal mass of the graph of T we then characterize all non-decreasing T for which the maximum mass of the endograph and the maximal mass of the graph coincide.

EO0571: **Some results on two-dimensional extreme-value copulas**
Presenter: **Manuela Schreyer**, University of Salzburg, Austria
Co-authors: Wolfgang Trutschnig, Juan Fernandez Sanchez
Working with Markov kernels (conditional distributions) and right-hand derivatives D^{+A} of Pickands dependence functions A we study some results of two-dimensional extreme-value copulas (EVCs) C_A. Underlining the usefulness of working directly with the right-hand derivatives D^{+A}, we give an alternative simple proof of the fact that EVCs with piecewise linear A can be expressed as weighted geometric mean of some EVCs whose dependence functions A have at most two edges and present a generalization of this result. Furthermore we sketch an elegant extension of the proof of the validity of the Hutchinson-Lai conjecture for EVCs to arbitrary, non-smooth A, and state a new conjecture on the precise region determined by all possible values of Kendall's τ and Spearman's ρ for EVCs.

EO0662: **CoVaR for asymptotically dependent losses**
Presenter: **Piotr Jaworski**, University of Warsaw, Poland
CoVaR (conditional Value at Risk) is a newly introduced risk measure which is oriented on systemic risk. If random variables X and Y are modelling our phenomena, say actuarial risks or losses from the investments, CoVaR of Y with respect to X is VaR of conditional Y. In more details $CoVaR_\beta(Y|X) = VaR_\beta(Y|X\in E)$, where E, the Borel subset of the real line, is modelling some adverse event concerning X. The basic properties of CoVaR will be presented. Especially, when the copula of the losses/risks belongs to one of the basic families modelling upper tail dependence: conic, extreme value or right truncation invariant.

EO469 Room 216 MODERN APPROACHES TO ANALYSIS OF CORRELATED DATA Chair: Hyokyoung Grace Hong

EO0250: **Conditional screening for survival data**
Presenter: **Hyokyoung Grace Hong**, Michigan State University, United States
Co-authors: Jian Kang, Yi Li
Identifying important biomarkers that are predictive for cancer patients'prognosis is key in gaining better insights into the biological influences on the disease and has become a critical component of precision medicine. The emergence of large-scale biomedical survival studies, which typically involve excessive number of biomarkers, has brought high demand in designing efficient screening tools for selecting predictive biomarkers. The vast amount of biomarkers defies any existing variable selection methods via regularization. The recently developed variable screening methods, though powerful in many practical setting, fail to incorporate prior information on the importance of each biomarker and are less powerful in detecting marginally weak while jointly important signals. We propose a new conditional screening method for survival outcome data by computing the marginal contribution of each biomarker given priorly known biological information. This is based on the premise that some biomarkers are known to be associated with disease outcomes a priori. Our method possesses sure screening properties and a vanishing false selection rate.The utility of the proposal is further confirmed with extensive simulation studies and analysis of a Diffuse large B-cell lymphoma (DLBCL) dataset.

EO0715: **Generalized principal component analysis: Dimensionality reduction through the projection of natural parameters**
Presenter: **Yoonkyung Lee**, Ohio State University, United States
Co-authors: Andrew Landgraf
Principal component analysis (PCA) is useful for a wide range of data analysis tasks. However, its implicit link to the Gaussian distribution can be undesirable for discrete data such as binary and multi-category responses or counts. We generalize PCA to handle various types of data using the generalized linear model framework. In contrast to the existing approach of matrix factorizations for exponential family data, our generalized PCA provides low-rank estimates of the natural parameters by projecting the saturated model parameters. Due to this difference, the number of parameters does not grow with the number of observations and the principal component scores on new data can be computed with simple matrix multiplication. We provide a computationally efficient algorithm for finding the principal component loadings and demonstrate the benefits of the proposed approach numerically.

EO0801: **Permutation based test on covariance separability**
Presenter: **Johan Lim**, Seoul National University, Korea, South
Covariance separability is an attractive feature of matrix variate data, which can improve and simplify many multivariate procedures. Due to its importance, testing separability has much attention in the literature. The procedures in the literature are based on the likelihood ratio (LR) test under normality, and aim to find a good approximation to its null distribution. We propose a new procedure, which is very different from existing ones. We rewrite the null hypothesis (the separability of the covariance matrix) into many sub-hypotheses (the separability of the sub-matrices of the covariance matrix), which are testable using a permutation-based procedure. We then combine the testing results of sub-hypotheses using the Bonferroni procedure. Our procedure has at least two advantages over the existing LR test under normality. Our procedure is based on the permutation procedure and inherently distribution free; thus it is robust to non-normality of the data. In addition, unlike the LR test, it is applicable to the data whose sample size is smaller than the number of unknown parameters in the covariance matrix. The numerical study and real data example show these advantages of our procedure over the existing LR test under normality.

EO1338: **Integrative association analysis of multiple heterogeneous data sources**
Presenter: **Irina Gaynanova**, Texas A and M University, United States
Co-authors: Gen Li
The growth of data collection and data sharing led to increased availability of multiple types of data collected on the same set of objects. As an example, RNASeq, miRNA expression and methylation data for the same tumor samples are publicly available through the Cancer Genome Atlas (TCGA). Due to the scale of the data, as well as its heterogeneity, it is typical to analyze each data type separately. We use penalized risk minimization framework as a building block for integrative association analysis of multiple heterogeneous data sources. By learning the sparse representation of underlying matrix decompositions, we are able to identify the patterns that are common across the data sources as well as source-specific patterns.

EO1330: **Rank-consistency of the generalized Bradley-Terry model with link misspecification**
Presenter: **Yongdai Kim**, Seoul National University, Korea, South
Estimation of ranks has been received much attention in the areas of information retrieval and online game design, and the Bradley-Terry model is popularly used. We show that the popularity of Bradley-Terry model is due to not only flexible modeling and easy computation but also some nice asymptotic properties even when the model is misspecified. We give sufficient conditions under which the Bradley-Terry model yields a consistent estimate of ranks with partial preference data when the true generative model of ranks belongs to the class of Thurstone model. By numerical experiments, we illustrate that the proposed sufficient conditions are important and practically useful.

EO145 Room 003 STATISTICAL METHODS FOR ACTUARIAL SCIENCES AND FINANCE Chair: Tim Verdonck

EO0285: **Regression modeling for the valuation of large variable annuity portfolios**
Presenter: **Emiliano Valdez**, University of Connecticut, United States
Variable annuities are life insurance policies that contain complex guarantees where there is no closed-form formula to calculate the value of the guarantees except for some special cases. Insurance companies rely heavily on the Monte Carlo simulation method, which is usually very time-consuming, to price the guarantees. It will take a computer that can process several cash flow projections per second about 2-3 days to finish the computation. As a result, valuation and risk management of such large portfolios are a big challenge to insurance companies. Metamodeling has been proposed to address this computational challenge which consists of two components: the experimental design used to select a small set of policies from a large portfolio and the metamodel used to approximate the fair market value (or similar quantities). Existing metamodels (e.g., kriging) assume that the fair market value of the guarantees follows a normal distribution, an assumption not appropriate for fair market value of the guarantees that are typically positively skewed. The skewness is due to the fact that the payoffs have a larger range than the fees of the guarantees. To address this problem, we propose to use GB2 (generalized beta of the second kind) distributions with four parameters to model the fair market values of VA guarantees. Compared to existing techniques, this method can better capture the skewness of the distribution of the fair market values.

EC1674: **Identifiability of plug-in age-period-cohort models of Lee-Carter type**
Presenter: **Eric Beutner**, Maastricht University, Netherlands
Co-authors: Simon Reese, Jean-Pierre Urbain
The basic Lee-Carter model is an age-period model that uses a non-linear parametrization of the logarithm of the central forces of mortality. To ensure identifiability constraints have to be added to the model. The same holds true for cohort extensions of the basic Lee-Carter model. To be

able to forecast with these age-period-cohort models a two step procedure is applied. In a first step the parameters of the models are estimated. In the second step a time series model that allows for forecasting is fitted to a subset of the estimated parameters. The models that result by replacing a subset of the parameters by a time series model are called plug-in Lee-Carter models. We will analyze whether these plug-in models are identifiable. It will be seen that, for instance, an identified age-period model combined with an identified time series model does not necessarily lead to an identified plug-in model.

EO1264: **Robust and Pareto optimality of insurance contracts**
Presenter: **Vali Asimit**, City University London, United Kingdom
Co-authors: Valeria Bignozzi, Ka Chun Cheung, Junlei Hu, Eun-Seok Kim
The optimal insurance problem represents a fast growing topic that explains the most efficient contract that an insurance player may get. The classical problem investigates the ideal contract under the assumption that the underlying risk distribution is known, i.e. by ignoring the parameter and model risks. Taking these sources of risk into account, the decision-maker aims to identify a robust optimal contract that is not sensitive to the chosen risk distribution. We focus on Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR)-based decisions, but further extensions are easily possible. The worst-case scenario and worst-case regret robust models are discussed, which have been already used in the investment portfolio literature. Closed-form solutions are obtained for the VaR Worst-case scenario case, while Linear Programming (LP) formulations are provided for all other cases. The Pareto optimality of the robust insurance contracts is also investigated and simple numerical methods are found for constructing insurance contracts that are Pareto and robust optimal. Our numerical illustrations have shown that marginal evidence in favour of our robust solutions is obtained for VaR-decisions, while our robust methods are clearly preferred for all other risk measures.

EO0273: **On CRDIV and banks' simultaneous defaults**
Presenter: **Marco Petracco**, Joint Research Centre of the European Commission, Italy
Co-authors: Francesca Di Girolamo, Andrea Pagano
The Capital Requirement Directive IV issues rules on the new regulatory standards for bank capital adequacy. Among others, it requires all instruments in the additional Tier 1 layer of a credit institution to be written down or converted into equity, as soon as the CET1 falls below 5.125% of risk weighted assets. Whether or not the new framework is making the banking sector more resilient, there is still one issue that regulators have never dealt with. What the Basel accord imposes to each bank is a minimum capital meant to cover unexpected losses as the banks were isolated entities. In reality, banks are exposed to common borrowers. A quantitative assessment of the effect of having a commonality among banking shocks is performed. The banks are supposed to be interconnected via a correlation and contagion structure and a micro simulation model is used to estimate the joint distribution of losses. Results show that the correlation increases systemic losses up to 5% or to 40% when adding second round effects. A series of rules are defined to annihilate the losses due to the commonality. The analysis provides evidence that the regulatory rule of requiring extra capital as soon as the CET1 falls down the 5.125% of risk weighted assets is more efficient than asking GSIBs or all banks to increase their capital, and may thus be an efficient macroeconomic tool to face banks' simultaneous defaults and help in dealing with the missing piece of the Basel framework.

EO1228: **A new approach for buffering stock returns in investment-linked annuity contracts**
Presenter: **Daniel Linders**, KU Leuven, Belgium
Co-authors: Servaas van Bilsen
The purpose is to introduce a new class of unit-linked insurance contracts providing a stream of future cash flows linked to stock market returns through the performance of a reference index. The innovation of the contract design lies in the absorption of financial shocks into future cash flows. Indeed, in order to temper the volatility in the cash flow stream, we only gradually adjust future cash flows to past financial shock. This is in contrast with the current unit-linked products where a financial shock is fully reflected in the subsequent cash flows. Our framework allows to tailor insurance contracts to the specific risk-profile of the policyholder. In order to build a realistic risk management framework, we consider a financial market with Levy distributed shocks. Whereas working in a Gaussian context might be mathematical tractable, it may be dangerous when the market exhibits extreme movements. Apart from the Gaussian distribution, our framework allows to use a broad spectrum of non-Gaussian distributions which incorporate stylized facts of stock market returns, such as heavy tails, skewness, kurtosis, etc. We show how to price and hedge the liabilities associated with our new insurance contract and provide numerical illustrations of the performance of the hedge.

EO053 Room 209 THEORETICAL FOUNDATIONS OF BIG DATA Chair: Guang Cheng

EO0353: **A distributed estimator based on a one-step approach**
Presenter: **Xiaoming Huo**, Georgia Institute of Technology, United States
Distributed statistical inference has recently attracted enormous attention. Many existing work focuses on the averaging estimator. We propose a one-step approach to enhance a simple-averaging based distributed estimator. We derive the corresponding asymptotic properties of the newly proposed estimator. We find that the proposed one-step estimator enjoys the same asymptotic properties as the centralized estimator. The proposed one-step approach merely requires one additional round of communication in relative to the averaging estimator; so the extra communication burden is insignificant. In finite sample cases, numerical examples show that the proposed estimator outperforms the simple averaging estimator with a large margin in terms of the mean squared errors. A potential application of the one-step approach is that one can use multiple machines to speed up large scale statistical inference with little compromise in the quality of estimators. The proposed method becomes more valuable when data can only be available at distributed ma-chines with limited communication bandwidth.

EO0688: **Divide and conquer in non-standard problems and the super-efficiency phenomenon**
Presenter: **Cecile Durot**, Univ. Paris Nanterre, France
The aim is to study how the divide and conquer principle — partition the available data into subsamples, compute an estimate from each subsample and combine these appropriately to form the final estimator — works in non-standard problems where rates of convergence are typically slower than square-root n, and limit distributions are non-Gaussian, with a special emphasis on the least squares estimator of a monotone regression function. We find that the pooled estimator, obtained by averaging non-standard estimates across the mutually exclusive subsamples, outperforms the non-standard estimator based on the entire sample in the sense of pointwise inference. We also show that, under appropriate conditions, if the number of subsamples is allowed to increase at appropriate rates, the pooled estimator is asymptotically normally distributed with a variance that is empirically estimable from the subsample-level estimates. Further, in the context of monotone function estimation we show that this gain in pointwise efficiency comes at a price — the pooled estimator's performance, in a uniform sense (maximal risk) over a class of models worsens as the number of subsamples increases, leading to a version of the super-efficiency phenomenon. In the process, we develop analytical results for the order of the bias in isotonic regression which are of independent interest.

EO0784: **Learning large directed acyclic graphs from high-dimensional data**
Presenter: **Qing Zhou**, UCLA, United States
Co-authors: Bryon Aragam, Arash Amini
A penalized likelihood estimation framework is developed to learn the structure of Bayesian networks, represented as directed acyclic graphs, from

high-dimensional data. In contrast to recent methods which accelerate the learning problem by restricting the search space, our main contribution is a fast algorithm for score-based structure learning which does not restrict the search space in any way. We provide a comprehensive comparison of our approach to several standard structure learning methods on graphs with up to 8,000 nodes, and show that our algorithm obtains higher sensitivity with comparable false discovery rates for high-dimensional data and scales efficiently as the number of nodes increases. We also establish nonasymptotic deviation bounds on the estimation error, sparsity bounds, and model selection consistency under concave regularization. These results apply to a wide variety of underlying covariance structures as well as many popular regularizers including the MCP, SCAD, ℓ_0 and ℓ_1.

EO1295: **Generalized cross-validation in divide-and-conquer**
Presenter: **Zuofeng Shang**, Binghamton University, United States

Tuning parameter selection is an important issue in smoothing splines. When data is huge, conventional methods such as generalized cross validation (GCV) face severe computational challenges. Specifically, the optimizer of the conventional GCV score function becomes "oracle", since it usually cannot be found in an acceptable time frame. We propose a new selection method that couples GCV with divide-and-conquer. The proposed method can efficiently complete the selection process and the selected tuning parameter performs equally well as the "oracle" tuning parameter. Both theoretical and computational results strongly support our findings.

EO1731: **Adaptive prediction in additive models**
Presenter: **Cun-Hui Zhang**, Rutgers University, United States

Penalized estimation is considered in additive models with a large number of mixed components including univariate linear effects, group effects and nonparametric effects of one or several variables. A prediction error bound, derived under a restricted eigenvalue or compatibility condition, provides rate optimality for the penalized estimator in various settings. In nonparametric additive models, the prediction error bound yields existing and new results under different smoothness and sparsity conditions. An adaptive estimator is constructed to unify these and some non-convex rate optimal methods.

EO081 Room 204 RARE AND EXTREME EVENTS IN CLIMATOLOGY Chair: Ivette Gomes

EO0453: **Extreme events of the mid-latitude atmospheric circulation: A dynamical systems perspective**
Presenter: **Davide Faranda**, CNRS, France

The link between rare recurrences of states in dynamical systems and the extreme value theory is explored. Statistics is used to infer important properties of the system such as the distribution of local dimensions (entropy) and the stability of all accessible states. We apply this framework to weather reanalysis data, the daily NCEP 1948-2015 sea level pressure (SLP) fields over the North Atlantic. We show that extremes in the local (daily) dimensions and in the stability of SLP fields are related to climate extremes such as historical storms and blocking.

EO0538: **Rare events point processes for chaotic dynamical systems**
Presenter: **Ana Freitas**, Universidade do Porto, Portugal

Stochastic processes are considered arising from dynamical systems by evaluating an observable function (which achieves a global maximum at a single point of the phase space) along the orbits of the system. We associate the existence of an Extremal Index less than 1 to the occurrence of a periodic phenomena. We show that, under certain conditions, in the absence of clustering, the point processes of exceedances converge to a standard Poisson process. In the presence of clustering, the point processes converge to a compound Poisson process, so instead of single exceedances, we have entire clusters of exceedances with a geometric distribution ruling its multiplicity.

EO0647: **Area and peaks over thresholds processes dynamically generated**
Presenter: **Jorge Freitas**, Universidade do Porto, Portugal

Stochastic processes arising from dynamical systems are considered by evaluating an observable function along the orbits of the system. We study marked point processes associated to extremal observations of such time series corresponding to exceedances of high thresholds. Each exceedance is marked by a quantity intended to measure the severity of the exceedance or simply to count them. In particular, we consider marked point processes measuring the aggregate damage by adding all the excesses over the threshold that mark each exceedance (AOT) or simply by adding the largest excesses in a cluster of exceedances (POT). We provide conditions to prove the convergence of such marked point processes to a compound Poisson process, with particular multiplicity distributions.

EO1183: **Extreme precipitation data in the North of Portugal: A spatial analysis application**
Presenter: **Manuela Neves**, ISA and CEAUL, Portugal

Quantifying and characterizing the behavior of extreme meteorological data is a topic of enormous importance due to the effects that their magnitude and frequency can have on human life, agricultural productivity and economic aspects, among others. Extreme Value Theory (EVT) is the theory of modeling and measuring events which occur with very small probability, but with large impact. A single meteorological extreme event may affect several locations and temporal independence is usually an unrealistic assumption. Luckily stationary and weekly dependent series have the same limit distribution as the independent ones, but parameters are influenced by dependence. However when data are collected in several locations their spatial dependence has to be appropriately taken into account. Classical Geostatistics, mostly based on Gaussian distribution is inappropriate for modeling tail behavior. So, after reviewing recent results on max-stable processes, a natural extension of multivariate extremes to the spatial set-up, an exploratory study of the annual maximum of monthly precipitation recorded in the northern area of Portugal from 1941-2006 at 32 locations is performed. Dependence measures are obtained, parameters are estimated and rainfall prediction maps are simulated.

EO0860: **Modeling full range rainfall intensities using sparse semi-parametric mixture models**
Presenter: **Patricia Tencaliec**, Universite Grenoble Alpes, France
Co-authors: Clementine Prieur, Anne Catherine Favre, Philippe Naveau

A semi-parametric approach is presented to model the entire range of rainfall intensities. Our model is based on the idea of extended generalized Pareto distribution (EGPD) introduced previously in the literature. The main purpose of EGPD is to incorporate additional information in the model (e.g. to better capture the bulk part of the distribution) without influencing the GPD-tail(s) behavior. Our work extends EGPD to the semi-parametric setting, in order to provide more flexibility in the model. Our approach is based on a l_1-norm penalized estimation procedure in sparse high-dimensional mixture models. We perform a simulation study to test the performances of our approach and then we apply it on a rainfall dataset. We compare our results with the ones obtained with the parametric EGPD procedure.

EO233 Room 205 BAYESIAN COMPUTATION WITH COMPLEX DATA Chair: David van Dyk

EO0461: **RMCMC: Safe, efficient updating of Bayesian models**
Presenter: **Din-Houn Lau**, Imperial College London, United Kingdom
Co-authors: Axel Gandy
The rolling MCMC (RMCMC) system is introduced that is able to update estimates from a sequence of probability distributions. The aim of the system is to quickly and efficiently produce estimates within a user-specified bound on the Monte Carlo error. The estimates are based upon weighted samples stored in a database. The stored samples are maintained such that the accuracy of the estimates and the quality of the samples are satisfied - in this sense, the system is self-monitoring and takes the appropriate course of action. This maintenance involves varying the number of samples and updating the weights. When required, new samples are generated by a Markov chain Monte Carlo algorithm - in this sense the system is efficient. The performance of the system is demonstrated for predicting the end of season ranks of a football league. Future work and directions to develop the RMCMC system are also discussed.

EO0506: **Probabilistic record linkage and deduplication after indexing, blocking, and filtering**
Presenter: **Jared Murray**, Carnegie Mellon University, United States
Probabilistic record linkage, the task of merging two or more databases in the absence of a unique identifier, is a perennial and challenging problem. It is closely related to finding duplicates in a single database, which can be cast as linking the database to itself. The number of possible links grows rapidly in the size of the databases under consideration, and in most applications it is necessary to first reduce the number of record pairs that will be compared. Spurred by practical considerations, a range of methods have been developed for this task. These methods go under a variety of names, including indexing and blocking, and have seen significant development. However, methods for inferring linkage structure that account for indexing, blocking, and additional filtering steps have not seen commensurate development. We present an overview of the implications of indexing, blocking and filtering for subsequent inference, and propose a new model to account for particular forms of indexing and filtering.

EO0636: **A geometric approach to Bayesian alignment of functional data**
Presenter: **Sebastian Kurtek**, The Ohio State University, United States
A Bayesian model is presented for pairwise nonlinear registration of functional data. We utilize the geometry of the space of warping functions to define appropriate prior distributions and sample from the posterior using importance sampling. A simple square-root transformation is used to simplify the geometry of the space of warping functions, which allows for computation of sample statistics, such as the mean and median, and a fast implementation of a k-means clustering algorithm. These tools allow for efficient posterior inference, where multiple modes of the posterior distribution corresponding to multiple plausible alignments of the given functions are found. We also show pointwise 95% credible intervals to assess the uncertainty of the alignment in different clusters. We validate this model using simulations and present multiple examples on real data from different application domains including biometrics and medicine.

EO0703: **Deep learning for demand**
Presenter: **Matt Taddy**, Microsoft Research and University of Chicago, United States
People make purchase decisions based upon preferences and prices. This leads to a complicated system where price and availability changes for one product can affect demand for all other products. We study such systems using contemporary machine learning and statistics ideas, particularly tools from natural language processing. Data applications include online purchase histories, cloud computing logs, and scanner data from grocery stores.

EO0729: **Parallelized Markov chain Monte Carlo algorithms utilizing GPUs with an application to RNAseq data analysis**
Presenter: **Jarad Niemi**, Iowa State University, United States
Co-authors: Will Landau
Markov chain Monte Carlo (MCMC) is the predominant tool used in Bayesian parameter estimation for hierarchical models. When the model expands due to an increasing number of hierarchical levels, number of groups at a particular level, or number of observations in each group, a fully Bayesian analysis via MCMC can easily become computationally demanding, even intractable. We illustrate how the steps in an MCMC for hierarchical models are predominantly one of two types: conditionally independent draws or low-dimensional draws based on summary statistics of parameters at higher levels of the hierarchy. Parallel computing can increase efficiency by performing embarrassingly parallel computations for conditionally independent draws and calculating the summary statistics using parallel reductions. During the MCMC algorithm, we record running means and means of squared parameter values to allow convergence diagnosis and posterior inference while avoiding the costly memory transfer bottleneck. We demonstrate the effectiveness of the algorithm on a model motivated by next generation sequencing data, and we release our implementation in R packages fbseq and fbseqCUDA.

EO249 Room 207 METHODS AND APPLICATIONS OF COPULA MODELING Chair: Yichen Qin

EO0507: **Factor copula models for replicated spatial data**
Presenter: **Pavel Krupskiy**, King Abdullah University of Science and Technology, Saudi Arabia
Co-authors: Raphael Huser, Marc Genton
A new copula model is proposed that can be used with replicated spatial data. Unlike the multivariate normal copula, the proposed copula is based on the assumption that a common factor exists and affects the joint dependence of all measurements of the process. Moreover, the proposed copula can model tail dependence and tail asymmetry. The model is parameterized in terms of a covariance function that may be chosen from the many models proposed in the literature, such as the Matern model. For some choice of common factors, the joint copula density is given in closed form and therefore likelihood estimation is very fast. In the general case, one-dimensional numerical integration is needed to calculate the likelihood, but estimation is still reasonably fast even with large data sets. We use simulation studies to show the wide range of dependence structures that can be generated by the proposed model with different choices of common factors. We apply the proposed model to spatial temperature data and compare its performance with some popular geostatistics models.

EO1174: **Extremes on river networks**
Presenter: **Peiman Asadi**, EPFL, Switzerland
Co-authors: Anthony Davison, Sebastian Engelke
Extreme value copulas are commonly used to model the tail dependence between rare events. We construct a class of extreme value copulas that arise as the finite dimensional distribution of a max-stable process on a river network. The extreme flows at two locations on a river network may be dependent because the locations are flow-connected or because of common meteorological events; such characteristics of network structure should be included explicitly in the copula. We introduce a max-stable process on the river network that allows flexible modeling of flood events and that enables risk assessment even at locations without a gauging station. Recent statistical methods for extreme value copulas are used to fit this process to a big data set from the upper Danube basin.

EO1099: **Copula modeling for data with ties**
Presenter: **Yang Li**, Renmin University of China, China
Co-authors: Yan Li, Yichen Qin, Jun Yan
The copula method is well discussed in the multivariate modeling in recent years. However, the ties in the dataset usually affect the estimate of copula qualitatively. Randomly breaking and average replacement are two common solutions for copula with ties but with limitations. It may deviate much from the true model with the increasing of the percentage of ties in data. An interval censoring method is proposed to deal with the data with ties for copula modeling. Numerical studies shows that the proposed method perform good on both estimation and hypothesis testing. An illustrative example for loss insurance data is discussed.

EO1002: **Nonparametric estimation for extreme-value copula functions via constrained spline regressions**
Presenter: **Yichen Qin**, University of Cincinnati, United States
Co-authors: Yang Li, Jun Yan, Siqi Xiang
A new nonparametric estimation procedure is introduced for extreme-value copulas using spline regressions. By fitting a shape constrained spline regression function to the data points obtained from the rank-based transformation of the original observations, the authors provide new estimates of the Pickands dependence functions of the extreme-value copula. In order to impose the shape constraints on the spline regression, a new set of basis functions which satisfies such constraints is proposed. Compared with existing methods, the method works well in simulation and in real data analysis.

EO0894: **Insurance experience rating using mixed D-vine copulas**
Presenter: **Peng Shi**, University of wisconsin - Madison, United States
Co-authors: Lu Yang
In non-life insurance, insurers use experience rating to adjust premium to reflect the policyholder's previous claim experience. Performing prospective experience rating can be challenging when the claim distribution is complex. We introduce a mixed vine pair copula construction framework for modeling semicontinuous longitudinal claims. Specifically, a two-component mixture regression is employed to accommodate the zero inflation and thick tails in claim distribution. The temporal dependence among repeated observations is modeled using a sequence of bivariate conditional copulas based on a mixed D-vine. In the application of government property insurance from the state of Wisconsin, we show that the proposed approach offers substantial opportunities for separating risks and identifying profitable business when compared with alternative experience rating methods.

EO203 Room 206 ADVANCES IN ROBUSTNESS FOR FUNCTIONAL AND COMPLEX DATA Chair: Graciela Boente

EO0530: **Robust simultaneous inference for the mean function of functional data**
Presenter: **Italo Raony Costa Lima**, Auburn University, United States
Co-authors: Nedret Billor, Guanqun Cao
Functional Data presents new challenges when studying outlier contaminated datasets. We will discuss robust simultaneous inference for the mean function, presenting robust estimators of the mean function, based on polynomial splines, together with robust simultaneous confidence bands. The asymptotic properties of the estimator are also discussed. An extension of the methods for the hypothesis testing of the difference of the mean functions of two populations is also presented. The performance of the proposed robust methods and their robustness are demonstrated with an extensive simulation study and real data examples.

EO0664: **Asymptotics for M-estimators in linear models with increasing dimension**
Presenter: **Ezequiel Smucler**, IMAS, CONICET, Argentina
Existing results on the asymptotic properties of M-estimators of regression in linear models with increasing dimension are reviewed. We prove consistency and asymptotic normality for a class of resdescending M-estimators of regression assuming $p/n \to 0$ and $p^3/n \to 0$, respectively. This class is wide enough to include several popular high breakdown point regression estimators such as S and MM-estimators.

EO0864: **Robust procedures for single index models with asymmetric errors**
Presenter: **Ana Maria Bianco**, IMAS, CONICET, Argentina
Co-authors: Claudio Agostinelli, Graciela Boente
A family of robust estimators is introduced for the link function and the single index parameter under a single index model with asymmetric errors. The proposal is based on a three-step procedure related to a profile approach that combines parametric and nonparametric techniques. The procedure controls high deviance or Pearson residuals through a bounded loss function. Its asymptotic behaviour is derived under mild assumptions. A numerical study illustrates the finite sample behaviour of the proposed procedures in different scenarios.

EO1358: **Robust variable screening for regression using factor profiling**
Presenter: **Yixin Wang**, University of Leuven, Belgium
Co-authors: Stefan Van Aelst
Sure Independence Screening (SIS) is a fast procedure for variable selection in ultra-high dimensional regression analysis. It satisfies the sure screening property which means that under certain model assumptions it can reduce the dimension to a large extent while keeping all the important predictors with an overwhelming probability. However, the performance of SIS deteriorates greatly with increasing dependence among the predictors. To solve this problem, Factor Profiled Sure Independence Screening (FPSIS) has been proposed. It assumes that the predictors can be represented by a few latent factors. The correlations can be profiled out by projecting the data onto the orthogonal complement of the subspace spanned by these factors. FPSIS can obtain coefficient estimation and variable selection consistency. However, none of the methods can handle potential outliers in the data. We propose a robust FPSIS method. We first apply the least trimmed squares estimator to estimate the factors and the profiled variables. Then, a robust regression method is used for screening. Both multivariate and component-wise versions are developed to deal with row-wise and component-wise contamination respectively. The results given by both the classical and the robust procedures with different types of outliers are compared.

EO1396: **Marginal estimation under a partially linear model with missing observations**
Presenter: **Ana Perez Gonzalez**, University of Vigo, Spain
Co-authors: Graciela Boente, Ana Maria Bianco, Wenceslao Gonzalez-Manteiga
A semiparametric partially linear regression model is considered where missing data occur in the response and in the covariates corresponding to the linear component. The aim is to estimate a marginal functional, such as the mean, the median or any $\alpha-$quantile. A missing at random (MAR) condition is assumed in order to prevent from bias in the estimation of the marginal measures under a non-ignorable missing mechanism. A three stepwise procedure is employed to estimate the parametric and nonparametric components, while different approaches are given for the estimation of the marginal functional of interest. Through a Monte Carlo study we compare the performance of the given proposals.

EO589 Room 210 NEW ADVANCES IN OPTIMAL DESIGN OF EXPERIMENTS AND STATISTICAL MODELLING Chair: Stefanie Biedermann

EO0585: **Optimal restricted-randomised response surface designs allowing for pure-error estimation of the variance components**
Presenter: **Kalliopi Mylona**, Carlos III University of Madrid, Spain
Co-authors: Steven Gilmour , Peter Goos
Textbooks on response surface methodology generally stress the importance of lack-of-fit tests. For lack-of-fit tests to be possible, experiments should allow for pure-error estimation. Therefore, they should involve replicated treatments. Moreover, very often, the experimental runs cannot be performed under homogeneous circumstances, in which case the experiment is blocked. On other occasions, the experiment involves hard-to-change factors or multiple steps in a process. In that case, the experiment generally is a split-plot experiment. We present a novel approach to design blocked and split-plot experiments. The key feature of the approach is that it ensures that the two variance components resulting from the blocked or split-plot nature of the experiment can be estimated from pure error, in addition to a precise estimation of the treatment effects and the response surface model. Our novel approach involves a new Bayesian compound D-optimal design criterion which pays attention to both the variance components and the fixed treatment effects.

EO0544: **Fitting fractional polynomial models to mixture experiments**
Presenter: **Rana Khashab**, University of Southampton, United Kingdom
Co-authors: Steven Gilmour, Stefanie Biedermann
In several practical experiments, the response depends on the proportions of certain components and not on the total amount of their mixture. Data from experiments with mixtures are usually modelled using Scheffe polynomials which, however, are not flexible enough in all situations. We propose a new class of nonlinear models for fitting to the data from such experiments, which are based on fractional polynomial models. We compare our models to the recently proposed class of general blending models and to several linear models, and demonstrate that our class is preferred in numerous practical examples.

EO1082: **Sensitivity analysis for informative censoring in parametric survival models: An evaluation of the method**
Presenter: **Panayiotis Bobotas**, University of Southampton, United Kingdom
Co-authors: Alan Kimber, Stefanie Biedermann
A sensitivity analysis was previously proposed for informative censoring in parametric survival analysis. More specifically, a parametric model was introduced that allows for dependence between the failure and censoring processes in terms of a parameter δ which can be thought of as measuring the size of the dependence between the two processes, and a bias function $B(t,\theta)$ (where θ is a parameter associated with the failure process) which can be thought of as measuring the pattern of this dependence. Based on this model, for small values of δ, simplified closed form expressions (approximations) were derived for the sensitivity analysis of the associated parameters of the model. First, some theoretical issues concerning the above approach are discussed. Then, the results of an extensive simulation study are reported, which indicate some shortcomings of the proposed sensitivity analysis, particularly in the presence of nuisance parameters.

EO1087: **Minimax optimal designs for approximately exponential-based proportional hazards models subject to Type-I censoring**
Presenter: **Alan Kimber**, University of Southampton, United Kingdom
Co-authors: Stefanie Biedermann, Maria Konstantinou
The exponential-based proportional hazards model is often assumed in time-to-event experiments but may only be approximately true. We consider deviations in a neighbourhood of this model that also includes other widely used parametric proportional hazards models and further assume that the data are subject to Type-I censoring. Minimax designs are then found explicitly based on criteria corresponding to the classical c- and D-optimality criteria. We therefore extend existing results on minimax optimal designs by incorporating proportional hazards models and censoring and also provide analytical characterisations of optimal designs with finite support. Our designs are then compared with the balanced design that is traditionally used in practice.

EO1640: **Optimal design and analysis of discrete choice experiments with partial profiles involving a no-choice option**
Presenter: **Roselinde Kessels**, University of Antwerp, Belgium
Co-authors: Daniel Palhazi Cuervo
The aim is to show how to optimally design and analyse a discrete choice experiment (DCE) with a no-choice option for estimating a nested logit model when partial profiles are used to study a large number of attributes. As a motivating example, we describe a DCE to identify and quantify the determinants that influence the competitive position of the coach bus as transport mode for medium-distance travel by Belgians. We measured the attractiveness of different bus services for different destinations (Lille, Amsterdam, Cologne, Paris and Frankfurt) by having participants choose their preferred bus trip out of two bus trips, while still allowing them to also choose not to take the bus but any other transport mode comprised by the no-choice option. Each bus trip is a combination of levels of seven attributes: price, duration, and comfort attributes including wifi, leg space, catering, entertainment and individual power outlet. Varying the levels of all seven attributes of the bus trips in the choice sets would be cognitively too demanding for respondents. To ensure sensible and manageable choice tasks, we present and compare new and existing optimal design and analysis approaches for a no-choice partial profile setting in which the levels of only a subset of the attributes vary within every choice set.

EO700 Room 208 SHRINKAGE ESTIMATORS Chair: Genso-Yuan Tsung Watanabe-Chang

EO0708: **Shrinkage estimators of Poisson means based on prior information and its applications to multiplicative Poisson models**
Presenter: **Genso-Yuan Tsung Watanabe-Chang**, Mejiro University, Japan
Co-authors: Nobuo Shinozaki
In estimating $p(p \geq 2)$ independent Poisson means, it has been previously given a class of estimators that shrink the unbiased estimators to the origin and have shown that the shrinkage estimators dominate the unbiased ones under the normalized squared error loss. We consider shrinking the unbiased estimators to the specified values only when observed values are greater than or equal to the specified ones, to their minimum, or more generally to some order statistics. We apply the proposed method to the simultaneous estimation of the means in multiplicative Poisson models and propose a class of multiple-shrinkage estimators which shrinks the MLE to the row-wise and column-wise order statistics.

EO0747: **Improvement of singular value shrinkage priors and block-wise Stein priors**
Presenter: **Takeru Matsuda**, University of Tokyo, Japan
Co-authors: Fumiyasu Komaki
Singular value shrinkage priors has been previously developed for the mean matrix parameters in the matrix-variate normal model with known covariance matrices. Singular value shrinkage priors are superharmonic and put more weight on matrices with smaller singular values. They are a natural generalization of the Stein prior. Bayes estimators and Bayesian predictive densities based on our priors are minimax and dominate those based on the uniform prior in finite samples. In particular, singular value shrinkage priors work well when the true value of the parameter has low rank. We develop priors that dominate singular value shrinkage priors in estimation and prediction. Our priors not only shrink singular values for each but also shrink singular values overall. We also develop priors that dominate block-wise Stein priors in estimation and prediction.

EO0814: **Shrinkage priors for nonparametric estimations**
Presenter: **Keisuke Yano**, The University of Tokyo, Japan
Co-authors: Fumiyasu Komaki
The estimation of the mean in the Gaussian infinite sequence model is considered when the parameter space is of the ellipsoidal form. For the estimation, we propose a shrinkage type prior and discuss the property of the Bayes estimator based on the proposed prior from the viewpoints of minimaxity and admissibility. Focusing on the scale ratio of the parameter and the noise, we show that the Bayes estimator is minimax up to a logarithmic factor when the scale ratio is large and that the Bayes estimator is nearly admissible on the parameter space.

EO1040: **Shrinkage priors for a class of Poisson regression models**
Presenter: **Fumiyasu Komaki**, The University of Tokyo, Japan
Bayesian prediction based on a class of Poisson regression models is investigated. Shrinkage priors for the Poisson regression models are introduced. It is shown that Bayesian predictive densities based on the shrinkage priors dominate those based on the Jeffreys priors under the Kullback-Leibler loss. The obtained results for the Poisson regression correspond to several known results for the Gaussian regression in the viewpoint of information geometry.

EO0943: **Harmonic Bayesian prediction under alpha-divergence**
Presenter: **Yuzo Maruyama**, The University of Tokyo, Japan
Bayesian shrinkage methods for constructing predictive distributions are investigated. We consider the multivariate Normal model with a known covariance matrix and show that the Bayesian predictive density with respect to Stein's harmonic prior dominates the best invariant Bayesian predictive density, when the dimension is not less than three. Alpha-divergence from the true distribution to a predictive distribution is adopted as a loss function.

EO045 Room 211 STATISTICAL DESIGN, MODELLING AND INFERENCE Chair: Apostolos Batsidis

EO0724: **Testing for the conditional variance in nonparametric regression models**
Presenter: **Maria Dolores Jimenez-Gamero**, University of Sevilla, Spain
Co-authors: Juan-Carlos Pardo-Fernandez
The problem of testing for the parametric form of the conditional variance is considered in the non-parametric regression model. A stochastic process based on the difference between the empirical characteristic function processes that are obtained from the standardized non-parametric residuals under the null hypothesis and the alternative is studied. A Cramer-von Mises test statistic is proposed. The finite sample properties of the proposed test are numerically investigated.

EO0278: **Estimation of multivariate distributions in censored gap times**
Presenter: **Luis Machado**, University of Minho, Portugal
In many longitudinal studies, subjects can experience recurrent events. This type of data has been frequently observed in medical research, engineering, economy and sociology. These studies involve repeated events, where a subject or sample unit may experience a well-defined event several times along his history. In medical research, the recurrent events could be multiple occurrences of hospitalization from a group of patients, multiple recurrence episodes in cancer studies, repeated heart attacks or multiple relapses from remission for leukemia patients. We introduce several estimators for the bivariate distribution function and for the bivariate survival function of the joint gap times and their extension to several gap times are also studied. Some related problems such as the estimation of the marginal distributions and the conditional distributions is also discussed. We study by simulation the behavior of the proposed estimators in finite samples. Real data illustration is included.

EO0714: **Goodness-of-fit in mixture cure models with interval-censoring**
Presenter: **Sylvie Scolas**, Universite catholique de Louvain, Belgium
Co-authors: Anouar El Ghouch, Catherine Legrand
In survival analysis, we study the time until the occurrence of an event. If the event is not observed during the study period, the observation is right-censored. In this case, it is generally assumed that the event will eventually occur. It may however be the case that some individuals are not susceptible to the event, that is, they are cured, and will actually never develop the event. Additionally, the event times can be interval-censored, i.e. the event is observed, but the true occurrence time is not known exactly, but belongs to an interval of time. The mixture cure model allows to take a cure fraction into account, but the presence of interval-censoring and right-censoring is challenging. Assuming a parametric distribution for the event times reduces the challenge, but may lead to erroneous conclusion if the parametric form is wrongly chosen. This is why it is of utmost importance to develop a goodness-of-fit test in our context. We show how the semi-nonparametric (SNP) representation can achieve this in a parametric mixture cure model, and present selected results of a simulation study.

EO0423: **A lifetime distribution with increasing failure rate**
Presenter: **Pedro Jodra-Esteban**, Universidad de Zaragoza, Spain
A new two-parameter probability distribution with increasing failure rate is studied. Some statistical measures are provided in closed-form such as the moments, the incomplete moments, the quantile function and the mean residual life, among others. In this regard, the upper incomplete gamma function and the Lambert W function play a central role. On the other hand, the parameter estimation problem is carried out by the maximum likelihood method. The good performance of this method is assessed by means of a Monte Carlo simulation study. Different real data sets illustrate that the proposed model may provide a better fit than other probability distributions.

EO0660: **Joint modelling of diagnosis and survival time**
Presenter: **Polychronis Economou**, University of Patras, Greece
A new model is proposed in order to jointly model the time of diagnosis of a disease and the survival time of a patient taking into account any available explanatory covariates. Since covariates such as gender, age, race etc are usually used as initial screening factor that result in accelerating or delaying the time of diagnosis, an accelerate lifetime model is assumed for the time of diagnosis. On the other hand, a proportional hazard model is adopted for the survival time. A likelihood ratio test is also presented in order to test if the survival time is independent or not on the time of diagnosis.

EO187 Room 110 APPLICATIONS OF TIMES SERIES IN ENERGY AND ECONOMICS Chair: Carolina Garcia-Martos

EO0771: **Assessing the properties of circulant SSA for signal extraction**
Presenter: **Eva Senra**, Universidad de Alcala, Spain
Co-authors: Juan Bogalo, Pilar Poncela
Singular Spectrum Analysis (SSA) is a powerful non parametric tool for signal extraction. Its circulant version provides an automatic procedure that strongly separates the elementary components by frequency, both for stationary and nonstationary time series. We check related aspects through some Monte Carlo experiments as the influence of the window lag in the magnitude and accuracy of the extracted signals and compare the performance of the method to other well-known procedures for signal extraction as Tramo-Seats. We also analyse a real data set of industrial production time series in France and Spain.

EO1130: **Estimation of 1-in-20 year national gas peak daily demand in some European countries**
Presenter: **Ricardo Bolado Lavin**, European Commission; DG-JRC; Directorate for Energy, Transport & Climate, Netherlands
Regulation 994/2010 on security of gas supply imposes the obligation on EU Member States (MS) to fulfill the infrastructure standard. This means that every MS has to be able to satisfy the 1-in-20 year national gas peak daily demand (Dmax) after the failure of its largest infrastructure. Proving the capacity to satisfy the standard involves obtaining an accurate estimate of Dmax. Three elements are considered in this estimation process: 1) the split of the demand in three user categories: industrial, power production and household plus commercial, 2) the spatial positive correlation of the daily demand in adjacent administrative units, and 3) the use of time series models and coupled with Monte Carlo simulations. Examples of estimates for some European regions with relevant differences in the market structure, role of natural gas in the energy mix and degree of supply diversification are provided. The obtained estimates are compared with the estimates delivered by the Competent Authorities of the countries in their Preventive Action Plans. Limitations and advantages of the proposed methodology are discussed in light of Regulation 994/2010 and of policy implications for security of gas supply.

EO1726: **Compositional analysis on econometric time series**
Presenter: **Robert Ortells**, London School of Economics, Spain
Co-authors: Mi Ortego, Juan Jose Egozcue
A composition is a multivariate vector with positive components where information is conveyed by the (log-)ratios between the components. Frequently, the treatment of econometric variables is a compositional problem. For instance, the 10 year Treasury yields paid by all sovereign countries in the world is a composition. The evolution of yields in time constitutes a compositional time series. If Spanish government is currently paying 1.06% per year on its 10 year sovereign bond this information is utterly meaningless depending on what other countries are currently yielding. For instance, the equivalent German bond is currently yielding 0.05% per year, and the equivalent Iraqi bond is yielding roughly 8% per annum. If tomorrow it is observed that the Spanish bonds jumps to 2%, we will be in the same situation: we will be unable to assess whether this is a high or low rate with only this information. Indeed, the true information does not lie on the absolute value of the variable (the yield) but on the relationship it has with the rest of parts of the composition. The consequence is that this time series needs a compositional treatment consisting of representing the yields as suitable log-ratios and analyzing them as a real multivariate time series.

EO1699: **Bankruptcy prediction by survival models based on current and lagged values of time-varying financial data**
Presenter: **Francesca Pierri**, University of Perugia, Italy
Co-authors: Chrys Caroni
Periods of economic crisis arouse interest in exploring the causes of firms closure, for preventive and predictive purposes. Failure prediction models are useful tools for bankers to measure the risk of lending and minimise losses, for firms wishing to evaluate their market position, and also for investors, asset managers and rating agencies. Quantitative methods to assess the performance of firms and to predict the bankruptcy event based on balance sheet indicators are widely used in the credit risk context. Logistic regression and survival analysis techniques based on hazard models are among the methods often employed. A large data set on capital companies in Italy from 2008 to 2013, including Business Register data supplying a complete picture of their legal situation, was used to develop survival models. Training ($n = 27286$) and holdout ($n = 7124$) samples were constructed for developing and testing models, respectively. Fixed and time-varying covariates were taken into account and macro-economic variables were included besides the firms individual financial indicators. Furthermore, we considered one- and two-year lagged values of each time-varying covariate. ROC curves that vary as a function of time and AUC up to a given time were used to compare models and obtain global concordance measures

EO1207: **Multivariate time series modeling with missing data: Application to the air quality of Madrid**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Mario Ramirez Jaen, Maria Jesus Sanchez Naranjo
Air pollution is an issue that affects every big city. The use of measurement networks has grown in the last decades. The aim is to use the data collected by those networks to understand and forecast the behavior of air pollution. Three approaches have been followed, using well-established methods in other fields: ARIMA modelling of daily data, univariate and dynamic factor models for hourly data, and Space State approach. The functionality of dealing with missing data has also been included, which is a typical empirical feature of this kind of data. It has been established the existence of a dominant $(1,0,0) \times (0,1,1)$ model and two separated geographic areas on either side of the most important avenue in Madrid: Paseo de la Castellana. Sort time forecasts for most scenarios have shown a MAPE of around 20%. However, a parallel approach for all series of hourly data can only achieve a good level forecasting. The two peripheral areas (Casa de Campo and Monte del Pardo) do not follow the estimated models as good as urban areas, being more accurate when multivariate series are considered. Regarding missing data, isolated ones do not affect forecasts, whereas consecutive ones may cause a decrease of forecasting accuracy. The feasibility of modeling and forecasting air pollution using time series is demonstrated in this way.

EO237 Room 214 RECENT APPLICATIONS AND METHODS IN DIRECTIONAL STATISTICS Chair: Eduardo Garcia-Portugues

EO0793: **Rank-regularized estimation of mixtures of multivariate von Mises**
Presenter: **Wouter Boomsma**, University of Copenhagen, Denmark
Co-authors: Eduardo Garcia-Portugues, Elizabeth Wood
An Expectation-Maximization (EM) algorithm is proposed for efficient fitting of mixtures of multivariate von Mises densities. The motivation comes from the need of summarizing large datasets of high-dimensional toroidal data in a rigorous, effective and succinct way. This kind of data is common in bioinformatics, for instance originating in protein structure simulations when parametrizing the backbone of a protein as a point in the hypertorus. The EM algorithm we present employs the pseudo-likelihood as a way of bypassing the intractability of the normalizing constant, performs automatic rank-regularization on the concentration matrices and makes use of analytic gradients for increased estimation efficiency. We illustrate the performance of the procedure both on synthetic data and on a large, 94-dimensional dataset of protein dihedral angles. Finally, we discuss the `C++` library that implements the contributed methods and its `R` and `Python` interfaces.

EO0881: **Simple optimal tests for reflective symmetry of circular data**
Presenter: **Jose Ameijeiras-Alonso**, University of Santiago de Compostela, Spain
Co-authors: Christophe Ley, Arthur Pewsey, Thomas Verdebout
Most statistical procedures for circular data are based on symmetric distributions (like the von Mises, wrapped Cauchy and cardioid). It is thus of paramount importance to be able to test whether a given data set is symmetric, and hence the classical tools can be used, or whether it is asymmetric, in which case more work is required. Efficient tests will be presented for circular reflective symmetry. Those tests are designed to be optimal against the broad class of sine-skewed alternatives. The focus will be on the case where the symmetry center is unspecified.

EO1097: **Bayesian inference and model selection in directional statistics**
Presenter: **Christopher Fallaize**, University of Nottingham, United Kingdom
The likelihood functions of many distributions used to model directional data, such as Bingham and matrix-Fisher distributions, contain normalising constants which are somewhat awkward to work with. This complicates Bayesian inference, since these constants depend on the parameter of interest, and hence are required when using traditional simulation-based techniques such as the basic forms of Markov chain Monte Carlo (MCMC) samplers. Such problems are termed doubly-intractable, since both the normalising constant of the likelihood as well as the marginal likelihood of the data are intractable. One possible strategy is to use an approximation to the normalising constant and then perform a standard analysis using MCMC. Here we show how, using recent advances in MCMC methodology for doubly-intractable problems, it is possible to perform exact (Monte Carlo) inference for the unknown parameters, in the sense that samples are drawn from the exact posterior distribution of the parameter of interest. The problem of choosing between competing models will also be discussed, where the use of traditional methods (such as Bayes factors) is again complicated by the presence of the awkward normalising constant. The methods will be illustrated on real data from applications in directional statistics.

EO1101: **A pairwise probabilistic model of protein sequence and structure evolution**
Presenter: **Michael Golden**, University of Oxford, United Kingdom
Co-authors: Eduardo Garcia-Portugues, Michael Sorensen, Kanti Mardia, Thomas Hamelryck, Jotun Hein
A generative evolutionary model is presented for protein sequence and structural conformation. Recently, there have been stochastic models of structural evolution, which have shown that the inclusion of structural information in addition to sequence leads to more reliable estimation of evolutionary parameters. We introduce a new pairwise evolutionary model which takes into account local dependencies between sequence and structural evolution. The structure of each protein in a pair is treated as a random walk in space through the use of a dihedral angle representation. The evolution of dihedral angles is described using a novel diffusion process. A coupling in our model is such that an amino acid change can lead to a change in diffusion process and a jump in dihedral angle conformation. This model is comparatively more realistic than previous stochastic models, since it allows for an improved understanding of the relationship between sequence and structural evolution. The generative nature of our model enables us to provide evidence of its validity and the ability to predict protein structure using the corresponding amino acid sequence and a homologous amino acid sequence or structure.

EO1070: **Local trigonometric moments for circular density estimation**
Presenter: **Stefania Fensore**, University of Chieti-Pescara, Italy
Co-authors: Marco Di Marzio, Agnese Panzera, Charles C Taylor
Estimating equations is proposed whose unknown parameters are the values taken by a circular density and its derivatives at a point. Specifically, we solve equations which relate local versions of population trigonometric moments with empirical ones. Major advantages of our approach are: higher order bias without asymptotic variance inflation, closed form for the estimators, and absence of numerical tasks. We provide theoretical results along with simulation experiments.

EO693 Room 215 BEYOND MEAN REGRESSION Chair: Andreas Mayr

EO0869: **The GAMLSS models: Past and future**
Presenter: **Dimitrios Stasinopoulos**, London Metropolitan University, United Kingdom
GAMLSS is a framework of statistical modelling introduced as a way to overcome some of the limitations associated with Generalised Linear Models (GLM) and Generalised Additive Models (GAM).In GAMLSS the exponential family distribution assumption for the response variable, y, is relaxed and replaced by a general distribution family, including highly skew and/or kurtotic continuous and discrete distributions. The systematic part of the model is expanded to allow modelling not only the mean (or location) but all the parameters of the distribution of y as linear and/or nonlinear parametric and/or additive non-parametric functions of explanatory variables and/or random effects. Hence GAMLSS is especially suited to modelling a response variable which does not follow an exponential family distribution, (e.g. leptokurtic or platykurtic and/or positive or negative skew response data, or overdispersed counts) or which exhibit heterogeneity (e.g. where the scale or shape of the distribution of the response variable changes with explanatory variables(s)). The latest developments of GAMLSS will be discussed using data examples.

EO1151: **Assessing the significance of effects in boosted distributional regression models**
Presenter: **Tobias Hepp**, Friedrich-Alexander-Universitaet Erlangen-Nuernberg, Germany
Co-authors: Matthias Schmid, Andreas Mayr
Generalized Additive Models for Location, Scale and Shape (GAMLSS) can be also estimated via a gradient boosting algorithm. It provides advantages like automatic variable selection and feasibility in high-dimensional settings with more predictors than observations. However, the implicit regularization that allows the shrinkage of effect estimates prevents the computation of standard errors. As a result, the construction of confidence intervals or significance tests is problematic. We discuss the performance of two potential solutions. The first option is based on permutation tests, where the variable of interest is replaced by regression residuals in order to remove possible correlations with covariates. Another option is to draw parametric bootstrap samples from the conditional distribution of the constrained model without the variable of interest. Then, the differences in the quality of fit between the full model and the constrained one are attributable only to the randomness of these samples and should be less distinctive than for the original data.

EO0533: **Stability selection for boosted generalized additive models for location scale and shape**
Presenter: **Janek Thomas**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Andreas Mayr, Matthias Schmid, Bernd Bischl, Benjamin Hofner
A new algorithm is proposed to incorporate stability selection for boosting generalized additive models for location, scale and shape (GAMLSS). In one application, a negative binomial hurdle model was fitted to handle excess zeros, overdispersion, non-linearity and spatiotemporal structures to investigate the abundance of seabirds in Massachusetts. In a second step, stability selection, an increasingly popular way to obtain stable sets of covariates while controlling the false discovery rate (FDR), was applied. The model is fitted repeatedly to subsampled data and variables with high selection frequencies are extracted. This lead to a fundamental problem with boosted GAMLSS, where in every boosting iteration, the algorithm sequentially selects the best fitting effect for each distribution parameter. Thus, it is currently not possible to stop fitting individual parameters as soon as they are sufficiently modeled. To solve this problem, we developed a new approach to fit boosted GAMLSS. Instead of updating all

distribution parameters in each iteration, we only update the parameter which leads to the biggest reduction in loss. With this modification, stability selection can be applied. Furthermore, optimizing the tuning parameters of boosting is reduced from a multi- to a one-dimensional problem. The performance of the algorithm is evaluated in a simulation study and the application is demonstrated for the seabirds data, selecting stable predictors while controlling the FDR.

EO0418: **Semiparametric bivariate conditional copula regression with binary and continuous marginals**
Presenter: **Thomas Kneib**, University of Goettingen, Germany
Co-authors: Nadja Klein, Giampiero Marra, Rosalba Radice
Copulas are a general and versatile tool for constructing multivariate distributions by combining a specific type of dependence structure specified via the copula with arbitrary marginal as long as these marginals are continuous. This allows for the construction of regression models for bivariate responses within the framework of distributional regression where regression predictors are placed on potentially all distributional parameters including the dependence parameter of the copula. We extend this framework by considering bivariate regression models where at least one of the responses is binary and therefore discrete. Based on the latent utility representation of binary regression models, we can still formulate a copula specification and combine the binary response part with a flexible specification of the dependence structure and the second marginal (unless this is also binary). We develop both penalized likelihood and Bayes inference and compare the results in an application on adverse birth outcomes where we combine a binary regression model for the presence/absence of low birth weight with a three parameter dagum specification for gestational age.

EO1570: **A Markov-Switching generalized additive model for compound Poisson processes**
Presenter: **Julien Hambuckers**, University of Gottingen, Germany
Co-authors: Thomas Kneib, Roland Langrock, Alexander Sohn
The behavior of random sums over time is modelled. Such models are particularly useful in insurance and finance, to describe the behavior of total losses (like operational losses) over time, and to correctly estimate tail-related risk indicators. To this end, we formulate a generalized additive Markov-switching compound process combining Poisson and Generalized Pareto distributions. This flexible model takes into account two important features: on the one hand, we follow the idea of generalized additive models for location, scale and shape (GAMLSS), and thereby allow all parameters of the considered distributions to depend on economic covariates. On the other hand, we allow this dependence to vary over time, via a hidden state variable. A simulation study indicates that, even in the case of a short time series, the model is easily and well estimated with a standard maximum likelihood procedure. Relying on this approach, we analyze a dataset of 817 losses resulting from frauds in the Italian bank UniCredit. We show that our model improves the estimation of the total loss distribution over time, compared to standard alternatives. In particular, this model provides estimations of the 99.9% quantile that are never exceeded by the historical total losses.

EO083 Room 203 STATISTICS FOR HILBERT SPACES Chair: Gil Gonzalez-Rodriguez

EO0900: **Dimension-reduced clustering of functional data via variance-penalized optimization**
Presenter: **Michio Yamamoto**, Kyoto University, Japan
In the clustering of functional data, assuming that cluster centers exist in a common low-dimensional subspace is needed to interpret the result of the clustering and obtain the graphical representation of the clustering structure. Although several clustering methods with dimension reduction have been proposed for this aim, there is one drawback: they do not work well when the data have an unfavorable structure for clustering, which is called a disturbing structure. To address this issue, a new clustering technique is proposed, which is based on the separate estimation of subspaces for the clustering and the disturbing structure. The loss function of the proposed method consists of a dimension reduction term with penalties for within- and between-cluster variances in the subspace, and subsumes some existing clustering techniques as special cases. Numerical examples show that the proposed method works well even if the disturbing structure exists. In addition, consistency of the proposed method is discussed.

EO0658: **Dynamic functional covariance models**
Presenter: **Xinghao Qiao**, London School of Economics, United Kingdom
Co-authors: Pallavi Basu, Chenlei Leng
Covariance estimation have attracted increasing attention in recent years, especially in settings involving high dimensional data. Some recent efforts have been devoted to modelling dynamic covariance matrix for independent but non-identically distributed scalar data. In this paper, we extend the covariance models concept to describe the covariance relationship among p random functions, each of which can be represented using a functional principal components expansion. We propose dynamic functional covariance models (DFCM) to describe the covariance dynamics, where principal component scores characterize the global covariance structure and principal component functions further lead to the dynamic representation of the covariance features. We then apply either entry-wise thresholding to the estimated dynamic covariance matrix or block-wise thresholding to the sample covariance matrix of estimated principal component scores for achieving uniform consistency results. Our theoretical results demonstrate the non-asymptotic error rates and support recovery properties of DFCM under both thresholding strategies even in large p small n scenario. Finally, we illustrate the sample performance of our proposed DFCM through a series of simulations and one real world data set.

EO0413: **Functional archetype and archetypoid analysis**
Presenter: **Irene Epifanio**, Universitat Jaume I, Spain
Archetype and archetypoid analysis can be extended to functional data. Each function is approximated by a convex combination of actual observations (functional archetypoids) or functional archetypes, which are a convex combination of observations in the data set. Well-known Canadian temperature data are used to illustrate the analysis developed. Computational methods are proposed for performing these analyses, based on the coefficients of a basis. Unlike a previous attempt to compute functional archetypes, which was only valid for an orthogonal basis, the proposed methodology can be used for any basis. It is computationally less demanding than the simple approach of discretizing the functions. Multivariate functional archetype and archetypoid analysis are also introduced and applied in an interesting problem about the study of human development around the world over the last 50 years. These tools can contribute to the understanding of a functional data set, as in the classical multivariate case.

EO0720: **Confidence regions for functional parameter estimates**
Presenter: **Matthew Reimherr**, Pennsylvania State University, United States
Co-authors: Hyunphil Choi
Functional data analysis, FDA, is a branch of statistics concerned with inference from samples of functions or curves. It is now a well-established discipline of statistics, with its core concepts and perspectives in place. Despite this, there are still fundamental statistical questions which have received relatively little attention. One of these is the systematic development of techniques for constructing confidence regions for functional parameters. New work will be presented concerned with developing, understanding, and visualizing such regions. Simulations will be presented as well as an application to Multiple Sclerosis and its connection to degeneration of the Corpus Callosum, a white matter tract connecting the two hemispheres of the brain.

EO1118: **Multivariate functional principal component analysis for data observed on different (dimensional) domains**
Presenter: **Sonja Greven**, LMU Munich, Germany
Co-authors: Clara Happ
Existing approaches for multivariate functional principal component analysis are restricted to data on a single interval. The presented approach focuses on multivariate functional data on different domains that may differ in dimension, e.g. functions and images. The theoretical basis for multivariate functional principal component analysis is given in terms of a Karhunen-Loeve Theorem. For the practically relevant case of a finite Karhunen-Loeve representation, a relationship between univariate and multivariate functional principal component analysis is established. This offers an estimation strategy to calculate multivariate functional principal components and scores based on their univariate counterparts. For the resulting estimators, asymptotic results are derived. The approach can be extended to finite expansions in general, not necessarily orthonormal bases. It is also applicable for sparse data or data with measurement error. The new method is shown to be competitive to existing approaches for data observed on a common one-dimensional domain. The motivating application is a neuroimaging study, where the goal is to explore how longitudinal trajectories of a neuropsychological test score covary with FDG-PET brain scans at baseline. An implementation is available in R packages fundata and MFPCA.

EO611 Room 217 STATISTICS IN SPORTS Chair: Marica Manisera

EO1081: **Goals in soccer: Factors that matter**
Presenter: **Jan Vecer**, Charles University, Czech Republic
The state-of-the-art statistical analysis for studying extensive soccer data that are recently becoming available is reviewed. Our data were obtained from OPTA in the case of the English Premier League and from the official websites of the English Premier League, the German Bundesliga and the World Cup 2014. The aim is to identify factors that play a major role in the goal creation. The major factors fall in 3 broad categories, namely the precision, the speed, and the discipline. The most interesting factors that influence the scoring are team specific and fall in the precision category. Some game situations depend partly (long balls, corners played into the box) or almost entirely (open play crossing) on luck and thus, the weaker teams are much better off if they use them in comparison to the stronger teams. As it turns out, the better teams are more efficient in scoring when they apply their skill rather than use game situations that depend on luck. In terms of the speed, the data suggest that the top speed (achieved in sprints) matters most in explaining the goal creation. The distance run while in sprint has a positive effect, but the distance run while in the fast play has a negative effect. Discipline (yellow or red cards) negatively impacts goals scored by the penalized team, but other various stoppages that fragment the game such as fouls, throw ins, corners or substitutions reduce goals of both teams.

EO0952: **A probability model for analyzing the points of a team in a football match: An application to the Spanish league**
Presenter: **Jose Maria Perez Sanchez**, Universidad de Las Palmas de Gran Canaria, Spain
Co-authors: Nancy Davila Cardenes, Emilio Gomez Deniz
In the last decades, many research papers applying statistical methods for analyzing sports data have been published. One of the most popular sport in these days is football, also called soccer, which is played in many countries in organized competition leagues. The aim is to develop a suitable probability model for studying the points achieved by a team in a football match, say a random variable X. For this purpose, we use weighted distributions to build a discrete probability distribution with truncated support in the set $\{0,1,3\}$. That is, if a discrete distribution $f(x/z)$ belongs to the exponential family of distributions, we build the new model by choosing $w(x)f(x/z)$, where $w(x)$ is a function of x. We also propose a regression model for studying the factors, which can explain the response variable X, then the points achieved by a team in a football match. We test its performance using data from the Spanish Football League during the session 2013-14.

EO0463: **Forecasting match results in European football competitions: New dynamic models and comparisons**
Presenter: **Rutger Lit**, VU Amsterdam, Netherlands
Co-authors: Siem Jan Koopman
A new dynamic multivariate model is developed for the analysis and the forecasting of football match results in national league competitions. The proposed dynamic model is based on the score of the predictive observation density for a large and weekly panel of match results. Given that our main interest is to forecast whether the match result is a win, a loss or a draw for each team, we investigate whether we should model the match result as a pairwise observation, or as a difference between the number of goals, or direct as a category (win, loss, draw). These different variables require different model specifications. Our interest focuses on which approach yields better forecast results. We show that our dynamic modelling framework outperforms relevant benchmark models in a large-scale forecasting study where we consider six football competitions in Europe and a forecasting period of seven years. The magnitude of our forecasting study allows us to draw strong conclusions with respect to the forecast performances of the considered models.

EO1116: **A new Markovian model for tennis matches**
Presenter: **Marco Ferrante**, Universita di Padova, Italy
Co-authors: Giovanni Fonseca
A generalisation of previously considered Markovian models for Tennis is presented that overcome the assumption that the points played are iid. Indeed, we postulate that in any game there are two different situations: the first 6 points and the, possible, additional points after the first deuce, with different winning probabilities. We are able to compute the winning probabilities and the expected number of points played to complete a game and a set in this more general setting and we test our results considering 57 matches between Novak Djokovic, Roger Federer and Rafael Nadal.

EO0874: **Big data analytics to model scoring probability in basketball: The effect of shooting under high-pressure conditions**
Presenter: **Paola Zuccolotto**, University of Brescia, Italy
Co-authors: Marica Manisera, Davide Verzeletti
In the last decades the idea of a statistical thinking in sports has gained a rapidly growing interest, as documented by the wide scientific production on this theme and also by the publication of some insightful collections of statistical analyses applied to data from a wide range of sports, including American football, baseball, basketball, and ice hockey. As concerns basketball, several statistical techniques have been applied to analyze data with a great variety of aims, ranging from simply depicting the main features of a game by means of descriptive statistics to the investigation of more complex problems, such as forecasting the outcomes of a game or a tournament, analysing players performance, or identifying optimal game strategies. Nevertheless, new research questions arise from the huge amount of large sets of play-by-play data available, combined with the absence of a sound theory explaining the relationships among the involved variables and the increasing computational power allowing the developing of data mining tools for big data treatment. We use data mining tools to analyze shooting performance under pressure or in stressful game situations, using two play-by-play datasets from the Italian Serie A2 Championship 2015/2016 and the Olympic Basketball Tournament Rio 2016.

EO253 Room 002 RECENT ADVANCES IN THE ANALYSIS OF LONG MEMORY TIME SERIES Chair: Rebecca Killick

EO1136: **Long range dependence, fractional renewal models and Bayesian inference**
Presenter: **Nicholas Watkins**, London School of Economics and Political Science, United Kingdom
Co-authors: Christian Franzke
Since the 1960s, long range dependence (LRD) as embodied by the fractional Gaussian noise and ARFIMA models, has been a well-studied mechanism for the origin of $1/f$ noise and the Hurst effect. Two new avenues of research will be discussed. The first concerns breakpoints. These have long been known to be a source of the Hurst effect, but recent research by one of us has shown that Mandelbrot had proposed a model with power law intervals between the breaks as early as 1963, and that by 1965-67 he was showing how this was an alternative non-ergodic model for $1/f$ noise, with consequences for model choice and time series interpretation that are increasingly becoming topical in physics and elsewhere. The second concerns Bayesian inference when an LRD model is plausible. We will discuss our recent work on a novel systematic Bayesian approach for joint inference of the memory and tail parameters in an ARFIMA model with heavy-tailed innovations.

EO0215: **Multivariate spurious long memory and a robust local Whittle estimator**
Presenter: **Christian Leschinski**, Leibniz University Hannover, Germany
Co-authors: Philipp Sibbertsen
For univariate time series it is well documented that low frequency contaminations generate spurious long memory. This analysis is extended to vector valued processes. A multivariate generalization of the random level shift process is introduced and its properties are derived. These results show, that spurious long memory is an issue for vector series as it is for univariate series. Therefore, a robust multivariate local Whittle (RMLW) estimator is derived that is robust to low frequency contaminations and asymptotically normal. A Monte Carlo study shows that the RMLW estimator has good finite sample properties. The standard Gaussian semiparametric estimator, on the other hand, is severely biased.

EO0645: **Long memory and changepoint models: A spectral classication procedure**
Presenter: **Ben Norwood**, Lancaster University, United Kingdom
Co-authors: Rebecca Killick
Time series within fields such as Finance and Economics are often modelled using long memory models. Alternative studies suggest that series may actually contain a 'changepoint' (a point within the time series where the data generating process has changed). These models have been shown to have elements of similarity, such as within their spectrum. Without prior knowledge this leads to an ambiguity between these two models, giving difficulty to assessing which model is most appropriate. This problem of appropriate identification is what we seek to address using a wavelet based classification approach. Simulation results are presented across a number of models followed by an application to US Price Inflation. The results indicate that the proposed classification outperforms an existing hypothesis testing approach on a number of models and performs well across others.

EO0737: **Strict stationarity, persistence and volatility forecasting in ARCH(∞) processes**
Presenter: **Xiaoyu Li**, University of Exeter, United Kingdom
Co-authors: James Davidson
A simple sufficient condition is derived for strict stationarity in the ARCH(∞) class of processes with conditional heteroscedasticity. The concept of persistence in these processes is explored, and is the subject of a set of simulations showing how persistence depends on both the pattern of lag coefficients of the ARCH model and the distribution of the driving shocks. The results are used to argue that an alternative to the usual method of ARCH/GARCH volatility forecasting should be considered.

EO0934: **Testing for neglected strong dependence in explosive models**
Presenter: **Christoph Wegener**, IPAG Business School, Germany
Co-authors: Robinson Kruse
A widely applied procedure has been developed recently to test against explosive behavior in time series econometrics. There is fast-growing empirical literature, which applies the test to detect bubbles in real estate, commodity, and stock prices. However, there is a gap in the literature concerning strong dependent innovation processes in this framework. We consider such a test under strong dependencies and find severe size distortions in case the residuals of the autoregressive model exhibit long memory. The unit root hypothesis is rejected far too often even for a mild form of strong dependence. We use the limit theory for mildly explosive models to uncouple the consistent estimation of the autoregressive coefficient and the degree of persistence of the residuals. Thus, we are able to test the unit root hypothesis against an explosive alternative in the presence of long-range dependencies. We are particularly interested in small sample properties of all methods in this context, because explosive behavior seems to be a temporal phenomenon. In order to apply this method to detect explosive behavior, we suggest a modification of a previous procedure to test against strong dependent innovations.

EO007 Room 201 MODEL SELECTION IN HIGH DIMENSIONS Chair: Malgorzata Bogdan

EO1219: **Multiple testing and adaptive estimation via the sorted L-one norm**
Presenter: **Weijie Su**, The Wharton School, University of Pennsylvania, United States
Co-authors: Malgorzata Bogdan, Ewout van den Berg, Chiara Sabatti, Emmanuel Candes
In many real-world statistical problems, we observe a large number of potentially explanatory variables of which a majority may be irrelevant. For this type of problem, controlling the false discovery rate (FDR) guarantees that most of the discoveries are truly explanatory and thus replicable. We propose a new method named SLOPE to control the FDR in sparse high-dimensional linear regression. This computationally efficient procedure works by regularizing the fitted coefficients according to their ranks: the higher the rank, the larger the penalty. Whenever the columns of the design matrix are not strongly correlated, we show empirically that SLOPE obtains FDR control at a reasonable level while offering substantial power. Although SLOPE is developed from a multiple testing viewpoint, we show the surprising result that it achieves optimal squared errors under Gaussian random designs over a wide range of sparsity classes. An appealing feature is that SLOPE does not require any knowledge of the degree of sparsity. This adaptivity to unknown sparsity has to do with the FDR control, which strikes the right balance between bias and variance. The proof of this result presents several elements not found in the high-dimensional statistics literature.

EO0777: **Some optimality properties of FDR controlling modifications of AIC and BIC in high dimensions**
Presenter: **Florian Frommlet**, Medical University Vienna, Austria
Co-authors: Malgorzata Bogdan
Penalized selection criteria like AIC or BIC are among the most popular methods for variable selection. Their theoretical properties have been studied intensively and are well understood in case of a moderate number of variables. However, these criteria do not work well in a high-dimensional setting under the assumption of sparsity. We will introduce different modifications of AIC and BIC which will allow to control the family wise error rate (mAIC, mBIC) or the false discovery rate (mAIC2, mBIC2), respectively, in terms of including false positive regressors in the model. Our modifications of BIC have a nominal level which is roughly indirect proportional to the square root of the sample size, whereas the

modifications of AIC have a fixed nominal level. We will discuss for which purpose one might prefer mAIC2 or mBIC2 and give the conditions under which they provide selection procedures which are asymptotically Bayes optimal under sparsity (ABOS). Finally we want to compare their performance in the context of GWAS data.

EO0397: **Model-free high-dimensional false discovery rate control with knockoffs**
Presenter: **Lucas Janson**, Stanford University, United States
A common problem in modern statistical applications is to select, from a large set of candidates, a subset of variables which are important for determining an outcome of interest. For instance, the outcome may be disease status and the variables may be hundreds of thousands of single nucleotide polymorphisms on the genome. For data coming from low-dimensional ($n \geq p$) linear homoscedastic models, the knockoff procedure recently introduced by Barber and Candès solves the problem by performing variable selection while controlling the false discovery rate (FDR). The present paper extends the knockoff framework to arbitrary (and unknown) conditional models and any dimensions, including $n < p$, allowing it to solve a much broader array of problems. This extension requires the design matrix be random (independent and identically distributed rows) with a covariate distribution that is known, although we show our procedure to be robust to unknown/estimated distributions. To our knowledge, no other procedure solves the variable selection problem in such generality, but in the restricted settings where competitors exist, we demonstrate the superior power of knockoffs through simulations. Finally, we apply our procedure to data from a case-control study of Crohns disease in the United Kingdom, making twice as many discoveries as the original analysis of the same data.

EO0640: **A scalable and consistent variable selection method for high dimensional logistic regression**
Presenter: **Naveen Narisetty**, University of Illinois at Urbana-Champaign, United States
Within the framework of Bayesian computation, we provide a novel variable selection method for logistic regression that adapts to both the sample size n and the number of potential covariates p with desirable features. We propose a Gibbs sampler called "Skinny Gibbs" whose computational complexity grows only linearly in p, but it attains the property of strong model selection consistency. In contrast with the standard Gibbs sampler, Skinny Gibbs is much more scalable to high dimensional problems, both in memory and in computational feasibility. We compare our proposed method with several leading variable selection methods through a simulation study to show that our proposed approach selects the correct model with higher probabilities than existing methods while being computationally appealing.

EO1240: **Bayesian dimensionality reduction with PCA using penalized semi-integrated likelihood**
Presenter: **Piotr Sobczyk**, Politechnika Wroclawska, Poland
Co-authors: Malgorzata Bogdan, Julie Josse
Despite of the importance of the problem of estimating the number of principal components in Principal Components Analysis (PCA), it comes as a surprise that there does not exist a coherent asymptotic framework which would justify different approaches depending on the actual size of the data set. We address this issue by presenting an approximate Bayesian approach based on Laplace approximation and introducing a general method for building the model selection criteria, called PEnalized SEmi-integrated Likelihood (PESEL). Our general framework encompasses a variety of existing approaches based on probabilistic models, like e.g. Bayesian Information Criterion for the Probabilistic PCA (PPCA), and allows for construction of new criteria, depending on the size of the data set at hand and additional prior information. Specifically, we apply PESEL to derive two new criteria for data sets where the number of variables substantially exceeds the number of observations, which is out of the scope of currently existing approaches. We also report results of extensive simulation studies and real data analysis, which illustrate good properties of our proposed criteria as compared to the state-of-the-art methods and very recent proposals. Specifically, these simulations show that PESEL based criteria can be quite robust against deviations from the probabilistic model assumptions. Selected PESEL criteria are implemented in R package varclust (https://github.com/psobczyk/varclust).

EP733 Room Hall POSTER SESSION II Chair: Marta Garcia Barzana

EP1469: **A bias-corrected estimator of competitive balance in sports leagues**
Presenter: **Sarah Kim**, Seoul National University, Korea, South
Co-authors: Younghoon Lee, Yongdai Kim
The ratio of the actual standard deviation to the idealized standard deviation of win percentages (RSD) is the conventional measure of competitive balance (CB). RSD is designed to control for the effect of season length on the sample standard deviation of win percentages (ASD). Theoretically, the RSD should be greater than one, but empirical values below one have been found in previous studies. We employ a mathematical statistics approach to evaluate the statistical properties of RSD and ASD. In doing so, we find that RSD is constructed by an invalid normalization approach and that ASD is biased. We also present a bias-corrected standard deviation (BCSD) as a new estimator of the standard deviation of true win probabilities. Results from the simulations confirm the following: (1) ASD is prone to underestimate CB levels when the number of games is small, (2) the RSD values become unreasonably large when the number of games is large, and (3) BCSD performs well with respect to mean bias and root mean squared errors. According to empirical analysis on the English Premier League (EPL) and the Korea Baseball Organization (KBO), BCSD shows that the KBO was more competitively-balanced than the EPL between 2000-2015, but the RSD implies that the two leagues were more or less equal.

EP1618: **A comparison of spatial Bayesian, empirical Bayesian and frequentist interpolators with large data**
Presenter: **Konstantin Krivoruchko**, Esri, United States
Several methods of spatial prediction recently developed for interpolation of large spatial datasets are compared, for example, where measurements numbers are several thousands, and one or more model parameters change in spatial extent. Specifically, we compare modified predictive processes, a stochastic partial differential equation approach, lattice kriging, fixed rank kriging, and empirical Bayesian kriging (EBK). The first four methods are available in R, while EBK is implemented in Esri ArcGIS Geostatistical Analyst. A large number of simulated and real data were used. The performance of each spatial interpolator is evaluated empirically using the validation datasets for each simulation experiment. Prediction quality is compared using prediction errors, continuous ranked probability score, the number of validation data inside the 90- and 95 percent prediction intervals, and the ability to predict the largest and the smallest observation values. Results show that empirical Bayesian kriging outperforms all other predictors, and increasingly so with larger data complexity. Also, prediction standard error maps produced by models implemented in R are largely functions of the observed data locations density, while the prediction standard error map created by EBK is a function of both points density and the data values.

EP1452: **A new method of measuring uncertainty propagation in the study of climate change on hydrology**
Presenter: **Gyuseung Baek**, Seoul National University, Korea, South
Co-authors: Yongdai Kim, Ilsang Ohn
Uncertainty in climate change impact assessments is composed of several stages consisting of emission scenarios, global climate models, downscaling techniques and hydrological models. To investigate the total uncertainty and the uncertainty propagation process, it is necessary to quantify uncertainty in a comprehensive manner. But existing methods that measure the uncertainty of each source separately are inappropriate to study uncertainty propagation, since they do not guarantee proper uncertainty propagation. We propose a new way of quantifying and decomposing the

total uncertainty by defining sequential uncertainty. The sequential uncertainty is defined as the difference of the uncertainty up to a given source and the uncertainty up to the stage just before a given source, which represents how much a given source propagates the total uncertainty in the total process of projection. We show that the sequential uncertainty provides proper uncertainty propagation, which implies that the uncertainty keeps increasing as the stage proceeds.

EP1505: **The multinomial logistic regression model to predict eye color: An application to the Portuguese case**
Presenter: **Helena Mourino**, Faculdade de Ciencias - Universidade de Lisboa, Portugal
Co-authors: Paulo Dario

DNA phenotyping research is one of the most emergent areas of forensic genetics. IrisPlex, an eye color prediction assay, revealed high prediction rates in European populations. However, this is less predictive in some non-European populations, maybe due to admixing. When compared to other European countries, Portugal has a relatively admixed population. The aim is to evaluate the utility of IrisPlex in the Portuguese population. Buccal swabs and blood samples were collected from 192 unrelated volunteers residing in Portugal, and under informed consent. This sample consisted of 95 males and 97 females with ages ranging from 16 to 72. Iris characteristics of each individual was classified into one of three categories: blue eyes; neither blue nor brown; brown eyes. We propose a Multinomial Logistic Regression Model to predict eye color, where the response variable is the Eye (iris) color. The covariates correspond to the classification of individuals based on population minor allele frequencies (obtained from the IrisPlex). The model is estimated by the maximum likelihood method: the BFGS algorithm. For prediction purposes, we develop accurate techniques to compute the AUC in the multinomial framework. The performance of the estimated model is also evaluated in the context of cross-validation.

EP0785: **Partially reduced-bias value-at-risk estimation**
Presenter: **Dinis Pestana**, FFCUL, Universidade de Lisboa e CEAUL, Portugal
Co-authors: Ivette Gomes, Frederico Caeiro, Fernanda Otilia Figueiredo, Ligia Henriques-Rodrigues

The value-at-risk (VaR) at a small level q, $0 < q << 1$, is the size of the loss that occurs with a probability q. Semi-parametric partially reduced-bias (PRB) VaR-estimation procedures based on the mean-of-order-p of a set of k quotients of upper order statistics, with p any real number, are put forward. After a reference to their asymptotic behaviour, these PRB VaR-estimators are altogether compared with the classical ones for finite samples, through a large-scale Monte-Carlo simulation study. A brief application to financial log-returns is also provided.

EP1455: **Semi-parametric logistic regression model with random forest**
Presenter: **Bo Young Kim**, Seoul National University, Korea, South
Co-authors: Yongdai Kim, Jaesung Hwang, Oh-Ran Kwon

In two-class classification problem, generalized linear model (glm) with logistic link function which estimates the probability of a binary response based on one or more independent variables is the most commonly accepted model. But glm has the limitation of classification performance in terms of sensitivity and specificity. So we propose the semi-parametric model called glm+rf model, which combines glm and random forest to improve its accuracy results. Since it is not easy to estimate the glm part and random forest part simultaneously, we adopt two stage method. In the first stage, we estimate the coefficients of the glm part with general logistic regression assuming that there is not random forest part. The second step is to estimate the random forest part suggesting two algorithms. One is to perform like the random forest does, which maximize the log likelihood in the left and right node. The other is to use negative gradient. Additionally, we select the bias correction term to reduce the bias of the random forest predictors. Throughout the simulation, we compare the prediction results of the new model using negative gradient approach, random forest approach and the existing glm model.

EP1465: **Comparison of sub-daily precipitation extremes in observed data and regional climate model simulations**
Presenter: **Jan Kysely**, Institute of Atmospheric Physics AS CR, Czech Republic

The study compares characteristics of observed sub-daily precipitation extremes with those simulated by HadRM3 and RCA4 regional climate models (RCMs) driven by reanalyses. The observed warm-season (May-September) maxima of short-duration (1 h, 2 h and 3 h) amounts show one diurnal peak in the afternoon, which is simulated reasonably well by RCA4, although the peak occurs too early in the model. HadRM3 provides an unrealistic diurnal cycle with a night-time peak and an afternoon minimum coinciding with the observed maximum for all three ensemble members, which suggests that convection is not captured realistically. Distorted relationships of the diurnal cycles of hourly precipitation to daily maximum temperature in HadRM3 further evidence that underlying physical mechanisms are misrepresented in this RCM. Goodness-of-fit tests indicate that the Generalized Extreme Value (GEV) distribution is an applicable model for both observed and RCM-simulated precipitation maxima. However, the RCMs are not able to capture the range of the shape parameter estimates of distributions of short-duration precipitation maxima realistically, leading to either too many or too few grid boxes in which the shape parameter corresponds to a heavy tail. This means the distributions of maxima of sub-daily amounts are distorted in the RCM-simulated data and do not match reality well.

EP1531: **Computing the distribution of collective risk models via numerical inversion the characteristic function**
Presenter: **Gejza Wimmer**, Mathematical Institute, Slovak Academy of Sciences, Slovakia
Co-authors: Viktor Witkovsky

A typical model for insurance risk is the collective risk model (CRM). The collective risk model mathematically describes the aggregate loss of an insurance portfolio in a certain period of time (e.g. 1 year). Insurance portfolio is regarded as a process that produces claims over time. The sizes of these claims are taken to be independent, identically distributed random variables independent also of the number of claims generated in this time period. We present the typical parametric collective risk models and their characteristic functions (CFs), and introduce a new MATLAB Toolbox called CRM for high precision calculation of probability density functions (PDF) and cumulative distribution functions (CDF) of the CRM distributions. Method of calculation is based on the numerical inversion of CRM's characteristic function. The suggested numerical approaches are based on the Gil-Pelaez inversion formulae for computing the probability distribution functions (PDF and/or CDF) of the univariate continuous random variables. Moreover, a non-parametric method based on inverting the empirical characteristic functions is presented and illustrated.

EC1508: **Significance of explanatory variables when model is estimated by S-weighted estimator**
Presenter: **Jan Amos Visek**, Charles University in Prague, Czech Republic

The S-weighted estimator is a generalization of the Least Weighted Squares as well as of S-estimator. It inherited all pros and removed restriction of LWS on the quadratic function. It depressed a potential high sensitivity to inliers and restriction on bounded objective function of S-estimator. The estimator is the argument which minimizes the scale of residuals under a constraint on the sum of products of weights and of order statistics of the squared residuals plugged into the objective function. So, it reaches a high flexibility of estimator and it can be tailored to the character and level of contamination of data. The consistency of new estimator was proved and the reliability of algorithm was demonstrated. Significance of the individual explanatory variable is a key diagnostic tool. It is surprising that there are nearly no results on this topics for the robust identification of model. Presumably it is due to the fact that - even if we start with i.i.d. framework - robustification transforms this framework to the framework with heteroscedasticity. The denominator of t-statistics is so the sum of independent random variables, each of them is the square of normal r.v.- but the r.v.'s have different variances. We obtain something like generalized chi-square-distribution and consequently generalized t-distribution. The distribution of generalized t-statistic has to be simulated.

EP1769: **The lattice generalized Von Mises distribution**

Presenter: **Anna Sjostrom**, Department of Statistics, Lund University, Sweden

The Lattice Generalized von Mises distribution is presented, which allows for modelling multimodal discrete circular data, where no further restrictions are put on data. In the case of circular data, a wide range of continuous distributions are described in the literature, whereas the discrete distributions presented are fewer. Nevertheless, the need for discrete circular distributions is not insignificant, since circular data are often observed on discrete outcome spaces, for example in the form of temporal or spatial data. For continuous circular data, the Generalized von Mises distribution (GvM) allows for modelling higher orders of modality. Due to this fact, the GvM is a natural starting point in the process of modelling unrestricted discrete circular multimodal data, leading to what we call the Lattice Generalized von Mises distribution. This is a flexible probability model which is able to handle multimodality and asymmetry between modes. The pdf of the continuous GvM typically includes types of modality in a sequential manner. In the proposed model, we allow for including some but not necessarily all types of modality of a sequence. We demonstrate the model for a data set concerning bird migration behavior, where data are measured as counts on circle sectors. We show an approximation to the continuous case when the grid of circle sectors becomes finer. Further, we describe an EM-algorithm for finding ML-estimators of the model parameters.

EP1770: **Recovery of Fisher's linear discriminant subspace by invariant coordinate selection (ICS) methods: Latest advances**

Presenter: **Radka Sabolova**, The Open University, United Kingdom

Co-authors: Hannu Oja, Germain Van Bever, Frank Critchley

It is a remarkable fact that, using any pair of scatter matrices, invariant coordinate selection can recover the Fisher linear discriminant subspace and its dimension without knowing group membership. The aim is to summarise and extend our body of work in this area. For any choice of scatter matrices satisfying mild assumptions, we have (a) the limiting distribution of the estimator of the signal space and (b) both asymptotic and bootstrap tests for its dimension. This latter test is shown to be useful for real data analysis. The finite sample adequacy of asymptotic results is examined with reference to a novel family of scatter matrices and reported in a designed simulation study.

Saturday 10.12.2016 14:25 - 16:05 Parallel Session I – CFE-CMStatistics

CO654 Room 104 FORECASTING AND VOLATILITY: R&M UNIVERSITY OF MESSINA PROJECT Chair: Edoardo Otranto

CO0263: **A multiplicative dynamic model for realized covariance matrices**
Presenter: **Luc Bauwens**, Universite catholique de Louvain, Belgium
Co-authors: Manuela Braione, Giuseppe Storti

A class of multiplicative dynamic models for realized covariance matrices is introduced. The multiplicative structure enables consistent three-step estimation of the parameters, starting by covariance targeting of a scale matrix. The dynamics of conditional variances and correlations are inspired by specifications akin to the consistent dynamic conditional correlation model of the multivariate GARCH literature, and estimation is performed by the quasi-maximum likelihood method, assuming a Wishart conditional distribution. Simulations show that in finite samples a three-step estimator has smaller bias and root mean squared error than a one-step estimator when the cross-sectional dimension increases. An empirical application illustrates the flexibility of the model in a low-dimensional setting, and another one illustrates its effectiveness and practical usefulness in high dimensional portfolio allocation strategies.

CO0487: **Estimation of a multiplicative covariance structure**
Presenter: **Christian Hafner**, UCL Louvain-la-Neuve, Belgium
Co-authors: Oliver Linton, Haihan Tang

A Kronecker product structure for large covariance matrices is considered, which has the feature that the number of free parameters increases logarithmically with the dimensions of the matrix. We propose an estimation method of the free parameters based on the log linear property of these structures, and also a Quasi-Likelihood method. We establish the rate of convergence of the estimated parameters when the size of the matrix diverges. We also establish a CLT for our method. We apply the method to portfolio choice for S&P500 daily returns and compare with alternative methods.

CO1019: **Estimating the frequency of a time series**
Presenter: **Walter Distaso**, Imperial College London, United Kingdom
Co-authors: Liudas Giraitis, Karim Abadir

A novel methodology is presented that allows us to estimate both the long memory parameter of a (possibly nonstationary) time series and the frequency which maximizes the spectral density. We derive limiting distributions, assess them through a simulation exercise and also provide a valid bootstrap scheme that uses both time domain and frequency domain information. We then illustrate the usefulness of our model in capturing the salient features of dynamics of different datasets.

CO1245: **More than just errors: Accounting for state dependent mean reversion in volatility forecasting**
Presenter: **Giampiero Gallo**, University of Florence, Italy
Co-authors: Edoardo Otranto, Fabrizio Cipollini

Recent results are revisited, where an improvement of forecasting realized volatility is attained by inserting an estimate of quarticity in a volatility model. In our approach, quarticity is recognized to be strongly correlated with volatility itself, therefore introducing a curvature effect in a volatility quadratic equation, whereby higher volatility induces a stronger mean reversion effect. Moreover, this state dependence can be linked to a low frequency component of volatility forecasting which is suitably captured by Smooth Transition of Markov Switching behavior. Our results on 30 DJ components reveal that when other forms of state dependent behavior in volatility modeling are accounted for, the quarticity term generally loses importance, showing that measurement errors are providing only incomplete explanation for this state dependent behavior.

CO425 Room 216 NONPARAMETRIC METHODS FOR NONSTATIONARY TIME SERIES Chair: Christopher Walsh

CO0270: **Empirical characteristic function-based inference for locally stationary processes**
Presenter: **Marco Meyer**, TU Braunschweig, Germany
Co-authors: Carsten Jentsch, Anne Leucht, Carina Beering

A kernel-type estimator is proposed for the local characteristic function (local CF) of locally stationary processes. Under weak moment conditions, we prove joint asymptotic normality for local empirical characteristic functions (local ECF). Precisely, for processes having a (two-sided) time-varying MA(∞) representation, we establish a central limit theorem under the assumption of finite absolute first moments of the process. Additionally, we prove process convergence of the local ECF. We apply our asymptotic results to parameter estimation of time-varying α-stable distributions. Furthermore, we extend the notion of distance correlation to locally stationary processes and provide asymptotic theory for local empirical distance correlations. Finally, we provide a simulation study on minimum distance estimation for α-stable distributions based on local ECF and illustrate the pairwise dependence structure over time of log returns of German stock prices via local empirical distance correlations.

CO1241: **Partially linear time series models with time-varying coefficients**
Presenter: **Lionel Truquet**, ENSAI, France

Partially linear models have been widely studied in Econometrics and Statistics. Their semiparametric nature is general enough for capturing complex nonlinearities and meanwhile it avoids the curse of dimensionality. In linear regressions with time-varying coefficients, partially linear models appear naturally when testing time-constancy of some of the parameters, or for the semiparametric inference in such models. We will discuss how to conduct inference in this case, using in particular Robinsons type estimates. In a second part, we will discuss an extension of partially linear models using time-varying coefficients as an interesting approach for avoiding the curse of dimensionality in nonparametric, time-varying regression models.

CO1092: **Analysing dynamic interactions in limit order book markets using nonparametric methods**
Presenter: **Christopher Walsh**, University of Vienna, Austria
Co-authors: Nikolaus Hautsch

A novel theoretical explanation has been recently provided for price impact in limit order book markets. The corresponding empirical analysis was conducted using static parametric regression models on data from the LOBSTER database, which contains high frequency limit order book data from NASDAQ. We go beyond the simple justification for the parametric forms via the visual comparison of model fits to some nonparametric fits. Specifically, we will construct a specification test for the proposed models and generalize the models by allowing for the presence of dynamics.

CO1762: **Determination of vector error correction models in higher dimensions**
Presenter: **Melanie Schienle**, Karlsruhe Institute of Technology, Germany
Co-authors: Chong Liang

A shrinkage type methodology is provided which allows for simultaneous model selection and estimation of vector error correction models (VECM). Model determination is treated as a joint selection problem of cointegrating rank and autoregressive lags. We show consistency of

the selection mechanism by the resulting Lasso-VECM estimator under sparsity in lags and cointegration relations. In contrast to existing two-step approaches based on information criteria, we also derive the asymptotic properties of the final estimator and point to estimation refinements. Moreover, with only linear computational complexity, the procedure remains computationally tractable also for higher dimensions. We demonstrate the effectiveness of the proposed approach by a simulation study and an empirical application to recent CDS data after the financial crisis.

CO473 Room 101 ALPHA-STABLE PROCESSES WITH APPLICATIONS IN FINANCIAL ECONOMETRICS Chair: Nourddine Azzaoui

CO0319: **Simulation of a general class of α-stable processes**
Presenter: **Malcolm Egan**, Universite Blaise Pascal, France
Co-authors: Nourddine Azzaoui, Gareth Peters

The heavy-tail and extremal dependence properties of α-stable processes have lead to their extensive use in fields ranging from finance to engineering. In these fields, the stochastic integral representation plays an important role both in characterizing α-stable processes as well as for the purposes of simulation and parameter estimation. In order use the stochastic integral representation, constraints on the random measure must be imposed. A key constraint is the independently scattered condition, where orthogonal increments of the random measure are independent. A key feature of the independently scattered condition is that the covariation is both left and right additive, which allows for simulation and estimation of this class of processes. Recently, a new generalization of the independently scattered condition has been introduced, which also preserves the left and right additivity of the covariation. This new generalization allows the characteristic function a wide class of α-stable processes to be determined by a bimeasure. We deal with the problem of simulating from the bimeasure characterization of α-stable processes. In particular, we prove conditions under which the bimeasure leads to a positive definite characteristic function for the case of a two-dimensional skeleton. Based on this result, we then propose a method to construct and simulate $n-$dimensional skeletons, for arbitrary $n > 2$.

CO0363: **Stable-GARCH models for financial returns: Fast estimation and tests for stability**
Presenter: **Marc Paolella**, University of Zurich, Switzerland

A fast method for estimating the parameters of a stable-APARCH not requiring likelihood or iteration is proposed. Several powerful tests for the (asymmetric) stable Paretian distribution with tail index $1 < \alpha < 2$ are used for assessing the appropriateness of the stable assumption as the innovations process in stable-GARCH-type models for daily stock returns. Overall, there is strong evidence against the stable as the correct innovations assumption for all stocks and time periods, though for many stocks and windows of data, the stable hypothesis is not rejected.

CO1329: **Spatial and temporal interpolation based on impulsive alpha stable processes.**
Presenter: **Nourddine Azzaoui**, Blaise Pascal University, France
Co-authors: Gareth Peters

Spatio-temporal modeling is a fundamental step for understanding the mechanisms that govern the evolution of natural phenomena and their underlying econometric and financial processes; It allows us to predict a studied variable at unobserved points of a given field based on measurements of this variable at some points of the domain. Many results and solutions have been found for second order stochastic processes, e.g. spatial kriging and spatio-temporal interpolation based on Guassian processes. We sketch a generalisation of these techniques to the case where the Gaussian spatial stochastic processes is replaced by a dependent heavy tailed alpha-stable processes. We show some application examples based on a simulated alpha-stable process.

CO1327: **New results in non-independent increment stable process specifications and characterizations**
Presenter: **Gareth Peters**, University College London, United Kingdom

New results will be presented on characterization of alpha-stable processes that generalize the representation and characterization of independent increment processes to non-independent increment processes. Ideas on characterization and construction will be the focus.

CO351 Room 102 REGIME SWITCHING, FILTERING, AND PORTFOLIO OPTIMIZATION Chair: Joern Sass

CO0330: **Expert opinions and utility maximization in a market with partially observable Gaussian drift**
Presenter: **Ralf Wunderlich**, BTU Cottbus-Senftenberg, Germany

A continuous-time financial market with partial information on the drift is considered, and utility maximization problems are solved, which include expert opinions on the unobservable drift. Stock returns are driven by a Brownian motion and the drift depends on a factor process which is an Ornstein Uhlenbeck process. Thus, the drift is hidden and has to be estimated from observable quantities. If the investor only observes stock prices then the best estimate is the Kalman filter. However, to improve the estimate, an investor may also rely on expert opinions providing a noisy estimate of the current state of the drift. This reduces the variance of the filter and thus improves the expected utility. That procedure can be seen as a continuous-time version of the classical Black-Litterman approach. For the associated portfolio problem with logarithmic utility, explicit solutions are available in the literature. We consider the case of power utility. We apply dynamic programming techniques and solve the corresponding dynamic programming equation for the value function. Diffusion approximations for high-frequency discrete-time experts allow to simplify the problem and to derive more explicit solutions. We illustrate our findings by numerical results.

CO0443: **Analyzing and testing arbitrage parities**
Presenter: **Julia Reynolds**, Vienna Graduate School of Finance, Austria
Co-authors: Leopold Soegner, Martin Wagner, Dominik Wied

New econometric tools are developed that allow for the monitoring of deviations from arbitrage parities. The econometric methodology is applied in order to test for the stability of the Covered Interest Rate Parity, the Triangular Arbitrage Parity, and the American Depository Receipt Parity. Economically relevant variables that possibly result in arbitrage parity violations are analyzed to connect our empirical results to the limits to arbitrage literature. In addition, we also check whether (possible) arbitrage parity deviations can be connected to historical events such as the failure of Lehman Bros. or to central bank interventions.

CO1091: **A joint model of firm failure and credit ratings**
Presenter: **Laura Vana**, WU Wirtschaftsuniversitaet Wien, Austria
Co-authors: Rainer Hirk, Kurt Hornik

Credit risk modeling including the measurement of credit quality has been intensively investigated by academics and practitioners over the past decades. We contribute by developing a framework for jointly modeling firm failures (e.g. bankruptcies) and ordinal credit ratings as outcomes. This model, unlike prior work, simultaneously incorporates failures and credit ratings and allows inference about the quantitative relationships between these outcomes by simultaneously making use of both sources of information. In addition the model does not require a balanced data set in the outcome variables, i.e. missing values in the outcomes are possible. We hypothesize that 1) the model alleviates the 'low failure portfolio' problem and outperforms univariate models of failure or credit rating models in terms of prediction accuracy; 2) useful insights into rating heterogeneity are gained; 3) information about the systematic rating patterns of the credit rating agencies is uncovered. The joint model will be formulated such that it takes the ordinal nature of the credit ratings into account and can therefore be incorporated in the class of multivariate

ordinal models. Failure, firm-level and stock price data for publicly traded North American companies as well as issuer credit ratings from the big three rating agencies (S& P, Moody's and Fitch) are collected and analyzed to illustrate the proposed framework.

CO0949: **Filter-based discrete-time EM algorithm with diffusion and point process observation**
Presenter: **Camilla Damian**, WU Vienna University of Economics and Business, Austria
Co-authors: Zehra Eksi-Altay, Ruediger Frey
The focus is on statistical inference in a dynamic, reduced-form, partial-information model for Eurozone sovereign credit spreads. The main assumption is that default intensities are driven by an unobservable finite-state Markov chain. Regarding methodology, an extension of the EM algorithm is involved: instead of pure diffusion information, both diffusive and point-process observations are considered. In the financial application, the point process represents default history of a given country. The goal is to estimate the model parameters, in particular the infinitesimal generator of the underlying Markov chain, making use of the robust discretization of continuous-time filters. Both a simulation analysis, which is essential to check performance, accuracy and stability of the algorithm, and an application to real data are presented.

CO291 Room 105 ONE-STEP-AHEAD IN DYNAMIC FACTOR MODELS Chair: Esther Ruiz

CO0384: **Identification and estimation of dynamic factor models**
Presenter: **Joerg Breitung**, University of Cologne, Germany
Co-authors: Matei Demetrescu
In empirical practice dynamic factor models are typically estimated by performing a Principal Component Analysis (PCA) on the static factor representation. A two-step PC estimator and a sequential least-squares approach is proposed to estimate the original dynamic factors that enter the model with a prespecified number of lags. The identification of the dynamic factors subject to the usual normalization restrictions is analyzed and the consistency of the estimator is established. Monte Carlo simulations suggest that the sequential least-squares estimator performs slightly better than the two-step PCA approach.

CO0826: **Markov-switching dynamic factor models in real time**
Presenter: **Gabriel Perez Quiros**, Bank of Spain, Spain
Co-authors: Maximo Camacho, Pilar Poncela
The Markov-switching dynamic factor model is extended to account for some of the specificities of the day-to-day monitoring of economic developments from macroeconomic indicators, such as ragged edges and mixed frequencies. We examine the theoretical benefits of this extension and corroborate the results through several Monte Carlo simulations. Finally, we assess its empirical reliability to compute real-time inferences of the US business cycle.

CO0976: **Dynamic functional principal components**
Presenter: **Marc Hallin**, Universite Libre de Bruxelles, Belgium
The problem of dimension reduction for functional time series is addressed. Such time series arise frequently, e.g. when a continuous time process is segmented into some smaller natural units, such as days, each observation representing one intraday curve. We argue that functional principal component analysis (FPCA), which is a key technique in the field, does not provide an adequate dimension reduction in a time series context. FPCA is a static procedure which ignores the essential serial dependence features of the data. Therefore, inspired by Brillingers theory of dynamic principal components, we propose a dynamic version of FPCA which is based on a frequency domain approach, and show that it provides the optimal dimension reduction. By means of a simulation study and an empirical illustration, we show the considerable improvement our method entails when compared to the usual (static) procedure.

CO1056: **Measuring the uncertainty of principal components in dynamic factor models**
Presenter: **Esther Ruiz**, Universidad Carlos III de Madrid, Spain
Co-authors: Irene Albarran, Javier de Vicente
Factors extracted from large macroeconomic are central for policy makers. Measuring the uncertainty associated with the estimated factors is a central issue for an adequate policy. One of the most popular factor extraction procedures is Principal Components which is nonparametric and computationally simple. The uncertainty associated with factors extracted using PC can be measured using their asymptotic distribution. However, the asymptotic distribution is a poor approximation to the finite sample distribution of the factors. The asymptotic intervals are too tiny. Alternatively, several authors propose using bootstrap methods in the context of PC factor extraction. We show that the available bootstrap methods are not adequate as they are based on bootstrapping the data as if it were iid, on bootstrapping from the marginal instead of the conditional distribution or on considering the estimated common component as if it were the true component. After a detailed analysis of the finite sample properties of the asymptotic and bootstrap factor prediction intervals, we propose a new bootstrap procedure with appropriate properties that mimic that of the original data. We show that the coverages of the new bootstrap intervals are close to the nominal regardless of the properties of the idiosyncratic noises as far as the factor is not close to the nonstationary region.

CO471 Room 107 SPARSE BAYESIAN MODELLING Chair: Helga Wagner

CO0458: **Using variable selection within Bayesian clustering to explore dependence between categorical variables**
Presenter: **Michail Papathomas**, School of Mathematics and Statistics, University of St Andrews, United Kingdom
Co-authors: Sylvia Richardson
Detecting interactions when analysing data sets created by large studies is becoming increasingly important in the Social sciences, Economics and Biostatistics. Investigating complex dependence structures within a linear modelling framework is not straightforward due to the difficulty in searching an unwieldy large space of competing models. One approach for reducing the dimensionality of the problem is to utilize a Bayesian modelling approach based on the Dirichlet process. We investigate the relation between the Dirichlet process and linear modelling, and discuss the utility of the Dirichlet process for the exploration of high order interactions, especially when sparse data are analysed.

CO1314: **Bayesian effect fusion with spike and slab prior**
Presenter: **Daniela Pauger**, Johannes Kepler University Linz, Austria
Co-authors: Helga Wagner
In many applications, especially in social or economic studies, potential covariates for a regression analysis are categorical, measured either on an ordinal or on a nominal scale. Including categorical variables in regression models can easily lead to a high-dimensional vector of effects. We present a method for sparse modelling of the effects of categorical covariates, where sparsity is achieved by excluding irrelevant predictors and/or by fusing levels which have essentially the same effect on the response. To encourage effect fusion, we construct a prior that is based on the specification of spike and slab priors on differences of level effects and hence allows for selective shrinkage of these level effects to each other. The proposed prior is designed mainly for fusion of effects but automatically excludes irrelevant predictors as well. Furthermore, the prior allows

to take into account any available information on the ordering of categories and to incorporate prior information which levels should not be fused. We demonstrate the performance of the developed method in simulation studies and in a real data example.

CO0713: **Effect fusion using sparse finite mixtures**
Presenter: **Gertraud Malsiner-Walli**, Johannes Kepler University Linz, Austria
Co-authors: Helga Wagner, Daniela Pauger, Bettina Gruen
In social studies, variables are often measured on a nominal scale with many categories. However, different classification scales using either a finer or a coarser grid are possible. If such a variable is included in a regression model, the inclusion of too many categories can lead to imprecise estimates of the effects. In contrast, if the categorization is too coarse, important effects could be missed. Therefore, it would be appealing to have a method available which achieves classification automatically. For clustering categories, we propose the specification of a modified standard spike and slab prior on the effects. The spike component at zero allows to capture categories with no effect. The slab distribution is a spiky location mixture distribution and allows to identify categories with similar effect size. Model-based clustering of the effects during MCMC allows to both detect categories which have the same effect size and identify variables with no effect at all.

CO1083: **Bayesian variable selection for latent class analysis using a collapsed Gibbs sampler**
Presenter: **Arthur White**, Trinity College Dublin, Ireland
Co-authors: Jason Wyse, Thomas Brendan Murphy
Latent class analysis is used to perform model based clustering for multivariate categorical responses. Selection of the variables most relevant for clustering is an important task which can affect the quality of clustering considerably. We consider a Bayesian approach for selecting the number of clusters and the best clustering variables. The main idea is to reformulate the problem of group and variable selection as a probabilistically driven search over a large discrete space using Markov chain Monte Carlo (MCMC) methods. Both selection tasks are carried out simultaneously using an MCMC approach based on a collapsed Gibbs sampling method, whereby several model parameters are integrated from the model, substantially improving computational performance. Post-hoc procedures for parameter and uncertainty estimation are outlined. The approach is tested on simulated and real data.

CO622 Room 110 CYCLICAL PROPERTIES OF FINANCIAL AND ECONOMIC DATA Chair: Jozef Barunik

CO0479: **Characterising the financial cycle: A multivariate and time-varying approach**
Presenter: **Yves Schueler**, t b d, Germany
Co-authors: Paul Hiebert, Tuomas Peltonen
A methodology is introduced to characterise financial cycles combining a novel multivariate spectral approach to identifying common cycle frequencies across a set of indicators, and a time varying aggregation emphasising systemic developments. The methodology is applied to 13 European Union countries as well a synthetic euro area aggregate, based on a quarterly dataset spanning 1970-2013. Results suggest that credit and asset prices share cyclical similarities, which, captured by a synthetic financial cycle, outperform the credit-to-GDP gap in predicting systemic banking crises on a horizon of up to three years. Financial cycles tend to be long, particularly in upswing phases and with important dispersion across country cases. Concordance of financial and business cycles is observed only 2/3 of the time. While a similar degree of concordance for financial cycles is apparent across countries, heterogeneity is high whereby a cluster of countries tends to exhibit a high synchronisation in their financial cycle phase.

CO0648: **Assessing the cross-country interaction of financial cycles: Evidence from the US and the UK**
Presenter: **Christian Proano**, Otto-Friedrich-University Bamberg, Germany
Co-authors: Till Strohsal, Juergen Wolters
New evidence is provided for the US and the UK on financial cycles as an international phenomenon and how financial cycles interact. Considering the properties of the data in both the time and the frequency domains, we find a strong relation between the financial cycles of the US and the UK. US financial cycles have a significant impact on the UK, but not the other way around. The relation is clearly most pronounced for cycles between 8 and 30 years, which is also the frequency range that explains almost all variation of the data.

CO0964: **The role of financial frictions in minor and major business cycles within the DSGE framework**
Presenter: **Federico Giri**, Polytechnic University of Marche, Italy
Co-authors: Marco Gallegati, Antonio Palestrini
Following the evidence on the business cycle features of the US economy, we argue that estimating DSGE models over different subsets of business cycle frequencies can give insights into the role of financial frictions for the real economy. After separating fluctuations corresponding lower (4 to 8 years) and higher (2 to 4 years) business cycle frequencies (NBER definition of 'minor' and 'major' cycles) we run alternative DSGE models, with and without financial frictions and banks, using US quarterly data from 1992:Q1-2011:Q4. Our findings support the hypothesis that there are substantial advantages for model estimation over different frequency bands and in particular over subsets of business cycle frequencies corresponding to 'major' and 'minor' cycles. Two results are noteworthy: first, we find that a number of parameter estimates vary substantially over different frequency bands, with the most notable differences pertaining to the estimated coefficients for wage and price rigidity, and those of the monetary policy rule. Second, when looking at the role of financial frictions over different subsets of frequencies we provide evidence suggesting that the financial cycle is not a short run phenomenon but it is better described by the frequency bands corresponding to 'major' cycles.

CO0745: **Quantile vector-autoregressions**
Presenter: **Tomas Krehlik**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
Quantile vector autoregressive processes are introduced for modeling rich dependence structures in economic and financial time series. A novel simulated Whittle-like minimum distance estimator of a general process is devised for estimation of parameters. An overview of construction of the estimator is provided accompanied by Monte Carlo simulations and theoretical results that underscore its properties. Moreover, we discuss methodological issues with the estimate, various potential applications of the methodology on both financial data and macroeconomic data that highlight usefulness of the model for an economist.

CO421 Room 106 NETWORK ANALYSIS AND HIGH DIMENSIONAL TIME SERIES MODELS Chair: Christian Brownlees

CO0497: **Simultaneous change-point and factor analysis for high-dimensional time series**
Presenter: **Matteo Barigozzi**, London School of Economics, United Kingdom
Co-authors: Haeran Cho, Piotr Fryzlewicz
A method is proposed for simultaneously analysing the factor structure of the data and detecting (possibly) multiple change-points in high-dimensional time series. Firstly, we introduce a piecewise stationary factor model that enables introducing and, consequently, detecting changes not only in loadings but also in factors and idiosyncratic component, which has not been explored in the existing literature. Next, it is shown that the common component estimated with an over-estimated factor number achieves consistency, which motivates our change-point detection methodology. Then, we propose to transform the data so that an existing panel data segmentation method is applicable to the problem of detecting multiple change-points in the factor structure, and consistency of such an approach is established in terms of the total number and locations of estimated change-points. Empirical performance of the proposed method is investigated on simulated datasets as well as macroeconomic and financial time series.

CO0705: **The impact of network connectivity on factorexposures, asset pricing and portfolio diversication**
Presenter: **Loriana Pelizzon**, Goethe University, Germany
Co-authors: Massimiliano Caporin, Roberto Panzica, Monica Billio
The classic factor-based asset pricing model is extended by including network linkages in linear factor models. We assume that the network linkages are exogenously provided. This extension of the model allows a better understanding of the causes of systematic risk. More specifically we show that (i) network exposures act as an inflating factor for systematic exposure to common factors; (ii) the power of diversification is reduced by the presence of network connections. Moreover, empirically, in the presence of network links a misspecified traditional linear factor model presents residuals thatare correlated and heteroskedastic. We support our claims with an extensive simulation experiment.

CO1232: **Detecting granular time series in large panels**
Presenter: **Geert Mesters**, Universitat Pompeu Fabra, Spain
Co-authors: Christian Brownlees
In large panels of economic and financial time series there are series that have a pervasive influence over the entire panel. We call such series "granular", in the sense that as the cross section dimension increases such series do not become negligible. We introduce methodology for detecting and testing granular series. We introduce a signaling device that ranks the series in order of granularity. The tests that we develop are designed to infer which of the series form the granular series. The tests have power against a variety of other popular structures for the covariance matrix, such as factor, sparse and block structures.The methodology is examined in a large Monte Carlo study as well as for empirical applications in macroeconomics and finance.

CO0917: **Inference in partially identified models with many moment inequalities using lasso**
Presenter: **Anders Kock**, Aarhus University and CREATES, Denmark
Inference is considered in a partially identified moment (in)equality model with many moment inequalities. We propose a novel two-step inference procedure that combines previous methods with a first-step moment inequality selection based on the lasso. Our method controls size uniformly, both in underlying parameter and data distribution. Also, the power of our method compares favorably with that of the corresponding two-step method in CCK14 for large parts of the parameter space, both in theory and in simulations. Finally, our lasso-based first step is straightforward to implement.

CO295 Room 109 ECONOMETRICS OF BOND PRICES Chair: Niklas Ahlgren

CO0671: **Finnish wood purchasing cartel: Damage estimates on pulpwood based on the VECM approach**
Presenter: **Leena Kalliovirta**, Natural Resources Institute Finland, Finland
Co-authors: Riitta Hanninen
Cartel effects in the highly persistent pulpwood prices in Finland are estimated by employing publicly available monthly data and the vector error correction approach (VECM). We generalize the before-and-after comparator-based approach into the nonstationary setting and show that the dummy variable approach yields a reliable estimate of the cartel damage even if the data is nonstationary. Further, the approach allows us to study the existence of a transition period after the cartel. We estimate a large economic model and its statistically sufficient sub-models to depict the pulpwood markets and compute the but-for prices implied by these VEC models. While operational the purchasing cartel was able to generate a price lowering deterministic trend in the pulpwood markets. During the cartel period, the economic model implies that the monthly prices were decreasing by 0.88% (birch), 0.50% (spruce), and 1.0% (pine) whereas the conservative, statistically adequate sub-models imply slower decreases. Thus, the damage accumulated along the duration of the cartel and the average but-for prices were between 12% to 52% (birch), 4% to 26% (spruce), and 18% to 63% (pine) higher than the observed prices.

CO1120: **A probabilistic approach to unit root testing in the presence of structural breaks**
Presenter: **Johan Lyhagen**, Uppsala University, Sweden
Co-authors: Lars Forsberg
The consequences of structural beaks on unit root tests are well known and depend crucially on the assumptions made. This have been the ground for a new class of unit root tests that are robust against structural change. By motivating the AR process using a probabilistic argument we derive a unit root test with higher power than previous test in the presence ofstructural breaks. The method is exemplified using Credit Default Swap data.

CO0884: **Dynamics of VIX term structure and macroeconomic news arrivals**
Presenter: **Juho Kanniainen**, Tampere University of Technology, Finland
Co-authors: Ye Yue, Kim Christensen
An approach is provided to incorporate scheduled macroeconomic news arrivals into a time-varying GARCH model to capture the empirical properties of VIX term structure. The traditional GARCH model is nested in our macro-news characterization, which makes it possible to examine how much VIX term structure is determined by macroeconomic fundamentals. Our model allows flexible specifications of news impact functional form and consequently variance term structure curves deduced by our model are more in comparison to the traditional GARCH model. By providing an empirical analysis with different macroeconomic news, we conclude that the VIX term structure is partially determined by macroeconomic fundamentals.

CO0667: **Pricing of risk in the long run with strong persistence in volatility**
Presenter: **Niklas Ahlgren**, Hanken School of Economics, Finland
Co-authors: Paul Catani
The credit spread and credit default swap (CDS) price are prices of risk in two markets. If the markets price risk equally in the long run, the prices

should be equal. The non-arbitrage relation is tested as an equilibrium relation in the cointegrated vector autoregressive (VAR) model. Empirical studies typically find that the prices are cointegrated for some but not all companies in a sample. Some theoretical reasons for rejecting cointegration between CDS prices and credit spreads are discussed in the finance literature. However, there are several issues related to the time series properties of the series. For a sample of ten large companies and daily data from 2009 to 2016, we show that there is strong persistence in the series, strong persistence in volatility, and the errors are skewed and heavy-tailed. The existence of fourth moments of the errors is not satisfied for some of the companies. Asymptotic and bootstrap tests become unreliable in the presence of conditional heteroskedasticity. Monte Carlo simulations show that wild bootstrap tests have size close to the nominal level even when the existence of fourth moments is not satisfied. The power of wild bootstrap tests is low. Obtaining high power requires time series of more than 1000 observations or more than four years of daily observations.

CO271 Room 112 BIG DATA AND ECONOMETRICS Chair: Matthew Harding

CO0712: **Poverty in HD: What can we learn about economic welfare using high resolution satellite imagery**
Presenter: **Jonathan Hersh**, Boston University, United States
Co-authors: David Newhouse, Ryan Engstrom
Sri Lankan data are used to investigate the ability of features derived from high spatial resolution satellite images to predict spatial variation in poverty across small areas. Object-based features include the number and density of buildings, shadow area (building height proxy), car counts, the density of paved and unpaved roads of various widths, the type of farmland, and roof material. Spectral and texture features include a vegetation index (NDVI), Speeded-Up Robust Features (SURF), Pantex, Fourier and Gabor transformations. These features are matched to predicted poverty rates, based on estimates of household per capita household consumptions imputed into the 2011 census, for 1,251 Gram Niladhari (GN) Divisions covering a total area of roughly 3,500 sq. km. High spatial resolution satellite features explain about sixty percent of the variation in the share of the GN population in the bottom forty percent of the national distribution of predicted per capita consumption. Measures of building density are particularly strong correlates of poverty. Within satellite scenes, car counts and SURF are strongly correlated with poverty in urban areas, while the share of roads that are paved and roof type are strong correlates in rural areas. Out-of-sample predictions are less accurate but tend to preserve rank. These results suggest that features extracted from high spatial resolution imagery hold considerable promise for contributing to small area poverty estimates.

CO1076: **Tensor autoregression in economics and finance**
Presenter: **Giuseppe Brandi**, LUISS University, Italy
The increase of data dimensionality poses a problem in its analysis. Standard methods try to reduce the multidimensional data in vectors or matrices and do analysis on this simpler objects. However, this approach has two main drawbacks. The first one in related to the fact that vectorize a multidimensional dataset destroys the interconnections between dimensions and secondly, it generates an exponentially increase of the parameters to be estimated. Take a dataset $\mathbf{A}$ consisting of three dimensions and that such dimensions are $30, 30, 50$. The vectorized form of statistical analysis would need to estimate $30 \times 30 \times 50 = 45000$ parameters. A common way to reduce the problem is to use PCA methods, but as it is known, PCA factors are then difficult to interpret. A way to overcome this problem is to rely on tensor analysis. Treating the dataset $\mathbf{A}$ as a tensor and fitting a model on it, requires to estimate "just" $30 + 30 + 50 = 110$ parameters. We employ a Tensor Autoregression model on multidimensional economic data trying to show the superiority of such method to the classical VAR.

CO1029: **A panel quantile approach to attrition bias in big data: Evidence from a randomized experiment**
Presenter: **Matthew Harding**, University of California, Irvine, United States
Co-authors: Carlos Lamarche
A quantile regression estimator is introduced for panel data models with individual heterogeneity and attrition. The method is motivated by the fact that attrition bias is often encountered in Big Data problems. For example, many users sign-up for the latest utility program but few remain active users several months later, making the evaluation of such interventions inherently very challenging. Building on an earlier work, we provide a simple identification strategy that leads to a two-step estimation procedure. In the first step, the coefficients of interest in the selection equation are consistently estimated using parametric or nonparametric methods. In the second step, standard panel quantile methods are employed on a subset of weighted observations. The estimator is computationally easy to implement in Big Data applications with a large number of subjects. We investigate the conditions under which the parameter estimator is asymptotically Gaussian and we carry out a series of Monte Carlo simulations to investigate the finite sample properties of the estimator. We explore an application to the evaluation of a recent Time-of-Day electricity pricing experiment.

CO1548: **The predictive power of Google searches in forecasting US unemployment**
Presenter: **Juri Marcucci**, Bank of Italy, Italy
Co-authors: Francesco D Amuri
The use of an index of Google job-search intensity is suggested as the best leading indicator to predict the monthly US unemployment rate. We perform a deep out-of-sample forecasting comparison among models thatadopt the Google Index, the more standard initial claims or alternativeindicators based on consumers' and employers' surveys, selecting the best model specification in sample using the BIC. Google-based model soutperform traditional ones, with their relative performance improving with the forecast horizon. Furthermore, quarterly predictions constructed using Google-based models provide more accurate forecasts than the Survey of Professional Forecasters, models based on labor force flows or standard nonlinear models. Google-based models seem to predict particularly well at the turning point taking place at the beginning of the Great Recession, while their relative predictive ability stabilizes afterwards.

CO729 Room 111 FORECASTING II Chair: Francesco Ravazzolo

CO0734: **Modeling networks of currency tick-data as continuous time graphical models**
Presenter: **Jonas Hallgren**, Royal Institute of Technology KTH, Sweden
Co-authors: Timo Koski
Continuous time Bayesian networks (CTBNs) are graphical representations of the dependence structures between continuous time random processes with finite state spaces. We propose a method for learning the structure of the CTBNs using a causality measure based on Kullback-Leibler divergence. We introduce the causality matrix which can be seen as a generalized version of the covariance matrix. We give a mixed radix representation of the process that much facilitates the learning and simulation. A new graphical model for tick-by-tick financial data is proposed and estimated. The model is inspired by speech recognition applications and computes the state of the process using transformations to frequency domain. The suggested approach is able to learn the graphical structure of both the tick-data and of a simulated example. Neuroscientists proposed Integrated Information Theory (IIT) as a way of measuring the amount of consciousness in a system. By the use of information geometry, we relate the proposed causality measure to Integrated Information Theory. The suggested approach allows us to work with tick-data in an easy way where it is not necessary to take every tick into account separately. Because of this and the promising results, we consider the CTBN-approach be relevant for studying tick-by-tick data.

CO0743: **Density forecasts of inflation and interest rates using Bayesian DCS models with dynamic conditional skewness**
Presenter: **Blazej Mazur**, Cracow University of Economics, Poland

The focus is on potential gains in density forecasting performance of univariate macroeconomic series using models that allow for dynamic changes in conditional asymmetry. The conditional distribution used is flexible as it allows for skewness and shape (tail) asymmetry, being a two-piece asymmetric generalized t. The benchmark model assumes the Dynamic Conditional Score mechanism for updating of conditional location and scale, with asymmetry/skewness parameters being time-invariant. It is then extended using two alternative parametrizations implying dynamic conditional skewness. Within the first approach the overall scale parameter is separated from the skewness coefficient, whereas the other one is based on left-scale and right-scale parameters (with overall scale and skewness coefficients being implicit). Within the two approaches diagonal and matrix versions of DCS models are compared. The matrix specification allows for dependencies in dynamics of the three features (location, scale and skewness or location, left-scale and right-scale). The aim is to identify optimal parametrization and to indicate crucial dynamic dependencies. Within the forecasting experiment, disaggregate US PCE inflation rates and interest rates are analyzed. Density forecasting performance is examined using CRPS and log-Score criteria, calibration is investigated using PITs; performance of linear pools is also considered.

CO0902: **Probabilistic forecasting of hierarchical time series data**
Presenter: **Souhaib Ben Taieb**, Monash University, Australia
Co-authors: James Taylor, Rob Hyndman

Time series can often be naturally represented in a hierarchical or grouped structure. For example, a manufacturing company can disaggregate total demand for their products by country of sale, retail outlet, product type, package size, and so on. As a result, there can be a large number of individual time series to forecast at the most disaggregated level, plus additional series to forecast at higher levels of aggregation. In order to allow consistent decisions over different levels of the hierarchy, the forecasts for the disaggregated series are usually required to add up exactly to the forecasts of the aggregated series, a constraint known as aggregate consistency. Computing mean aggregate consistent forecasts involves fitting a linear regression model where the design matrix has one column for each of the series at the most disaggregated level. We will discuss some algorithms to generate probabilistically aggregate consistent forecasts with an application to electricity demand forecasting using smart meter data.

CO0847: **Optimal Value-at-Risk reports: Reducing capital requirement with state dependent forecasting**
Presenter: **Patrick Schmidt**, Heidelberg Institute for Theoretical Studies, Goethe University Frankfurt, Germany

The optimal risk reporting strategy for Value-at-Risk forecasts is investigated under the three-zone approach of capital requirement and back-testing as currently suggested by the Bank for International Settlements. We find that it is optimal to report overly conservative in times of low volatility and understate risk in times of high volatility. We characterize the class of state-dependent loss functions, which are consistent for the optimal strategy. In a simulation study, we compare the performance of commonly applied estimation strategies with the optimized estimation - based on the new class of loss functions - in terms of capital requirement.

CO321 Room 103 EMPIRICAL MACROECONOMICS Chair: Daniel Kaufmann

CO1138: **Large Bayesian VAR forecasting in a national accounting environment**
Presenter: **Alexander Rathke**, ETH Zurich, KOF Swiss Economic Institute, Switzerland
Co-authors: Samad Sarferaz

Large Bayesian VARs are used, including national accounting identities from the expenditure side and the production side, i.e. the demand and supply side. Hence, we produce forecasts for GDP that fully rely on the data generating process. We show that by using national accounting identities, we outperform smaller VARs that only use a final GDP measure, instead of the aggregated one. We further show how to implement expert based priors on the expenditure and production side using soft and hard conditions. For the expenditure side we use expert based knowledge and for the production side we use business tendency surveys conducted in different sectors to inform our priors. Given that the forecasting exercise results in two different forecasts for GDP, we further combine these two estimates using different weighting schemes, which can differ across different frequencies and horizons. We show that a combination of these two estimates is beneficial. Interestingly, the optimal weight for the combination of the two GDP measures gives the expenditure side more in the short-run and the production side more weight the medium-term.

CO0773: **FAVAR revisited: A sparse dynamic factor approach**
Presenter: **Simon Beyeler**, University of Bern, Switzerland
Co-authors: Sylvia Kaufmann

Extracting relevant information from many different time series measuring different aspects of an economy and compressing it using factor analysis is a neat way to handle large amounts of data and to circumvent the curse of dimensionality problem without ignoring possibly important features. To do so, we combine the FAVAR framework with recently developed estimation and identification procedures for sparse dynamic factor models. Our estimation procedure allows us to explicitly discriminate between zero and non-zero factor loadings. This provides one solution to the identification common to all factor models. Further, the identified unobserved factors get a meaningful economic interpretation due to the structure of non-zero loadings. An additional distinction to traditional factor models is that we work with correlated factor innovations allowing us to implement different strategies to identify structural shocks used in the literature and to perform traditional structural VAR type analysis. Applying our methodology to US macroeconomic data (FRED QD) reveals indeed a high degree of sparsity in the data. The proposed identification procedure yields seven unobserved factors that together account for about 55 percent of the variation in the data. All of them are clearly related to an economic concept.

CO1102: **Seasonal adjustment and data inefficiency: Evidence from simulation and real-world data**
Presenter: **Ronald Indergand**, State Secretariat for Economic Affairs, Switzerland

Seasonal adjustment based on moving-average filters is suboptimal from a forecast efficiency perspective. First, subsequent revisions to the data are comparatively large. Second, initial announcements are of mean-reverting character, that is, large initial growth rates tend to be revised downward. Both features increase the difficulty of assessing the dynamics of a variable and lead to suboptimal signals for policy makers. It is shown by simulation that model-based seasonal adjustment reduces the first feature of the data and eliminates the second if the data generating process is known. This results in more efficient preliminary estimates allowing for a more accurate assessment of the current state of an economic variable. Using GDP data from nine countries we demonstrate in a real-world setting how seasonal adjustment produces mean-reverting preliminary GDP releases. Overall, seasonal adjustment may account for the bulk of the results regarding mean-reverting data revisions that have been found by numerous studies.

CO0233: **On the cost of deflation: Evidence from noisy historical data**
Presenter: **Daniel Kaufmann**, ETH Zurich, Switzerland

The aim is to show that methodological deficiencies of 19th century retrospective CPI estimates spuriously increase the volatility of inflation as well as the number of deflationary episodes, and they lower inflation persistence. Against this backdrop, the link between real activity and deflation is revisited for the US from 1834 to 1945. First, an improved measure of inflation is estimated based on a factor model. Second, the results are confirmed by an IV-regression approach that uses an independent noisy proxy to instrument the error-ridden retrospective CPI. A significant association emerges between lower real activity and deflation, which is not limited to the Great Depression.

CP001 Room Hall POSTER SESSION Chair: Stella Hadjiantoni

CP1405: **Simultaneous event multivariate point process models with application to causality analysis of financial markets**
Presenter: **Daisuke Kurisu**, University of Tokyo, Japan
Co-authors: Naoto Kunitomo
In economic financial time series, big events are sometimes observed which are relatively rare, but often have significant influences, not only on a financial market, but also on several different markets and macro economics. By using simultaneous event Hawkes type models, which allows co-jumps, we analyze the causal effects of big events in the sense of the Granger-non- causality. In our data analysis, we investigate the causality among the financial markets of Tokyo, New York, London and other markets. We have found several important findings among financial markets.

CP1519: **Dynamic bayesian models for panel data with dynamic skewness**
Presenter: **Hasan Uri**, The University of Sheffield, United Kingdom
Time series models often rely on symmetry of the innovation process to carry out estimation. There are several alternatives in the literature extending some of the most commonly used models to include skewness. I propose a further extension, allowing for the skewness in the innovations to vary with time. This can be useful in, for instance, defining new measures of convergence in economics and finance. I use the reciprocal skewing mechanism and propose an autoregressive gamma process to model the path of skewness. This is embedded in an linear AR(1) model for panel data from a Bayesian perspective, so I provide a hierarchical prior structure suitable for most applications envisaged and discuss sensitivity to prior parameters. The methodology is illustrated using data on EU economic growth.

CP1599: **Model selection tests in the Markov switching models**
Presenter: **Anna Czapkiewicz**, AGH University of Science and Technology, Poland
Co-authors: Antoni Leon Dawidowicz
Markov switching models are discussed in two cases: models with constant transition matrix and models with time-varying probabilities in the transition matrix. We can not use the classical tests to compare these models since some regular conditions are not fulfilled. We propose approaches of comparing discussed models in two cases. Firstly, we compare models with different number of regimes. Secondly, we compare a model with a constant transition matrix and a model with time-varying transition matrix. For testing purposes we study asymptotic properties of maximum likelihood estimators.

CP1588: **Common cycles in risk**
Presenter: **Lucie Kraicova**, Charles University, Faculty of Social Sciences, Czech Republic
Co-authors: Jozef Barunik
The relationship is studied between conditional quantiles of returns and the long-, medium- and short-term volatility in a portfolio of financial assets. We argue that the combination of quantile panel regression and wavelet decomposition of the volatility time series provides us with new insights into the pricing of risk and increases the accuracy of our estimates of return quantiles. Our results contribute to the literature on the risk-return relationship with an emphasis on portfolio management under various investment horizons. Moreover, the analytical framework that we introduce should be applicable to a wide range of problems outside of our research area.

CC1516: **Consistent estimation for true fixed effects stochastic frontier models**
Presenter: **Luca Grassetti**, University of Udine, Italy
Co-authors: Ruggero Bellio
Inference is considered for true fixed effects stochastic frontier models, which are employed in panel data settings to separate time-invariant heterogeneity from efficiency. Such models have a noteworthy theoretical appeal, yet the estimation of their structural parameters is hindered by the incidental parameter problem, which may be severe for settings with a large number of short panels. We propose an estimation approach that exploits the equivariance property of maximum likelihood estimation to obtain a simple marginal maximum likelihood estimator of the structural parameters. This results in root-n consistent estimation, and it can be applied to a wide array of distributional specifications without the need for simulation techniques. Another important feature of the proposal is its computational simplicity, and we illustrate the use of the TMB R package for automatic differentiation to obtain a scalable and efficient implementation.

CC1485: **Spatial variability of the relationship between carbon dioxide and temperature**
Presenter: **Umberto Triacca**, University of L Aquila, Italy
Co-authors: Francesca Di Iorio
Several recent studies have established that there is a clear relationship between the amount of carbon dioxide (CO2) in the atmosphere and global temperature. We propose a new statistical approach to analyze the spatial variability of this relationship based on a distance measure between pairs of vector autoregressive (VAR) models. In particular, we estimate a bivariate VAR model (in the variables CO2 and temperature) for eight latitude bands and we compute the proposed distance between any pair of models obtaining a matrix of distances. Our analysis by multidimensional scaling of this matrix shows that, despite CO2 being a globally well-mixed gas, the 'fingerprints' of CO2 forcing are regionally very different. In the latitude belts surrounding the equator (0 - 24S and 0 - 24N) the estimated VAR models are very close. Another cluster is given by the VAR models concerning the lower mid-latitudes of both hemispheres (24S - 44S and 24N -44N) and the 44S - 64S latitude belt. However, the main result of our analysis is that the VAR models for the polar regions seem to be 'isolated points'. In particular, it is interesting to underline that the largest distance is that between the VAR models for Arctic and Antarctic. There is a clear different response of the Arctic and Antarctic to anthropogenic forcing. Also the phenomenon referred to as Arctic warming amplification seems to be confirmed by our analysis.

CP1754: **Robust monetary policy in a linear model of the Polish economy**
Presenter: **Mariusz Gorajski**, University of Lodz, Poland
In the majority of central banks the short-term rate paths are smooth and only gradual changes can be observed. Optimal monetary policy models in the linear-quadratic framework produce high variability of interest rates, and are hence inconsistent with the data. One can obtain gradual behaviour of optimal monetary policy by adding an interest rate smoothing term to the central bank objective. This heuristic procedure has not much substantiation in the central bank's targets and raises the question: What are the rational reasons for the gradual movements in the monetary policy instrument? In this paper we determine optimal monetary policies in a VAR model of the Polish economy with parameter uncertainty. We prove that there exists a structure of the multiplicative uncertainty in the optimal linear-quadratic model that explains the central bank's behaviour. Thus proving that parameter uncertainty can be the rationale for 'timid' movements in the short-interest rate dynamics. Finally, we show that there is trade of between parameter uncertainty and the interest rate smoothing incentive.

CP1768: **The market price of skewness and limits to arbitrage**
Presenter: **Paola Pederzoli**, University of Geneva Swiss Finance Institute, Switzerland
A new measure of skewness is provided which estimates the third moment of asset returns while being independent from the first, second and fourth moment. A numerical study assesses the precision of the methodology. Our skewness measure can be traded directly on American options

in a form of skewness swap, in which the investor exchanges the risk neutral skewness of the asset (Q skewness) with the realised skewness (P skewness). We implement the skewness swaps on all the constituents of the SP100 separately and we find that there is a positive and significant skewness risk premium in the single stock equity market, in particular after the 2007-2009 financial crisis. We find a strong dependence between the Q skewness and the ratio between the trading volumes of the out-of-the-money calls and out-of-the-money puts. We find that market makers increase the prices of the most traded category of options because they are exposed to more risks and costs in that category, in accordance to a limit to arbitrage hypothesis.

EI675 Room Graduation hall MULTIVARIATE EXTREMES AND APPLICATIONS Chair: Armelle Guillou

EI0433: **D-norms and the max-characteristic function: Offsprings of multivariate extreme value theory**
Presenter: **Michael Falk**, University of Wuerzburg, Germany
Co-authors: Gilles Stupfler

Multivariate extreme value theory (MEVT) is the proper toolbox for analyzing several extremal events simulaneously. Its practical relevance in particular for risk assessment is, consequently, obvious. But on the other hand MEVT is by no means easy to access, a fil rouge is not visible. Writing the 'angular measure' in MEVT in terms of a random vector, however, provides the missing fil rouge: Every result in MEVT, every relevant probability distribution, be it a max-stable one or a generalized Pareto distribution, every relevant copula, every tail dependence coefficient etc. can be formulated using a particular kind of norm on multivariate Euclidean space, called D-norm. Max-characteristic functions (max-CFs) are introduced as an offspring of D-norms. A max-CF characterizes the distribution of a random vector in $\mathbb{R}^d$, whose components are nonnegative and have finite expectation. Pointwise convergence of max-CFs is shown to be equivalent with convergence with respect to the Wasserstein metric. The space of max-CFs is not closed in the sense of pointwise convergence. An inversion formula for max-CFs is established.

EI0508: **Conditional extreme value models: Fallacies and pitfalls**
Presenter: **Holger Drees**, University of Hamburg, Germany
Co-authors: Anja Janssen

Classical multivariate extreme value theory deals with the behavior of random vectors when at least one component is large. Sometimes, however, one is interested in the behavior of the vector when not an arbitrary, but a pre-specified component is large. To this end, the so-called conditional extreme value (CEV) models were introduced and their relation to classical multivariate extreme value models was discussed. We first recall concepts of marginal standardization in CEV models from the literature. We will illustrate by counterexamples that the standardization of the not necessarily extreme components is not well adapted to the particular features of the CEV models. From the above interpretation, it seems plausible that (under mild extra conditions ruling out degenerate cases) the classical multivariate extreme value model for a random vector implies that the assumptions of all CEV models which suppose that one specific component is large are fulfilled, and vice versa. Unfortunately, the precise relationship between the CEV models and the classical multivariate extreme value model turns out to be much more intricate than intuition suggests. In particular, we give counterexamples to several claims about this relation which can be found in the literature. Time permitting, we finally discuss alternative modeling approaches which avoid some of the drawbacks of CEV models, but exhibit other shortcomings.

EI1119: **Empirical likelihood based testing for multivariate regular variation**
Presenter: **John Einmahl**, Tilburg University, Netherlands
Co-authors: Andrea Krajina

Multivariate regular variation is an important property of a multivariate probability distribution that can be very useful in extreme value statistics. Therefore it is desirable to check this property based on a multivariate random sample. We construct a hypothesis test for multivariate regular variation based on localized empirical likelihood and establish its limiting null distribution. We also investigate the finite sample performance of the test using simulated and real data.

EO151 Room 208 COMPUTATIONAL TOOLS FOR THE ANALYSIS OF UNCONVENTIONAL DATASETS Chair: Jochen Luedering

EC0173: **Monetary policy on Twitter and its effect on asset prices: Evidence from computational text analysis**
Presenter: **Peter Tillmann**, Justus-Liebig University Giessen, Germany
Co-authors: Jochen Luedering

The public debate about the future course of monetary policy is dissected and the effects of selected topics of this discourse on U.S. asset prices are traced. We focus on the "taper tantrum" episode in 2013, a period with large revisions in expectations about Fed policy. Based on a novel data set of 90,000 Twitter messages ("tweets") covering the entire debate of Fed tapering on Twitter we use Latent Dirichlet Allocation, a computational text analysis tool to quantify the content of the discussion. Several estimated topic frequencies are then included in a VAR model to estimate the effects of topic shocks on asset prices. We find that the discussion about Fed policy on social media contains price-relevant information. Shocks to shares of "tantrum"-, "QE"- and "data"-related topics are shown to lead to significant asset price changes. We also show that the effects are mostly due to changes in the term premium of yields consistent with the portfolio balance channel of unconventional monetary policy.

EO0174: **Forward or backward looking: The economic discourse and the observed reality**
Presenter: **Jochen Luedering**, University of Giessen, Germany
Co-authors: Peter Winker

The question whether the academic research is anticipating economic shake-ups or merely reflecting the past is considered. Exploiting the corpus of articles published in the Journal of Economics and Statistics for the years 1949 to 2010, this pilot study proposes a quantitative framework for addressing these questions. The framework comprises two steps. First, methods from computational linguistics are used to identify relevant topics and their relative importance over time. In particular, Latent Dirichlet Analysis is applied to the corpus after some preparatory work. Second, for some of the topics which are closely related to specific economic indicators, the developments of topic weights and indicator values are confronted in dynamic regression and VAR models. The results indicate that for some topics of interest, the discourse in the journal leads developments in the real economy, while for other topics it is the other way around.

EO0391: **Words are the new numbers: A newsy coincident index of business cycles**
Presenter: **Leif Anders Thorsrud**, Norwegian Business School BI, Norway

A daily business cycle index based on quarterly GDP and textual information contained in a daily business newspaper is constructed. The newspaper data is decomposed into time series representing newspaper topics using a Latent Dirichlet Allocation model. The business cycle index is estimated using the newspaper topics and a time-varying Dynamic Factor Model where dynamic sparsity is enforced upon the factor loadings using a latent threshold mechanism. We show that both contributions, the usage of newspaper data and the latent threshold mechanism, contribute towards the qualities of the derived index: It is more timely and accurate than commonly used alternative business cycle indicators and indexes, and, it provides the index user with broad based high frequent information about the type of news that drive or reflect economic fluctuations.

EC1439: **Sparse dimension reduction of multiple data on same observations**
Presenter: **Nickolay Trendafilov**, Open University, United Kingdom
Co-authors: Purificacion Galindo
It has become common to perform multiple analyzes on the same set of observations, e.g. patients. This requires analyzing simultaneously several data sets each involve different variables measured on same set of observations. Such type of data is a special case three-way data array also called tensor data. However the third mode does not really play ant role. We review the available methods for analysis of such data, and show that they are either to simplistic or too complicated for analysis of large data. We propose some ways for improvement and possibly new methods for analysing multiple data matrices measured on same objects.

EO179 Room 214 HIGHER MOMENTS IN MULTIVARIATE ANALYSIS Chair: Nicola Loperfido

EO0180: **Third cumulant for multivariate aggregate claim models**
Presenter: **Krzysztof Podgorski**, Lund University, Sweden
Co-authors: Nicola Loperfido, Stepan Mazur
The third moment cumulant for the aggregated multivariate claims is considered. A formula is presented for the general case when the aggregating variable is independent of the multivariate claims. It is discussed how this result can be used to obtain a formula for the third cumulant for a classical model of multivariate claims. Two important special cases are considered. In the first one, multivariate skewed normal claims are considered and aggregated by a Poisson variable. The second case is dealing with multivariate asymmetric generalized Laplace and aggregation is made by a negative binomial variable. Due to the invariance property the latter case can be derived directly leading to the identity involving the cumulant of the claims and the aggregated claims. There is a well established relation between asymmetric Laplace motion and negative binomial process that corresponds to the invariance principle of the aggregating claims for the generalized asymmetric Laplace distribution. We explore this relation and provide multivariate continuous time version of the results.

EO0191: **Tail behavior and dependence structure in the APARCH model**
Presenter: **Farrukh Javed**, Orebro University, Sweden
Co-authors: Krzysztof Podgorski
The APARCH model attempts to capture asymmetric responses of volatility to positive and negative news shocks the phenomenon known as the leverage effect. Despite its potential, the model's properties have not yet been fully investigated. While the capacity to account for the leverage is clear from the defining structure, little is known how the effect is quantified in terms of the models parameters. The same applies to the quantification of heavy-tailedness and dependence. To fill this void, we study the model in further detail. We study conditions of its existence in different metrics and obtain explicit characteristics: skewness, kurtosis, correlations and leverage. Utilizing these results, we analyze the roles of the parameters and discuss statistical inference. We also propose an extension of the model. Through theoretical results we demonstrate that the model can produce heavy-tailed data. We illustrate these properties using S&P500 data and country indices for dominant European economies.

EO0255: **Optimal tests for elliptical symmetry against skew-elliptical alternatives**
Presenter: **Christophe Ley**, University of Ghent, Belgium
Co-authors: Marc Hallin, Laetitia Gelbgras
The majority of statistical procedures involving multivariate data assumes elliptical symmetry of the data at hand. This assumption, however, is often violated, especially in finance. A very popular alternative are the skew-elliptical distributions, obtained by multiplying elliptically symmetric densities with a skewing function. We shall show how to build tests for elliptical symmetry that are designed to be optimal against skew-elliptical distributions. Starting from optimal parametric tests (in the spirit of score tests), we shall render them semi-parametric and hence valid in the entire class of elliptical distributions, yet they shall inherit optimality properties from their parametric antecedents. One particular aspect of our tests is that they will not require the typical constraint of finite high-order moment assumptions. This is particularly useful when dealing with financial data, which are often heavy-tailed. We shall illustrate the benefits of these new tests by analyzing data from major stock markets.

EO0335: **MultiSkew: An R package for skewness-based data analysis**
Presenter: **Cinzia Franceschini**, Chieti-Pescara University, Italy
Co-authors: Nicola Loperfido
Multivariate skewness plays an important role in several areas of multivariate analysis: normality testing, projection pursuit, independent component analysis, model-based clustering, portfolio optimization and density approximation. MultiSkew is an R package purported to describe, testing and visualize multivariate skewness. In particular, it incorporates state-of-the-art algorithms for computing linear projections which either maximize or minimize absolute skewness. The former are useful to detect data structures, the latter in removing them.

EO559 Room Board meeting room II LIFETIME MODELING AND ESTIMATION Chair: Maria Luz Gamiz

EO0187: **Testing model effect on projected mortality indicators**
Presenter: **Ana Debon**, Universitat Politecnica de Valencia, Spain
Co-authors: Steven Haberman, Francisco Montes, Edoardo Otranto
The parametric model introduced by Lee and Carter in 1992 for modelling mortality rates in the USA was a seminal development in the forecasting of life expectancies, and has been widely used since then. Different versions and extensions of this model, using different hypotheses about the data, constraints on the parameters and fitting methods have led to improvements in the model's fit to historical data and ability to forecast the future. The main objective is to evaluate if differences between three extensions of the Lee-Carter model are reflected in the forecasts of different mortality indicators. Considering three different extensions of the Lee-Carter model, forecasts of three different mortality indicators (period life expectancy, modal age of death and the Gini index of the life table) are obtained. The parameter error in these forecasts is allowed for using the block-bootstrap technique. Furthermore, the projected mortality indicators are compared using ANOVA tests based on functional data.These techniques are applied to Spanish mortality data. We find that there are significant differences between the models in the projection of all the mortality indicators considered. Hence, we find model effect on mortality tendency and dispersion.

EO0306: **Degradation and maintenance modeling using the inverse Gaussian process**
Presenter: **Bo Henry Lindqvist**, Norwegian University of Science and Technology, Norway
Co-authors: Odd Eirik Farestveit
Models for degradation and maintenance of items based on first passage times of stochastic processes have proven useful in diverse applications such as production machines or pipelines. We consider items that are inspected and maintained at sequentially determined random times, with degradation modeled by increasing stochastic processes. The inverse Gaussian process is considered as a concrete example. Two threshold levels will be considered for the deterioration process, a lower one corresponding to a degraded state, for which a preventive maintenance action may be performed, and a higher one corresponding to the failure state at which the item is replaced. A simulation based algorithm is developed to calculate

long-run expected costs with an aim to arrive at optimal inspection and maintenance procedures. An interesting part of the algorithm involves bridge sampling from the degradation process.

EO0513: **Residual plots for the modelling of covariates in accelerated failure time models**
Presenter: **Jan Terje Kvaloy**, University of Stavanger, Norway
Co-authors: Bo Henry Lindqvist, Stein Aaserud
In parametric accelerated failure time regression models one issue of concern is proper modelling of the functional form of the covariates. We will discuss residual plots to reveal and check the functional form of the covariates in such models. Two approaches for handling censoring will be discussed; adjusting censored residuals by adding a residual time, and using nonparametric exponential regression on unadjusted censored residuals. It will in particular be demonstrated how residuals can be used to infer the correct functional form for misspecified covariates. Applications to real and simulated data will be shown.

EO0848: **Best one-sided cross-validation for hazards**
Presenter: **Maria Luz Gamiz**, University of Granada, Spain
Co-authors: Lola Martinez-Miranda, Jens Perch Nielsen
A practical implementation is presented for local linear hazard estimation adding a multiplicative bias correction. It includes asymptotic theory for standard bandwidth selection methods. Although double one-sided cross-validation (do-validation) is recommended when implementing local linear hazards, when looking through details of numerical examples, it turns out that often one of the two one-sided validation approaches does not do well. This seems to be related to situations where exposure is highly varying, for example around boundaries. While do-validation seems to be "the lucky one" doing well in most cases, it still leaves a somewhat unrobust impression to add aggregate a well-working one-sided cross-validation with its failing opposite to end up well by something looking like a coincidence. These insights turn out to be more important when introducing bias correction leading us to the new principle of best one-sided cross-validation (bo-validation) where we propose using a simple measure at every single point to determine to go right or to go left when validating. Bo-validation does as well as do-validation in almost all cases and it is particular powerful when bias correcting, where it is better than do-validation in all the finite-sample cases considered.

EO123 Room 213 NON- AND SEMIPARAMETRIC INFERENCE FOR COMPLEX TIME TO EVENT DATA Chair: Markus Pauly

EO0197: **Two-stage modelling of additive Gamma frailty models for survival data, case-control and ascertainment**
Presenter: **Thomas Scheike**, University of Copenhagen, Denmark
A simple modification of the addtive gamma frailty model is presented that has several computational advantages and can be used for even very large data. The motivation for these developments are to study gene-environement effects, or more generally familial aggregation. Such data is often sampled in different ways and we therefore also discuss howto deal with case-control sampling in the context of survival data. In addition we show how to deal with data that is "ascertained" due the identification of all cases (or a random sample of these). It is crucial to distinguish ascertainment from case-control sampling. We believe that data is often ascertained rather than case-control data when obtained due to the recruitment of cases at a particular clinic. We also here suggest combining ascertainment of the cases with a random sample of controls or the the cohort that in this setting improve the performance considerably. We point out that ascertainment is particular easy to deal with when based on a pairwise composite likelihood approach where only pairs are ascertained rather than families.

EO0224: **Nonparametric estimation of state occupation and transition probabilities in non-Markov multistate models**
Presenter: **Jan Beyersmann**, Ulm University Institute of Statistics, Germany
Co-authors: Arthur Allignol, Carina Mueller
The Aalen-Johansen estimator generalizes the Kaplan-Meier estimator for independently left-truncated and right-censored survival data to estimating the transition probability matrix of a time-inhomogeneous Markov model with finite state space. The Markov assumption enables use of martingale methods to investigate properties of the Aalen-Johansen estimator. It has been noted that the Aalen-Johansen estimator, standardized by a consistent estimator of the initial distribution of the multi-state model, consistently estimates the state occupation probabilities of a non-Markov model if censoring is entirely random. We extend this result to random left-truncation, improve on martingale arguments used to establish these results and discuss how transition probabilities may subsequently be estimated.

EO0420: **Inference for the concordance index for possibly right-censored and tied data**
Presenter: **Dennis Dobler**, University of Ulm, Germany
Co-authors: Markus Pauly
In a two-sample survival setting with independent survival variables T_1 and T_2 and independent right-censoring, the concordance index $p = P(T_1 > T_2) + \frac{1}{2}P(T_1 = T_2)$ is an intuitive measure for discriminating two survival distributions. Comparing two treatments, the case $p > 1/2$ suggests the superiority of the first. Nonparametric maximum likelihood estimators based on normalized Kaplan-Meier estimators naturally handle tied data, which are omnipresent in practical applications. Studentizations allow for asymptotically accurate inference for p. For small samples, however, coverage probabilities of confidence intervals are considerably enhanced by means of bootstrap and permutation techniques. The latter even yields finitely exact procedures in the situation of exchangeable data. Simulation results support all theoretic properties under various censoring and distribution set-ups.

EO0546: **Regression models for the restricted residual mean life**
Presenter: **Giuliana Cortese**, University of Padua, Italy
Co-authors: Stine Holmboe, Thomas Scheike
Hazard functions are typically studied via Cox's regression survival models. The resulting hazard ratios are difficult to interpret and hard to be translated into clinical benefits in terms of increased survival time. The objective is often to study survival functions over a time period. Therefore, there is increasing interest in summary measures based on the survival function, which are easier to interpret than the hazard ratio. For the survival time T, we consider the residual mean time $E(T-t|T \geq t)$, which has recently received increasing interest. This quantity represents a partial area under the survival function and is interpreted as the residual life expectancy of individuals who had survived up to a certain time t. However, due to the presence of right censoring, the tail of the survival distribution is often difficult to be correctly estimated. As a solution, we propose to study the restricted residual mean time $E(min(T,\tau)-t|T \geq t)$, for any $\tau > 0$. We present regression models for this new measure, based on weighted estimating equations and the inverse probability of censoring weighted estimator to model potential right censoring. We also show how to extend the models to deal with delayed entries. Estimation performance is investigated by simulation studies. Using real data about the Danish Monitoring Cardiovascular Risk Factor Survey, we illustrate an application of these regression models when the link function is identity or exponential.

EO511 Room 207 RECENT DEVELOPMENTS ON STATISTICAL MACHINE LEARNING Chair: Yuying Xie

EO0212: **A computationally and theoretically efficient framework for sparse additive modeling**
Presenter: **Noah Simon**, UW Biostatistics, United States

Our ability to collect data has exploded over recent years. Across science and business, we now collect thousands of measurements on each person. It is often of interest to use these measurements to predict some response (eg. likelihood of getting a disease, or probability of clicking a link). Prediction methods must balance 3 objectives: predictive performance, computational tractability, and, in many applications, interpretability. We will discuss a broad class of models which balance these objectives: Sparse additive models induced by combining a structural semi-norm and sparsity penalty. These are more flexible than the standard linear penalized model, but maintain its interpretability and computational tractability. We will show when these penalties can and cannot be combined to induce the desired structure and sparsity. We will give an efficient algorithm for fitting a wide class of these models. And we will give asymptotic upper and lower bounds for prediction accuracy (under a functional incoherence assumption). Time permitting we will also brush on calculating p-values for "significance of features" in these models. With not-too-restrictive assumptions, rates of convergence give us hope that an approach involving targeted maximum likelihood (or a one step correction) could be successful.

EO0349: **Joint estimation of multiple dependent Gaussian graphical models with applications to mouse genomics**
Presenter: **Yuying Xie**, Michigan State University, United States

Gaussian graphical models are widely used to represent conditional dependence among random variables. Thus, we propose a novel estimator for data arising from a group of Gaussian graphical models that are themselves dependent. A motivating example is that of modeling gene expression collected on multiple tissues from the same individual: here the multivariate outcome is affected by dependencies acting not only at the level of the specific tissues, but also at the level of the whole body; existing methods that assume independence among graphs are not applicable in this case. To estimate multiple dependent graphs, we decompose the problem into two graphical layers: the systemic layer, which affects all outcomes and thereby induces cross-graph dependence, and the category-specific layer, which represents graph-specific variation. We propose a graphical EM technique that estimates both layers jointly, establish estimation consistency and selection sparsistency of the proposed estimator, and confirm by simulation that the EM method is superior to a simple one-step method. We apply our technique to mouse genomics data and obtain biologically plausible results.

EO0403: **Sufficient dimension reduction via principal L_q support vector machine**
Presenter: **Yuexiao Dong**, Temple University, United States

Principal L_q support vector machines are introduced as a unified framework for linear and nonlinear sufficient dimension reduction. By noticing that the solution of L_1 support vector machine may not be unique, we set $q > 1$ to ensure the uniqueness of the solution. The asymptotic distribution of the proposed estimators are derived for $q > 1$. We demonstrate through numerical studies that the proposed L_2 support vector machine estimators improve existing methods in accuracy, and are less sensitive to the tuning parameter selection.

EO0449: **Penalized principal logistic regression for sparse sufficient dimension reduction**
Presenter: **Seung Jun Shin**, Korea University, Korea, South
Co-authors: Andreas Artemiou

The sufficient dimension reduction (SDR) is a successive tool for reducing the dimensionality of predictors by finding the central subspace, a minimal subspace of predictors that preserves all the regression information. When the predictor dimension is large, it is often assumed that only a small number of predictors is informative. In this regard, the sparse SDR is desired to achieve variable selection and dimension reduction simultaneously. We propose a principal logistic regression (PLR) as a new SDR tool and extend it to a penalized version for the sparse SDR. Asymptotic analysis shows that the penalized PLR enjoys the oracle property. Numerical investigation supports the advantageous performance of the proposed method compared to the existing methods.

EO247 Room 217 RECENT ADVANCES ON ANALYSIS OF HIGH-DIMENSIONAL DATA Chair: Xin Zhang

EO0226: **Parsimonious tensor response regression with applications to neuroimaging analysis**
Presenter: **Xin Zhang**, Florida State University, United States

Aiming at abundant scientific and engineering data with not only high dimensionality but also complex structure, we study a regression problem with a multidimensional array (tensor) response and a vector predictor. Applications include, among others, comparing tensor images across groups after adjusting for additional covariates, which is of central interest in neuroimaging analysis. We propose a parsimonious tensor response regression adopting a generalized sparsity principle. It models all voxels of the tensor response jointly, while accounting for the inherent structural information among the voxels. It effectively reduces the number of free parameters, leading to feasible computation and improved interpretation. We achieve model estimation through a nascent technique called the envelope method, which identifies the immaterial information and focuses the estimation based upon the material information in the tensor response. We demonstrate that the resulting estimator is asymptotically efficient, and it enjoys a competitive finite sample performance. We also illustrate the new method on two real neuroimaging studies.

EO1217: **Valid inference on semiparametric estimators with regressors generated by high dimensional regularization**
Presenter: **Shaojun Guo**, Institute of Statistics and Big Data, China
Co-authors: Xinghao Qiao

In many situations, estimation on the parameter of interest often involves covariates, which are not directly observable, but could be estimated from data in the first step. These so-called generated covariates appear in numerous applications, including two-stage nonparametric regression and censored regression models. We focus on the problem where covariates are generated through high dimensional regularization. It turns out that the regularization step has a very serious effect for valid inference on parameters of interest. Our primary interest is to develop a novel regularized approach to generate covariates. The proposed estimator can be shown to be asymptotically normal. To illustrate, we provide several examples to demonstrate the superiority of the proposed approach. This approach is also applicable to linear or nonlinear functionals in other sparse nonparametric high dimensional regression models such as additive or varying coefficient models.

EO1361: **Testing the presence of significant covariates through conditional marginal regression**
Presenter: **Emre Barut**, George Washington University, United States
Co-authors: Huixia Judy Wang, Yanlin Tang

In many statistical applications, researchers may have a priori information on relative importance of the predictors; this information is then used to screen out a majority of the covariates. In these approaches, an important question is whether any of the discarded covariates have predictive powers when the “most relevant” predictors are included in the model. We consider the problem of testing whether none of the discarded covariates is significant conditional on some pre-chosen covariates. We propose a test statistic and show that the proposed statistic has a non-standard asymptotic distribution that can be estimated through bootstrap, giving rise to the conditional adaptive resampling test. We prove the consistency of the test procedure under very general assumptions. Moreover, we illustrate how the suggested test can be used as a stopping rule in forward regression. We

show, through simulation, that the proposed test provides adequate control of family-wise error rate with competitive power; although the asymptotic theory used to calibrate the test assumes a fixed dimension, the test is very efficient even in high-dimensional cases. In further simulation studies, we establish the advantage of the proposed forward regression when covariates are heavily correlated. Finally we demonstrate the effectiveness of the proposed approach through an analysis of an eQTL dataset.

EC1688: **A feature-splitting distributed algorithm for generalized linear models under generalized and group lasso penalties**
Presenter: **Seyoon Ko**, Seoul National University, Korea, South
Co-authors: Donghyeon Yu, Joong-Ho Won

Penalized regression problems with a modest sample size and a large number of variables arise frequently in various applications such as bioinformatics and neuroscience. In this kind of settings, solving the regression problem in a distributed fashion such that each compute node handles a subset of variables is an effective way to handle large-size data. The focus is the generalized linear models with generalized lasso or possibly overlapping group lasso penalties. While for the vanilla lasso and non-overlapping group lasso a feature-splitting distributed algorithm can be easily devised via popular ADMM approaches, for the generalized lasso (including the fused lasso) and the overlapping group lasso, dependencies between the split variables make such a development difficult. A fixed-point iteration algorithm suitable for the assumed distributed setting is proposed. This algorithm does not require matrix inversion. The algorithm is implemented in Apache Spark, which supports fast iterative MapReduce distributed computation. Scalability of the algorithm is demonstrated with neuroscience data, in which logistic regression with the three-dimensional fused lasso penalty is applied to a brain imaging dataset with more than 210,000 variables and less than 900 subjects.

EO225 Room 203 INFERENCE FOR FUNCTIONAL DATA, WITH LIFE SCIENCES APPLICATIONS Chair: Laura Sangalli

EO0247: **Modeling heterogeneity in motor learning using heteroskedastic functional principal components**
Presenter: **Jeff Goldsmith**, Columbia University, United States
Co-authors: Daniel Backenroth, Jeff Goldsmith, Michelle Harran, Juan Cortes, John Krakauer, Tomoko Kitago

A novel method is proposed for estimating population-level and subject-specific effects of covariates on the variability of functional data. We extend the functional principal components analysis framework by modeling the variance of principal component scores as a function of covariates and subject-specific random effects. In a setting where principal components are largely invariant across subjects and covariate values, modeling the variance of these scores provides a flexible and interpretable way to explore factors that affect the variability of functional data. The motivation arises from a novel dataset concerning an experiment assessing upper extremity motor control, and quantifies the reduction in motion variance associated with skill learning.

EO0440: **Detecting synchronisation in EEG- and EMG-signals via boosted functional historical models**
Presenter: **David Ruegamer**, LMU Munich, Germany
Co-authors: Sarah Brockhaus, Kornelia Gentsch, Klaus Scherer, Sonja Greven

In the field of cognitive affective neuroscience, synchronisation analysis of time series data have become very important. With the aim to better understand the link between different psychophysiological measures such as electroencephalography (EEG) and facial electromyography (EMG), our motivating study recorded brain activity (EEG) as well as facial muscle activity (EMG) of 24 participants who were playing a gambling task. The resulting data set yields study setting- and subject-specific time series for each of the two recorded signals, for the analysis of which statistical methods with efficient computational implementation are required. We apply function-on-function regression models to the given time series and estimate the underlying relationship by a component-wise gradient boosting algorithm. Previous attempts to investigate the synchronisation of EEG and EMG signals are, for example, based on coherence or time series analysis, which in many aspects do not provide the flexibility of function-on-function regression approaches. To model the theoretically assumed relationship of both signals in this application, we extend the regression approach from simple historical models to models including random, factor-specific as well as factor-specific random historical effects.

EO0575: **Analysis of data from experiments in human movement science**
Presenter: **Kevin Hayes**, University of Limerick, Ireland

The analysis of high-dimensional high-throughput data collected is considered from experimental investigations examining human movement. In the main we focus on elite standard athletes. Recent advances in technology have opened up opportunities for the simultaneous measurement of muscle activations, force generation and movement patterns. Hitherto, these outcomes have been studied separately, and limitations in sensor technologies have curtailed the ability to measure deep muscle activation. The available data from these studies are any combination of (a) kinematic parameters obtained from either 3D video capture and digitising, or wearable triaxial accelerometers; (b) muscle activity as measures by wearable EMG sensors; (c) high-precision force plate recordings. All these data types lend themselves amiable to statistical modelling using functional data analysis (FDA). The FDA methodologies required for modelling these data types are discussed. The ultimate aim is to better understand how muscle function impacts on biomechanics response and force generated. Also off particular concern is the assessment of coordination between various biomechanical activities.

EO0928: **Partial differential interval-wise testing for the functional data analysis of tongue profiles**
Presenter: **Alessia Pini**, Politecnico di Milano, Italy
Co-authors: Lorenzo Spreafico, Simone Vantini, Alessandro Vietti

The spatiotemporal functional data analysis of a data set of tongue profiles is presented. We test differences in articulation between three syllables sharing the same structure (plosive - rhotic - vowel), for a study on Tyrolean, a German dialect spoken in South Tyrol. The data set is composed by time-varying observations of the sagittal profiles of the tongue during 200ms. The analysis is based on the Partial Differential Interval-Wise Testing (PD-IWT), a non-parametric inferential procedure for functional data. The PD-IWT tests differences between the functional data, and selects the intervals of space and time presenting significant differences between the syllables. The first and second order partial derivatives with respect to space and time are jointly tested, to provide complete information in shape (i.e. tongue curvature and concavity) and time-dynamics (i.e. tongue velocity and acceleration). PD-IWT based comparisons result in an informative and detailed representation of the differences in articulation between the three syllables.

EO079 Room 201 ADVANCES IN TIME SERIES ANALYSIS Chair: Gregory Rice

EO0252: **Testing for randomness in a random coefficient autoregression model**
Presenter: **Lorenzo Trapani**, Cass Business School, United Kingdom
Co-authors: Lajos Horvath
A test is proposed to discern between an ordinary autoregressive model, and a random coefficient one. To this end, we develop a full-fledged estimation theory for the variances of the idiosyncratic innovation and of the random coefficient, based on a two-stage WLS approach. Our results hold irrespective of whether the series is stationary, nonstationary, or on the boundary between the two regimes. In addition to deriving strong rates of convergence, we also, as a technical by-product, present a complete set of results for the boundary case, providing an almost sure lower bound for the growth rate of the series in this case. Building on these results, we develop a randomised test statistic for the null that the coefficient is random, as opposed to the alternative of an AR(1) model.

EO0235: **On the CLT for the discrete Fourier transforms of functional time series**
Presenter: **Clement Cerovecki**, Univ libre de Bruxelles, Belgium
Co-authors: Siegfried Hormann
Functional data often arise by segmenting a continuous time process into natural units, such as days. Then a certain degree of dependence between the observed curves $X_1, \ldots, X_T$ is well expected and, consequently, a thorough statistical investigation requires time series methodology. During recent years functional time series (FTS) analysis has seen an upsurge in the scientific community and diverse related practical and theoretical problems have been addressed. Some of the latest publications are devoted to frequency domain topics for FTS. Motivated by this fact, we study the key ingredient for the frequency domain approach: the discrete Fourier transform (DFT) of the function valued random process. We derive its weak convergence to a complex Gaussian (functional) random element under very mild assumptions.

EO0425: **Dating structural breaks in functional data without dimension reduction**
Presenter: **Ozan Sonmez**, University of California, Davis, United States
Co-authors: Gregory Rice, Alexander Aue
An estimator for the time of a break in the mean of stationary functional data is proposed that is fully functional in the sense that it does not rely on dimension reduction techniques such as functional principal component analysis (fPCA). A thorough asymptotic theory is developed for the estimator of the break date for fixed break size and shrinking break size. The main results highlight that the fully functional procedure performs best under conditions when analogous fPCA based estimators are at their worst, namely when the feature of interest is orthogonal to the leading principal components of the data. The theoretical findings are confirmed by means of a Monte Carlo simulation study in finite samples. An application to one-minute intra-day cumulative log-returns of Microsoft stock data highlights the practical relevance of the proposed fully functional procedure.

EO1142: **Robust multivariate change point analysis based on data depth**
Presenter: **Shojaeddin Chenouri**, University of Waterloo, Canada
Co-authors: Gregory Rice
Modern methods for detecting changes in the scale or covariance of multivariate distributions rely primarily on testing for the constancy of the covariance matrix. These depend on higher order moment conditions, and also do not work well when the dimension of the data is large or even moderate relative to the sample size. We propose a nonparametric change point test for multivariate data using rankings obtained from data depths. As the data depth of an observation measures its centrality relative to the sample, changes in data depth may signify a change of scale of the underlying distribution, and the proposed test is particularly responsive to detecting such changes. We provide a full asymptotic theory for the proposed test statistic under the null hypothesis that the observations are stable, and natural conditions under which the test is consistent. The finite sample properties are investigated by means of a Monte Carlo simulation, and these along with the theoretical results confirm that the test is robust to heavy tails, skewness, and high dimensionality.

EO223 Room 205 CAUSAL INFERENCE IN THEORY AND PRACTICE I Chair: Jonas Peters

EO0266: **Marginalization and reduction of structural causal models**
Presenter: **Stephan Bongers**, University of Amsterdam, Netherlands
Co-authors: Jonas Peters, Bernhard Scholkopf, Joris Mooij
Structural Causal Models, also known as (Non-Parametric) Structural Equation Models, are widely used for causal modelling purposes. One of their advantages is that they allow for cycles, i.e., causal feedback loops. In this work, we give a rigorous treatment of Structural Causal Models. Two different types of variables play a role in SCMs: "endogenous" variables and "exogenous" variables (also known as "disturbance terms"). We define a marginalization operation on SCMs that effectively removes a subset of the endogenous variables from the model. This operation can be seen as projecting the description of a full system to the description of a subsystem. We show that this operation preserves the causal semantics. We also show that in the linear case, the number of exogenous variables can be reduced so that only a single one-dimensional disturbance term is needed for each endogenous variable. This "reduction" can reduce the model complexity significantly and offers parsimonious representations for the linear case. In general, we show that it is always possible to reduce to only one single one-dimensional disturbance term, however this comes at the price of loosing dimensional information. We show that under certain additional conditions such a reduction is not possible.

EO0933: **Causal transfer in machine learning**
Presenter: **Mateo Rojas Carulla**, Cambridge / Max Planck for Intelligent Systems, Germany
Co-authors: Bernhard Scholkopf, Richard Turner, Jonas Peters
Traditional methods for machine learning assume that the training and test data are drawn independently from the same probability distribution. Transfer learning aims to go beyond this paradigm and considers settings in which the data at training and testing can come from different probability distributions, referred to as tasks. Methods of transfer learning try to combine knowledge from several related tasks (or domains) to improve performance on a test task. Inspired by causal methodology, we relax the usual covariate shift assumption and assume that it holds true for a subset of predictor variables: the conditional distribution of the target variable given this subset of predictors is invariant over all tasks. We prove that in an adversarial setting using this subset for prediction is optimal if no examples from the test task are observed. If examples from the test task are available, we provide a method to transfer knowledge from the training tasks and exploit all available predictors for prediction. We introduce a practical method which allows for automatic inference of the above subset. We present results on synthetic data sets and a gene deletion data set.

EO1163: **Identifiability of causal effects from subsampled and mixed-frequency time series**
Presenter: **Ali Shojaie**, University of Washington, United States
Causal inference in multivariate time series is confounded by subsampling in time between the true causal scale and the observed data sampling rate. In practice, this presents challenges for inferring causal interaction between time series due to differences in sampling rates across time series and generally low sampling rates due to technological limitations. Building on recent developments for identifiability of structural vector autoregressive (SVAR) models, we show that when the underlying noise, or shocks, to the system are non-Gaussian, both the parameters of the

true model and its causal structure are identifiable from subsampled and mixed frequency data. Existing results are generalized to the setting where SVAR includes instantaneous interactions, and different time series may be observed at different frequencies.

EO1399: **Model parameterisations using causal quantities**
Presenter: **Robin Evans**, University of Oxford, United Kingdom
Co-authors: Vanessa Didelez
Many causal parameters of interest are marginal quantities: that is, they are formed by averaging over a real or hypothetical population. Several authors have noted the practical difficulties of dealing with such quantities. This is due to the apparent incompatibility of a marginal parameterisation involving the causal quantity of interest and conditional parametric models used for dealing with confounding (either observed or unobserved). In some cases, the so-called g-null paradox implies that it is logically impossible for the conditional models and the marginal null hypothesis to hold simultaneously. This means that even simulating from the null model to test new methods is not always possible, let alone performing likelihood-based inference. We adapt a previous marginal parameterisations to causal models, allowing us to parameterise a wide range of causal models including marginal structural models (MSMs), Cox MSMs and structural nested models. This makes it easy to simulate from and fit models; to introduce possibly high-dimensional individual-level covariates; and to include additional assumptions such as stationarity or symmetry. Our approach also avoids the null paradox where possible, and makes transparent when it is unavoidable by assumption.

EO723 Room 212 DEPENDENCE MODELS AND COPULAS III Chair: Wolfgang Trutschnig

EO0352: **Inference of elliptical copula generators**
Presenter: **Alexis Derumigny**, ENSAE-CREST, France
Co-authors: Jean-David Fermanian
When one chooses elliptical copulae to model the dependence of a given distribution, there is two parameters of interest : the correlation matrix and the generator of the corresponding elliptical distribution. In the general case, both have to be estimated. The correlations can be estimated directly from the pseudo-sample. We propose nonparametric estimators for the generator, by iterating estimators specific to elliptical distributions, or by using M-estimation. We give also some empirical results and theorical properties.

EO0472: **A semiparametric and location shift copula based mixture model**
Presenter: **Gildas Mazo**, Universite Catholique de Louvain, Belgium
Modeling of distributions mixtures has rested on Gaussian distributions and or a conditional independence hypothesis for a long time. Only recently have researchers begun to construct and study broader generic models without appealing to such hypotheses. Some of these extensions use copulas as a tool to build flexible models, as they permit to model the dependence and the marginal distributions separately. But this approach also has drawbacks. First, it greatly much the number of choices the practitioner has to make, and second, marginal misspecification may loom on the horizon. This paper aims at overcoming these limitations by presenting a copula based mixture model which is semiparametric. Thanks to a location shift hypothesis, semiparametric estimation, also, is feasible, allowing for data adaptation without any modeling effort.

EO0157: **Building conditionally dependent parametric factor copulas**
Presenter: **Nathan Uyttendaele**, Catholic University of Louvain, Belgium
So far, factor copulas induce conditional independence with respect to a latent factor. These factor copulas are extended through two representations allowing for a varying conditional dependence structure. Estimation issues are discussed, as well as how to distinguish between conditionally independent and conditionally dependent factor copulas, thanks to a novel statistical test which does not assume any parametric form for the conditional dependence structure.

EO1221: **Estimating non-simplified vine copulas using penalized splines**
Presenter: **Christian Schellhase**, Bielefeld University, Germany
Co-authors: Fabian Spanhel
Vine copulas (or pair-copula constructions) have become an important tool for high-dimensional dependence modeling. Typically and without any consideration, so called simplified vine copula models are estimated where bivariate conditional copulas are approximated by bivariate unconditional copulas. We present the first non-parametric estimator of a non-simplified vine copula that allows for varying conditional copulas using penalized hierarchical B-splines. Throughout the vine, we test for the simplifying assumption in each knot, establishing a data driven non-simplified vine copula.To overcome the curse of dimensionality, we approximate conditional copulas with more than one conditioning argument by a conditional copula with the first principal component as conditioning argument. An extensive simulation study is conducted, showing a substantial improvement in the out-of-sample Kullback-Leibler divergence if the variation in the conditional copula is not negligible, especially for non-simplified vine copulas. As application we present a classification of a banknote authentication data set, demonstrating the potential benefit that can be achieved when conditional copulas are modeled.

EO595 Room 210 DIMENSION REDUCTION FOR REGRESSION Chair: Stephane Girard

EO0470: **Student sliced inverse regression**
Presenter: **Florence Forbes**, INRIA, France
Co-authors: Alessandro Chiancone, Stephane Girard
Sliced Inverse Regression (SIR) has been extensively used to reduce the dimension of the predictor space before performing regression. SIR is originally a model free method but it has been shown to actually correspond to the maximum likelihood of an inverse regression model with Gaussian errors. This intrinsic Gaussianity of standard SIR may explain its high sensitivity to outliers as observed in a number of studies. To improve robustness, the inverse regression formulation of SIR is therefore extended to non-Gaussian errors with heavy-tailed distributions. Considering Student distributed errors it is shown that the inverse regression remains tractable via an Expectation-Maximization (EM) algorithm. The algorithm is outlined and tested in presence of outliers both on simulated and real data, showing improved results in comparison to a number of other existing approaches.

EO0719: **Piecewise constant models: From** $\log n$ **to** $\log\log n$
Presenter: **Fang Han**, University of Washington, United States
Co-authors: Chao Gao, Cun-Hui Zhang
The piecewise constant normal mean estimation problem is considered. The observations are two-fold. First, assuming k pieces in the model, when there is no additional shape constraint, we show a phase transition exists: When k increases from 2 to higher, the minimax rate of convergence increases from log log n to $k\log(en/k)$. Such a result is absent in change point detection and nonparametric regression literature, and motivates retrospect on the "sparsity assumption" popular in high dimensional statistics. Secondly, we show, when some additional shape constraints are posed, the minimax rate can be further improved. In particular, adding isotonic constraint to piecewise constant models gives an improved $k\log\log(16n/k)$ minimax rate. The obtained result is in contrast to Kiefer's Theorem, and generalizes these in a recent work.

EO0412: **Robust Variable Section with the sparse shooting S**
Presenter: **Ines Wilms**, KU Leuven, Belgium
Co-authors: Christophe Croux
In multiple regression analysis, a response variable is predicted based on a set of many predictor variables. We are particularly interested in high-dimensional multiple regression analysis where a large number of predictors is available. The lasso has become a popular estimator to reduce the dimensionality of such high-dimensional regression models by imposing sparsity on the estimated regression parameters. As such, the lasso performs variable selection since it only keeps a few predictors and discards the remaining predictors by setting their respective parameter estimates to zero. The lasso is, however, not a robust estimator. Nevertheless, outliers, i.e. atypical observations, frequently occur in high-dimensional data sets. Therefore, we propose a cellwise robust lasso estimator, the sparse shooting S. This estimator can deal with cellwise contamination, where many cells of the design matrix of the predictor variables may be outlying. Moreover, the sparse shooting S is computable in high-dimensional settings with more predictors than observations and it gives sparse parameter estimates. We compare its performance to several other sparse and/or robust regression estimators.

EO1108: **Variable importance assessment in sliced inverse regression for variable selection**
Presenter: **Jerome Saracco**, ENSC - Bordeaux INP, France
Co-authors: Ines Jlassi
The focus is on treating the relationship between a dependent variable y and a $4p$-dimensional covariate x in a semiparametric regression model. Since the purpose of most social, biological or environmental science research is the explanation, the determination of the importance of the variables is a major concern. It is a way to determine which variables are the most important when predicting y. Sliced inverse regression (SIR) methods allows us to reduce the space of the covariate x by estimating the directions that form an effective dimension reduction (EDR) space. The aim is to propose a computational method based on importance variable measure (only relying on the EDR space) in order to select the most useful variables in SIR model. The numerical behavior of this approach, implemented in R, is studied on a simulation study. An illustration on a real data is also provided.

EO551 Room 007 SMOOTHING METHODS FOR COMPLEX DATA Chair: Maria Xose Rodriguez-Alvarez

EO0501: **Statistics of the human face**
Presenter: **Liberty Vittert**, University of Glasgow, United Kingdom
Co-authors: Adrian Bowman, Stanislav Katina
Three-dimensional surface imaging, through laser-scanning or stereo-photogrammetry, provides high-resolution data defining the surface shape of objects. Using a human face as this object, each image corresponds to an observation, a manifold, represented by a triangulated point cloud. In an anatomical setting this can provide invaluable quantitative information. Particular applications vary widely including success or failure of cosmetic/reconstructive plastic surgery, facial recognition, facial asymmetry, concepts of sexual dimorphism, and even the survival of mussels (food we consume) given climate change. However, the initial challenge is to characterize these complex surfaces, without laborious manual intervention. Surface curvature provides the key information in doing this, allowing for the creation of a surface mask replicable throughout all these objects. Once the full surface representation has been obtained, the new issue arises of how to best characterize and visualize the differences in shape. The issues involved with analysis of this data and multiple visualization methods will be discussed and illustrated.

EO0963: **Fast estimation of adaptive P-spline models**
Presenter: **Maria Durban**, Universidad Carlos II de Madrid, Spain
Co-authors: Maria Xose Rodriguez-Alvarez, Dae-Jin Lee, Paul Eilers
In many applications it is desirable and needed to adapt smoothness locally to the data, and adaptive P-splines have been suggested. However, the existing estimation procedures can be very slow or even unstable. We extend a previous method, and generalize the Separation of Anisotropic Penalties (SAP) algorithm to deal with the proposed adaptive penalty in one or more dimensions. The practical performance of the algorithm is evaluated by means of simulations, and comparisons with alternative methods are made on the basis of the mean square error criterion and the computing time. Finally, we illustrate our proposal with the analysis of two datasets: one corresponds to photon counts of diffracted x-ray radiation as a function of the angle of diffraction, and the other in which we consider a spatio-temporal adaptive penalized for modelling the firing rate of visual neurons.

EO1060: **Disaggregation of spatial and spatio-temporal counts in epidemiology: A penalized composite link model approach**
Presenter: **Dae-Jin Lee**, BCAM - Basque Center for Applied Mathematics, Spain
Co-authors: Diego Ayma, Maria Durban, Paul Eilers
Mortality data provide valuable information for the study of the spatial distribution of mortality risk, in disciplines such as spatial epidemiology and public health. However, they are frequently available in an aggregated form over irregular geographical units, hindering the visualization of the underlying mortality risk. Also, it can be of interest to obtain mortality risk estimates on a finer spatial resolution, such that they can be linked to potential risk factors that are usually measured in a different spatial resolution. We propose the use of the penalized composite link model and its mixed model representation. This model considers the nature of mortality rates by incorporating the population size at the finest resolution, and allows the creation of mortality maps at a finer scale, thus reducing the visual bias resulting from the spatial aggregation within original units. We also extend the composite link model to the spatio-temporal case where counts are aggregated spatially and temporally. We illustrate our proposal with several data.

EO1011: **Non-linear regularized regression in multi-omic applications**
Presenter: **Mar Rodriguez-Girondo**, Leiden University Medical Center, Netherlands
In the last decade, much attention has been devoted to build linear prediction models for health traits based on omics data, such as transcriptomics, proteomics, and metabolomics. Incorporating omic sets of predictors into prediction of traits is challenging due to high dimensionality and existing correlations between features within omic datasets. Linear regularized regression models have been broadly used for building prediction models in high-dimensional settings. However, it is unclear if the linearity assumption is realistic in omic applications. Moreover, it is common nowadays that clinical studies include several sets of such omics markers available for each patient, measuring different levels of biological variation. As a result, main challenges in predictive research are: 1) integration of different sources of omic biomarkers for the prediction of health traits and 2) relaxing linearity assumption to handle more complex relations among the available predictors. We review several approaches for the combination of omic markers, all based on double cross-validation and flexible regularized regression models, based on penalized splines. We illustrate the methods through the analysis of transcriptomics and metabolomics as predictors of obesity using data from the Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study, a population-based cohort, from Finland.

EO505 Room S24 RECENT ADVANCES IN MIXED EFFECTS MODELLING Chair: Estelle Kuhn

EO0537: **On BIC for mixed-effects models**
Presenter: **Maud Delattre**, AgroParisTech, France
Co-authors: Marie-Anne Poursat
The Bayesian Information Criterion (BIC) is widely used for variable and random effects selection in mixed effects models. However, its expression is unclear in typical situations of mixed effects models, where simple definition of the sample size is not meaningful. An appropriate BIC expression is derived, that is consistent with the random effect structure of the mixed effects model. The behavior of the proposed criterion is illustrated through simulation experiments and a case study. The use of the new criterion is recommended as an alternative to various existing BIC versions that are implemented in available softwares.

EO0740: **Testing variance components in nonlinear mixed effects models: Application to plant growth modelling**
Presenter: **Charlotte Baey**, CentraleSupelec, France
Co-authors: Paul-Henry Cournede, Estelle Kuhn
Mixed-effects models are commonly used in a large variety of disciplines to account for and describe the inter-individual variability of a population. A key question when adjusting a model to a population in this framework, is to understand and identify the parameters carrying the variability, i.e. those that can be considered constant in the population, referred to as 'fixed-effects', and those that vary among individuals, referred to as 'random effects'. This is particularly the case when the underlying model is a mechanistic model, for which parameters have a biological or physical meaning, for example. From a statistical point of view, this can be rephrased as a hypothesis testing problem, where one is interested in testing if the variances of a subset of the random effects are non-null. The issue of variance components testing in the context of linear mixed models has been addressed by several authors, and relates more generally to constrained statistical inference and hypothesis testing. We consider a testing procedure based on the likelihood ratio test statistics in the context of non linear mixed effects models. We establish theoretical results for the asymptotic distribution of the test statistics. Several special cases are detailed as well as general methods for a practical computation of the test.The methodology is applied to a nonlinear mixed model of plant growth, and tested on simulated and real data.

EO0834: **Wavelet-based shape invariant estimation in functional mixed-effects modelling using warping functions**
Presenter: **Joyce Madison Giacofci**, IRMAR - Universite Rennes, France
Co-authors: Gerda Claeskens, Irene Gijbels, Maarten Jansen
Nowadays scientific studies increasingly yield functional data, in which the ideal units of observation are curves. We consider data that consist of set of curves recorded on individuals. In a multi-individual context, curves are often recorded on subject-specific non-regular time grids and variations both in phase and in amplitude are commonly encountered. While such variations are widely studied separately in the functional setting, only little attention has been devoted to the study of simultaneous amplitude and phase variations. We propose a functional warped mixed-model that accounts for individual deviations in amplitude around a main functional pattern by considering functional random effects, whereas modeling of individual phase deviations is done through nonlinear warping functions. A wavelet representation of both the main pattern and individual amplitude variations, adapted to non-homogeneous functions is considered. Non-linear subject-specific phase variations are modeled through a monotone Hermite spline decomposition, allowing a close relation with the landmark structure. Estimation is done through a maximum-likelihood method using a EM-type algorithm. The non-linearity of the mixed model yields the E-step intractable and a stochastic approach is used for predictions of individual amplitude deviation coefficients and individual warping parameters. The performance of the procedure is investigated through a simulation study and illustrated on real data examples.

EO1124: **Parametric inference for discrete observations of diffusions with mixed effects**
Presenter: **Catherine Laredo**, INRA, France
Co-authors: Valentine Genon-Catalot, Maud Delattre
Independent stochastic processes are considered defined by a stochastic differential equation with mixed effects effects in the drift term and in the diffusion coefficient. Each process is discretely observed on a fixed time interval with length T at n regularly spaced discrete times. We consider parametric distributions for the random effects and estimate the unknown parameters from the observations as both the number of sample paths N and the number of observations per sample path n goes to infinity. We study cases that give rise to explicit approximations of the likelihood functions: (1) Multidimensional Gaussian random effect in the drift and fixed effect in the diffusion coefficient; (2) Fixed effect in the drift and random effect in the diffusion coefficient with inverse Gamma distribution. In each case, we characterize the asymptotic behaviour of the associated estimators.

EO061 Room 211 CHANGE POINT DETECTION IN TIME SERIES Chair: Herold Dehling

EO0567: **A Wilcoxon-type test for distinguishing between long-range dependence and short-range dependence with a change in mean**
Presenter: **Carina Gerstenberger**, Ruhr-Universitaet Bochum, Germany
For some applications it is necessary to know whether a time series is short- or long-range dependent, and quite a number of test statistics discriminate between these dependence structures. However, most of these tests spuriously reject the hypothesis of short-range dependence when change-points are present. We propose a Wilcoxon-type testing procedure that is able to distinguish between long-range dependence and short-range dependence with a change in mean. This procedure is based on an previous work which used the CUSUM statistic. We derive the asymptotic distribution of the Wilcoxon-type testing procedure under the hypothesis of L_1 near epoch dependent random variables with a change in mean at unknown time. This procedure is less sensitive to outliers in the data compared to the CUSUM-type testing procedure. We also introduce a robust Wilcoxon-type change-point estimator and derive its consistency and the rate of consistency. Simulation results show that the Wilcoxon-type testing procedure and the Wilcoxon-type change-point estimator behave as good as their CUSUM-type counterparts under normal conditions but in contrast are nearly not affected by outliers.

EO1017: **Subsampling-based change-point detection in LRD time series**
Presenter: **Annika Betken**, Ruhr-Universitat Bochum, Germany
Co-authors: Martin Wendler
A robust change-point test is considered based on the Wilcoxon two-sample rank statistic to identify changes in the mean of dependent data. The scaling needed to ensure convergence of the corresponding test statistic to a non-degenerate limit usually depends on unknown parameters. Estimation of an unknown standardization may be avoided by an application of self-normalized test statistics. However, under long-range dependence the asymptotic distribution of self-normalized statistics still depends on parameters that characterize the intensity of dependence in the data. Approximating the limit distribution of self-normalized statistics by subsampling procedures bypasses estimation of these quantities. It is shown that the so-called sampling window method is valid for general statistics applied to long-range dependent subordinated Gaussian processes which satisfy mild regularity conditions. Furthermore, we investigate the finite sample performance of subsampling-based change point tests in a simulation study. We compare it to the performance of the usual testing procedure which generates test decisions on the basis of critical values that arise from the asymptotic distribution of the respective statistic.

EO1047: **Structural break detection using Fourier methods**
Presenter: **Max Wornowizki**, TU Dortmund University, Germany
Co-authors: Herold Dehling, Roland Fried

A new class of tests for checking the constancy of a distributional parameter such as the variance for a sequence of random variables is presented. The corresponding test statistics are based on a Fourier type transformation of blockwise estimators of this parameter. Different weight functions result in different test statistics, which are all given by simple explicit formulae. Assuming independence and piecewise identical distributions, testing is conducted applying the permutation principle or using asymptotic results. To prove the latter the test statistics are viewed as U-statistics constructed from the blockwise parameter estimators. Since the distribution of these estimators depends on the sample size, a new LLN and a CLT are proven. The asymptotic and the permutation test are compared to other tests for constant variance in extensive Monte Carlo experiments and illustrated in an application. In comparison to their competitors the new methods offer good power particularly in the case of multiple structural breaks and estimate the positions of the structural breaks adequately.

EO1351: **Estimation methods for the long memory parameter under a change in the mean**
Presenter: **Ieva Zelo**, TU Dortmund, Germany
Co-authors: Aeneas Rooch, Roland Fried

Time series analysis is often based on the assumption of stationarity. However, the estimation of any time-constant parameter is affected heavily by the presence of a change in the mean. When analyzing time series which are supposed to exhibit long memory, a basic issue is the estimation of the long memory parameter, for example the Hurst parameter H. Conventional estimators of H easily lead to spurious detection of long memory if the time series includes a shift in the mean. This defect has fatal consequences in change-point problems: Tests for a level shift rely on H, which needs to be estimated before, but this estimation is distorted by the level shift. We investigate techniques to adapt estimators of H to the case that the time series includes a jump and compare them via simulations. Based on our results, we recommend an overlapping blocks approach: If one uses a consistent estimator, the adaption will preserve this property and it performs well in simulations. The blocks technique is also useful for the estimation of other parameters, such as the variance, as will be illustrated.

EO646 Room S22 TESTING AND INFERENCE IN NONPARAMETRIC MODELS OF PRODUCTION Chair: Paul Wilson

EO0577: **Quality of European universities: Model uncertainty, endogeneity and testing of unobserved heterogeneity**
Presenter: **Cinzia Daraio**, University of Rome La Sapienza, Italy
Co-authors: Leopold Simar, Paul Wilson

Up-to-date econometric techniques are developed to test for the relevance of latent heterogeneity factors and to assess their impact on the efficient boundary of the production set. We extend a previous framework to account for endogeneity and latent heterogeneity. Our approach is implemented in the context of the quantitative assessment of European Higher Education systems. Higher Education Institutions (HEIs) carry out a complex production process. Multiple activities, such as teaching, research and third mission are realized by combining different inputs (or resources): human capital, financial stocks and infrastructures; to produce heterogeneous outputs, such as: undergraduate degrees, PhD degrees, scientific publications, citations, service contracts, patents, spin off and so on, within an heterogeneous environment in which size and subject mix play also an important role. Within this process, the quality of HEIs, which is an unobserved factor, play a crucial role. The main objective is to estimate this latent factor and assess its impact on the European Higher Education system.

EO0600: **Bandwidth selection issues for nonparametric conditional efficiency estimation**
Presenter: **Luiza Badin**, Bucharest University of Economic Studies, Romania
Co-authors: Cinzia Daraio, Leopold Simar

Recent theoretical and empirical studies involving nonparametric conditional frontier models stress the importance of conditional efficiency measures as the only way to treat appropriately the presence of environmental factors in a production process. Conditional efficiency measures provide a flexible tool to evaluate the impact of external-environmental variables in a completely nonparametric framework, without any restrictive assumptions on the data generating process. The bandwidths for the conditioning variables play a crucial role in the process of estimating these efficiency measures, since they tune the localization for computing the conditional efficiencies (FDH and/or DEA). We present and compare, through a Monte-Carlo simulation study, the performances of several different approaches, all asymptotically equivalent, to calculate the bandwidths necessary to estimate the conditional efficiency scores. The main objective is to advice on the best performing approach in terms of the resulting estimates of conditional efficiency and to provide practical suggestions for a simple implementation for practitioners.

EO0665: **Cross-section dependence and latent heterogeneity: A conditional nonparametric frontier analysis**
Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy

A unified non parametric framework is proposed for accommodating simultaneously the problem of model specification uncertainty, potential endogeneity and cross-section dependence in modelling technical efficiency in frontier models. We emphasize the importance to take into account the observed and observed endogeneity in assessing the economic performance across countries and to evaluate the process of technological catching up towards the global production frontier . By using a dataset of 40 countries over 1970-2007, we estimate the global frontiers. We extend existing methodological tools - robust frontier in non parametric location-scale models - to examine these interrelationships. We use a flexible nonparametric two-step approach on conditional efficiencies to eliminate the dependence of production inputs/outputs on common factors. We emphasize the usefulness of pre-whitened inputs/outputs to obtain more reliable measure of productivity and efficiency to better investigate the impact of human capital on the catching-up productivity process. Then, we take into account the problem of unobserved heterogeneity and endogeneity in the analysis of the influence of human capital on the production process by extending an instrumental nonparametric approach to account also for cross section and time dependence.

EO0891: **On cross-validation in regression**
Presenter: **Paul Wilson**, Clemson University, United States
Co-authors: Byeong Park, Leopold Simar, Valentin Zelenyuk

It is well known that nonparametric regression and density estimates are sensitive to the choice of bandwidth. Least-squares cross-validation (LSCV) is frequently used to optimize bandwidths, but the LSCV function is often rough, with perhaps many local minima. In addition, the search for a minimum must often be constrained by a lower bound to avoid a minimum at 0. These issues are especially problematic in Monte Carlo and simulation settings. Moreover, the problems are compounded when two or more bandwidth values must be selected. We examine the efficacy of various rules for choosing among multiple minima and setting lower bounds for the search that have been proposed in the literature and implemented in widely-used software. In addition, we propose new strategies for selecting bandwidth values via LSCV and examine their performance in Monte Carlo experiments.

EO259 Room 206 ROBUST COVARIANCE MATRIX ESTIMATION OF HIGH-DIMENSIONAL DATA Chair: Esa Ollila

EO0661: **Fitting generalized multivariate Huber losses**
Presenter: **Ami Wiesel**, Hebrew University, Israel
Co-authors: Eli Peker
A class of generalized multivariate Huber (GMH) loss functions is considered. Our goal is computationally efficient parameter estimation in linear models contaminated by non Gaussian yet correlated noise. We define a class of convex GMH loss functions with structured covariance matrices and flexible regularization parameters. This framework includes the classical weighted least squares and Hubers function as special cases. Next, we assume access to a secondary dataset of independent noise realizations, and we use this data to choose the best GMH function associated with the data. We use this fitted function to perform parameter estimation in a linear model by solving a simple convex optimization problem. We demonstrate the advantages of our proposed technique in heavy tailed correlated noise distributions.

EO1188: **Robust RMT for covariance matrix estimation and applications to signal processing**
Presenter: **Frederic Pascal**, CentraleSupelec, France
Recent advances are discussed on robust covariance matrix estimation of multivariate high-dimensional observations. We consider the case where the sample size n is of the same order than the dimension p of the data set. More precisely, the eigenvalues distribution of robust covariance matrix estimates is analyzed in the asymptotic regime where both n and p tends to infinity with a ratio p/n tending to a constant $c \in (0,1)$. Robust estimation is essential partly because outliers are more difficult to glean from high-dimensional data sets, but also due to an increase of impulsive measurement environments and outliers in practical sensing systems. On the other hand, to tackle the high dimensionality of the data, we build on recent results from the RMT to derive the asymptotics of the robust covariance matrix estimates. This leads to many significant and profound differences in asymptotic analysis. More precisely, this approach allows us to obtain closed-form expressions for regularized robust estimators, which is of particular interest for optimizing the regularized (penalty) parameters. Moreover, it will be shown that the RMT results provide similarities with the ones obtained in the classical asymptotic regime. To illustrate theoretical derivations, applications will focus on signal detection and source localization through direction of arrival estimation.

EO0812: **Automatic diagonal loading for Tyler's robust covariance estimator**
Presenter: **Teng Zhang**, University of Central Florida, United States
Co-authors: Ami Wiesel
An approach of regularizing Tyler's robust M-estimator of the covariance matrix is proposed. We also provide an automatic choice of the regularization parameter in the high-dimensional regime. Simulations show its advantage over the sample covariance estimator and Tyler's M-estimator when data is heavy-tailed and the number of samples is small. Compared with the previous approaches of regularizing Tyler's M-estimator, our approach has a similar performance and a much simpler way of choosing the regularization parameter automatically.

EO0808: **Simultaneous penalized M-estimation of covariance matrices using geodesically convex optimization**
Presenter: **Esa Ollila**, Aalto University, Finland
Co-authors: David Tyler, Ilya Soloveychik, Ami Wiesel
A common assumption when sampling p-dimensional observations from K distinct group is the equality of the covariance matrices. We propose two penalized M-estimation approaches for the estimation of the covariance or scatter matrices. The first approach begins by generating a pooled M-estimator of scatter based on all the data, followed by a penalised M-estimator of scatter for each group, with the penalty term chosen so that the individual scatter matrices are shrunk towards the pooled scatter matrix. In the second approach, we minimize the sum of the individual group M-estimation cost functions together with an additive joint penalty term which enforces some similarity between the individual scatter estimators, i.e. shrinkage towards a mutual center. In both approaches, we utilize the concept of geodesic convexity to prove the existence and uniqueness of the penalized solution under general conditions. We consider three specific penalty functions based on the Euclidean, the Riemannian, and the Kullback-Leibler distances. In the second approach, the distance based penalties are shown to lead to estimators of the mutual center that are related to the arithmetic, the Riemannian and the harmonic means of positive definite matrices, respectively. A penalty based on an ellipticity measure is also considered which is particularly useful for shape matrix estimators. Fixed point equations are derived penalty and regularized discriminant analysis application is considered.

EO015 Room 215 JOINT MODELLING Chair: Thomas Kneib

EO0691: **Boosting joint models**
Presenter: **Andreas Mayr**, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
Co-authors: Elisabeth Waldmann
Joint models are commonly estimated via likelihood based expectation maximization or Bayesian approaches. Drawbacks of these frameworks are that they do not allow for automated variable selection and are not feasible in high-dimensional data situations. We propose a predictor-wise gradient boosting algorithm overcoming these shortcomings being able to simultaneously estimate and select predictors for joint models in potentially high-dimensional data. The new algorithm cycles through the different additive predictors and applies simple base-learners for each candidate variable while including in each iteration only the best-performing base-learner. When the algorithm is stopped before convergence, base-learners that have never been selected are effectively excluded from the final model, leading to data-driven variable selection.

EO0689: **Combining distributional regression with joint modelling**
Presenter: **Elisabeth Waldmann**, Friedrich-Alexander-Universitaet Erlangen-Nuernberg, Germany
When modelling repeated measurement and time to event data simultaneously, analysis is often based on combining random effects regression on the mean with survival analysis. The assumptions made by this model, however, are not necessarily met, when analysing real life data sets. Especially the assumption of equidispersion is too strict in many cases. In simple longitudinal setups this issue is often dealt with distributional regression, which introduces separate predictors for all parameters of the distribution. The aim is to combine the setup of a Bayesian distributional model with a survival model, linking both, the conditional mean and conditional variance to the risk of event and thus developing a new approach towards distributional joint modelling.

EO0586: **Joint modeling of multivariate longitudinal and survival data**
Presenter: **Christel Faes**, Hasselt University, Belgium
Co-authors: Carmen Cadarso Suarez, Ipek Guler
Joint modeling of a longitudinal biomarker with a survival outcome has become very popular in recent years, as the longitudinal profile can inform us about a subjects risk of a (later) survival outcome. Several approaches have been investigated in literature, including hierarchical modeling of the longitudinal and survival process via shared latent random effects. As typically not a single biomarker, but multiple biomarkers are being investigated, it is worthwhile to investigate whether several biomarkers together can (better) inform us about the survival process. Already with

two or three biomarkers, computational difficulties arise, and alternative estimation methods to the maximum likelihood methods need to be investigated. A two-stage pairwise likelihood approach was investigated for this purpose.

EO0429: **Joint modelling of multivariate longitudinal and time-to-event data**
Presenter: **Graeme Hickey**, University of Liverpool, United Kingdom

Research into joint modelling methods has grown substantially over recent years. Previous research has predominantly concentrated on the joint modelling of a single longitudinal outcome and a single time-to-event outcome. In clinical practice, the data collected will be more complex, featuring multiple longitudinal outcomes and/or multiple, recurrent or competing event times. Harnessing all available measurements in a single model is advantageous and should lead to improved and more specific model predictions. Notwithstanding the increased flexibility and better predictive capabilities, the extension of the classical univariate joint modelling framework to a multivariate setting introduces a number of technical and computational challenges. These include the high-dimensional numerical integrations required, modelling of multivariate unbalanced data, and proper estimation of standard errors. Consequently, software capable of fitting joint models to multivariate data is lacking. Building on recent methodological developments, we extend the classical joint model to multiple continuous longitudinal outcomes, and describe how to fit it using a Monte Carlo Expectation-Maximization algorithm with antithetic simulation for variance reduction. The development of an R package will be discussed. An application to a recent clinical trial dataset will be presented.

EO553 Room 204 HIGH-DIMENSIONAL INFERENCE IN SPATIAL MODELS AND PANEL DATA Chair: Carles Breto

EO0768: **Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension**
Presenter: **Abhimanyu Gupta**, University of Essex, United Kingdom
Co-authors: Peter Robinson

Pseudo maximum likelihood estimates are developed for higher-order spatial autoregressive models with increasingly many parameters, including models with spatial lags in the dependent variables and regression models with spatial autoregressive disturbances. We consider models with and without a linear or nonlinear regression component. Sufficient conditions for consistency and asymptotic normality are provided, the results varying according to whether the number of neighbours of a particular unit diverges or is bounded. Monte Carlo experiments examine finite-sample behaviour.

EO0381: **A spatio-temporal Markov switching model for the detection of influenza outbreaks**
Presenter: **David Conesa**, Universitat de Valenci, Spain
Co-authors: Ruben Amoros, Antonio Lopez-Quilez, Miguel Martinez-Beneito

Influenza dispersion is related with climate variables and spreads person to person, which suggests a spatio-temporal evolution of the incidence. We present a spatio-temporal extension of the Bayesian Markov switching model over the differentiated rates for the detection of influenza epidemic outbreaks previously suggested. The variable of the Markov switching model represents the epidemic and non-epidemic states. The non-epidemic state differentiated rates only share a common mean for each time. The rates on the epidemic state are spatially and temporally related through Gaussian Markov random fields. This new proposal has been compared with the previous one and offers better scores in terms of the continuous rank probability score by approximate cross-validatory predictive assessment.

EO1008: **Real-time surveillance for abnormal events: The case of influenza outbreaks**
Presenter: **Yao Rao**, The University of Liverpool, United Kingdom
Co-authors: Brendan McCabe

A method of surveillance using deviations from probabilistic forecasts is introduced. Realised observations are compared with probabilistic forecasts and the "deviation" metric is based on low probability events. If an alert is declared, the algorithm continues to monitor until an all-clear is announced. Specifically, the problem of syndromic surveillance for influenza (flu) is addressed with the intention of detecting outbreaks, due to new strains of viruses, over and above the normal seasonal pattern. The syndrome is hospital admissions for flu like illness and hence the data are low counts. In accordance with the count properties of the observations, an integer valued autoregressive process is used to model flu occurrences. Monte Carlo evidence suggests the method works well in stylised but somewhat realistic situations. An application to real flu data indicates that the ideas may have promise. The model estimated on a short run of training data, did not declare false alarms, when used with new observations deemed in control, ex post. The model easily detected the 2009 H1N1 outbreak.

EO1157: **Panel data analysis via mechanistic models**
Presenter: **Carles Breto**, The University of Michigan, United States
Co-authors: Edward Ionides, Aaron King

Panel or longitudinal analysis of dynamic systems is becoming more widespread thanks to the increasing availability of data and to recent advances in statistical methodology that aim at dispensing with linearity and Gaussianity assumptions. Examples of such statistical tools are iterated filtering algorithms. However, these algorithms have so far been developed with a multivariate time series framework in mind, without explicitly considering a panel setting. Panel settings can be useful both to disentangle within-individual and between-individual features and to deal with parameter biases or weak identifiabilities through partial pooling of data. In the context of nonlinear, non-Gaussian panels, we extend existing methodology with a novel panel iterated filtering algorithm and apply it to model infectious disease dynamics.

EO005 Room 006 STATISTICAL NETWORK MODELING Chair: Harry Crane

EO0893: **Edge exchangeability: A new foundation for network modeling**
Presenter: **Walter Dempsey**, University of Michigan, United States
Co-authors: Harry Crane

Exchangeable models for vertex labeled graphs cannot replicate the large sample behaviors of sparsity and power law degree distributions observed in many network datasets. Out of this mathematical impossibility emerges the question of how network data can be modeled in a way that reflects known empirical behaviors and respects basic statistical principles. To address this issue, we introduce the principle of edge exchangeability, which is more natural for most applications and admits models for networks with sparse and/or power law structure. The vertices in an edge exchangeable network arrive in size-biased order according to their degree, further explaining why vertex exchangeability is an untenable assumption for many applications.

EO1030: **Estimating the number of communities in networks by spectral method**
Presenter: **Can Minh Le**, University of California, Davis, United States
Co-authors: Liza Levina

Community detection is a fundamental problem in network analysis with many methods available to estimate communities. Most of these methods assume that the number of communities is known, which is often not the case in practice. We propose to estimate the number of communities using

spectral properties of the deformed Laplacian. We show that the method performs well under several models and a wide range of parameters. It is also more accurate and computationally efficient than several existing methods for estimating the number of communities.

EO0649: **COSTNET, networks and detecting epistatic selection**
Presenter: **Ernst Wit**, University of Groningen, Netherlands
Statistical Network Science is one of the hot topics of this moment. Network phenomena are widespread and a multitude of network data formats appear in a variety of applications. Often the applied scientist has only a limited awareness of the available modelling and inference techniques available. We will describe how recombinant Inbred Lines (RILs) derived from divergent parental lines can display extensive segregation distortion and long-range linkage disequilibrium (LD) between distant loci on same or different chromosomes. The reconstruction of these interaction networks from observations of pair-wise marker-marker correlations or pair-wise genotype frequency distortions is challenging as multiple testing approaches are under-powered and true long-range LD is difficult to distinguish from drift, particularly in small RIL panels. We develop an efficient method for reconstructing an underlying network of genomic signatures of high-dimensional epistatic selection from multi-locus genotype data.

EO1752: **Community detection in degree-corrected block models**
Presenter: **Anderson Zhang**, Yale University, United States
Co-authors: Chao Gao, Harrison Zhou, Zongming Ma
Community detection is a central problem of network data analysis. Given a network, the goal of community detection is to partition the network nodes into a small number of clusters, which could often help reveal interesting structures. The present paper studies community detection in Degree-Corrected Block Models (DCBMs). We first derive asymptotic minimax risks of the problem for a misclassification proportion loss under appropriate conditions. The minimax risks are shown to depend on degree-correction parameters, community sizes, and average within and between community connectivities in an intuitive and interpretable way. In addition, we propose a polynomial time algorithm to adaptively perform consistent and even asymptotically optimal community detection in DCBMs.

EO121 Room 202 RECENT ADVANCES IN STATISTICAL DEPTH Chair: Davy Paindaveine

EO0988: **A topologically valid definition of depth for functional data**
Presenter: **Heather Battey**, Imperial College London and Princeton University, United Kingdom
Co-authors: Alicia Nieto-Reyes
An axiomatic definition of depth for functional data is presented which recognises topological features such as continuity, smoothness and contiguity. Functional depth constructions adhering to our axioms intrinsically address the delicate challenge of partial observability of functional data, providing a minimal guarantee on the performance of their empirical counterparts beyond the idealised and practically infeasible case of full observability. As an incidental product, functional depths satisfying our definition achieve a robustness that is commonly ascribed to depth, despite the absence of a formal guarantee in the multivariate definition of depth. Some functional depth proposals are discussed with reference to the aforementioned axioms.

EO1152: **Shape depth**
Presenter: **Germain Van Bever**, The Open University, United Kingdom
Co-authors: Davy Paindaveine
In many problems from multivariate analysis (principal component analysis, testing for sphericity, etc.), the parameter of interest is not the scatter matrix but the so-called shape matrix, that is, a normalised version of the corresponding dispersion matrices. We propose, under elliptical assumptions, a depth concept for shape. If shape matrices are normalised to have determinant one, our shape depth results from a previous parametric depth construction. For other normalisations, however, defining a proper shape depth requires a semiparametric extension of this construction, which is likely to have applications in other contexts. We show that the proposed shape depth does not depend on the normalisation adopted and is affine-invariant. We also establish consistency, in the sense that shape depth is maximised at the true shape value. Finally, we consider depth-based tests for shape, and investigate their finite-sample performances through simulations.

EO0709: **Asymptotics for extreme depth-based quantile region estimation**
Presenter: **Yi He**, Monash University, Australia
Consider the small-probability quantile region in arbitrary dimensions consisting of extremely outlying points with nearly zero data depth value. Since its estimation involves extrapolation outside the data cloud, an entirely nonparametric method often fails. Using extreme value statistics, we extend a previous semiparametric estimation procedures to incorporate various depth functions. Under weak regular variation conditions, both consistency and asymptotic normality are derived. Refined asymptotics are derived particularly for the half-space depth. The finite-sample coverage probabilities of our asymptotic confidence sets are evaluated in a simulation study for the half-space depth and the projection depth. We use our method for risk management by applying it to financial data.

EO0841: **Approximate computation of data depths that satisfy the projection property**
Presenter: **Rainer Dyckerhoff**, University of Cologne, Germany
Co-authors: Pavlo Mozharovskyi
Data depth is a concept in multivariate statistics that measures the centrality of a point in a given data cloud in R^d. If the depth of a point can be represented as the minimum of the depths with respect to all unidimensional projections of the data, then the depth satisfies the so-called projection property. Such depths form an important class that includes many of the depths that have been proposed in literature. For depths that satisfy the projection property an approximate algorithm can easily be constructed since taking the minimum of the depths with respect to only a finite number of unidimensional projections yields an upper bound for the depth with respect to the multivariate data. Such an algorithm is particularly useful if no exact algorithm exists or if the exact algorithm has a high computational complexity, as is the case with the halfspace depth or the projection depth. To compute these depths in high dimensions, the use of an approximate algorithm with better complexity is surely preferable. Several strategies for selecting the univariate projections are proposed and the performances of the respective algorithms are compared.

EO555 Room 209 STATISTICAL TECHNIQUES FOR THE ANALYSIS OF COMPLEX BIOMEDICAL SIGNALS Chair: Anna Maria Paganoni

EO1190: **Analyzing dependence for multivariate functional data**
Presenter: **Francesca Ieva**, Universita degli Studi di Milano, Italy
Co-authors: Juan Romo
A matrix description of dependence is proposed for multivariate functional observations. The extension of the covariance matrix from multivariate data to multivariate functional curves is not straightforward. We use an alternative nonparametric measure of dependence, the Spearman correlation coefficient, in order to build a matrix expressing dependence among components of the multivariate functional data. A simulation study aimed at testing the ability of the index in correctly detecting the dependence is provided. Then the Spearman coefficient is used over two different samples of multivariate curves (specifically, electrocardiographic signals from healthy and not healthy people) in order to check if the patterns of dependence among components are different in the two cases.

EO1026: **Modeling mandibular shape variation using functional ANOVA models**
Presenter: **Luigi Ippoliti**, University G.d'Annunzio Chieti-Pescara, Italy
Co-authors: Lara Fontanella, Pasquale Valentini
The shape changes of human mandible observed in a longitudinal study is addressed. A classical evaluation of growth and shape changes in the human mandible is performed by means of conventional cephalometric analyses of lateral radiographs of the craniofacial complex. In general, conventional metrical approaches have proved to be insufficient for the analysis of size and shape changes of complex anatomical forms such as the human mandible. For example, lines and angles measured by traditional methods are not able to provide information about where the growth and shape change has occurred. It will be shown that functional data analysis represents a natural approach for representing the shape variations of a mandible and that functional linear models provide an elegant statistical framework for functional variability decomposition. First, the data and the problem of curve registration are described. Then, a Bayesian functional ANOVA model is introduced to describe the different sources of variability as well as to summarize the main modes of variation of the mandibles. Finally, some computational issues are addressed and fits of the model to the data are discussed.

EO0972: **Interval-wise testing for dependent functional data**
Presenter: **Helle Sorensen**, University of Copenhagen, Denmark
Interval-wise testing is an inference procedure used to test hypotheses about functional data. The procedure provides an adjusted p-value function, correcting for multiple testing, which is used to identify intervals of the domain with statistical significance. Interval-wise testing was originally developed for independent functional observations. The focus is now on dependent data. More specifically, consider situations where several treatments are tested on the same subjects. In a comparison of treatments, the dependence between functions from the same subjects must be taken into account. There are several alternatives: One is to only allow permutations that keep observations from the same subject together; another is to make permutations of residual curves from a model with random subject effect. Different strategies are compared using simulated data, and the method is used to compare the gait pattern of horses under different circumstances.

EO1129: **Numerical and computational challenges in ML estimation of mixture models for analyzing diffusion MRI data**
Presenter: **Aymeric Stamm**, Politecnico di Milano, Italy
Co-authors: Olivier Commowick, Simone Vantini, Simon Keith Warfield
Maximum likelihood estimators (MLE) are widely used for estimating parameters of a statistical model. In effect, MLEs have appealing properties such as consistency, asymptotic normality and efficiency and their computation is conceptually trivial since it boils down to likelihood maximization. The focus is on obtaining MLEs of mixture models from diffusion magnetic resonance imaging (dMRI) data. The scope is to track water in the brain subject to random diffusion motion. Mixture models are a natural choice since water is compartmentalized in impermeable substrates with homogeneous diffusion. Complexity of dMRI data is then two-fold: (i) mixture models are non-linear and thus require an optimization step for numerically approaching their ML estimator and (ii) signals are observed on a 3-dimensional grid with millions of pixels, on each of which the numerical estimation needs to be performed. The optimization step has mainly been neglected in dMRI, which focuses on model specification and acquisition. We emphasize that it is in fact essential as it completely determines the computational burden, which is a critical aspect when patients lives are at stake, and it is key to guaranteeing that the numerical estimator benefits from the properties of the targeted ML estimator. Moreover, we will show how the optimization problem can be efficiently solved by exploiting analytical properties of the likelihood rather than blindly feed it into an optimizer.

EO037 Room S23 SOFTWARE DEVELOPMENTS TOWARD HIGH FREQUENCY DATA Chair: Nakahiro Yoshida

EO1360: **ctsem: R software for continuous time structural equation modelling**
Presenter: **Charles Driver**, Max Planck Institute for Human Development, Germany
The ctsem software allows for the analysis of multiple subjects with repeated measurements of multivariate normal or binomial data, using Ornstein-Uhlenbeck type continuous time dynamic models (linear stochastic differential equations) coupled with a measurement model. The major benefits to such an approach in comparison to the more common and accessible discrete time vector autoregressive moving average type models are that measurements need not be equally spaced in time, more parsimonious specifications of complex dynamics are possible, and measurement error and structure can be accounted for. The models are estimated either with a frequentist structural equation modelling framework using the OpenMx software, in which case random effects are possible only for intercept variables, or in a hierarchical Bayesian framework using the Stan software, allowing for subject specific parameters drawn from an estimated population distribution. The effect of time varying covariates on the latent processes may be assessed, as may the effect of time invariant covariates on subject level parameters.

EO0542: **Stepwise estimation and assessment of Levy driven SDE**
Presenter: **Hiroki Masuda**, Kyushu University, Japan
Estimation and assessment of Levy driven SDE observed at high frequency is considered, building on some quasi-likelihoods made through small-time distributional approximation of the underlying statistical experiments. We present computer implementation by YUIMA package in R, as well as a review of recently developed theoretical backgrounds. Some of them extend reach of existing methodologies, providing us with a higher and deeper perspective for the underlying mechanisms of why those methodologies work well.

EO1063: **Point process regression models in YUIMA project**
Presenter: **Lorenzo Mercuri**, University of Milan, Italy
Co-authors: Nakahiro Yoshida
Estimation and simulation procedures of a point process regression model introduced recently and implemented in yuima package are reviewed. The point process regression model can be seen as a generalization of a self-exciting point process, since it is possible to consider external covariates in the intensity process. In financial the point processes have been used for modeling limit order book (LOB). The LOB is the central system for aggregation of all buy and sell orders issued by traders. The description of LOBs dynamics using high-frequency data requires advanced math-

ematical tools. In this context the point process regression model seems to be adequate to capture non-synchronicity, lead-lag and microstructure in high-frequency data using the self-exciting structure of the point process and reproducing exogenous effects arising from the presence of external covariates. After presenting the classes and methods available in yuima for the point process regression model, we apply them on a real LOB problem.

EO1058: **The Yuima framework for simulation and inference of stochastic processes and its GUI**
Presenter: **Stefano Iacus**, University of Milan, Italy
The Yuima package will be presented. It is a system of S4 classes and methods for the simulation and inference of stochastic processes including stochastic differential equation with or without jumps, fractional Brownian motion, Poisson and general point processes, CARMA and COGARCH processes. Yuima is a collaborative project and includes several simulation schemes as well statistical tools for quasi-maximum likelihood estimation, model selection, hypotheses testing, change point analysis. It also includes methods of asymptotic expansion. Recently, the Yuima package has been coupled with a graphical user interface, namely the YuimaGUI, which simplifies the usage of the package and allows for a complete flow of analysis: from data ingestion, to model selection and/or estimation, and estimation.

EO547 Room Board meeting room I FLEXIBLE MODELS FOR MODERN ECONOMETRIC AND STATISTICAL ANALYSIS Chair: Hao Zhang

EO1498: **Network vector autoregression**
Presenter: **Rui Pan**, Central University of Finance and Economics, China
Co-authors: Xuening Zhu, Guodong Li, Yuewen Liu, Hansheng Wang
A large-scale social network is considered with a continuous response observed for each node at equally spaced time points. The responses from different nodes constitute an ultra-high dimensional vector, whose time series dynamic is to be investigated. In addition, the network structure is also taken into consideration, for which we propose a network vector autoregressive (NAR) model. The NAR model assumes each node's response at a given time point as a linear combination of (a) its previous value, (b) the average of its connected neighbors, (c) a set of node-specic covariates, and (d) an independent noise. The corresponding coefficients are referred to as the momentum effect, the network effect, and the nodal effect respectively. Conditions for strict stationarity of the NAR models are obtained. In order to estimate the NAR model, an ordinary least squares type estimator is developed, and its asymptotic properties are investigated. We further illustrate the usefulness of the NAR model through a number of interesting potential applications. Simulation studies and an empirical example are presented.

EO1709: **Combining subsample estimators in quantile regression**
Presenter: **Keith Knight**, University of Toronto, Canada
The aim is to explore the extent to which the classical regression quantile estimator can be "improved" by combining estimators from bootstrap samples or subsamples. Such estimators can have the same asymptotic efficiency as the classical estimator while being more robust to small changes in the data.

EO0596: **Causality networks: Estimation and combination**
Presenter: **Roberto Panzica**, Goethe University House of finance, Italy
Co-authors: Giovanni Bonaccolto, Massimiliano Caporin
Granger causality test is commonly used in finance for estimating the causality relationship-based network among financial assets. Nevertheless, for investigating the comovement of assets, causality in the sense of Granger might not be the most appropriate tool as it focuses on mean relationships. The econometrics literature includes generalizations of the Granger causality concept for investigating the causality across quantiles of two distributions. However, this approach has never been considered to estimate a financial network. Moreover, the relationship in term of contribution of these two forms of causality within a more general equilibrium model is unclear. Previous works are extended by introducing a combination various causality networks within a standard multifactor model. The various networks will be considered as part of a more general multiplex network. Our advantage is to provide a measure of the relevance of each layer within the assets interconnectedness. In the empirical analysis we estimate the layers of the multiplex network by using the standard Granger causality and quantile Granger causality tests, using both parametric and non-parametric methods. We point at showing both the relevance of quantile-based causality in network estimation and the role played by alternative networks in a multifactor model accounting for asset interconnections.

EO1725: **Robust dynamic conditional score models**
Presenter: **Mauro Bernardi**, University of Padova, Italy
Dynamic conditional score models have been introduced as observation driven alternatives to more complex parameter driven state space models which are optimal in the sense of reducing the local Kullback-Liebler divergence between the true and the model implied conditional density. We apply the density power divergence approach of (1988) to robustify the dynamic conditional score equation to incidental influential observations and outliers. The density power divergence approach minimises an appropriately modified divergence where an additional parameter controls for the efficiency-robustness trade-off. Our approach combines the optimality of the score dynamics with the robustness properties of the class of M-estimators and it is a viable alternative to the Student-t approach. The generality of the proposed approach is illustrated through several examples of linear, non-linear and quantile regression.

EO431 Room 002 RECENT ADVANCES IN STATISTICAL GENETICS Chair: Wesley Thompson

EO0792: **A novel Bayesian model for the local false discovery rate**
Presenter: **Wesley Thompson**, Institute of Biological Psychiatry, Denmark
Classical multiple-comparison procedures tend to be underpowered in large-scale hypothesis testing problems. Procedures that control false discovery rate are more powerful, yet treat all hypothesis tests as exchangeable, ignoring any auxiliary covariates that may influence the distribution of the test statistics. A novel Bayesian semi-parametric two-group mixture model is proposed and a Markov chain Monte Carlo fitting routine for a covariate-modulated local false discovery rate (cmfdr) is developed. The probability of non-null status depends on the covariates via a logistic function and the non-null distribution is approximated as a linear combination of B-spline densities, where the weight of each B-spline density also depends on the covariates. We illustrate our proposed methods on a schizophrenia genome wide association study. In particular, we demonstrate that cmfdr dramatically improves power. We also show that the new approach fits the data closely, performing better than our previously proposed parametric gamma model for the non-null density.

EO1071: **Advanced methods in molecular quantitative trait loci mapping**
Presenter: **Alfonso Buil**, Mental Health Center Sct. Hans, Denmark
The use of molecular phenotypes as gene expression, methylation or metabolic measurements to understand the biology underlying genetic association results has increased largely in the last years. In this context, molecular Quantitative Trait Loci (QTL) mapping methods have become widely needed. These methods are computationally challenging because they have to perform millions of test and millions of permutations to assess statistical significance. This challenge is even harder when the samples include related individuals (as in family studies) where we have to

take into account the non-independence between the samples. We present two methods to address these two challenges: 1) fastQTL is a fast an accurate approach to map molecular phenotypes that uses a beta distributions trained from a few permutations to obtain adjusted p-values at any level of significance with little computational cost; 2) lmmQTL is an approximate linear mixed model approach that allows fast and accurate eQTL mapping in related samples. It uses a transformation of the data based on the spectral decomposition of the relationship matrix among samples to move them to a space where they are independent and can be analyzed with the standard QTL mapping approach.

EO1325: **Statistical characterisation of diffusion-based approaches in biological networks**
Presenter: **Sergio Picart-Armada**, Universitat Politecnica de Catalunya, Spain
Co-authors: Francesc Fernandez-Albert, Wesley Thompson, Alfonso Buil, Alexandre Perera-Lluna
Network analysis in computational biology pursues understanding experimental data in the context of known interactions between genes, proteins or metabolites, aiming to unravel new interactions, find molecular signatures or characterise new mechanisms. The analysis of diffusion processes in these networks quantifies how a perturbation starting at some seed nodes (e.g. downregulated genes) propagates to the rest of biological entities - ultimately identifying affected subnetworks. Diffusion-based approaches are robust to the noisy nature of experimental data and the presence of spurious associations in the knowledge model, but they are inherently related to the topology of the network. We analyse the statistics of diffusion scores through the introduction and characterisation of null models. We explore (1) the influence of topology, (2) the impact that the distribution of seed nodes has on the diffusion scores, (3) possible hypotheses testing on single nodes or whole subnetworks and (4) the definition of biologically sound null models. We find that the network architecture leverages the diffusion states and that the topological features of the nodes are reflected in their null distribution. Furthermore, nodes with correlated null distributions are prone to share biological functions. On the other hand, the success of hypotheses testing or subnetwork selection also depends on the seed nodes distribution and the formulation of the null model.

EP734 Room Hall POSTER SESSION III Chair: Stella Hadjiantoni

EP1378: **A comparative study on time series forecasting in telecommunications**
Presenter: **Irene Castro-Conde**, Optare Solutions S.L., Spain
Co-authors: Marta Cousido Rocha, Javier Roca Pardinas, Antonio Vidal Vidal
Telecommunications service providers have large quantities of historical data that could assist them to manage their business. In this sense, time series forecasting can assist decision making in marketing, planning or network management. In particular, we analyse real mobile daily data of a middle size operator corresponding to the number of people who left the company (churn) over the last four years. This is a problem of great interest since it is much cheaper to retain an existing customer than to acquire a new one and therefore retention strategies have to be developed in advance to avoid the churn. The time series under analysis presents multiple challenging characteristics such as multiple seasonalities and the influence of holidays and business cycle. The aim is to select the most suitable model in order to predict the next month occurrences. To this end we compare different methodologies which range from more classical approximations like ARIMA or Generalized Additive Models (GAM) to more general regression purpose methods like Recursive Partitioning Trees and Neural Networks. In order to do feature selection, parameter tuning and evaluate the results, three different error measures (MAE, RMSE and MASE) will be computed in a moving window validation scheme.

EP1388: **Another look at the value of individual labor income**
Presenter: **Jan Voelzke**, University of Muenster, Germany
Co-authors: Till Weigt
Human capital is considered as relevant in many asset pricing models, and for most people the individual human capital is by far the largest asset they possess. However, there is only little effort done to empirically determine the value of individual human capital. Avant-garde research is taken as starting point and price intervals for cash-flows created by individual labor income are provided. The aforementioned research is developed further in two points. Firstly, instead of pricing from the view of the individual and focusing on an individual stochastic discount factor, we use a national consumption-based market-wide stochastic discount to analyze prices that could be expected on financial markets for a corresponding human capital contract. Secondly, we use the substantial gain loss ratio instead of Sharpe-ratio to calculate good deal bounds. The proposed procedure is applied to German and US data.

EP1473: **Marginal expected shortfall as a systemic risk measure: The Czech PX index case study**
Presenter: **Radek Hendrych**, Charles University, Czech Republic
The systemic risk is undoubtedly an important concept employed in the framework of modern risk regulatory systems as are Basel III in finance or Solvency II in insurance. We primarily concentrate on a particular quantitative approach to measuring the systemic risk, which seems to be a significant risk in today's financial world (not solely in banks and insurance companies). The marginal expected shortfall measure is based on the well-known concept of the expected shortfall. More specifically, it can be regarded as a conditional version of the expected shortfall in which the global returns exceed a given market drop. We shall demonstrate that the marginal expected shortfall is a useful risk measure, when the systemic risk is analysed by studying the Czech data represented by the Prague Stock Exchange index and its constituents. The corresponding modelling scheme is introduced and discussed. It is supplemented in such a way that one can describe time-varying dependencies using the multivariate GARCH modelling class. Moreover, such an econometric approach enables to forecast the capital shortfall over a potentially long period (e.g. a quarter or six months), which might be truly useful in financial and insurance practice.

EP1628: **Modeling financial returns in different regions with a regime switching regular vine copula.**
Presenter: **Artur Machno**, AGH University of Science and Technology, Poland
Co-authors: Henryk Gurgul
The dependence structure of financial returns can be defined in different ways. The multivariate distribution of the univariate residuals is analyzed. Those distributions are described by multivariate vine copulas. Dependences are analyzed in terms of time and geography. Time variability is introduced by Markov regime switching component. Additionally, results for different regions are compared. Expected shortfall and value at risk are investigated in addition to differences between copulas and regime variables between regions.

EP1662: **Modelling seasonality with long seasonal period**
Presenter: **Jiri Prochazka**, University of Economics, Prague, Czech Republic
Co-authors: Samuel Flimmel
The main aim is the discussion of different methods used for modeling seasonality with long seasonal periods. A special focus is given on procedures based on the representation of the seasonal pattern using basis functions. Advantages and disadvantages of the various methods are discussed. Specifically, the most commonly used methods for modeling seasonality with short seasonal periods are shown not to be appropriate in the case of long seasonal periods since they can suffer from overparametrization if a sufficient number of observations is not available for the model estimation. On the other hand, the decomposition of the seasonal pattern using different types of basis functions is argued to provide a potential

remedy for modeling seasonality with long seasonal periods in the case of a small number of observations. The study is supported with illustrations that make use of simulated as well as real-life data sets.

CO273 Room 104 REAL-TIME DATA ANALYSIS Chair: Dean Croushore

CO0216: **Improving model-based near-term GDP forecasts by subjective forecasts: A real-time exercise for the G7 countries**
Presenter: **Jasper de Winter**, De Nederlandsche Bank, Netherlands
The aim is to investigate to what extent it is feasible to improve model-based near-term GDP forecasts by combining them with judgmental (quarterly) forecasts by professional analysts (Consensus survey) in a real-time setting. Our analysis covers the G7 countries over the years 1999-2013. We consider as combination schemes the weighted average and the linear combination. Incorporating subjective information delivers sizable gains in forecasting ability of statistical models for all countries except Japan in 1999-2013, even when subjective forecasts are somewhat dated. Accuracy gains are much more pronounced in the volatile period after 2008 due to a marked improvement in predictive power of Consensus forecasts. Since 2008, Consensus forecasts are superior at the moment of publication for most countries. For some countries Consensus forecasts can be enhanced by model-based forecasts in between the quarterly release dates of the Consensus survey, as the latter embody more recent monthly information.

CO0350: **The effects of CBS revisions on CPB forecasts**
Presenter: **Adam Elbourne**, CPB Netherlands Bureau for Economic Policy Analysis, Netherlands
Co-authors: Kasia Grabska, Henk Kranendonk, Jason Rhuggenaath
The uncertainty caused by data revisions is put into perspective by asking how different CPBs forecasts would have been based on revised data. Over the period 2004-2014 GDP growth (quarter on quarter) was typically revised upwards on average 0.1 percentage points with a standard deviation of 0.2 percentage points from the 1st to the second estimate of the quarterly growth rates. Secondly, we look how a typical revision would change our published forecasts. Our measure of a typical revision takes into account the observed cross correlation between the components of GDP, for example when GDP is revised upwards consumption is also revised upwards. A typical GDP revision in the last quarter of this year would increase our GDP forecast for next by 0.23 percentage points. Finally, we use our measure of the sensitivity of our published forecasts to typical revisions to get an idea of the importance of uncertain national accounts data for our overall forecast error. For the period 2004-2014 the root mean square forecast error for our CEP forecast would decline from 0.65 percentage points to about 0.51 percentage points if our forecasts were based on definitive rather than provisional estimates of the national accounts. Thus, the majority of our forecast error can be attributed to other sources of uncertainty.

CO0373: **Improving GDP measurement further: Data revisions with news-noise measurement errors**
Presenter: **Samad Sarferaz**, ETH Zurich, Switzerland
Co-authors: Jan Jacobs, Jan-Egbert Sturm, Simon van Norden
Recently a new measure of historical U.S. GDP growth (GDPplus) was proposed adopting a measurement-error perspective. By distinguishing news and noise measurements errors and allowing for data revisions, we propose a new measure for U.S. GDP growth and compare its dynamic properties to GDPplus.

CO1697: **Fiscal forecasts at the FOMC: Evidence from the Greenbooks**
Presenter: **Dean Croushore**, University of Richmond, United States
New data set of fiscal policy forecasts and estimates prepared for the FOMC are used to understand how they have influenced U.S. monetary policy. We find limited evidence of forecast bias and that the forecasts contain useful information beyond that in the CBO's forecasts. Forecast errors for the fiscal variables have been only weakly correlated with forecast errors for inflation and output growth, but those for the budget surplus are much more highly correlated with those for the unemployment rate and the output gap. Some fiscal variables can also account for a significant fraction of the "exogenous" changes in the federal funds rate target previously studied, which is consistent with the Board's statements on the importance of fiscal policy.

CO501 Room 105 MANAGING LIQUIDITY RISK Chair: Gaelle Le Fol

CO0258: **Event-studies and (endogenous) zero returns**
Presenter: **Fabrice Riva**, Universite Paris-Dauphine, France
Co-authors: Laurent Deville, Juan Raposo
Zero returns are a pervasive phenomenon on stock markets. Failure to account for their potential endogeneity leads to (OLS) market model parameter estimates that are systematically biased. Taking the view that zero returns endogenously arise from implicit transaction costs, we estimate the parameters of stocks true return-generating process using a latent variable estimation technique. We test the usefulness of our approach in the context of event-studies. Based on a previous simulation framework, we show that our new estimates lead to better specification and increased power for statistical tests of abnormal returns.

CO0604: **A self-exciting model of mutual fund flows**
Presenter: **Ran Sun**, University of Paris-Dauphine, France
Dynamics of mutual funds flows are studied with a unique private database provided by several french asset management firms. This database covers a large range of funds with different characteristics. Compared to the other public available databases of fund flows, our database gives more information which allow us to understand better the mutual funds clients investing decision making. One important characteristic is the frequency of fund flows. We use a discrete counting process to model this flow frequency. Some stylized facts of the data are discovered; among them the self exciting property is a substantial one. In line with literature, this self exciting property indicates the correlation nature of funding liquidity. In a financial point of view, we give an analysis of client risk of mutual funds; in a statistical point of view, our model estimates more accurately tail risk of fund flows.

CO1446: **Financial market liquidity: Who is acting strategically**
Presenter: **Gaelle Le Fol**, Universite Paris - Dauphine, and CREST, France
In a new environment where liquidity providers as well as liquidity consumers act strategically, understanding how liquidity flows and dries-up is key. We propose a model that specifies the impact of information arrival on market characteristics, in the context of liquidity frictions. We distinguish short-lasting liquidity frictions, which impact intraday prices, from long-lasting liquidity frictions, when information is not fully incorporated into prices within the day. We link the first frictions to the strategic behavior of intraday liquidity providers and the second to the strategic behavior of liquidity consumers, i.e. long-term investors who split up their orders not to be detected. Our results show that amongst 61% of the stocks facing liquidity problems, 57% of them point up liquidity providers as the sole strategic market investor. Another 27% feature long-term investors as the single strategic player, while both liquidity providers and liquidity consumers act strategically in the remaining 16%.

This means that 43% of these stocks are actually facing a slow-down in the information propagation in prices, which thus results in a significant decrease of (daily) price efficiency due to long-term investors' strategic behavior.

CO1576: **Multivariate Hawkes processes: A microscope for high-frequency order book dynamics**
Presenter: **Marcello Rambaldi**, CMAP - Ecole Polytechnique, Palaiseau, France, France
Co-authors: Emmanuel Bacry, Fabrizio Lillo
The aim is to show that multivariate Hawkes processes coupled with the recent non-parametric estimation procedure previously proposed can be successfully used to study complex interactions between the time of arrival of orders and their size, observed in a limit order book market. We apply this methodology to high-frequency order book data of futures traded at EUREX. Specifically, we demonstrate how this approach is amenable not only to analyse interplay between different order types (market orders, limit orders, cancellations) but also to include other relevant quantities, such as the order size, into the analysis, showing also that simple models assuming the independence between volume and time are not suitable to describe the data.

CO297 Room 112 MACROECONOMETRICS AND TIME SERIES Chair: Matteo Barigozzi

CO0301: **Quantile factor models**
Presenter: **Jesus Gonzalo**, University Carlos III de Madrid, Spain
Co-authors: Juan jose Dolado, Liang Chen
A novel concept is introduced: Quantile Factor Models (QFM), where a few unobserved common factors may affect all parts of the distributions of many observed variables in a panel data set of dimension NxT . When the factors affecting the quantiles also affect the means of the observed variables, a simple two-step procedure is proposed to estimate the common factors and the quantile factor loadings. Conditions on N and T ensuring uniform consistency and weak convergence of the entire quantile factor loadings processes differ from standard conditions in factor-augmented regressions with smooth object functions. Based on these results, we show how to make inference on the quantile factor loadings in a location-scale shift factor model. When factors affecting the quantiles differ from those affecting the means of the observed variables, we propose an iterative procedure to estimate both factors and factor loadings at a given quantile. Simulation results confirm a satisfactory performance of our estimators in small to moderate sample sizes. In particular, it is shown that the iterative procedure can consistently estimate common factors that cannot be captured by PC estimators. Empirical applications of our methods to stocks and mutual fund returns are considered.

CO0646: **A general inversion theorem with applications to integrated processes**
Presenter: **Paolo Paruolo**, European Commission Joint Research Centre, Italy
Co-authors: Massimo Franchi
The aim is to present generalizations of Granger's Representation Theorem for I(1) processes and of Johansen's Representation Theorem for I(2) processes to processes integrated of any order. These generalizations exploit novel results on the inversion of matrix functions that are singular at a given point (labelled extended local rank factorization, ELRF) to provide extensions and a unifying approach to alternative representations of linear systems integrated of any order. In particular, for linear systems integrated of any order, the ELRF gives explicit expressions for the coefficients of the vector AutoRegressive and Equilibrium Correction representations in terms of their Moving Average representation. Vice-versa, for vector AutoRegressive processes, it gives explicit expressions for the coefficients of the Moving Average and the Common Trends representations. The elrf encompasses earlier inversion results for I(1) and I(2) autoregressive systems, and it is shown to be related to the Jordan pairs, Jordan chains and the local Smith form; this allows us to provide explicit links to previous representations.

CO0655: **Impulse response estimation by smooth local projections**
Presenter: **Christian Brownlees**, UPF, Spain
Co-authors: Regis Barnichon
Vector autoregressions (VAR) and local projections (LP) are well established methodologies for the estimation of impulse responses (IR). These techniques have complementary features: the VAR approach is more efficient when the model is correctly specified whereas the LP approach is less efficient but is more robust to model misspecification. We propose a semi-parametric impulse response estimator, called smooth local projections (SLP), that attempts to strike a balance between these two extremes. The procedure consists of using local projections under the constraint that the IR is a smooth function of the forecast horizon. Inference is carried out using semi-parametric techniques based on B-splines, and the IR can be estimated by standard (ridge) regression. We also show how SLP may be used in conjunction with common identification schemes such as timing restrictions or instrumental variables to recover structural IRs. We apply our technique to study the effects of monetary shocks and show using out-of-sample validation criteria that smooth local projections provide more precise IR estimates than VAR and LP.

CO1261: **Sentiments in the times of crisis**
Presenter: **Antonio Conti**, Universite Libre de Bruxelles, Italy
Co-authors: Matteo Barigozzi, Fabrizio Venditti
The role of sentiment shocks on US economy are evaluated focusing on the impact on both financial and real variables.To understand whether these effects are constant over time we estimate a Structural VAR model with kernel smoothing to allow for time-variation in the transmission of structural shocks, providing the proper framework for analyzing the Great Financial Crisis of 2007-09. Identification relies on the maximization variance method which allows for disentangling sentiment shocks from financial ones, the latter modeled as stock prices, credit or house prices shocks. We show that sentiment shocks (i) display indeed time-varying effects on consumption, investment and output (ii) raise their effects during periods of crisis (iii) are at least as relevant as financial shocks for the US business cycle and (iv) they account for a large share of output and stock prices dynamics in the Global Financial Crisis.

CO347 Room 106 TIME SERIES MODELLING OF CHALLENGING ECONOMIC PHENOMENA Chair: Jeanne Diesteldorf

CO0374: **Global impact of US and euro area unconventional monetary policies: A comparison**
Presenter: **Marco Lombardi**, Bank for International Settlements, Switzerland
Co-authors: Feng Zhu
The domestic and cross-border effects of US and euro area unconventional monetary policy measures are compared on 24 major advanced and emerging economies, based on an estimated global vector error-correction model (GVECM). Unconventional monetary policies are measured using shadow interest rates recently developed. Monetary policy shocks are identified using sign restrictions. The GVECM impulse responses suggest that US unconventional monetary policy generally has stronger domestic and cross-border impacts than euro area non-standard measures. Its spillovers to other economies are estimated to be more sizeable and persistent, especially in terms of output growth and inflation. There is evidence of diverse responses in the emerging economies in terms of exchange rate pressures, credit growth as well as monetary policy. In addition, the strength of cross-border transmission channels to the emerging economies appears to differ for US and euro area policies.

CO0379: **The impact of speculators on agricultural commodity futures prices**
Presenter: **Jeanne Diesteldorf**, Westfalische Wilhelms-Universitat Munster, Germany
Co-authors: Martin T Bohl
Departing from the lively discussion about the Masters' hypothesis, the aim is to examine whether increasing activity of speculators in commodity futures markets impacts on price movements. In particular, we investigate the effect of the activity of speculators that are not commodity index traders on market outcomes. To this end, we use data for the five most liquid agricultural futures markets compiled by the CFTC to implement Granger causality tests and GARCH models. We conclude that speculators do not impact on price movements in the markets under scrutiny.

CO0382: **The speculative component in Chinese agricultural commodity futures**
Presenter: **Claudia Wellenreuther**, Westfalische Wilhelms-Universitat Munster, Germany
Co-authors: Martin T Bohl, Pierre Siklos
The aim is to investigate empirically whether speculative activity in Chinese futures markets for agricultural commodities destabilizes futures prices. To capture speculative activity we use a ratio, defined as daily trading volume divided by end-of-day open-interest. Applying a GARCH-model we first analyse the influence of the speculation ratio on the conditional volatility of five heavily traded Chinese futures contracts, namely soybeans, soybean meal, corn, sugar and cotton. Furthermore, we try to gain insight into the lead-lag-relationship between speculative activity and price volatility by using a VAR-model in conjunction with Granger causality tests, impulse-response analyses and variance decompositions.

CO1466: **What makes the market jump**
Presenter: **Chardin Wese**, ICMA Centre, University of Reading, United Kingdom
Co-authors: Marcel Prokopczuk
Using intraday transaction prices and a non-parametric jump test, we show that jumps in the S& P 500 and VIX are low-probability, high-impact events. Extant research investigating the causes of jumps primarily focuses on scheduled macro-announcements. However, we find that unscheduled news, which has so far received little attention, triggers twice as many jumps and accounts for a larger proportion of the jump variation than scheduled news. Intriguingly, we show that close to 50% of jumps are not explained by fundamental news, revealing the presence of excess jumps in financial markets.

CO343 Room Graduation hall FINANCIAL TIME SERIES FORECASTING Chair: Jeroen Rombouts

CO0489: **Sparse change-point and Markov-switching HAR models for realized volatility**
Presenter: **Arnaud Dufays**, Laval University, Canada
Co-authors: Jeroen Rombouts
Change-point and Markov-switching time series specifications constitute flexible models that capture unknown 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 switches in all parameters. This implies that all parameters have to change when a break happens. We introduce sparse change-point and Markov-switching 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 regime to another. Additionally, we derive a Gibbs sampler for inferring the parameters of these processes. Relying on this new framework, we study the stability of the HAR model using realized volatilities series of eleven international indices between January 2000 and August 2015.

CO0932: **Robust time series forecasting with exponential smoothing methods**
Presenter: **Ruben Crevits**, KU Leuven, Belgium
Co-authors: Christophe Croux
Simple methods like exponential smoothing are very popular for forecasting univariate time series. The R-package for forecasting with exponential smoothing has been downloaded numerous times. The method chooses whether or not to include a (damped) trend or seasonality effects, both of which often occur in real time series. We provide a robust alternative for the exponential smoothing forecaster. For each variation of exponential smoothing we present a robust alternative. The robust method is developed by robustifying every aspect of the original exponential smoothing variant. We compare the standard non-robust version with our robust proposal in a simulation study. The methodology is applied to data from the M3 competition, which includes time series from microeconometrics, macroeconometrics, demographics, finance and industry.

CO0193: **Dynamics of variance risk premia, inverstors' sentiment and international return predictability**
Presenter: **Jeroen Rombouts**, ESSEC Business School, France
Co-authors: Francesco Violante, Lars Stentoft
A flexible approach is proposed for retrieving the variance risk premia (VRP) which delivers more refined, precise and realistic estimates of the market price of risk. We define a class of structural time series models that isolates as structural components the dynamics of the physical variance and, by embedding its expectations into the model, the price attached by the market to the variance risk (i.e. the VRP). In fact, by doing this we deconstruct the mechanism of formation of the variance expectations under the risk neutral measure. Given the latent nature of the variables of interest of which only imprecise approximations are observable (i.e. high frequency return based variance measures and option implied risk neutral variance expectations), we advocate the use of methodologies based on signal extraction techniques. We advocate the inclusion of interactions and discontinuities, with emphasis on structural breaks, extreme events, uncertainty due to heteroskedasticity, correlations and spillovers, as being essential to replicate complex dynamics and interdependencies between the physical variance and its risk neutral expectation. In an empirical application to the S&P500, we address the excess return puzzle by disentangling the predictability stemming from the part of the variance risk premium associated with normal sized price fluctuations from that associated with extreme tail events, i.e. tail risk.

CO405 Room 102 QUANTILES, PREDICTABILITY, AND HEAVY-TAILED DISTRIBUTIONS Chair: Richard Luger

CO0702: **Point-optimal sign-based tests for stock return predictability**
Presenter: **Kaveh Salehzadeh Nobari**, Durham University, United Kingdom
Co-authors: Abderrahim Taamouti, Jean-Marie Dufour
Simple point-optimal sign-based tests are proposed for inference on linear and nonlinear regression models in the presence of stochastic regressors. The motivation is to build sign-based tests for linear and nonlinear predictability of asset returns. The most popular predictors of stock returns (e.g. dividend-price ratio, earning-price ratio, etc.) are known to be stochastic. The proposed sign-based tests are exact, distribution-free, and robust to heteroskedasticity of unknown form. They may be inverted to build confidence regions for the parameters of the regression function. Since the point-optimal sign tests depend on the alternative hypothesis, an adaptive approach based on split-sample techniques is suggested in order to choose the appropriate alternative. We present a Monte Carlo study to assess the performance of the proposed quasi-point-optimal sign test by comparing its size and power to those of some common tests which are supposed to be robust against heteroskedasticity. The results show that our procedures are superior. Finally, an empirical application using real data is considered to illustrates the proposed quasi-point-optimal sign tests.

CO0778: **Quantiles and inequality indices estimation from heavy-tailed distribution**
Presenter: **Emmanuel Flachaire**, Aix-Marseille University, France
Co-authors: Arthur Charpentier
Quantiles and inequality indices are estimated from a nonparametric density estimation based on transformed data. A parametric cumulative distribution function is initially used to transform the data into values over the unit interval, from which a non-parametric density estimation is obtained. Finally, an estimation of the density of the original sample is obtained by back-transformation. This approach may be particularly useful to estimate heavy-tailed distributions. We discuss its implementation and its finite sample properties for density estimation, and for estimation and inference with quantiles and inequality indices.

CO0969: **Exact inference in predictive quantile regressions**
Presenter: **Sermin Gungor**, Bank of Canada, Canada
Co-authors: Richard Luger
An exact simulation-based procedure is developed to test for quantile predictability at several quantile levels, jointly. The approach proceeds by combining the quantile regression t-statistics and uses Monte Carlo resampling techniques to control the overall significance level in finite samples. As a by-product our procedure also yields an exact distribution-free confidence interval for the persistence parameter of a first-order autoregressive model, assumed for the predictor variable. We employ the new procedure to test the ability of many commonly used variables to predict the quantiles of excess stock returns.

CO0973: **Exact and heavy-tail robust inference in GARCH models**
Presenter: **Richard Luger**, Laval University, Canada
A procedure is developed for building exact confidence intervals in GARCH models without any restrictions on tail heaviness. The considered GARCH models may even be subject to variance targeting. The approach uses profile quasi-likelihood ratios with Monte Carlo resampling techniques to obtain exact bounds tests. These level-exact tests are then inverted to produce conservative confidence intervals for the model parameters. The endpoints of the confidence intervals can be found quickly under certain conditions by a combination of bisection and grid search. Daily returns on major stock market indices are used to illustrate the exact inference procedure in a variety of GARCH specifications.

CO634 Room 111 NONSTATIONARITY IN TIME SERIES Chair: Lionel Truquet

CO1023: **Moment bounds and central limit theorems for non stationary Gaussian subordinated arrays**
Presenter: **Jean Marc Bardet**, University Paris Pantheon-Sorbonne, France
A general moment bound for sums of products of Gaussian vector's functions extending a previous moment bound is established. A general central limit theorem for triangular arrays of nonlinear functionals of multidimensional non-stationary Gaussian sequences is proved. This theorem extends the previous results. A Berry-Esseen-type bound in the above-mentioned central limit theorem is derived. Two applications of the above results are discussed. The first one refers to the asymptotic behavior of a roughness statistics for continuous-time Gaussian processes and the second one is a central limit theorem satisfied by long memory locally stationary process.

CO1259: **Robust econometric inference in systems of cointegrating and predictive regressions**
Presenter: **Tassos Magdalinos**, University of Southampton, United Kingdom
Co-authors: Peter CB Phillips
Econometric methodology of inference in systems of cointegrating and predictive regressions is extended to accommodate time series with multiple persistence rates of unknown order. It is well known that conventional approaches to estimating cointegrating regressions fail to produce even asymptotically valid inference procedures when the regressors are nearly integrated, and substantial size distortions can occur in econometric testing. The new framework enables a general approach to inference that is robust to the persistence characteristics of the regressors, making it suitable for general practical application. Endogenously generated mildly integrated instruments are employed which eliminate the endogeneity problems of conventional cointegration methods and robustify inference to uncertainty over the nature of the integration in the system. The use of mildly integrated instruments also provides a mechanism for linking the conventional treatment of endogeneity in simultaneous equations with the econometric methodology for cointegrated systems. The methods are easily implemented, widely applicable and help to alleviate practical concerns about the use of cointegration methodology.

CO1045: **Gradient-based structural change detection for non-stationary time series M-estimation**
Presenter: **Weichi Wu**, University College London, United Kingdom
Co-authors: Zhou Zhou
Structural change testing is considered for a wide class of time series M-estimation with non-stationary predictors and errors. Flexible predictor-error relationships, including exogenous, endogenous and autoregressive regressions and their mixtures, are allowed. New uniform Bahadur representations are established with nearly optimal approximation rates. A CUSUM-type test statistic based on the gradient vectors of the regression is considered. A simple bootstrap method is proposed and is proved to be consistent for M-estimation structural change detection under both abrupt and smooth non-stationarity and temporal dependence. Our bootstrap procedure is shown to have certain asymptotically optimal properties in terms of accuracy and power. A public health time series data set is used to illustrate our methodology, and asymmetry of structural changes in high and low quantiles are found.

CO1651: **Asymptotic theory for time-varying regression models**
Presenter: **Wei Biao Wu**, University of Chicago, United States
Parameter estimation, hypothesis testing and variable selection is considered for partially time-varying coefficient models. Our asymptotic theory has the useful feature that it can allow dependent, nonstationary error and covariate processes. With a two-stage method, the parametric component can be estimated with a root n-convergence rate. A simulation-assisted hypothesis testing procedure is proposed for testing significance and parameter constancy. We further propose an information criterion that can consistently select the true set of significant predictors. Our method is applied to autoregressive models with time-varying coefficients. Simulation results and a real data application are provided.

CO269 Room 101 PORTFOLIO OPTIMIZATION AND PERFORMANCE EVALUATION Chair: Chulwoo Han

CO1054: **Improving the naive portfolio strategy**
Presenter: **Chulwoo Han**, Durham University, United Kingdom
Four approaches are considered to improve the naive strategy, being one of which newly proposed. The new approach attempts to minimize the difference from a reference portfolio while satisfying risk and return targets. Other well-known strategies such as variance minimization are also compared. Empirical results show that the new approach consistently outperforms the naive strategy. However, the best performance comes from the minimum variance portfolios. It is also worth noting that optimal strategies appear to perform well for factor portfolios, which implies careful construction of asset universe could have a large impact on portfolio optimization performance.

CO1126: **Portfolio optimization with analyst forecasts**
Presenter: **Sunyoung Kim**, KAIST (Korea Advanced Institute of Science and Technology), Korea, South
Co-authors: Chulwoo Han
The aim is to examine the forecast power of analysts' target prices and their utlization in portfolio optimization. Abnormal returns of stocks are predicted from analyst forecasts, which are then used for expected return estimation. Stocks with non-zero abonormal return forecasts are selected and combined with an market index to form an optimal portfolio. Tested on S& P500, we find that analyst target prices are significant in forecasting future returns but other forecasts such as earnings per share do not help predict future returns. When the assets with nonzero abnormal returns are optimally combined with the market index, the resulting optimal portfolio turns out to outperform the index.

CO1158: **ETFs and market coverage**
Presenter: **Junqi Li**, Durham University, China
Co-authors: Chulwoo Han
ETFs' role is investigated as an investment alternative to stocks. While ETFs are considered an effective investment vehicle, it is not clear whether ETFs are diverse enough to cover the feasible region of the stocks. We find through ex-post efficient frontier analysis that ETFs based on equities are diverse, but not as much to cover the efficient frontier of stocks. However, when the ETFs on other assets such as commodities are also included, the efficient frontier of the ETFs are extended to cover most of the efficient frontier of stocks. Portfolio optimization results only support using ETFs instead of individual stocks for portfolio construction.

CC0227: **Tactical asset allocation with binary regression trees andforests**
Presenter: **Juan Laborda**, Carlos III, Spain
Co-authors: Ricardo Laborda
Given the statistical characteristics of financial series, we propose a methodology that allows us to accurately forecast the relative returns of financial assets and to implement tactical asset allocation (TAA) strategies, as part of a basic portfolio management model (equities, bonds and cash). We will focus on that research topic which, with the implication of predictability in series of returns, allows us, on the basis of fundamental variables, to exploit the set of information that these offer in order to segment and classify homogeneous areas, on the basis of which we can predict higher returns from one asset relative to another. The initially proposed technique is single binary regression trees (classification and regression trees) that naturally extends to ensembles trees, applying methods like random forest. Once the forecast relative returns are calculated, a TAA system is developed from which we derive the structure of the optimal aggressiveness factors of the various tactical strategies, which allows us, using a benchmark portfolio, to calculate the weightings to hold in each of the assets.

CO731 Room 107 NOWCASTING AND FORECASTING MACROECONOMIC TRENDS II Chair: Gian Luigi Mazzi

CO1111: **Use of panel VAR models for nowcasting GDP**
Presenter: **Dan Rieser**, European Commission, Eurostat, Luxembourg
Co-authors: Gian Luigi Mazzi, James Mitchell
The application of Panel Vector Auto-Regressive (PVAR) models to macroeconomic nowcasting is examined. Alternative model specifications proposed in the literature such as large scale Bayesian VAR models, Global VAR models and Dynamic Factor Models are reviewed. Their characteristics and performance for both forecasting and nowcasting is assessed by nowcasting GDP growth for 8 European countries and the Euro area. A mixed frequency, cross-country dataset is implemented for this purpose. This is a novelty as the use of mixed-frequency data (quarterly and monthly) in a PVAR model has previously been untried. In the past, nowcasts have been produced ignoring cross-country dependencies. The results based on six alternative panel VAR models point and density estimates suggest that the use of panel VAR model, with panel priors, can be helpful when forecasting GDP growth. However, when nowcasting based on the panel structure of the dataset and estimating a panel VAR model with panel priors, panel VAR models do not seem to be the preferred option. The most accurate nowcasts are obtained by shrinking the VAR model to a univariate model. This can be achieved by deploying the Minnesota prior that ignores the panel structure of the dataset and by not motivating the prior with reference to possible dynamic interdependencies, static interdependencies and cross sectional heterogeneities that one might expect in a panel dataset.

CO0945: **Nowcasting US GDP in real time: A Bayesian mixed-frequency latent-threshold model with stochastic volatility**
Presenter: **Boriss Siliverstovs**, KOF ETHZ, Switzerland
A novel approach is proposed for short-term forecasting of economic variables sampled at heterogeneous frequencies by adapting a Bayesian Latent-Threshold Dynamic model to the mixed-frequency setting. As argued before, introducing latent thresholds into dynamic multiple-regressor models helps handling the associated curse of dimensionality by inducing dynamic sparsity in estimated coefficients. This is especially important in models with mixed-frequency data where blocking of higher-frequency data into their lower-frequency counterparts leads to parameter inflation rapidly growing with the frequency mismatch. As a matter of fact, this approach can be considered as a useful alternative to commonly used Bayesian shrinkage applied to large mixed-frequency models. We illustrate the usefulness of the proposed model by nowcasting US GDP growth using historical real-time data vintages.

CO0201: **A quasi-Bayesian nonparametric approach to time varying parameter VAR models**
Presenter: **Katerina Petrova**, Queen Mary University London, United Kingdom
A quasi-Bayesian local likelihood (QBLL) estimation methodology is established for a multivariate model with time varying parameters. The validity of the resulting quasi-posterior distributions of the drifting parameters is proven in general and, in the special case of a Gaussian VAR model, a closed form Normal-Wishart expression for the quasi-posterior distribution of the QBLL estimator is provided. In addition, several Gibbs algorithms are developed, which can sample from a VAR model with a mixture of time varying and time invariant parameters. The proposed estimators differ from existing state space approaches to VAR models in that they estimate parameter time variation nonparametrically without imposing assumptions on the stochastic processes of the parameters. The QBLL estimators exhibit good finite sample properties and their performance compares well to existing parametric state space models, as illustrated by a Monte Carlo exercise. In addition, we demonstrate that the QBLL approach provides a remedy to the 'curse of dimensionality' by accommodating large dimensional VAR systems and delivers improvements in the out-of-sample forecasts of key macroeconomic variables. Finally, an empirical contribution to the literature on changing macro dynamics in the US is made, presenting evidence of a fall in inflation persistence and volatility during the Great Moderation period, in line with previous results.

CO0721: **Nowcasting payroll employment with traditional media content**
Presenter: **Thomas Renault**, IESEG, France
Co-authors: Clement Bortoli, Stephanie Combes
Flash payroll employment statistics in France are published quarterly, with a delay of 45 days after the end of the quarter. In order to "predict the present", forecasters mainly rely on business tendency/consumer confidence surveys. Building on findings from other fields of research that identify that value-relevant information can be extracted from content published on traditional newspapers, we contemplate media content as an

complementary source of data to improve the forecast of French employment at different time horizon. Features are extracted from a large sample of 1,354,100 articles published in the newspaper "Le Monde" between 1990 and 2016. We first adopt a simple "bag-of-words" representation using the frequency of occurrence of each word in the content published during a month. However, a simple bag-of-words approach fails to capture words polysemy and synonymy, as the context in which a word is used is not taken into account. In order to solve this issue, we also consider more advanced text analytics methods using a continuous bag-of words approach and probabilistic topics identification approach (Latent Dirichlet Allocation). Eventually, penalized regressions are mobilized to select the most relevant features with respect to forecasts performances.

CO357 Room 103 FINANCIAL MODELLING Chair: Genaro Sucarrat

CO1150: **On the unbiasedness of financial volatility proxies**
Presenter: **David Kreiberg**, Uppsala University, Sweden
Co-authors: Steffen Groenneberg, Genaro Sucarrat

Models for conditional volatility play an important role in asset pricing and in risk analysis. Since the conditional volatility is an unobserved quantity, a problem is how to undertake forecast comparisons. A common way to handle this problem is to use some form of volatility proxy such as realized volatility or a range-based volatility measure. However, such proxies may not represent unbiased measures of the true volatility. We device simple tests to evaluate hypotheses concerning the unbiasedness of these proxies. The proposed tests hold under mild assumptions and accommodate various hypotheses, depending on the form and nature of the observed volatility proxy. Simulations confirm that the empirical size corresponds well to the theoretical size, and that it has power under the alternative. The results are further illustrated in an empirical application.

CO0854: **The short-run and long-run components of financial market volatility**
Presenter: **Norbert Metiu**, Deutsche Bundesbank, Germany
Co-authors: Giovanni Motta

An Evolutionary Latent Factor GARCH (ELF-GARCH) model is introduced which captures the volatility common to a large panel of financial variables by a small number of common volatilities. Each common volatility is decomposed into the product of two components which display different degrees of persistence. The first component reflects short-run dynamics, while the second component captures persistent movements in volatility. We develop a semi-parametric estimation theory, and the finite-sample performance of our estimators is investigated considering a variety of simulation scenarios. Our empirical illustration shows that the bulk of the variation in 157 monthly U.S. financial variables is captured by a single common factor. The short-run volatility of this factor gauges widely used proxies of risk and uncertainty in financial markets, such as credit spreads and the VIX index, while its long-run volatility reaches elevated levels around major financial crises. We embed the two volatility components into a standard macro-financial vector autoregression. An unexpected rise in short-run volatility is followed by dynamic responses of output and employment that closely resemble those typically associated with an adverse uncertainty shock. Long-run volatility Shocks trigger a significant decline in output and employment at the 2-5 year horizon, which highlights the role of long-run risks in shaping business cycle fluctuations.

CO1061: **Models of financial return with time-varying zero probability**
Presenter: **Genaro Sucarrat**, BI Norwegian Business School, Norway
Co-authors: Steffen Groenneberg

The probability of an observed financial return being equal to zero is not necessarily zero. This can be due to price discreteness or rounding error, liquidity issues (e.g. low trading volume), market closures, data issues (e.g. data imputation due to missing values), characteristics specific to the market, and so on. Moreover, the zero probability may change and depend on market conditions. In standard models of return volatility, however, e.g. ARCH, SV and continuous time models, the zero probability is zero, constant or both. We propose a new class of models that allows for a time-varying zero probability, and which can be combined with standard models of return volatility: They are nested and obtained as special cases when the zero probability is constant and equal to zero. Another attraction is that the return properties of the new class (e.g. volatility, skewness, kurtosis, Value-at-Risk, Expected Shortfall) are obtained as functions of the underlying volatility model. The new class allows for autoregressive conditional dynamics in both the zero probability and volatility specifications, and for additional covariates. Simulations show parameter and risk estimates are biased if zeros are not appropriately handled, and an application illustrates that risk-estimates can be substantially biased in practice if the time-varying zero probability is not accommodated.

CC1506: **Towards a computationally-tractable maximum entropy principle for non-stationary financial time series**
Presenter: **Anna Marchenko**, Universita della Svizzera italiana, Switzerland
Co-authors: Patrick Gagliardini, Illia Horenko

Statistical analysis of financial time-series of equity returns is hindered by various unobserved/latent factors, resulting in non-stationarity of the overall problem. Parametric methods approach the problem by restricting it to a certain (stationary) distribution class through various assumptions, which often result in misspecification when the problem does not belong to this predefined class. Non-parametric methods are more general, but lead to ill-posed problems and computationally-expensive numerical schemes. We present a non-parametric methodology addressing these issues in a computationally-tractable way by using key concepts like the maximum entropy principle for non-parametric density estimation and a Lasso regularization technique for numerical identification of redundant parameters. In the context of volatility modeling, the presented approach identifies optimal number of regimes with different levels of volatility and their non-parametric switching behaviour. Using historical return data for an equity index we demonstrate that despite viewing the data as conditionally-independent, our methodology leads to identification of robust models being superior to the standard conditional heteroschedasticity models when compared with respect to the Akaike and Bayesian Information Criteria.

CO393 Room Board meeting room I ANIMAL SPIRITS, COMMUNICATION AND REGIME SHIFTS Chair: Ekkehard Ernst

CO1431: **Central bank signaling matters: Evidence from the sensitivity of financial variables to macroeconomic news**
Presenter: **Gabriele Ciminelli**, International Monetary Fund and University of Amsterdam, United States

Two strands of the literature study how financial variables react to (i) macroeconomic news and (ii) unconventional monetary policy announcements. We go a step further and investigate whether the sensitivity of domestic financial variables to US labor market news changed following unconventional monetary policy announcements by the Federal Reserve. The results indicate that this was indeed the case. Following the introduction of calendar-based forward guidance, data indicating higher than expected job creation led to a significant steepening of the yield curve and a marked increase of inflation expectations. Conversely, after Bernanke's Taper Tantrum the same data caused a hump-shaped reaction of the yield curve and no movement in expected inflation. We interpret these findings as suggestive evidence pointing to the existence of a signaling channel of monetary policy. Understanding how this channel works is crucial to explain the reaction of financial variables to news.

CO1530: **Financial frictions and housing collateral contraints in a macro model with heuristics**
Presenter: **Corrado Macchiarelli**, Brunel University, United Kingdom
Co-authors: Paul De Grauwe

The role of household debt on the real activity has attracted considerable attention recently mostly in the light of the observed increases in property

prices and the increase of household indebtedness prior to the 2008 bust in many countries. The relevant literature on housing points to a number of mechanism being likely to trigger or amplify real estate cycles (including bubbles). We focus on the interaction between banks and real estate developments, in particular assessing the implications of changing property prices on consumption decisions. We build on a previously described framework to introduce a real estate sector, accounting in itself for an explicit balance sheet dimension for consumers. The model thus results in an economy where - on the demand side - a collateral constraint limits households ability to borrow against the value of real estate, and - on the supply side - loan supply is constrained by capital. This allows studying the interactions of these two limits by drawing a stark distinction between the supply and demand for credit. While lending constraints are not a new feature of this framework, we take a step further and analyse the implications of lending constraints in a bounded rationality framework.

CO1456: **Central Bank communication: A comparative assessment.**
Presenter: **Rossana Merola**, ILO International Labour Office, Switzerland
Co-authors: Ekkehard Ernst
In the last decades, many central banks have started using communication as an instrument to manage agents expectations. Most of the literature so far has focused on the effects of central banks communication about interest rates, while limited research has been conducted on communication about other macroeconomic aspects. We fill this gap by analysing how central banks disclose information not only concerning interest rates, but also about other macroeconomic conditions. We empirically compare goals and strategies across several central banks (the US FED, the ECB, the Bank of Japan, the Bank of England, the Bank of Canada, the Reserve Bank of Australia and the South African Reserve Bank) over the period 1997-2016. Using the BIS collection of speeches by central bank senior executives, we select a list of keywords to estimate speech intensity in six different fields: monetary conditions, fiscal conditions, financial stability, external competitiveness, economic activity and labour and social issues. Our analysis is threefold. First, we assess whether announcements made by senior executives are mirrored in the policy decisions adopted by their respective central banks. Second, we identify which central banks are more likely to use communication to shape expectations. Third, we investigate whether communication strategy has changed during the crises. The comparison across the selected central banks highlights some insightful differences in strategies.

CC1622: **Endogenous fiscal multipliers**
Presenter: **Ekkehard Ernst**, International Labour Organization, Switzerland
Estimates of fiscal multipliers show a large variation both in terms of size and magnitude. Differences in research strategies are in part responsible for the estimation heterogeneity but so are economic and fiscal conditions. For instance, estimated fiscal multipliers tend to be larger in economic downturns and when fiscal positions are initially strong. Nevertheless, even when controlling for such factors, a significant heterogeneity in estimates remain. We argue that even under similar economic circumstances, differences in social dynamics and communication can lead to differences in fiscal policy impacts. To this effect, we make a distinction between Ricardian and non-Ricardian households, borrowed from the recent New-Keynesian Philips curve literature. Then, in highly segmented markets where Ricardian households interact which each other frequently, fiscal policy interventions loose effectiveness very rapidly. Also, in environments with regular, repeated shocks, fiscal policy interventions are less effective than following sudden, rare events. Finally, we demonstrate that policy communication can lift the effectiveness of interventions even in those cases where the Ricardian effect dominates, depending on the type and structure of communication.

CO1700: **Fiscal targets: A guide to forecasters**
Presenter: **Javier J Perez**, Bank of Spain, Spain
Co-authors: Gabriel Perez Quiros
Should rational agents take into consideration government policy announcements? A skilled agent (an econometrician) could set up a learning model to combine the following two pieces of information in order to anticipate the future course of fiscal policy in real-time: (i) the ex-ante path of policy as published/announced by the government; (ii) incoming, observed data on the actual degree of implementation of ongoing plans. We formulate and estimate empirical models for a number of EU countries (Germany, France, Italy, and Spain) to show that government (consumption) targets convey useful information about ex-post policy developments when policy changes significantly (even if past credibility is low) and when there is limited information about the implementation of plans (e.g. at the beginning of a fiscal year). In addition, our models are instrumental to unveil the current course of policy in real-time. Our approach complements a well-established branch of the literature that finds politically-motivated biases in policy targets.

CO359 Room Board meeting room II FUNDS PERFORMANCE MEASUREMENT Chair: Spyridon Vrontos

CO1471: **Features and determinants of risk in investment choices by private equity funds**
Presenter: **Barbara Chizzolini**, Universita Bocconi, Italy
Co-authors: Leonella Gori, Stefano Gatti
A measure of riskiness of Private Equity (PE) assets is proposed which is an alternative to the CAPM derived beta coefficient usually suggested in the literature on performance of PE Funds. We assume that at any given point in time there exist alternative investment opportunities that can be classified into a limited number of types, and that Funds manage their Portfolio optimally, within the range of investments allowed by their Placement Memorandum. We first estimate a discrete choice model of the Fund Managers investment decisions by type of PE investment, as a function of the observed characteristics of the Fund, of the Deal and of the Portfolio Company, as well as of the year when the deal is closed. Given the chosen type of investment, we then estimate the probability of negative returns of each deal in each investment class. These predicted probabilities together with the historical expected shortfalls by investment type, yield the Expected Loss by deal, the measure of pure risk we propose. We find that it is possible to identify the idiosyncratic features of each investment type and that the patterns and degrees of riskiness differ quite significantly among them.

CO1443: **Robust funds performance evaluation**
Presenter: **Spyridon Vrontos**, University of Essex, United Kingdom
One of the most controversial questions arising in fund management is whether fund managers create additional value to 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. We consider different classes of robust regression models that are designed to capture the special characteristics of funds returns. In this way we examine the existence or not of fund managerial skill and if the high fees are justified by their performance.

CC1254: **The dynamics of equity capital flows by global investment funds to emerging economies**
Presenter: **Andreas Savvides**, Cyprus University of Technology, Cyprus
Co-authors: Duc Nguyen, Gazi Salah Uddin
The dynamics of flows of equity capital by global investment funds to a group of emerging market economies (EMEs) are investigated for the period 1998-2013. Equity capital flows by foreign funds respond readily to changes in financial conditions and are thought to be a good gauge of investor sentiment towards EMEs. Our methodology uses the quantile regression approach for each emerging country separately and the dynamic panel

quantile regression approach for the 23 emerging countries as a group. The advantage of the quantile regression approach is to capture possible asymmetric responses of equity capital flows to internal vs external shocks, subject to various cycles (boom, burst, normal). Our study looks at: (i) the dynamic response of capital flows to the recent unconventional monetary policy; (ii) how uncertainty in global financial system is transmitted to capital flows; we pay particular attention to account for periods of risk on/risk off in how these factors determine capital inflows into EMEs; (iii) the responsiveness of equity capital to changes in global imbalances; and (iv) the extent to which foreign equity capital flows respond to domestic monetary developments as these are manifested in changes in the differential between domestic and foreign interest rates.

CG322 Room 109 CONTRIBUTIONS ON BUSINESS CYCLES Chair: Simon van Norden

CC0218: **Revisiting non-linearities in business cycles around the world**
Presenter: **Artur Silva Lopes**, ISEG - Universidade de Lisboa, Portugal
Co-authors: Gabriel Florin Zsurkis
First differenced logged quarterly series for the GDP of 29 countries and the euro area are used to assess the need to resort to nonlinear models to describe business cycle dynamic behaviour. Our approach is model (estimation)-free, based on testing only. Our aim is to maximize power to detect non-linearities and, simultaneously, to purport avoiding the pitfalls of data mining. The evidence that is found does not support some descriptions because the presence of significant non-linearities is observed for 2/3 of the countries only. Linear models cannot be simply dismissed as they are frequently useful. Contrarily to common knowledge, nonlinear business cycle variation does not seem to be an universal, undisputable and clearly dominant stylized fact. This finding is particularly surprising for the U.S. case. Some support for nonlinear dynamics for some further countries is obtained indirectly, through unit root tests, but this marginal to our study, based on indirect methods only and can hardly be invoked to support nonlinearity in classical business cycles. However, it is relevant from the output gap perspective.

CC0967: **Trend dominance in macroeconomic fluctuations**
Presenter: **Katsuyuki Shibayama**, University of Kent, United Kingdom
The multivariate Beveridge-Nelson decomposition of key macro aggregate data is investigated. We find (a) inflation seems to be dominated by its trend component, and, perhaps as a result of this, the short-term interest rate is also trend dominated; and (b) consumption also seems to be dominated by its trend component perhaps as the permanent income hypothesis suggests. What is new is that, although the difficulty of rejecting a unit root for these variables has been long recognized, we show that these unit root processes account for a large share of the variable fluctuations. This result raises a concern about the convention that the non-stationary data is detrended in standard DSGE-type structural estimation, in the sense that a significant portion of data variation actually may come from the trend components.

CC0733: **Financial cyclical factors and growth: Insights from an augmented stochastic Solow growth model**
Presenter: **Giulia Livieri**, Scuola Normale Superiore, Italy
Co-authors: Michael Donadelli, Antonio Paradiso
An augmented stochastic version of the Solow neoclassical growth model is presented to examine whether financial factors expressed as deviations from their trend represent important business cycle drivers. Our novel framework is used to study the dynamics of the US growth over the period 1890-2013. We find that financial cyclical factors played an important role in explaining output fluctuations in the US over the last century. By comparing different model specifications, we show that the role of each specific financial factor in explaining growth changes over time generating models instability. Taken together, our results have implications for the effectiveness of medium-term policy interventions. Accounting for such cyclical factors is thus relevant for policymakers.

CC1436: **Anchoring countercyclical capital buffers: The role of market liquidity**
Presenter: **Minh Doan**, Deakin University, Melbourne, AUstralia, Australia
The aim is to develop a market illiquidity measure (MI) based on a floating lookback put option of the market portfolio for the build-up and release of capital buffers across financial cycles according to the countercyclical capital buffer (CCB) schemes of Basel III. We find that MI outperforms the credit-to-GDP ratio, a primary indicator recommended by the Basel Committee for Banking Supervision. We show that MI provides timely signals to accumulate capital buffers prior to a crisis and release them as a crisis unfolds. In contrast, the credit-to-GDP ratio generates reliable signals only in the build-up phase and only for large countries such as Germany, the UK and the US. Our findings suggest that MI can be useful as an anchor for CCB schemes especially for releasing capital buffers.

EO725 Room 212 DEPENDENCE MODELS AND COPULAS IV Chair: Wolfgang Trutschnig

EO0155: **On tail dependence coefficients 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 shown. In particular we focus on multivariate Archimedean copulas. We first calculate multivariate transformed tail dependence coefficients when the generator of the considered transformed copula exhibits some regular variation properties, and we investigate the behaviour of these coefficients in cases that are close to tail independence. We obtain new results under specific conditions involving regularly varying hazard rates of components of the transformation. These results are also valid for non-transformed Archimedeans copulas. Secondly, we deal with some hyperbolic transformations. 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.

EO1064: **Seasonal vine copula model for a glacier in King George island**
Presenter: **Mario Gomez**, Universidad Carlos III de Madrid, Spain
Co-authors: Concepcion Ausin, Carmen Dominguez
The study of the mass balance in glaciers is vital to understand their future behaviour. One component of this mass balance is the loss of water produced by melting, known as glacier discharge. A vine copula structure is proposed to model the multivariate and nonlinear dependence among the discharge and meteorological variables. The multivariate joint distribution function is defined as a mixture of four components according to the presence or not of positive discharge and/or precipitation. Each mixture component is modelled with a vine copula. The conditional probability of not having discharge, given the meteorological conditions, is obtained from the proposed joint distribution. Moreover, the structure of the vine copula allows us to derive the conditional distribution function of the discharge given the meteorological values. Three different methods are proposed and compared to obtain predictions of the future discharge values based on three expected loss functions. This framework is applied to a large data base provided by the association GLACKMA who have been collecting data since 2002 from a measurement station located in the King George Island (Antarctica). The seasonal effect is considered by using different parameters for each season. We have found that both, the probability of positive discharge and the prediction of the discharge amount obtained with our proposed vine copula model outperforms the standard results of the classical degree-day models.

EO1645: **Directional multivariate extremes in environmental phenomena**
Presenter: **Rosa Lillo**, Universidad Carlos III de Madrid, Spain
Co-authors: Raul Andres Torres Diaz, Henry Laniado Rodas, Carlo De Michele
Several environmental phenomena can be described by different correlated variables that must be considered jointly in order to be more representative of the nature of these phenomena. For such events, identification of extremes is inappropriate if it is based on marginal analysis. Since there are many references in the literature that propose extremes detection based on copula models, the copula method is generalized by introducing the directional approach. It allows to analyze the data considering all the variables implied in the phenomena, as well as look at the data in interesting directions that can better describe an environmental catastrophe. Advantages and disadvantages of the non-parametric proposal that we introduce and the copula methods are provided. We show with simulated and real data sets how by considering the first principal component direction we can improve the visualization of extremes.

EO1012: **Multivariate CoVaR and hazard scenarios**
Presenter: **Fabrizio Durante**, Free University of Bozen-Bolzano, Italy
A quantitative approach to managing the risk is gaining ground in recent years in geo- and environmental sciences, since practitioners need to identify correctly the risk in order to mitigate major damages due to extreme events like floods, droughts, etc. Recently, various risk measures have been also introduced to quantify the risk of extreme realizations of a single variable given that another variable, also associated with the phenomenon of interest, is larger (respectively, smaller) than some prescribed level. Examples are provided by the Conditional Return Periods and by the Conditional Value-at-Risk (CoVaR). However, the multi-dimensional nature of risk poses the natural question of generalizing such measures to a multivariate setting. A novel methodology is presented to quantify the effects on a target variable Y of a set of covariates that assume values in an extreme multidimensional region. In particular, these latter risky regions are related to the AND, OR, Kendall and Survival Kendall hazard scenarios that are commonly used in environmental literature. Several closed formulas are considered together with an application to spatial risk analysis involving real data.

EO149 Room 214 COMPUTATIONAL METHODS FOR INFERENCE FROM MULTIVARIATE TIME SERIES Chair: Anna Staszewska-Bystrova

EO0161: **Order invariant evaluation of multivariate density forecasts**
Presenter: **Jonas Dovern**, Heidelberg University, Germany
Co-authors: Hans Manner
New tests are derived for proper calibration of multivariate density forecasts based on Rosenblatt probability integral transforms. These tests have the advantage that they i) do not depend on the ordering of variables in the forecasting model, ii) are applicable to densities of arbitrary dimensions, and iii) have superior power relative to existing approaches. We furthermore develop adjusted tests that allow for estimated parameters and, consequently, can be used as in-sample specification tests. We demonstrate the problems of existing tests and how our new approaches can overcome those using Monte Carlo Simulation as well as two applications based on multivariate GARCH-based models for stock market returns and on a macroeconomic Bayesian vector autoregressive model.

EO0370: **New joint confidence bands for structural impulse response functions**
Presenter: **Stefan Bruder**, University of Zurich, Switzerland
Co-authors: Michael Wolf
Constructing joint confidence bands for structural impulse response functions based on a vector autoregression is a difficult task, because of the non-linear nature of such functions. We propose a new joint confidence band that covers the entire true structural impulse response function up to a chosen maximum horizon with a prespecified probability $(1-\alpha)$, at least asymptotically. This confidence band is based on a certain bootstrap procedure from the multiple-testing literature. We compare the finite-sample properties of our method with those of existing methods via extensive Monte Carlo simulations. Additionally, we investigate the effect of endogenizing the lag order in our bootstrap procedure on the finite-sample properties. Furthermore, an empirical application to a real data set is provided.

EO1149: **Density forecasts with endogenous coefficient sparseness**
Presenter: **Sven Schreiber**, Macroeconomic Policy Institute IMK and Free U Berlin, Germany
Frequentist density forecasts with sparse VARs (possibly non-Gaussian) are considered. Sparseness or regressor selection is a form of shrinkage improving the forecasting performance. The forecast error density based on the innovation uncertainty alone can be simulated using resampling methods (residual-based bootstrap). This method can encompass parameter uncertainty by extending the bootstrap to the specification stage. A sparse VAR implies exclusion restrictions, hence the parameter uncertainty will take into account the probability that a coefficient estimate might be zero. The overall sampling distribution of a coefficient in such a model is therefore mixed discrete-continuous with positive mass at zero. However, the degree of sparseness must be chosen, affecting the bootstrap distribution and the forecast density. We propose to calibrate the sparseness by fitting the forecast densities in an evaluation sample. That is, we increase the sparseness of the VAR to minimize the deviation of the probability integral transform (PIT) of the bootstrapped density forecasts from the uniform ideal. For simplicity we use a backward stepwise search, and higher sparseness means a higher cutoff. Other selection tools could also be used. The estimated degree of sparseness implies a unique sampling distribution, allowing probabilistic statements about the sparseness structure, for example the maximal set of coefficients that can be jointly restricted to zero with a certain confidence.

EO0987: **Confidence intervals and bands for impulse responses in structural vector autoregressions with long-run restrictions**
Presenter: **Anna Staszewska-Bystrova**, University of Lodz, Poland
Co-authors: Helmut Luetkepohl, Peter Winker
There is evidence that structural vector autoregressive (VAR) analysis based on long-run restrictions can have quite different implications for the precision of estimated impulse responses than analysis based on short-run restrictions. This issue is investigated in a systematic comparison of confidence intervals and bands for impulse response functions derived using short-run and long-run restrictions. It is found in a Monte Carlo study that the width of confidence bands can differ quite substantially depending on the type of restrictions used, but neither identification scheme is uniformly better in terms of estimation precision.

EO267 Room 209 ADVANCES IN MODEL SPECIFICATION TESTS IN REGRESSION Chair: Tatyana Krivobokova

EO0198: **Exact post model selection inference: Confidence sets based on the lasso**
Presenter: **Ulrike Schneider**, Vienna University of Technology, Austria
Co-authors: Karl Ewald
In a linear regression model with fixed dimension, we construct confidence sets for the unknown parameter vector based on the Lasso estimator in finite samples as well as in an asymptotic setup, thereby quantifying estimation uncertainty of this estimator. In finite samples with Gaussian errors and asymptotically in the case where the Lasso estimator is tuned to perform conservative model selection, we derive formulas for computing the minimal coverage probability over the entire parameter space for a large class of shapes for the confidence sets, thus enabling the construction of valid confidence sets based on the Lasso estimator in these settings. The choice of shape for the confidence sets and comparison with the confidence ellipse based on the least-squares estimator is also discussed.

EO0344: **Bump detection in heterogeneous Gaussian regression**
Presenter: **Farida Enikeeva**, University of Poitiers, France
Co-authors: Axel Munk, Frank Werner
We consider the problem of the bump detection problem of a signal in a heterogeneous Gaussian regression model. We allow for a simultaneous change in mean and in variance of the signal and specify its impact on the difficulty to detect the null signal against a single bump. We derive lower and upper bounds of testing that lead to explicit characterizations of the detection boundary in several subregimes depending on the asymptotic behavior of the bump heights in mean and variance. In particular, we explicitly identify those regimes, where the additional information about a simultaneous bump in variance eases the detection problem. This effect is made explicit in the constant and the rate, appearing in the detection boundary. We also discuss the case of an unknown bump height and provide an adaptive test in that case.

EO1098: **Multiscale tests for shape constraints in linear random coefficient models**
Presenter: **Fabian Dunker**, University of Goettingen, Germany
Co-authors: Konstantin Eckle, Katharina Proksch, Johannes Schmidt-Hieber
A popular way to model unobserved heterogeneity is the linear random coefficient models $Y_i = \beta_{i,1}X_{i,1} + \beta_{i,2}X_{i,2} + \ldots + \beta_{i,d}X_{i,d}$. We assume that the observations $(\mathbf{X}_i, Y_i)$, $i = 1,\ldots,n,$, are i.i.d. where $\mathbf{X}_i = (X_{i,1},\ldots,X_{i,d})$ is a d-dimensional vector of regressors. The random coefficients $\beta_i = (\beta_{i,1},\ldots,\beta_{i,d})$, $i = 1,\ldots,n$ are unobserved i.i.d. realizations of an unknown d-dimensional distribution with density f_β independent of $\mathbf{X}_i$. We propose and analyze a nonparametric multiscale test for shape constraints of the random coefficient density f_β. In particular we are interested in confidence sets for slopes and modes of the density. The test uses the connection between the model and the d-dimensional Radon transform and is based on Gaussian approximation of empirical processes.

EO0909: **Exact, adaptive tests for polynomial regression against a general alternative**
Presenter: **Paulo Serra**, University of Amsterdam, Netherlands
Co-authors: Farida Enikeeva, Tatyana Krivobokova
Goodness-of-fit tests play an important role in the design and validation of parsimonious models. We consider a regression model with correlated noise, and test if the regression function is a polynomial versus a non-parametric alternative. The tests are based on a criterium for estimating parameters of penalised splines and the distribution of the test statistics under the null is known. We will explain the intuition behind the tests and address their consistency. We will also show some numerical results where our approach is compared with some competing methods from the literature. Our approach is also applied to a dataset of growth curves for stem cells.

EO443 Room 202 STATISTICAL GENOMICS FOR TUMOR HETEROGENEITY AND EVOLUTION Chair: Wenyi Wang

EO0267: **Molecular archaeology of cancer**
Presenter: **Peter Van Loo**, The Francis Crick Institute, United Kingdom
The cancer genome carries within it an archaeological record of the tumours past. We will introduce molecular archaeology of cancer algorithms to disentangle the subclonal architecture and life history of cancer from whole-genome sequencing data. These methods and their underlying statistics allow quantitative measurements of the timing of events in the tumours history and of the subclonal composition of the tumour. When multiple samples for the same patient are available, over multiple time points, through multi-region sequencing, or across multiple metastases, such molecular archaeology approaches allow in-depth insights into carcinogenesis and cancer evolution. In lethal metastatic prostate cancer, we for the first time observed polyclonal seeding in human cancers, challenging the commonly accepted notion that each metastasis is formed by one disseminating cancer cell. In an on-going pan-cancer analysis of 2,800 tumours with single biopsies, we are gaining insights into the extent of heterogeneity in different tumour types, the order of acquisition of genetic aberrations and the evolution of mutational signatures. Complementary to these broad-scale single sample bulk sequencing approaches, we are performing smaller-scale in-depth studies of tumour bulk sequencing in combination with single-cell sequencing, aiming to gain deep insight into intra-tumour heterogeneity.

EO0260: **A novel method for estimating somatic mutations cellular prevalence in cancer using whole genome haplotype phasing**
Presenter: **Donatien Chedom Fotso**, University of Oxford, United Kingdom
Co-authors: Christopher Yau, Ahmed Ahmed
Cancer exhibits extensive intra-tumor heterogeneity with multiple sub-populations of tumor cells containing both common and private genomic alterations. Tumor heterogeneity is now seen as one of the major obstacles for precision medicine and the development of effective treatment strategies. Revealing the proportion of cells affected by particular somatic changes is emerging as a key factor for designing targeted therapies, characterizing tumor evolution and understanding chemotherapy-resistance mechanisms. Determining the proportion of cells harboring a particular mutation (the so called cellular prevalence) is a difficult task as it is confounded by several variables as normal cell contamination, local copy number and the temporal relationship between the mutation and the copy number alteration. We have developed a novel method, which uses a combination of side, and haplotype phase information obtained from synthetic long read sequencing to more accurately compute mutational prevalence. We present theoretical examples of known genomic alterations and show using simulation that our method was able to accurately estimate the cellular prevalence compared to standard methods. We also apply our method to a real data set obtained from dense long fragment read sequencing of a single ovarian cancer and demonstrate the highly accurate estimation of cellular prevalence.

EO1258: **Comparing methods for cellularity estimation using different genomic data types in prostate cancer**
Presenter: **Svitlana Tyekucheva**, Dana-Farber Cancer Institute, United States
Co-authors: Syed Haider , Jaeil Ahn, Wenyi Wang, Paul Boutros, Massimo Loda
Reliable information about tumor content in samples subjected to genomic profiling is critical for the analysis of the data form different omics platforms, subsequent integration and interpretation of the results. Variable tumor content may dramatically reduce power of the inference on the genomic alterations and mutations determined from the DNA-based assays and may lead to spurious discoveries and biases in supervised and unsupervised analysis of the gene expression from the RNA-based assays. We performed a careful reassessment of the tumor cellularity in

TCGA prostate adenocarcinoma specimens by a team of highly experienced genitourinary pathologists, as well as estimated tumor content using nine different computational methods designed to process DNA, RNA, miRNA and methylation data. We found better agreement within DNA- and RNA-based methods and less agreement between these groups. Major discrepancies were observed between pathology and computational estimates of the tumor cellularity. We will discuss comparisons between computational and pathology assessment of tumor cellularity and potential sources of the discrepancies.

EO1386: **Heterogeneous tumor expression deconvolution: DeMix-Bayes**
Presenter: **Jaeil Ahn**, Georgetown University, United States
Co-authors: Wenyi Wang, Svitlana Tyekucheva, Giovanni Parmigiani, Ying Yuan
Tumor samples contaminated with neighboring healthy tissues can lead to underestimation of gene expression signatures associated with cancer prognosis or response to treatment. This natural phenomenon of admixing inevitably influences gene expression profiling commonly performed under microarray and RNAseq based transcriptomics. Physical extraction of pure tumor compartments is often not feasible, therefore, in silico dissection of mixed tumor samples is a promising alternative that needs to be precedent to analyze gene expression. We propose a two stage Bayesian deconvolution model (DeMix-Bayes) tenable for both RNAseq and microarray gene expression profiling. DeMix-Bayes mainly addresses two challenges: 1) estimation of tumor proportions; and 2) estimation of expression profiles for both tumor and non-tumor tissues. We first validate DeMix-Bayes through cell-line data, laser-capture micro dissection data. We demonstrate the utility of our approach in TCGA PanCancer RNAseq datasets, in particular, we focus on renal cell carcinoma data where deconvolved expressions offer new clusters that associated with poor and better survival outcomes. Then we compare copy number variations, biomarker-specific mutations, and pathways among the new clusters

EO019 Room 215 RECENT DEVELOPMENTS IN LARGE-SCALE INFERENCE Chair: Guang Cheng

EO0325: **Computationally efficient nonparametric testing**
Presenter: **Guang Cheng**, Purdue Univ, United States
A recent trend of big data problems is to develop computationally efficient inferences that embed computational thinking into uncertainty quantification. A particular focus is two new classes of nonparametric testing that scales well with massive data. One class is based on randomized sketches which can be implemented in one computer, while another class requires parallel computing. Besides introducing these two new methods, our theoretical contribution is to characterize the minimal computational cost that is needed to achieve the minimax optimal testing power.

EO0678: **Testing many moment inequalities**
Presenter: **Kengo Kato**, University of Tokyo, Japan
Co-authors: Victor Chernozhukov, Denis Chetverikov
The problem of testing many moment inequalities is considered when the number of moment inequalities, denoted by p, is possibly much larger than the sample size. There are variety of economic applications where the problem of testing many moment inequalities appears; a notable example is a previous market structure model. We consider the test statistic given by the maximum of Studentized (or t-type) statistics, and analyze various ways to compute critical values for the test statistic. Specifically, we consider critical values based upon (1) the union bound combined with a moderate deviation inequality for self-normalized sums, (2) the multiplier and empirical bootstraps, and (3) two-step and three-step variants of (1) and (2) by incorporating selection of uninformative inequalities that are far from being binding and novel selection of weakly informative inequalities that are potentially binding but do not provide first order information. We prove validity of these methods, showing that under mild conditions, they lead to tests with error in size decreasing polynomially in n while allowing for p being much larger than n.

EO1281: **A direct approach for sparse quadratic discriminant analysis**
Presenter: **Binyan Jiang**, The Hong Kong Polytechnic University, China
Quadratic discriminant analysis (QDA) is a standard tool for classification due to its simplicity and flexibility. Because the number of its parameters scales quadratically with the number of the variables, QDA is not practical, however, when the dimensionality is relatively large. To address this, we propose a novel procedure named QUDA for QDA in analyzing high-dimensional data. Formulated in a simple and coherent framework, QUDA aims to directly estimate the key quantities in the Bayes discriminant function including quadratic interactions and a linear index of the variables for classification. Under appropriate sparsity assumptions, we establish consistency results for estimating the interactions and the linear index, and further demonstrate that the misclassification rate of our procedure converges to the optimal Bayes risk, even when the dimensionality is exponentially high with respect to the sample size. An efficient algorithm based on the alternating direction method of multipliers (ADMM) is developed for finding interactions, which is much faster than its competitor in the literature. The promising performance of QUDA is illustrated via extensive simulation studies and the analysis of two datasets.

EO1415: **DECOrrelated feature space partitioning for distributed sparse regression**
Presenter: **Xiangyu Wang**, Duke University, United States
Co-authors: Chenlei Leng, David Dunson
Fitting statistical models is computationally challenging when the sample size or the dimension of the dataset is huge. An attractive approach for down-scaling the problem size is to first partition the dataset into subsets and then fit using distributed algorithms. The dataset can be partitioned either horizontally (in the sample space) or vertically (in the feature space). While the majority of the literature focuses on sample space partitioning, feature space partitioning is more effective when $p >> n$. Existing methods for partitioning features, however, are either vulnerable to high correlations or inefficient in reducing the model dimension. We solve these problems through a new embarrassingly parallel framework named DECO for distributed variable selection and parameter estimation. In DECO, variables are first partitioned and allocated to m distributed workers. The decorrelated subset data within each worker are then fitted via any algorithm designed for high-dimensional problems. We show that by incorporating the decorrelation step, DECO can achieve consistent variable selection and parameter estimation on each subset with (almost) no assumptions. In addition, the convergence rate is nearly minimax optimal for both sparse and weakly sparse models and does NOT depend on the partition number m. Extensive numerical experiments are provided to illustrate the performance of the new framework.

EO261 Room 002 STATISTICAL MODELLING IN INSURANCE Chair: Vali Asimit

EO0355: **Capital allocation for insurance portfolios with non-linear risk aggregation**
Presenter: **Tim Boonen**, University of Amsterdam, Netherlands
Co-authors: Andreas Tsanakas, Mario Wuethrich

Existing risk capital allocation methods, such as the Euler rule, work under the explicit assumption that portfolios are formed as linear combinations of random loss/profit variables, with the firm being able to choose the portfolio weights. This assumption is unrealistic in an insurance context, where arbitrary scaling of risks is generally not possible. Here, we model risks as being partially generated by Levy processes, capturing the non-linear aggregation of insurance risk. The model leads to non-homogeneous fuzzy games, for which the Euler rule is not applicable. For such games, we seek capital allocations that are in the core, that is, do not provide incentives for splitting portfolios. We show that the Euler rule of an auxiliary linearized game (non-uniquely) satisfies the core property and, thus, provides a plausible and easily implemented capital allocation. In contrast, the Aumann-Shapley allocation, does generally not belong to the core. For the non-homogeneous fuzzy games studied, Tasche's criterion of suitability for performance measurement is adapted and it is shown that the proposed allocation method gives appropriate signals for improving the portfolio underwriting profit.

EO0780: **Robust bootstrap procedures for claims reserving using GLM**
Presenter: **Tim Verdonck**, KU Leuven, Belgium
Co-authors: Kris Peremans, Pieter Segaert, Stefan Van Aelst

Insurers are faced with the challenge of estimating the future reserves needed to handle historic and current claims that are not fully settled. Settlement delays may occur to long legal trials or medical complications. The future reserves may be estimated using generalized linear models using so called run-off triangles. However due to the specific nature of these run-off triangles it is typically difficult to derive analytic expressions for the standard deviation of the resulting reserve estimates. A popular alternative for obtaining standard deviations is then to use the bootstrap technique. Traditional bootstrap procedures are however very sensitive to the possible presence of outliers. Even when bootstrapping a robust estimator, breakdown may occur as a bootstrap sample may contain a higher percentage of outliers than the original sample. Therefore we discuss and implement several robust bootstrap procedures in the claims reserving framework and we investigate and compare their performance on both simulated and real data.

EO1143: **Structural constraints on multi-population mortality models**
Presenter: **Pietro Millossovich**, Cass Business School, United Kingdom
Co-authors: Andres Villegas, Valeria D Amato

Multi-population mortality models have been proposed in the literature to compare and project the mortality of a group of two or more related populations. The benefits of these multivariate models are multiple, ranging from the measurement of mortality differentials), the production of more consistent forecasts of small populations mortality through the experience of larger populations, the assessment of basis risk in standardized hedges. We seek to identify structural constraints to be imposed on multi-population models which jointly specify the entire population and its exhaustive parts. Such approach is appealing because the mortality dynamics of each individual subpopulation take advantage from the reliable estimate of parameters obtained at the entire population level. However, it raises the issue of the internal consistency of such approach. We review different models proposed in the literature under different assumptions and establish under what conditions this consistency can be achieved. Finally, we investigate if the constraints are compatible with the coherence of a forecasting model and discuss implications on mortality projections.

EO1279: **Using risk factors in P& C insurance pricing: A data driven strategy with GAMs, regression trees and GLMs**
Presenter: **Katrien Antonio**, University of Amsterdam and KU Leuven, Belgium
Co-authors: Roel Verbelen, Roel Henckaerts, Maxime Clijsters

A fully data driven strategy is presented to deal with geographical information and continuous risk factors in an insurance tariff. First, a flexible generalized additive model is fit to the claims data. This model contains categorical risk factors, smooth effects of continuous risk factors and a spatial effect that captures geographical information. In practice, actuaries prefer to work with categorical risk factors in combination with generalized linear models. The goal is therefore to bin the continuous risk factors and the spatial effect in order to transform them to categorical risk factors and incorporate them in a GLM. Six different binning methods are compared to bin the spatial effect from which the Fisher-Jenks algorithm gives the best results. In a next step, the smooth effects of continuous risk factors, both single and interaction effects, are binned. Decision trees are proposed to perform the binning because these models result in consecutive bins for the risk factors. Evolutionary trees, combining the framework of decision trees with genetic algorithms, are used because of the extra flexibility of these models compared to classical regression trees. After binning the continuous risk factors and spatial effect, a GLM is fitted with the resulting categorical risk factors. The resulting GLM approximates the original GAM very closely thanks to the data driven binning strategy and even attains lower AIC and BIC values thanks to the reduction in number of parameters.

EO143 Room 213 FLEXIBLE REGRESSION MODELS IN SURVIVAL ANALYSIS Chair: Ingrid Van Keilegom

EO0356: **Estimation of extreme quantiles of a conditional survival distribution**
Presenter: **Jean-Francois Dupuy**, INSA de Rennes, France

The estimation of extreme quantiles of a conditional survival distribution with right-censoring is addressed. We assume that conditionally on covariates, the distribution of the event time is heavy-tailed. First, we propose a weighted kernel version of Hill's estimator of the extreme-value index of the conditional survival distribution. The role of the weighting term is to correct for censoring. Asymptotic normality of the proposed estimator is established. Then, a Weissman-type estimator of conditional extreme quantiles is constructed. A simulation study is conducted to assess finite-sample behaviour of the proposed estimators.

EO0445: **Spline backfitted kernel estimation of a nonparametric additive model with censored data**
Presenter: **Samuel Maistre**, Universite de Strasbourg, France
Co-authors: Anouar El Ghouch, Ingrid Van Keilegom

Nonparametric additive models have been studied widely since the 80's. They provide a good compromise between parametric and fully nonparametric modelings. As each function of interest is one-dimensional, it can be pictured and therefore interpreted easily by practicians. Different techniques aiming to avoid the curse of dimensionality have been proposed, including penalized regression splines and kernel backfitting. A relatively recent method combines the ideas of these two methods, namely Spline Backfitted Kernel (SBK) estimation. We propose to adapt it when the response variable is right censored. This method includes two steps: firstly, estimate the additive model using splines for which the number of knots leads to undersmoothing; secondly, use the previous estimates to perform a univariate kernel smoothing to estimate each function of interest. We show that when we use synthetic data, the asymptotic results of this procedure are similar to those of the uncensored case.

EO0696: **Large sample properties of nonparametric copula estimators under bivariate censoring**
Presenter: **Paul Janssen**, Hasselt University, Belgium
Co-authors: Candida Geerdens, Noel Veraverbeke
A new nonparametric estimator is proposed for the copula function of a bivariate survival function for data subject to right-censoring. We consider two censoring models: univariate and copula censoring. We show strong consistency and we obtain an i.i.d. representation for the copula estimator. In a simulation study we compare the new estimator to a previous estimator recently proposed.

EO0918: **Bayesian inference and identification issues in a flexible promotion time cure model**
Presenter: **Philippe Lambert**, Universite de Liege / Universite catholique de Louvain, Belgium
The promotion time model is a special case of cure survival models where an unidentified proportion of subjects will never experience the event of interest whatever the duration of the follow-up. We focus our interest on a flexible formulation where the baseline hazard for the susceptible subjects is specified using Bayesian *P*-splines. Identification issues arise under an insufficiently long follow-up when the same covariates enter the model specification for the probability to be cured and the time-to-event part. An original proposal to automatically identify such situations will be described and illustrated on real data. Inference will be made in a Bayesian framework using MCMC and compared with results based on Laplace approximations.

EO620 Room 203 NON- AND SEMI-PARAMETRIC APPROACHES IN FUNCTIONAL STATISTICS Chair: Enea Bongiorno

EO0371: **On the perfect reconstruction of partially observed (non-)sparse functional data**
Presenter: **Dominik Liebl**, University Bonn, Germany
Co-authors: Alois Kneip
A new prediction procedure is proposed that allows to reconstruct functional data from their fragmental observations. Similarly to the context of sparse functional data, it is assumed that only noisy discretization points of the random functions are observable. By means of a double asymptotic, we derive the uniform rates of consistency of our functional PCA based estimator for all cases from (moderately) sparse to densely sampled discretization points per function. Furthermore, we derive verifiable a condition, under which our prediction procedure allows for a perfect reconstruction (i.e., without any prediction error) of the unobserved functional fragments. Finite sample properties are investigated through simulations. Applicability of our prediction method is demonstrated by a real data study in which we seek to reconstruct electricity price functions.

EO0992: **Variable selection in functional additive Regression Models**
Presenter: **Manuel Febrero-Bande**, University of Santiago de Compostela, Spain
Co-authors: Wenceslao Gonzalez-Manteiga, Manuel Oviedo de la Fuente
Functional Additive Regression Models (FARM), i.e. regression models involving additive contributions (not necessarily linear) of functional covariates, have been received a lot of attention in the last years mostly focusing in extending the catalogue of multivariate models (parametric, nonparametric or semiparametric) to the functional framework. The problem of variable selection in these kind of models is considered when the covariates can be of different nature: scalar, multivariate or functional. We propose a forward stepwise algorithm that in every step selects the best covariate to be included in the model, updates the regression model seeking the optimal way the contribution of the selected covariate and finally, update the information of the covariates not included in the model given the response and the covariates in the model. Finally, a simulation study and two applications are provided showing the good performance of the method.

EO0982: **Inference on the means of functional data based on a generalized Mahalanobis distance**
Presenter: **Andrea Ghiglietti**, Universita degli Studi di Milano, Italy
Co-authors: Anna Maria Paganoni, Francesca Ieva
The problem of testing about the means of two independent samples of functional data is considered. Several inferential tools used in the classical multivariate analysis are not useful in the infinite dimensional framework of the functional data analysis. For instance, all the procedures based on the popular Mahalanobis distance are not allowed since this metrics is not well-defined in L^2. Then, the problem is typically solved by reducing the inference on a finite dimensional space given by few principal components, but this approach is not always satisfactory, since it may lead to a loss of information and some of its properties are in contrast with the Mahalanobis approach. For this reason, we propose a distance that generalizes the Mahalanobis metrics in infinite dimensional Hilbert space of square integrable functions defined on a compact interval, without any truncation on the number of components considered in the distance. We present a convergence result and a central limit theorem for the generalized distance between the means of two independent samples of curves without specifying the probability distribution of the processes generating the data. These asymptotic results allow us to construct critical regions to test the means of two-samples and to compute analytically the relative power. Finally, these inferential procedures are applied to a case study concerning samples of ECG curves.

EO1220: **Functional principal component for concentration curves**
Presenter: **Enea Bongiorno**, Universita del Piemonte Orientale, Italy
Co-authors: Aldo Goia
Concentration curves are widely used in economic studies (inequality, poverty, differentiation, etc). From a model point of view, such curves can be seen as constrained functional data that refer to the objects oriented data analysis literature. In fact, the family of concentration curves lacks of very basilar structures like the vectorial one and, hence, should be treated with ad hoc methods. The aim is to take care of such lacks providing a rigorous functional framework for concentration curves where it is possible to define Functional Principal Component Analysis (FPCA). The latter technique is then implemented and used to explore functional dataset.

EO077 Room 211 ADVANCES IN MIXED EFFECTS MODELING Chair: Subir Ghosh

EC0414: **Efficient estimation of variance components in nonparametric mixed-effects models with large samples**
Presenter: **Nathaniel Helwig**, University of Minnesota, United States
Linear mixed-effects (LME) regression models are a popular approach for analyzing correlated data. Nonparametric extensions of the LME regression model have been proposed, but the heavy computational cost makes these extensions impractical for analyzing large samples. In particular, simultaneous estimation of the variance components and smoothing parameters poses a computational challenge when working with large samples. To overcome this computational burden, we propose a two-stage estimation procedure for fitting nonparametric mixed-effects regression models. Our results reveal that, compared to currently popular approaches, our two-stage approach produces more accurate estimates that can be computed in a fraction of the time.

EO0376: **Simulation-based estimation in mixed effects models with measurement error and missing data**
Presenter: **Liqun Wang**, University of Manitoba, Canada
A simulation-based approach is proposed for parameter estimation in generalized linear mixed models where some covariates are measured with error and some response data are missing at random. This joint issue is deal with by combining the instrumental variables and probability weighting

method. The proposed estimator is consistent and asymptotically normal under general conditions. The finite sample performance of this estimator is investigated through simulation studies.

EO0639: **Near optimum estimation of variance components in mixed effects models**
Presenter: **Subir Ghosh**, University of California, United States
The variance component estimation for a mixed effects model is considered for analyzing the data obtained from an experiment or observational study. The optimum estimation method yielding the uniformly minimum variance quadratic unbiased estimator (UMIVQUE) is often impossible to carry out. A near optimum estimator (NOPE) is proposed in such situations under some distributional assumptions for the data. Examples of NOPE are presented in an example. A NOPE satisfies the unbiasedness property. The performance comparison of NOPE is made with the maximum likelihood estimator (MLE) and restricted maximum likelihood estimator (REMLE) by simulation for a designed experiment.

EC1557: **Assesing skewness, kurtosis and normality in linear mixed models**
Presenter: **Alexandra Soberon**, Universidad de Cantabria, Spain
Co-authors: Winfried Stute
Linear mixed models provide a useful tool to fit continuous longitudinal data, and the random effects and error term are commonly assumed to have normal distribution. However, this restrictive assumption can result in a lack of robustness and needs to be tested. We propose a very simple tests for skewness, kurtosis and normality based on moment conditions of generalized least squares (GLS) residuals. To do it, estimating higher order momens is necessary and an alternative estimation procedure is developed. Regadless to other procedures in the literature, this is a simpler method that provides a closed-form expression even for the third and fourth order moments. In addition, no further distributional assumptions on neither random effects or error term are needed to show the consistency of the proposed estimators and tests statistics. Their finite sample performance is examined in a Monte Carlo study and the methodology is used to examine changes in the life expectancy and maternal and infant mortality rate of a sample of OECD countries.

EO067 Room 204 MULTIVARIATE EXTREMES Chair: Michael Falk

EO0475: **Risk modelling by max-linear models on graphs.**
Presenter: **Claudia Klueppelberg**, Technical University of Munich, Germany
A new structural equation model is considered, where all random variables can be written as a max-linear function of their parents and independent noise terms. We assume that the dependence structure of the corresponding vector can be modeled by a directed acyclic graph. We show that the multivariate distribution is max-linear and characterize all max-linear models, which are generated by a structural equation model. We investigate the properties of our new model like minimal representations. We also discuss various applications and statistical estimation procedures.

EO0481: **Eigenvalues of sample covariance matrices of heavy-tailed stochastic volatility models**
Presenter: **Anja Janssen**, University of Copenhagen, Denmark
Co-authors: Thomas Mikosch, Mohsen Rezapour, Xiaolei Xie
Stochastic volatility models have become a standard tool for the modeling of both univariate and multivariate time series, in particular with respect to financial applications. We analyze the behavior of a class of multivariate heavy-tailed stochastic volatility models with a focus on the asymptotic behavior of the largest eigenvalues and corresponding eigenvectors of the sample covariance matrix as the number of observations tends to infinity and the number of dimensions is kept fixed. We show that the extremal behavior of the sample covariance matrix depends crucially on the particular choice of model and we analyze both the case where either the innovations or the volatility process are the source of the heavy-tailedness of the time series. In particular, we show that for heavy-tailed innovations the sample covariance matrix behaves asymptotically like in the case of i.i.d. observations (i.e. when there is neither temporal nor componentwise dependence) while the behavior becomes much more complex if the volatility sequence is heavy-tailed.

EO0897: **On some features of the skewed families of max-stable processes**
Presenter: **Boris Beranger**, University of New South Wales, Australia
Co-authors: Simone Padoan, Clement Dombry
Environmental phenomena are processes which are spatial by nature as a single extreme event (heat waves, floods, storms, etc.) often have repercussions at multiple locations. For risk management purposes it is important to have a good understanding of the dependence structure that links such events in order to make predictions on future phenomena, that can have a major impact on real life. Moreover available data at different sites can exhibit asymmetric distributions proving the necessity for max-stable processes that can handle skewness. The extremal-skew-t process, constructed from a non-stationary skew-normal process is a solution that recently appeared in the literature. A generalisation of the discussion on the asymptotics of skewed families of extreme-value processes is provided, including skew-elliptical processes as well as the derivation of a skewed version of the well-known Gaussian extreme-value process. The challenge of simulating these processes is then tackled using the conditional simulation framework and simulated examples are provided to illustrate the method.

EO0799: **Regularly varying Markov trees**
Presenter: **Johan Segers**, Universite catholique de Louvain, Belgium
Co-authors: Gildas Mazo
Extreme values of regularly varying Markov chains can be described in terms of the limiting conditional distribution of the normalized chain given that it is large at a particular time instant. The limit distributions are called forward and backward tail chains, according to the time direction considered. Viewing a chain as a tree consisting of a single, long branch, it is natural to seek for generalizations to general Markov trees, i.e. random vectors whose dependence structure is governed by a tree representing a set of conditional independence relations together with a collection of bivariate distributions along the tree edges. As for Markov chains, extremal dependence of such Markov trees can be described in terms of a collection of tail trees, each tree describing the limit distribution of the rescaled Markov tree given that its value at a particular node is large. Moreover, the time-change formula for tail chains generalizes to a relation between these tail trees. Tail trees can be used to compute quantities such as the number of nodes in the graph affected by a shock at a particular node or the probability that a particular part of the graph will be affected by a shock in another part of the graph. Moreover, specifying the graph structure and the bivariate distributions along the edges provides a construction method for max-stable models.

EO155 Room 201 STATISTICAL METHODS FOR BLIND SOURCE SEPARATION Chair: Sara Taskinen

EO0512: **Convergence of FastICA algorithms in performing temporal ICA for fMRI data**
Presenter: **Jari Miettinen**, University of Jyvaskyla, Finland
Co-authors: Klaus Nordhausen, Sara Taskinen, Klaudius Kalcher, Roland Boubela
The most popular method in independent component analysis (ICA) is called FastICA, which has two classical versions with different approaches to find the independent components. In deflation-based FastICA the components are extracted one by one and in symmetric FastICA simultaneously. Both versions suffer from convergence problems when the sample size is small, which is one reason why temporal ICA has remained less popular than spatial ICA in the analysis of fMRI data. A small modification of the deflation-based algorithm, that is implemented in a function of R package fICA, improves the convergence remarkably. We have tested the convergence of FastICA algorithms in performing temporal ICA for 50 task fMRI data sets with 274 time points.

EO0570: **Tensor-based convolutive independent component analysis**
Presenter: **Frederik Van Eeghem**, KU Leuven, Belgium
Co-authors: Lieven de Lathauwer
Independent component analysis (ICA) tries to find the statistically independent components making up a mixture. One class of algorithms for instantaneous ICA uses tensors, which have favorable properties to tackle this problem. These instantaneous tensor-based algorithms are well established. However, in applications such as digital telecommunications and speech separation, a convolutive mixture model is more appropriate than an instantaneous one. In an attempt to port the properties of tensors to convolutive mixture models, several methods for convolutive ICA have been presented. However, many of these fail to exploit the available structure arising from the convolution. We will show how this structure can be exploited, which will lead to fast algorithms.

EO0857: **Likelihood-based non-Gaussian and Gaussian component analysis**
Presenter: **Benjamin Risk**, SAMSI and University of North Carolina, Chapel Hill, United States
Co-authors: David Matteson, David Ruppert
Independent component analysis (ICA) is popular in many applications, including cognitive neuroscience and signal processing. Due to computational constraints, principal component analysis is used for dimension reduction prior to ICA (PCA-ICA), which could remove important information. The problem is that interesting independent components (ICs) could be mixed in several principal components that are discarded and then these ICs cannot be recovered. We formulate a linear non-Gaussian component model with Gaussian noise components. To estimate this model, we propose likelihood component analysis (LCA), in which dimension reduction and latent variable estimation is achieved simultaneously. Our method orders components by their marginal likelihood in a manner that parallels the ordering of components by variance used in principal component analysis (PCA). We present a parametric LCA using the logistic density and a semi-parametric LCA using tilted Gaussians with cubic B-splines. Our algorithm is scalable to datasets common in applications (e.g. hundreds of thousands of observations across hundreds of variables with dozens of latent components). In simulations, latent components are recovered that are discarded by PCA-ICA methods. We apply our method to an fMRI experiment from the Human Connectome Project and identify artifacts missed by PCA-ICA.

EO1286: **On complex valued ICA**
Presenter: **Pauliina Ilmonen**, Aalto University School of Science, Finland
Co-authors: Niko Lietzen, Klaus Nordhausen
In the independent component (IC) model, the elements of a p-variate random vector are assumed to be linear combinations of the elements of an unobservable p-variate vector with mutually independent components. In independent component analysis (ICA) the aim is to recover the independent components by estimating an unmixing matrix that transforms the observed p-variate vector to the independent components. Complex random signals play an increasingly important role in the field of ICA. The complex IC model is used for example in magnetic resonance imaging or antenna array signal processing for wireless communications and radar applications. In the context of complex valued ICA, we consider estimation based on simultaneous use of two complex valued scatter matrices and comparison of different estimates based on complex version of the minimum distance (MD) index.

EO521 Room 210 RECENT ADVANCES OF STATISTICAL GENETICS AND GENOMICS Chair: Lihong Qi

EO0514: **Learning about the genetic architecture of complex traits across populations**
Presenter: **Hua Tang**, Stanford University, United States
Genome-wide association studies (GWAS) have become a standard approach for identifying loci influencing complex traits. However, GWAS in non-European populations are hampered by limited sample sizes and are thus underpowered. Can GWAS results in one population be exploited to boost the power of mapping loci in another population? We introduce an empirical Bayes approach, which improves the power of mapping trait loci relevant in a specific minority population through adaptively leveraging multi-ethnic evidence. Extending this approach, we discuss methods for genetic risk prediction.

EO0580: **Adaptive surrogate confounding adjustment in genomic mediation analysis and an application to GTEx data**
Presenter: **Lin Chen**, University of Chicago, United States
Recent studies of the effects of genetic variation on expression of distal genes (trans-eQTLs) have revealed that many trans-eQTL effects are mediated by a local (cis-) gene transcript near the eQTL. When conducting mediation analyses in the genome, millions of mediation trios consisting of SNP, cis-gene and trans-gene transcripts are considered and each trio may suffer from different confounding effects. Instead of adjusting for the same set of confounders or surrogate variables in all mediation tests, we propose an adaptive surrogate confounder adjustment algorithm that adaptively selects surrogate confounders for different mediation trios. We applied the proposed methods to a multitissue expression data from National Institute of Health Common Funds Genotype-Tissue Expression (GTEx) program, which has collected transcriptome data in a wide variety of tissue types and genetic data from post-mortem donors. We examined the patterns of cis-gene mediating trans-effect within each tissue type and across different tissue types, with the proposed adaptive surrogate confounding adjustment.

EO1024: **Joint random forest for the simultaneous construction of multiple related networks**
Presenter: **Francesca Petralia**, Icahn School of Medicine at Mount Sinai, United States
The focus is on characterizing common and different coexpression patterns among RNAs and proteins in breast cancer tumors. To address this problem, we introduce Joint Random Forest (JRF), a novel nonparametric algorithm to simultaneously estimate multiple coexpression networks by effectively borrowing information across protein and gene expression data. The performance of JRF was evaluated through extensive simulation studies using different network topologies and data distribution functions. Advantages of JRF over other algorithms that estimate class-specific networks separately were observed across all simulation settings. JRF also outperformed a competing method based on Gaussian graphic models. We then applied JRF to simultaneously construct gene and protein coexpression networks based on protein and RNAseq data from CPTAC-TCGA

breast cancer study. We identified interesting common and differential coexpression patterns among genes and proteins. This information can help to cast light on the potential disease mechanisms of breast cancer.

EO0882: **Construct tumor specific gene regulatory network based on expression data from samples with tumor purity heterogeneity**
Presenter: **Pei Wang**, Icahn School of Medicine at Mount Sinai, United States

Tumor tissue samples often contain an unknown fraction of normal cells. Tumor purity heterogeneity (TPH), which is prevalent in tumor samples, is recently recognized as a severe issue in omic studies. Specifically, if TPH is ignored in the gene co-expression network analysis based on samples with diverse tumor purities, edges are likely to be estimated among genes with mean shifts between normal and tumor samples instead of gene pairs regulating each other in tumor cells. To address this issue, we propose a new method, TSNet, to construct tumor-cell specific gene/protein regulatory networks based on gene/protein expression profiles of tumor tissues. TSNet treats observed expression profiles as mixture of expressions from the normal and tumor cells and explicitly models tumor purity percentage in each tumor sample. Advantage of TSNet over existing methods ignoring TPH are illustrated through extensive simulation examples. In the end, we apply TSNet the analyze the TCGA breast cancer gene expression data set and identify novel co-expression modules specific to tumor cells.

EO1349: **Bayesian modeling to evaluate multiple gene-gene and gene-environment interactions**
Presenter: **Ivan Gorlov**, Dartmouth College, United States
Co-authors: Christopher Amos

Large scale genetic analyses have identified thousands of variants influencing susceptibility to complex diseases but in most cases have failed to explain much of the inter-individual variability in risk. Gene-gene and gene-environment interactions are thought to explain additional risk but modeling these interactions presents challenges in model selection and parameter estimation. We developed Bayesian analytical approaches using a slab and spike prior and using stochastic search variable selection to characterize gene-gene and gene-environment interactions. Constraints were applied to control inclusion and estimation of the interaction parameters. Approaches studied included not constraining the interactions, versus only allowing interaction parameters in the analysis if either one or both of the interactions were included. Simulation studies showed that a model in which interactions were included when either main effect was present best controlled excess parameter inclusion without excess shrinkage. Application to data from lung cancer and melanoma studies supported this general approach.

EO113 Room 006 INFERENCE AND ASSESSMENT OF STOCHASTIC PROCESS MODELS Chair: Hiroki Masuda

EO0525: **Hybrid estimators for discretely observed small diffusion processes**
Presenter: **Masayuki Uchida**, Osaka University MMDS CREST JST, Japan

We consider parametric inference for both drift and volatility parameters of small diffusion processes based on high frequency data. The adaptive maximum likelihood (ML) type estimator and the adaptive Bayes type estimator are proposed. By using the polynomial type large deviation inequality for the statistical random field and the Ibragimov-Has'minskii-Kutoyants program recursively, it is proved that the estimators have asymptotic normality and convergence of moments as the sample size tends to infinity and the small dispersion parameter goes to zero. Furthermore, we study the hybrid estimator defined by the adaptive ML type estimation with the initial Bayes type estimator and show that the estimator has asymptotic normality and convergence of moments. In order to investigate the asymptotic behavior of the estimators from the viewpoint of numerical analysis, we give an example and simulation results for the adaptive estimator and the hybrid estimator. This is a joint work with Ryosuke Nomura.

EO0899: **Time-varying lead-lag effect**
Presenter: **Yuta Koike**, Tokyo Metropolitan University, Japan

A new framework for modeling lead-lag effects in high frequency financial markets is proposed. The model can be accommodated to non-synchronous trading and market microstructure noise as well as the intraday heterogeneity of the lead-lag effects, which are essential for empirical applications. A simple statistical methodology for analyzing the proposed model is presented as well. In particular, a statistical test for the absence of the time heterogeneity of the lead-lag effects can be performed in the proposed framework.

EO0941: **Applications of the quasi-likelihood analysis for point processes to high frequency data**
Presenter: **Nakahiro Yoshida**, University of Tokyo, Japan

Some recent applications of point processes to high-frequency data are discussed. With count data, we estimate certain latent structures such as lead-lag relationships, dependency of the data to covariates, etc. Recently constructed quasi-likelihood analysis (QLA) for point processes is applied to parameter estimation problems. Nonparametric estimation of the correlation between the latent intensity processes of point processes is another application of point processes.

EO1418: **Diffusion processes for the forecast of wind turbine energy production and LAN property**
Presenter: **Alexandre Brouste**, Universite du Maine, France

The management of a wind farm is highly dependent on the forecasting of wind speed and direction. Efficient and safe use of the wind farm with storage, efficient trading and optimal maintenance operations are important. Two diffusion processes (CIR and marginal Weibull diffusion process) are presented and evaluated for short-term forecasting. They can be an entry for stochastic control problems appearing in operational management. Statistical experiments for calibrate the models are studied in terms of the local asymptotic normality property of the likelihoods.

EO575 Room S22 OBJECT ORIENTED DATA ANALYSIS Chair: Benjamin Eltzner

EO0527: **Elastic shape analysis of neuronal trees**
Presenter: **Anuj Srivastava**, Florida State University, United States

There is a great interest neuronal morphology, which seeks tools for comparing neuronal shapes, modeling variability within and across neuron populations, and for clustering/testing/classifying neurons. The challenges include: (1) tremendous variability in size and shape of the main branch (axon), and (2) different numbers, sizes, and shapes of the side branches. That is, the neurons differ in both geometry and topology, and that makes it difficult to model their shapes. An important sub-problem is the registration of points across neurons. An elastic framework for shape analysis of neuronal tree is presented. It is an extension of elastic functional data analysis, where one compares individual curves while being invariant to certain shape-preserving transformations. We define a shape space of these tree representations and impose an elastic Riemannian metric on it to compare different trees. The resulting geodesic paths between neurons show the main branch of one tree deforming into the main branch of the other, while optimally deforming/sliding/creating/destroying the side branches of one into the side branches of other. Using this metric, we define sample mean and variances, and perform principal component analysis of shape data. Furthermore, we cluster and classify neurons into wild types and mutations using this approach. We present some preliminary results using axonal trees taken from the Neuromorpho database.

EO0878: **Random domain decomposition for object oriented spatial statistics**
Presenter: **Piercesare Secchi**, Politecnico di Milano, Italy

Object Oriented Spatial Statistics (O2S2) addresses a variety of application-oriented statistical challenges where the atoms of the analysis are complex data points spatially distributed. The object oriented viewpoint consists in considering as building block of the analysis the whole data point, whether it is a curve, a distribution or a positive definite matrix, regardless of its complexity. When data are observed over a spatial domain, an extra layer of complexity derives from the size, the shape or the texture of the domain, posing a challenge related to the impossibility, both theoretical and practical, of employing approaches based on global models for capturing spatial dependence. A powerful non-parametric line of action is obtained by splitting the analysis along an arrangement of neighborhoods generated by a random decomposition of the spatial domain. The local analyses produce auxiliary new data points which are then aggregated to generate the global final result. We will illustrate these ideas with a few examples where the target analysis is dimensional reduction, classification or prediction.

EO1229: **Inferring accents from geo-localized acoustic speech recordings**
Presenter: **Shahin Tavakoli**, University of Cambridge, United Kingdom
Co-authors: John Aston, Davide Pigoli

The goal is the modelling the variations of speech and accents within England, using acoustic speech recordings from the Audio Edition of the British National Corpus. We propose a non-isotropic spatial functional mean and covariance regression model, in which sound recordings of a particular vowel of interest –encoded by their Mel-Frequency Cepstral Coefficients (MFCC)– are modelled as a spatial mean plus an error term with varying spatial covariance. Due to the particular form of the recordings available, where multiple recordings are available for a same location, we are able not only to estimate the spatial mean MFCC, but also the spatial covariance field of the error terms, without isotropic assumptions. The mean field is fitted by a local linear smoother, and the covariance fields are fitted using a local linear smoother in the tangent plane. Both are fitted using a variable bandwidth parameter chosen in a data driven way.The output of the estimation procedure can then be visualized using an appropriate dimension reduction, and can also be resynthesized into sounds, which can provide qualitative insight into the proposed model.

EO1319: **Nested subspace asymptotics with application to stem cell analysis**
Presenter: **Benjamin Eltzner**, University Goettingen, Germany
Co-authors: Stephan Huckemann

Backward nested subspaces are a valuable tool for dimension reduction of manifold or stratified space valued data. Strong consistency and asymptotics of such estimators are essential for the validity of such methods. To prove strong consistency and standard asymptotics of the full set of nested subspaces, the usual conditions for M-estimators have to be complemented with new arguments dealing with nested random sequences and nested random projection maps of the dimension reduction method as well as with M-estimation under constraints. Interpretation of the asymptotics in terms of a suitable class of splitting charts yields asymptotic normality of each single subspace element. In particular, it is shown that the latter result pertains to principal nested spheres (PNS) analysis. We apply this new method to inferential analysis on early stem cell differentiation.

EO577 Room 216 ADVANCED STATISTICAL MODELING OF NEUROIMAGING DATA Chair: Timothy Johnson

EO0529: **Multi-scale factor analysis of high dimensional time series**
Presenter: **Hernando Ombao**, University of California Irvine, United States

A method is presented for modeling dependence between components of a high-dimensional time series in the context of functional magnetic resonance imaging (fMRI) and electroencephalogram (EEG) data. The primary challenge here is the high dimensionality of these brain signals. The common approach is to first divide the brain volume into anatomically-determined regions; summarize brain activity in each region by spatial averaging and finally compute dependence between the ROI-specific averages. We develop an alternative rigorous approach via multi-scale factor analysis (MSFA). In our framework, we first reduce dimensionality by principal components analysis (PCA) within each ROI. This produces localized summaries according to the PCA criterion. Connectivity between ROIs is characterized through these summaries. The proposed procedure gives a representation of localized brain activity that is an optimal solution to the PCA criterion; captures the multi-scale dependence structure at both local (within-ROI) level and global (between ROIs) level; and achieves dimension reduction therefore can efficiently handle the massive brain data. The novel MSFA approach is used to study functional connectivity in resting-state fMRI and EEG data, which reveals interesting modular and hierarchical structure of human brain networks.

EO0554: **Functional data modeling of dynamic PET data**
Presenter: **Todd Ogden**, Columbia University, United States
Co-authors: Yakuan Chen, Jeff Goldsmith

One major goal of dynamic positron emission tomography (PET) imaging, with particular relevance to the study of mental and neurological disorders, is the estimation of the special density of specific proteins throughout the brain. Current analysis strategies involve applying parametric models that require fairly strong assumptions, reducing information for each subject and each voxel/region into a single scalar-valued summary, and modeling each subject and each voxel/region sequentially. We describe extensions of the analysis in three different directions: a nonparametric approach to the modeling of the observed PET data; a functional data analytic (FDA) approach to modeling the impulse response function; and the ability to consider observed PET data from multiple subjects in a single function-on-scalar regression model. We demonstrate the application of this approach and compare the results with those derived from standard parametric approaches.

EO0623: **A Bayesian approach to the joint modeling of image-based and coordinate-based neuroimaging meta-analysis data**
Presenter: **Silvia Montagna**, University of Warwick, United Kingdom
Co-authors: Tor Wager, Timothy Johnson, Thomas Nichols

Now over 20 years old, functional MRI has a large and growing literature that is best synthesised with meta-analytic tools. Studies entering into the meta-analysis can either share whole-brain statistic images (image-based meta-analysis (IBMA) data) or only report the (x, y, z) locations of peak activations (foci) in published papers (coordinate-based meta-analysis (CBMA) data). We propose a Bayesian hierarchical model for the joint analysis of IBMA and CBMA data to identify areas of consistent activations across studies. For CBMA, we model the foci from each study as a Cox process, whereas for IBMA we regard the statistic image from each study as error-prone measurements of an underlying smooth surface. Both the study-specific log intensity functions and smooth surfaces are characterised as a linear combination of a (common) high-dimensional basis set, and a sparse representation is guaranteed through latent factor modelling of the basis coefficients. Within our framework, it is possible to account for the effect of study-level covariates (meta-regression) and perform model-based predictions of task type or cognitive process for new studies (reverse inference), significantly expanding the capabilities of the current neuroimaging meta-analysis methods available.

EO1226: **Functional domain selection for neuroimaging**
Presenter: **AhYeon Park**, University of Cambridge, United Kingdom
Co-authors: John Aston, Frederic Ferraty, Ah Yeon Park

Motivated by increasing trends of relating brain images to a clinical outcome of interest, we propose a functional domain selection (FuDoS) method that effectively selects subregions of the brain associated with the outcome. View each individual's brain as a 3D functional object, the statistical

aim is to distinguish the region where a regression coefficient $\beta(t) = 0$ from $\beta(t) \neq 0$, where t denotes spatial location. FuDoS is composed of two stages of estimation. We first segment the brain into several small parts based on the correlation structure. Then, potential subsets are built using the obtained segments and their predictive performance are evaluated to select the best subset, augmented by a stability selection criterion. We conduct extensive simulations both for 1D and 3D functional data, and evaluate its effectiveness in selecting the true subregion. We also investigate predictive ability of the selected stable regions. To find the brain regions related to cognitive ability, FuDoS is applied to the ADNI's PET data. Due to the induced sparseness, the results naturally provide more interpretable information about the relations between the regions and the outcome. Moreover, the selected regions from our analysis show high associations with the expected anatomical brain areas known to have memory-related functions.

EO626 Room S23 DOUBLY STOCHASTIC COUNTING PROCESSES Chair: Paula Bouzas

EO0595: **Hawkes process modeling and estimation for crime and conflict data**
Presenter: **George Mohler**, IUPUI, United States
The aim is to provide an overview of self and mutually exciting Hawkes processes with applications to crime and conflict data. We discuss both nonparametric and Bayesian inference methods, including estimation of doubly stochastic Hawkes process models. Results from several case studies and field trials of predictive policing will be presented.

EO0782: **Goodness-of-fit test for Cox processes and generalizations**
Presenter: **Nuria Ruiz-Fuentes**, University of Jaen, Spain
Co-authors: Paula Bouzas
Assessing whether a sample path of an observed counting process follows a Cox process model is a challenge, especially if any fixed structure of the intensity process is assumed. A goodness-of-fit test is derived for a Cox process in different situations, when the intensity of the null hypothesis is known or when it has to be estimated. Considering a compound Cox process, it is usual that the interest lies in the points with certain given marks. These points are proven to be a Cox process and the new intensity can be calculated or estimated using the representation theorem of the compound Cox process. Therefore, the goodness-of-fit test can be extended to this new process. The derived hypothesis tests are illustrated with several simulations of Cox and compound Cox processes. The simulated sample paths were tested directly and after disrupting them in different ways to check the performance of the test. The high percentage of correct decisions proves the accuracy of the proposed tests.

EO0877: **Forecasting a compound Cox process**
Presenter: **Carmen Montes-Gijon**, Universidad de Granada, Spain
Co-authors: Paula Bouzas, Nuria Ruiz-Fuentes
The compound Cox process generalizes the Cox process and models several variations of the latter. A Cox process with random deletions, one with simultaneous occurrences, or a time-space Cox process can all be modeled by a compound Cox process. In any of these cases, the points with a given specific mark can be of interest and, thus, the counting process that is formed is equally of interest. This new counting process has been shown to be another Cox process whose statistics are derivable, therefore closed form expressions of the counting and time statistics are given here. Most of these statistics can be expressed in terms of the mean process of the compound Cox process. Using principal components prediction models, the mean, and hence the statistics, can be predicted. In particular, it is possible to forecast the mean number of points with specific marks as well as the mode or the cumulative distribution function. Additionally, the probability of having a new point within an interval of time can also be forecast under weak assumptions, as proven. Simulations of several examples of compound Cox processes illustrate the results.

EO1480: **Matérn thinned Cox point processes**
Presenter: **Ina Trolle Andersen**, Aarhus University, Denmark
Co-authors: Ute Hahn
Matérn thinned Cox point processes form a new class of models that combine short range repulsion with medium range clustering. They are obtained from Cox point processes, i.e. Poisson point processes with a random intensity measure. The realizations of a Cox point process appear clustered, and there is no lower bound for the distance between neighbouring points. Such a lower bound is introduced by Matérn's thinning algorithm, which originally was defined for homogeneous Poisson point processes only. Points that come closer than a predefined hard core distance are removed following a random decision rule. Matérn thinned Cox point processes are thus triply stochastic point processes. We present general expressions for the intensity and for second-order characteristics of Matérn thinned Cox processes. For shot noise Cox processes, an exact formula for the intensity is derived from the Palm distribution. For other cases, we give an approximation that simplifies calculation. We illustrate the applicability of these models with an example from pathology, where the degree to which cells in bone marrow form clusters is used as a diagnostic criterion.

EO235 Room 208 NONPARAMETRIC AND SEMIPARAMETRIC METHODS FOR DIRECTIONAL DATA Chair: Thomas Verdebout

EO0641: **On the estimation of the density of a directional data stream**
Presenter: **Baba Thiam**, LEM Universite Lille 3, France
Co-authors: Aboubacar Amiri, Thomas Verdebout
Many directional data, such as wind directions, can be collected extremely easily so that experiments typically yield to a huge number of data points that are sequentially collected. To deal with such big data, the traditional nonparametric techniques rapidly require a lot of time to be computed and therefore become useless in practice if real time or online forecasts are expected. We propose a recursive version of the kernel density estimator for directional data which (i) can be updated extremely easily when a new set of observations is available and (ii) keeps asymptotically the nice features of the same classical estimator. Our methodology is based on Robbins-Monroe stochastic approximations ideas. We show that our estimator outperforms the traditional techniques in terms of computational time while being extremely competitive in terms of efficiency with respect to its competitors in the sequential context considered here. We obtain expressions for its asymptotic bias and variance together with an almost sure convergence rate and an asymptotic normality result. Our technique is illustrated on a wind dataset collected in Spain. A Monte Carlo study confirms the nice properties of our recursive estimator with respect to its non-recursive counterpart.

EO0730: **Testing uniformity on high-dimensional spheres against symmetric and asymmetric spiked alternatives**
Presenter: **Christine Cutting**, Universite Libre de Bruxelles, Belgium
Co-authors: Thomas Verdebout, Davy Paindaveine
The problem of testing uniformity on high-dimensional unit spheres is considered. We are primarily interested in non-null issues and focus on spiked alternatives. We show that such alternatives lead to two Local Asymptotic Normality (LAN) structures. The first one is for a fixed spiked direction theta and allows to derive locally asymptotically optimal tests under specified theta. The second one relates to the unspecified-theta problem and allows us to identify locally asymptotically optimal invariant tests. Interestingly, symmetric and asymmetric spiked alternatives lead to very different optimal tests, based on sample averages and sample covariance matrices, respectively. Most of our results allow the dimension p

to go to infinity in an arbitrary way as a function of the sample size n. We perform Monte Carlo studies to illustrate our asymptotic results and we treat an application related to testing for sphericity in high dimensions.

EO1000: **Nonparametric density estimation with directional data under rotational symmetry**
Presenter: **Eduardo Garcia-Portugues**, Carlos III University of Madrid, Spain
Co-authors: Christophe Ley, Thomas Verdebout

Rotational symmetry is a recurrent methodological assumption in directional statistics. Most of the classical distributions for directional or axial data are rotationally symmetric (Fisher-von Mises-Langevin, Watson, Wrapped Cauchy, etc) and the roots of recently developed statistical methods rely on this property. Rotational symmetry is exploited for constructing a constrained semiparametric Kernel Density Estimator (KDE). The estimator is obtained by means of a new operator, termed *rotasymmetrizer*: applied to a KDE, it ensures that the resulting estimator, the RKDE, is rotationally symmetric. The operator is based on the tangent-normal decomposition and connects the RKDE with an adapted KDE in the domain $[-1,1]$. The main properties of the RKDE are derived (bias, variance, asymptotic normality, error measurement), being the most relevant the variance order $(nh)^{-1}$ for arbitrary dimension. These properties hold with the axis of rotational symmetry either known or estimated $\sqrt{n}$-consistently. The improvement in performance with respect to the KDE is checked empirically in a simulation study for fixed and data-driven bandwidths. Finally, some applications of the RKDE in testing are discussed.

EO1214: **Localized spherical deconvolution**
Presenter: **Thanh Mai Pham Ngoc**, University Paris Sud Orsay, France

A new algorithm is provided for the treatment of the deconvolution problem on the sphere which combines the traditional SVD inversion with an appropriate thresholding technique in a well chosen new basis. We establish upper bounds for the behavior of our procedure for any Lp loss. It is important to emphasize the adaptation properties of our procedures with respect to the regularity (sparsity) of the object to recover as well as to inhomogeneous smoothness. We also perform a numerical study which proves that the procedure shows very promising properties in practice as well.

EO231 Room 206 R AND ROBUSTNESS AND OFFICIAL STATISTICS Chair: Valentin Todorov

EO0839: **Robust SNP-based estimation of heritability with R**
Presenter: **Vanda Lourenco**, Faculty of Sciences and Technology - New University of Lisbon, Portugal
Co-authors: Paulo Canas Rodrigues, Ana Maria Pires

Trait heritability, which is the proportion of phenotypic variation in the population that is caused by genetic differences among individuals, plays an important role in plant breeding studies reflecting the breeding potential of a trait. Recently, there has been growing interest in marker-based estimation of heritability which is done via the fit of a linear mixed model, considering individual SNPs as fixed effects and an unobserved polygenic effect in the random effects component of the model. In these circumstances, and assuming marker information is complete, both the coefficient of determination and broad-sense heritability (H2) are equivalent. Since the former is known to be biased when in particular the normality assumption of the error distribution is violated, estimation of H2 may therefore be compromised. Given that plant phenotypic data are prone to contamination and that data contamination is a major contributor to the violation of the normality premise, we propose a robust derivative-free restricted maximum likelihood (DF-REML) framework for the linear mixed model and a robust coefficient of determination for the SNP-based estimation of H2. The performance of the new approach is compared to the classical counterpart and an example of application to a real maize dataset is presented. The robust DF-REML method is implemented in the R software.

EO1238: **Robust estimation in surveys using the conditional bias approach and implementation with R**
Presenter: **Anne Ruiz-Gazen**, Toulouse School of Economics, France
Co-authors: David Haziza, Jean-Francois Beaumont

The classical tools of robust statistics have to be adapted to the finite population context and several robust estimation methods already exist. Recently, a unified approach for robust estimation in surveys has been introduced. It is based on an influence measure called the conditional bias that allows to take into account the particular finite population framework and the sampling design. The main properties of the conditional bias and how it can be used for robust estimation of a total will be recalled for the design-based approach. The implementation in R will also be presented with some functions that estimate the conditional bias and calculate the proposed robust estimators for some particular sampling designs. Some recent advances using the conditional bias approach will be summarized.

EO0985: **Web scraping, text mining and machine learning experiences at Istat with R and other open source systems**
Presenter: **Tiziana Tuoto**, Italian National Institute of Statistics (Istat), Italy
Co-authors: Giulio Barcaroli

Since 2013 Istat is experimenting the use of new sources of data for statistical production, in particular Internet data sources (Internet queries, social networks, websites). One of the most interesting pilots regards the survey on ICT for enterprises, where texts scraped from websites are processed making use of text mining and machine learning techniques in order to produce a subset of the target estimates (characteristics of websites used by enterprises to present their business, for instance, if the website offers e-commerce facilities or job vacancies) of the survey. The sample of surveyed data is the training set used to fit the model to be applied to the generality of websites. The usefulness of such an approach is twofold: (i) to enrich the information available in the Business Register, (ii) to increase the quality of the estimates produced by the survey. All the software used in this application is open. In particular, a large use of R packages as tm, RTextTools, caret, e1071, randomForest, nnet, adabag, rpart, ROCR, rattle has been made. First results of this combined use of survey data and data from the Internet are illustrated, and an evaluation of R packages usefulness is reported.

EO1423: **Exploring outliers in compositional data with structural zeros**
Presenter: **Karel Hron**, Palacky University, Czech Republic
Co-authors: Matthias Templ, Peter Filzmoser

The analysis of multivariate observations carrying relative information (aka compositional data) using the log-ratio approach is based on ratios between variables (compositional parts). Zeros in the parts thus cause serious difficulties for the analysis. This is a particular problem in presence of structural zeros, resulting from a structural process rather than from imprecision of a measurement device. Therefore, they cannot be simply replaced by a non-zero value as it is done, e.g. for values below detection limit or missing values. Instead, zeros have to be incorporated into further statistical processing. We lay the focus on exploratory tools for identifying outliers in compositional data sets with structural zeros. For this purpose, robust Mahalanobis distances are estimated; computed either directly for subcompositions determined by their zero patterns or by using imputation to improve the efficiency of the estimates. We proceed to the subcompositional and subgroup level. For this approach, new theory is formulated that allows to estimate covariances for imputed compositional data and to apply estimations on subgroups using parts of this covariance matrix. Moreover, the zero pattern structure is analyzed using PCA for binary data to achieve a comprehensive view of the overall multivariate structure of zeros. The proposed tools are applied to large-scale data from official statistics, where the need for an appropriate treatment of zeros is obvious.

EO095 Room 207 CHANGE-POINT ANALYSIS Chair: Ivan Kojadinovic

EO0953: **Estimating non-simultaneous changes**
Presenter: **Daniela Jaruskova**, Czech Technical University, Czech Republic

At time points $i = 1,\ldots,n$ we observe a sequence of independent two-dimensional vectors $\{(X_1(i),X_2(i))\}$ such that $corr(X_1(i),X_2(i)) = \rho$ where ρ is known. We suppose that both series have changed once, the goal of statistical inference is to estimate the changepoints (i.e. one change point in the first coordinate, the second change in the second coordinate). Supposing that the true change points differ we show the different behavior in case that both change-points are estimated simultaneously with the case when they are estimated separetely. Finally, we study the situation when the both true change points coincide.

EO0700: **Applications of the spatial sign covariance matrix**
Presenter: **Daniel Vogel**, University of Aberdeen, United Kingdom
Co-authors: Alexander Duerre, Roland Fried, David Tyler

The spatial sign covariance matrix (SSCM) is a highly robust, easy-to-compute multivariate scatter estimator. We explore its use for various aspects of multivariate data analysis. A central research question concerning the SSCM is how its eigenvalues are related to those of the covariance at elliptical distributions. In dimension two, this relation is known explicitly, which can be used to construct a robust pairwise correlation estimator. This estimator has a variety of nice properties. It is fast to compute, distribution-free within the elliptical model, as efficient as similarly robust estimators, and its asymptotic variance admits an explicit form, which facilitates inferential procedures. We further study the SSCM in higher dimensions and propose a change-point test for multivariate dependence based on the SSCM. The asymptotic distribution under the null is derived for stationary, short-range dependent sequences without any moment assumption.

EO0711: **Detecting changes in climate records**
Presenter: **Philippe Naveau**, CNRS-IPSL, France

The focus is on detecting changes in temperature and precipitation over long time periods, either from instrumental or climate model records. The term extremes refers to record events in annual maxima of daily maxima time series. Records are often the most important events in terms of impact and choosing records as the event of interest can improve the inference of the ratio of two small probability events. The gain is particularly substantial if we leverage the tools and hypotheses used in Extreme Value Theory, especially from records properties. To illustrate our approach, the Fraction of Attributable Risk, an often used indicator in event attribution studies, is modified and tailored to handle records. We discuss the advantages of our method through theoretical results, simulation studies, observational time series and outputs from numerical climate models. Comparisons with other non-parametric techniques will be presented.

EO0978: **robts - An R-package for robust time series analysis and change-point detection**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Alexander Duerre, Tobias Liboschik, Jonathan Rathjens

Our progress on the R-package robts, which is available in an early version from R-Forge, are reported. Our package provides different techniques for robust estimation of autocorrelations, partial autocorrelations and spectral densities, for fitting autoregressive time series models, for model diagnostics and prediction. Since many time series models assume second order stationarity, we include robust tests for checking the hypothesis of a stationary mean, of a stationary variance or of stationary autocovariances. Our tests are based on suitable U-statistics or U-quantiles and work under the assumption of short range (near epoch) dependence. The advantages of such tests are illustrated when testing for a constant mean function, whereas the tests for constant autocovariances are currently still under investigation.

EO638 Room 217 MODEL-BASED CLUSTERING OF HIGH DIMENSIONAL DATA Chair: Cristina Tortora

EO1051: **Clustering of high dimensional categorical data via dimension reduction**
Presenter: **Alfonso Iodice D Enza**, Universita di Cassino e del Lazio Meridionale, Italy
Co-authors: Michel van de Velden, Angelos Markos

Conventional clustering algorithms do not scale well in case of high dimensional data because of the well-know curse of dimensionality: the distance of any two data points tends to converge to a same quantity as the dimensionality increases. It is therefore hard to identify clusters. To overcome this problem, one possible approach is to use dimension reduction (DR) techniques (such as principal components analysis) prior to cluster analysis. The sequential application of DR and clustering, however, may suffer the so-called cluster masking problem, that is, the subspace found in the first step may fail at reproducing the cluster structure, or even hide it away. Therefore, several methods have been proposed that jointly optimize the objectives of DR and clustering steps and tackle the cluster masking problem, for continuous and categorical data. The focus is on joint DR and clustering for categorical data: the methods in question combine (multiple) correspondence analysis and K-means. Experiments are presented to evaluate the cluster recovery performances of such methods, also with respect to the sequential approach. All the considered methods, as well as the counterparts for continuous data, are implemented in the R package **clustrd**, which will also be presented.

EO1006: **Using copulas for model based clustering**
Presenter: **Dimitris Karlis**, Athens University of Economics & Business, Greece
Co-authors: Ioannis Kosmidis

The majority of model-based clustering techniques is based on multivariate Gaussian models and their variants. The framework of copula-based finite mixture models is introduced and studied for clustering applications. In particular, the use of copulas in model-based clustering offers several advantages over current methods including the ability to obtain a range of shapes for the clusters, the choice of very flexible marginal models, the ability to jointly model different types of data, just to name few. The basic ideas are presented. The focus is on issues of estimation and usage for different types of data. Estimation in the general case can be performed using standard EM, and, depending on the mode of the data, more efficient procedures can be used that can fully exploit the copula structure. For higher dimensions we use a composite likelihood approach to facilitate the estimation. Computational issues will be discussed. The exposition of the methodology will be accompanied by the analysis of real and artificial data.

EC1432: **Mixture-model based clustering of high-throughput sequencing data**
Presenter: **Anjali Silva**, University of Guelph, Canada
Co-authors: Steven Rothstein, Sanjeena Subedi

High-dimensional data of discrete and skewed nature is commonly encountered in high-throughput sequencing (HTS) studies. Model-based clustering (MBC) techniques for such data are investigated. MBC utilizes mixture models and can account for heterogeneity inherent in a population. Current MBC models for HTS data employ Poisson mixture models with the aim of discovering groups of co-expressed genes in order to shed light on biological functions of genes with unknown functions. As the multivariate extension of Poisson distribution is computationally expensive, many methods assume independence and ignore correlations present between genes. However, implementation of such models can lead to over-fitting of clusters or result in spurious clusters. An alternate using the multivariate Poisson log-normal distribution (MPLN) is also proposed. MPLN allows

for the specification of a covariance structure and can support over-dispersed data. Methodology and computational challenges of these methods will be outlined. Performance is demonstrated using experimental and simulated HTS datasets.

EO1057: **From individual to group predictions**
Presenter: **Ole-Petter Moe Hansen**, Norwegian School of Economics, Norway
Co-authors: Gernot Doppelhofer, Ingvild Almas
The aim is to analyze how groups make economic predictions in an experiment, where members of the group have individually undertaken the same prediction in advance. First, we show that observed group predictions yield significantly higher profit compared to any simple statistical weighting of individual predictions. Second, we quantify the influence each group member has in the group prediction using a structural model. In particular, we find that well prepared males and females with higher grades (GPA) and previous experience with investments are influential in groups. Unprepared females have, on the other hand, far less sway over the group decision. Finally, we find that optimal profits in the model can explain less than 40% of realized group profits, pointing to sizeable group dynamics.

EO708 Room S24 VARIABLE SELECTION METHODS Chair: Camino Gonzalez

EO1121: **Statistical approaches to sensitivity analysis in structural mechanics**
Presenter: **Javier Cara**, Universidad Politecnica de Madrid, Spain
Co-authors: Camino Gonzalez, Jose Mira
The applications of sensitivity analysis in structural mechanics are numerous: inverse analysis, numerical optimization, reliability analysis, modal identification,... Probably the most recent problem is the so-called Finite Element Model Updating, where researchers and engineers try to adjust certain parameters of a mathematical model of a real structure (bridges, buildings,..) taking into account the values recorded using sensors. These mathematical models (obtained using the finite element method) sometimes have thousands of variables, so it is very difficult to perform sensitivity analysis analytically, and many numerical methods have been used instead. In this sense, the performance in this kind of problems of statistical methods such as analysis of variance, linear regression and tree-based methods is analysed.

EO0919: **Variable selection to predict electricity price categories**
Presenter: **Camino Gonzalez**, Technical University of Madrid, Spain
Co-authors: Jose Mira, Julia Garcia Lezana
A legal framework has been recently implemented in the Spanish electricity market under which residential consumers pay according to the hourly price of electricity. For consumers it could be more useful to know, rather than the exact value of the hourly electricity price, the interval or category where this price belongs, for instance high, medium and low, in such a way that the consumers can also implement consumption shift strategies depending on the price signals received at home. Both design of experiments and data mining techniques have been applied to identify the variables with the highest influence on threshold values of price categories as well as optimal settings for these thresholds.

EO0925: **A pruning algorithm for mining maximal length frequent itemsets**
Presenter: **Sina Lessanibahri**, Politecnico Di Milano, Universidad Politecnica de Madrid, Italy
Co-authors: Camino Gonzalez, Luca Gastaldi
Association rule mining is one of the most popular exploratory data mining techniques to discover interesting and previously unknown correlations from datasets. The main goal of association rules algorithms is to find the most frequent set of variables, and then find the correlations between the frequent items. Current algorithms for association rule mining are computationally expensive, especially for very large datasets. Moreover, the large number of discovered frequent itemsets hinders the applications of the algorithms in many real-world datasets. Usually frequent sets with larger length are more interesting and finding the set of maximal length itemsets is useful for many applications. We introduce a novel algorithm, called Width-Sort that efficiently discovers the maximal length frequent itemsets. In Width-Sort, dataset is partitioned based on the transactions lengths to reflects over the additional information hidden in them. Lemmas are developed to estimate an upper bound for the maximal length of the frequent itemsets as well as to prune the items that cannot be part of the maximal length frequent itemsets. The efficiency of the algorithm is tested using both simulated and real-world datasets.

EO1005: **Variable selection with random forests of the influence of road geometry on accident severity**
Presenter: **Jose Mira**, Universidad Politecnica de Madrid, Spain
Co-authors: Bahar Dadashova, Karen Dixon, Blanca Arenas, Francisco Aparicio
Road accidents are highly complex processes very relevant in developed and developing countries, and large resources are devoted to understand them in more depth and thus improve decision-making. Decision trees have been applied in their ensembles of trees version of random forests, which are based on bootstrapping of the data and on random selection of input variables. The data analyzed correspond to the Madrid-Irun corridor in Spain, for 5 different collision types: head-on, angle collisions, sideswipe collisions, rear-end collisions and multiple vehicle collisions. The results for the different collision types are basically coincident as far as the set of most relevant inputs, although some differences are observed in the rankings within the set.

EO609 Room 007 ADVANCES IN BIOSTATISTICS Chair: Maria De Iorio

EO1250: **Covariate dependent clustering via the dependent generalised Dirichlet process**
Presenter: **William Barcella**, University College London, United Kingdom
Co-authors: Maria De Iorio, Stefano Favaro, Gary Rosner
A novel Bayesian nonparametric prior is proposed for collections of random discrete distributions based on the Generalised Dirichlet Process. The weights of the process are modelled using the Beta regression framework allowing to introduce dependence among the random distributions. This strategy allows covariate dependent clustering of the observations. Advantages of the proposed approach include wide applicability, ease of interpretation and efficient MCMC algorithms. The methodology is illustrated through a real data application involving acute lymphoblastic leukaemia and dyslipidemia.

EO1088: **On graphical models for studying the metabolic signature of a disease**
Presenter: **Anna Gottard**, University of Firenze, Italy
The analysis of the metabolome is potentially relevant for many diseases and disorders connected with both genetic and environmental factors. Several studies have detected well-defined metabonomic signatures of diseases. Metabolic signatures can be obtained, for example analysing high-resolution nuclear magnetic resonance spectroscopy (H NMR). Changes in the metabolic spectrum due to appearance/disappearance or concentration of metabolites can provide early evidences of a specific disease or disorder. Graphical models can be useful for a better understanding of this signature in terms of conditional independence between metabolites to investigate the complex connections between the singular elements of a biological system that underlies complex diseases. However, the assumptions of multivariate Gaussianity, linearity in the dependence structure and sparsity are evidently erroneous for this kind of data. We want to compare several alternatives on the lack of linear Gaussian assumption among

which the nonparanormal distribution and the additive noise model. Moreover, we propose some extensions to better meet the peculiarity of this kind of data.

EO0984: **Bayesian autoregressive semiparametric models for gap times of recurrent events**
Presenter: **Alessandra Guglielmi**, Politecnico di Milano, Italy
Co-authors: Maria De Iorio
Autoregressive Bayesian semi-parametric models is proposed for waiting times between recurrent events. The aim is two-fold: inference on the effect of possibly time-varying covariates on the gap times and clustering of individuals based on the time trajectory of the recurrent event. Time-dependency between gap times is taken into account through the specification of an autoregressive component either for the response itself (i.e. the length of the gap time between two successive events), or for the random effect parameters influencing the response at different times. The order of the autoregression may be assumed unknown and object of inference and we consider two alternative approaches to perform model selection under this scenario. The specification of a Dirichlet process on the autoregressive parameters leads to a flexible model strategy and induces a data-driven clustering of the individuals in the sample. Covariates may be easily included in the regression framework and censoring and missing data are easily accounted for. As the proposed methodologies lies within the class of Dirichlet process mixtures, posterior inference can be performed through efficient MCMC algorithm. We illustrate the models through simulations and medical applications involving recurrent hospitalizations of chronic heart failure patients and successive urinary tract infections.

EO1429: **Bayesian nonparametric models in comparative effectiveness research**
Presenter: **Gary Rosner**, Johns Hopkins University, United States
Co-authors: Chenguang Wang
Comparative effectiveness research focuses on evaluations of treatment strategies in real-world settings. The use of Bayesian nonparametric models may help learn about characteristics and sources of variation that contribute to heterogeneities of treatment outcomes across a broad population of patients. We discuss the application of a Bayesian nonparametric model to data from a large (18,000 patients) randomized clinical trial involving community hospitals in the U.S. We also discuss an analysis that considers a registry of patients who did not participate in the clinical trial but who underwent the same procedures.

EO529 Room 205 CAUSAL INFERENCE IN THEORY AND PRACTICE II Chair: Marloes Maathuis

EO1355: **Bayesian structure learning in causal graphical models**
Presenter: **Jack Kuipers**, ETH Zurich, Switzerland
DAGs are the underlying representation of Bayesian networks, a class of probabilistic graphical models widely used in causal inference. Learning the underlying graph from data is a way of gaining insights about the structural properties of a domain. Structure learning forms one of the inference challenges of statistical graphical models. MCMC methods, notably structure MCMC, to sample graphs from the posterior distribution given the data are probably the only viable option for Bayesian model averaging. Score modularity and restrictions on the number of parents of each node allow the graphs to be grouped into larger collections, which can be scored as a whole to improve the chain's convergence. We present an algorithm which employs the underlying combinatorial structure of DAGs to define a new grouping. As a result convergence improves compared to structure MCMC, while still retaining the property of producing an unbiased sample. We discuss applications to causal estimation from psychological survey data and model based clustering of genomic mutation data. For large networks, however, inference remains computationally intensive. By pruning the search space with constraint-based conditional independence tests, the complexity can be significantly reduced allowing the MCMC scheme to be extended to higher dimensional inference.

EO1440: **Learning causal models from time series data in the presence of latent variables**
Presenter: **Daniel Malinsky**, Carnegie Mellon University, United States
Algorithms which search for causal graphical models from observational data can be used to infer causal relations and inform decisions about interventions. While most applications of these methods have been in the i.i.d. domain, it is in many cases straightforward to adapt these algorithms to time series data. We discuss modeling dynamic systems with ancestral graph Markov models, which are well-suited to domains with possible unmeasured confounders (i.e. latent variables) since ancestral graphs accurately represent causal relations and conditional independence facts even when some variables have been marginalized out. We give an overview of these ancestral graphical models for time series with an emphasis on their relationship to structural vector autoregressions (SVARs). We present constraint-based and score-based procedures for learning equivalence classes of (dynamic) ancestral graphs and explore the prospects of relaxing assumptions like linearity, as well allowing for some kinds of heterogeneity.

EO0628: **Characterizing and constructing adjustment sets**
Presenter: **Emilija Perkovic**, ETH Zurich, Switzerland
Co-authors: Johannes Textor, Markus Kalisch, Marloes Maathuis
A graphical criterion is presented for covariate adjustment that is sound and complete for four different classes of causal graphical models: directed acyclic graphs (DAGs), maximum ancestral graphs (MAGs), completed partially directed acyclic graphs (CPDAGs), and partial ancestral graphs (PAGs). Our criterion unifies covariate adjustment for a large set of graph classes. Moreover, we define an explicit set that satisfies our criterion, if there is any set that satisfies our criterion. We also give efficient algorithms for constructing all (minimal) sets that fulfill our criterion, implemented in the R packages dagitty and pcalg.

EO1390: **Incorporating knowledge into structural equation models using auxiliary variables**
Presenter: **Bryant Chen**, UCLA, United States
Co-authors: Judea Pearl, Elias Bareinboim
A new framework, called auxiliary variables (AVs), is introduced. It extends graph-based identification methods for linear models by allowing background knowledge in the form of externally evaluated parameters. Such information could be obtained, for example, from a previously conducted randomized experiment, from substantive understanding of the domain, or even from another identification technique. To incorporate such information systematically, AVs are constructed so that certain paths will be conveniently cancelled. This cancellation allows the AVs to help conventional methods of identification (e.g. single-door criterion, instrumental variables, instrumental sets) and model testing (e.g. d-separation, over-identification). Moreover, by iteratively alternating steps of identification and adding AVs, we can improve the power of existing identification and model testing methods, even without additional knowledge. We have operationalized this general approach for instrumental sets (a generalization of instrumental variables) and show that the resulting procedure subsumes the most general identification method for linear systems known to date. We further discuss the application of this new operation in the tasks of model testing and z-identification.

Sunday 11.12.2016 08:40 - 10:20 Parallel Session L – CFE-CMStatistics

CI683 Room Graduation hall COMBINING FORECASTS Chair: Ekaterini Panopoulou

CI0296: **Density combinations for large data sets in economics and finance**
Presenter: **Francesco Ravazzolo**, Free University of Bozen/Bolzano, Italy
Co-authors: Roberto Casarin, Stefano Grassi, Herman van Dijk

A Bayesian semi-parametric dynamic model combination is proposed in order to deal with a large set of predictive densities. It extends the mixture of experts and the smoothly mixing regression models by allowing combination weight dependence between models as well as over time. It introduces an information reduction step by using a clustering mechanism that allocates the large set of predictive densities into a smaller number of mutually exclusive subsets. The complexity of the approach is further reduced by making use of the class-preserving property of the logistic-normal distribution that is specified in the compositional dynamic factor model for the weight dynamics with latent factors defined on a reduced dimension simplex. The whole model is represented as a nonlinear state space model that allows groups of predictive models with corresponding combination weights to be updated with parallel clustering and sequential Monte Carlo filters. The approach is applied to predict Standard & Poor's 500 index using more than 7000 predictive densities based on US individual stocks and finds substantial forecast and economic gains. Similar forecast gains are obtained in point and density forecasting of US real GDP, Inflation, Treasury Bill yield and employment using a large data set.

CI0328: **Quantile forecast combinations**
Presenter: **Ekaterini Panopoulou**, University of Kent, United Kingdom
Co-authors: Loukia Meligkotsidou, Ioannis Vrontos, Spyridon Vrontos

Whether it is possible to improve point, quantile and density forecasts via quantile forecast combinations is tested. The models we employ are quantile autoregressive and mean regression models augmented with a plethora of macroeconomic and financial variables. Complete subset combinations of both linear and quantile forecasts enable us to efficiently summarize the information content in the candidate predictors. We also develop a recursive algorithm that selects, in real time, the best complete subset for each predictive regression quantile. We provide two forecasting applications; one related to stock market return forecasting and the second on realised volatility forecasting. We show that our approach delivers statistically and economically significant out-of-sample forecasts relative to both the historical average/autoregressive benchmark and the complete subset regression approach.

CO0582: **Combining multiple frequencies in multivariate volatility forecasting**
Presenter: **Alessandra Amendola**, Department of Economics and Statistics - University of Salerno, Italy
Co-authors: Vincenzo Candila, Giuseppe Storti

In a multivariate volatility framework, several options are available to estimate the conditional covariance matrix of returns. Some models, like the multivariate GARCH (MGARCH) ones, rely on daily returns while others exploit the additional information provided by intra-daily prices, like the realized covariance (RC) specifications. A first question arises in the choice of a superior model, for a given period and dataset. An additional source of uncertainty is related to the selection of the frequency at which the intradaily returns, used to construct the RC matrices, are observed. In order to overcome these issues, we propose a prediction strategy based on the combination of multivariate volatility forecasts coming from different model structures, estimated using information at various frequencies. More specifically, we adopt a Model Confidence Set (MCS) procedure applied on rolling forecasts under different loss functions (LFs). Some of the selected LFs rely on the economic evaluation criteria while other belong to the class of robust statistical LFs. The conditional covariances of those models entering the MCS are then equally or proportionally weighted, in order to obtain the combined predictor. Empirical findings give evidence that in an out-of-sample perspective the combined predictor always belongs to the set of superior models.

CO499 Room 104 MACROECONOMETRIC FORECASTING Chair: Shaun Vahey

CO0204: **Multistep prediction error decomposition in DSGE models: Estimation and forecast performance**
Presenter: **Simon Price**, University of Essex, United Kingdom
Co-authors: George Kapetanios, Theodoridis Konstantinos

DSGE models are of interest because they offer structural interpretations, but are also increasingly used for forecasting. Estimation often proceeds by methods which involve building the likelihood by one-step ahead ($h = 1$) prediction errors. However in principle this can be done using different horizons where $h > 1$. Using a well-known model, for $h = 1$ classical ML parameter estimates are similar to those originally reported. As h extends some estimated parameters change, but not to an economically significant degree. Forecast performance is often improved, in several cases significantly.

CO0683: **Asymmetries and unemployment rate forecasts**
Presenter: **Simon van Norden**, HEC Montreal, Canada
Co-authors: John Galbraith

Asymmetries in unemployment dynamics have been observed in the historical time series of unemployment rates in a number of countries, including the United States. Patterns in the corresponding unemployment rate forecasts are studied. We consider conditions under which optimal forecasts will display asymmetry, and the patterns that we would expect to see in forecasts or forecast errors at increasingly long horizons. Using data from the U.S. Survey of Professional Forecasters, we examine the patterns associated with informational rigidity and test the hypothesis, suggested by a simple model of asymmetric response to an underlying causal factor, that asymmetry in forecast errors will be non-decreasing with forecast horizon.

CO1037: **Cognitive dissonance and forecaster overconfidence in a model of inflation expectations with distributional inaccuracies**
Presenter: **Shaun Vahey**, Warwick University, United Kingdom

Professional forecasters and central bankers often describe and quantify the asymmetric risks associated with macroeconomic variables. In contrast, nearly all extant models of inflation expectations assume that agents know exactly the distribution of elliptical disturbances. An alternative mechanism is explored in which agents form expectations of inflation in the presence of distributional inaccuracies. Agents learn the true distribution and inflation dynamics by fitting a Gaussian copula model to the univariate time series. The resulting predictive distributions for inflation are generally asymmetric and dependence is time-varying even though the underlying inflation dynamics are linear. The Gaussian copula model produces superior forecasts to recursive least squares in large samples but not necessarily in small samples. The implications of copula learning for expectations are explored in an application using quarterly US inflation (GDP deflator) data. We examine two specific characteristics of the agents' forecasts relative to recursive least squares. Namely, large and persistent forecasting errors at the mean (cognitive dissonance) and insufficient uncertainty (overconfidence). Nevertheless, with exposure to extreme inflation events, agents better fit the tails of the unknown error distribution, consistent with the notion of adaptive learning.

CO0951: **The real-time properties of the Bank of Canada's staff output gap estimates**
Presenter: **Rodrigo Sekkel**, Bank of Canada, Canada
The revision properties of the Bank of Canada's staff output gap estimates since the mid-1980s are studied. Our results suggest that the average staff output gap revision has decreased significantly over the last 15 years, in line with recent evidence for the U.S. Alternatively, revisions from purely statistical methods to estimate the gap have not experienced the same drop in magnitude. We then examine the usefulness of real-time gap estimates for forecasting in inflation and find no deterioration in forecast performance when inflation projections are conditioned on real-time rather than final estimates of the gap.

CO491 Room 107 MODELLING AND FORECASTING INFLATION AND INFLATION EXPECTATIONS Chair: Gian Luigi Mazzi

CO0205: **The demise of the treaty of Detroit and (dis)inflation dynamics**
Presenter: **Jae Sim**, Federal Reserve Board, United States
Co-authors: Isabel Cairo
The implications for inflation dynamics of the decoupling of real wage growth and productivity growth since 1980s is explored through the lens of New Keynesian monetary economics. A canonical New Keynesian Phillips curve predicts that the current inflation rate is determined as the expected present value of future marginal cost, which is nothing but the ratio of real wage relative to productivity in the long run. We study the ability of the monetary authority to achieve its inflation target when the real wage growth is predicted to be stalled for a considerable amount of time. Using a simple New Keynesian model with search and matching frictions, we consider two cases: one in which the monetary authority observes the drivers of the natural rate of unemployment and correctly readjusts its assessment of the natural rate in real time; an alternative environment in which the monetary authority observes only noisy signals and fails to update the natural rate in a timely manner. We show that the misperception of the natural rate in the latter environment may create a systematic linkage among seemingly unrelated tendencies: misperception in the natural rate; disinflationary pressure; rising income inequality; and growing financial instability.

CO0390: **The anchoring of inflation expectations in the short and in the long run**
Presenter: **Till Strohsal**, Freie Universitaet Berlin, Germany
Co-authors: Dieter Nautz, Aleksei Netsunajev
Structural VAR analysis is introduced as a tool for investigating the anchoring of inflation expectations. We show that U.S. consumers' inflation expectations are anchored in the long run because macro-news shocks are long-run neutral for long-term inflation expectations. The identification of structural shocks helps to explain why inflation expectations deviate from the central bank's target in the short run. Our results indicate that the recent decline of long-term inflation expectations does not result from deanchoring macro-news but can be attributed to downward adjustments of consumers' expectations about the central bank's inflation target.

CO0624: **Risk-adjusted expectations of inflation**
Presenter: **Marcello Miccoli**, Banca dÍtalia, Italy
Co-authors: Marco Casiraghi
A new way is proposed to compute market-based risk-adjusted measures of inflation expectations. We focus on rates of inflation swap contracts, a financial contract widely used to hedge inflation and the paramount proxy of market-based inflation expectations used by central banks. Borrowing from the finance literature, we study the ex-post excess return on inflation swap contracts (the difference between the swap rate at a given maturity and the realized inflation rate over the same horizon) which is an unbiased proxy of risk premia under the rational expectations hypothesis. Using data from the euro area and the US, preliminary analysis shows that the risk premia on inflation swap rates at short-to-medium maturities can be predicted by macroeconomic and financial variables that are present in agents information set at the time the contract is signed. The estimated risk premia are increasing in economic activity and decreasing in financial market volatility. This econometric analysis is then used to construct a measure of risk-adjusted inflation expectations so as to assess the role of risk premia in determining inflation swap rates.

CO0767: **Censored fan charts of inflation**
Presenter: **James Mitchell**, University of Warwick, United Kingdom
The aim is to consider the interpretation, and present methods for the production and evaluation, of fan charts or density forecasts of inflation explicitly distinguishing between risk (knowable uncertainty) and (unknowable or radical) uncertainty. An application to the fan charts published by the Bank of England's Monetary Policy Committee demonstrates the importance of this distinction; and considers its modelling implications by extending the two-piece normal and t distributions to either censored or modelled extreme values.

CO293 Room 110 SVAR ANALYSIS Chair: Gregor von Schweinitz

CO0213: **News shocks: Different effects in boom and recession**
Presenter: **Fischer Sarah**, University of Bern, Switzerland
Co-authors: Maria Bolboaca
The nonlinearity in the effects of news shocks about technological innovations is investigated. In a maximally flexible logistic smooth transition vector autoregressive model, state-dependent effects of news shocks are identified based on medium-run restrictions. We propose a novel approach to impose these restrictions in a nonlinear model using the generalized forecast error variance decomposition. We compute generalized impulse response functions that allow for regime transition and find evidence of state-dependency. The results also indicate that the probability of a regime switch is highly influenced by the news shocks.

CO0495: **On the the Swiss exchange rate shock**
Presenter: **Gregor von Schweinitz**, Halle Institute for Economic Research, Germany
Co-authors: Manuel Buchholz, Lena Tonzer
Following a depreciation of the Euro due to turmoil in the Euro area and the announcement of further monetary expansion of the European Central Bank, the Swiss National Bank (SNB) suddenly abolished the exchange rate peg with the Euro in January 2015. The unexpected step led to immediate disruptions in stock markets and uncertainty about the consequences for the future growth pattern in Switzerland. We first ask whether the Swiss exchange rate shock caused significant stock market reactions. Second, we test whether this effect has been heterogeneous across sectors like banks or non-financial firms. We exploit that the removal of the one-sided cap of the exchange rate occurred unexpectedly. Making use of this surprise component, we have an exogenous shock that allows identifying effects in a synthetic matching framework. Third, results from a sign-restricted smooth transition VAR show that the exchange rate did not have any medium- to long-run effects despite the massive short-run market disturbances.

CO0559: **An analysis of monetary policy during the great moderation**
Presenter: **Makram El-Shagi**, Henan University, China
Co-authors: Logan Kelly
An empirical framework is developed to show the importance of money during the Great Moderation, while accounting for the fact that monetary policy was exclusively conducted through interest rates. We estimate the impulse response functions and forecast error variance decomposition derived from the LASSO-SVAR with a LASSO based lag selection. The variance decomposition suggests that a substantial component of macroeconomic variation has been driven by shocks to the money market, which were not only unintended by the Federal Reserve, but worse passed unnoticed allowing those shocks to accumulate over time.

CO0576: **Structural VAR modelling with independent shocks**
Presenter: **Helmut Herwartz**, Georg-August-University Goettingen, Germany
Structural shocks in multivariate dynamic systems are hidden and often identified with reference to a-priori economic reasoning. Based on a non-Gaussian framework of independent shocks, an approach is provided to discriminate between alternative identifying assumptions on the basis of dependence diagnostics. Relying on principles of Hodges-Lehmann estimation, we suggest a decomposition of reduced form covariance matrices that yields implied least dependent (structural) shocks (LDS). A Monte Carlo study illustrates power and consistency of the proposed identification strategy in distinguishing competing ad-hoc specifications of transmission from the structural to the reduced form model. Applying the approach to a macroeconomic model of the Euro area, independent shocks conform with features of demand, supply and monetary policy shocks. A tightening of monetary policy is found to exert a dampening effect on asset prices which tapers off after six quarters.

CO642 Room 103 ADVANCES IN FINANCIAL DATA MODELING Chair: Matthias Hartmann

CO0239: **Forecast performance, disagreement, and heterogeneous signal-to-noise ratios**
Presenter: **Matthias Hartmann**, Heidelberg University, Germany
An imperfect information model for the expectations of forecasters is proposed that explains differences in average disagreement levels across forecasters by means of cross sectional heterogeneity in the variance of private noise signals. We show that the forecaster-specific signal-to-noise ratios determine both the average individual disagreement level and an individuals' forecast performance: forecasters with very noisy signals deviate strongly from the average forecasts and report forecasts with low accuracy. We take the model to the data by empirically testing for this implied correlation. Evidence based on data from the Surveys of Professional Forecasters for the US and for the Euro Area supports the model for short- and medium-run forecasts but rejects it based on its implications for long-run forecasts.

CO0380: **Testing for an omitted long-term component in multiplicative GARCH models**
Presenter: **Christian Conrad**, Heidelberg University, Germany
A misspecification test is developed for the multiplicative two-component GARCH-MIDAS model. 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 an explanatory variable. We suggest a Lagrange Multiplier statistic for testing the null hypothesis that the 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 derive the asymptotic theory for our test statistic and investigate its finite sample properties by Monte-Carlo simulation. The usefulness of our procedure is illustrated by an empirical application to S&P 500 return data.

CO0410: **Stylized facts for extended HEAVY models: The importance of asymmetries, power transformations and long memory**
Presenter: **Menelaos Karanasos**, Brunel University, United Kingdom
Co-authors: Starvoula Yfanti
The High frequency bAsed VolatilitY (HEAVY) model is studied and extended. Our main contribution is the enrichment of the model with asymmetries, power transformations and long memory (fractionally integrated or hyperbolic) through the hyperbolic (double) asymmetric power (HYDAP) formulation. The conclusion that the lagged realized measure does all the work at moving around the conditional variance of stock returns while it holds in the benchmark specification it does not hold once we allow for asymmetric, power and long memory effects. Our main findings are as follows. First, the power transformed conditional means (of the squared returns and the realized measure) are significantly affected by both the lagged power transformed realized measure and absolute negative returns. Second, fractional integration fits the HEAVY model of the returns better, whereas hyperbolic long memory is more suitable for modelling the conditional mean of therealized measure. Third, the overnight trading activity indicator affects positively the daily volatility and lowers its persistence, whereas it has a trivial impact on the intra-daily conditional variance. Fourth, the HEAVY framework applied to the Garman Klass (GK) volatility gives results very similar to the ones from the stock returns.

CO0236: **Semiparametric estimation of multivariate GARCH models**
Presenter: **Claudio Morana**, Universita di Milano-Bicocca, Italy
A new simple semiparametric estimator of the conditional variance-covariance and correlation matrix (SP-DCC) is introduced. While sharing a similar sequential approach to existing dynamic conditional correlation (DCC) methods, SP-DCC has the advantage of not requiring the direct parameterization of the conditional covariance or correlation processes, therefore also avoiding any assumption on their long-run target. In the proposed framework, conditional variances are estimated by univariate GARCH models, for actual and suitably transformed series, in the first step; the latter are then nonlinearly combined in the second step, according to basic properties of the covariance and correlation operator, to yield nonparametric estimates of the various conditional covariances and correlations. Moreover, in contrast to available DCC methods, SP-DCC allows for straightforward estimation also for the non-symultaneous case, i.e., for the estimation of conditional cross-covariances and correlations, displaced at any time horizon of interest. A simple ex-post procedure, to ensure well behaved conditional covariance and correlation matrices, grounded on nonlinear shrinkage, is finally proposed. Due to its sequential implementation and scant computational burden, SP-DCC is very simple to apply and suitable for the modeling of vast sets of conditionally heteroskedastic time series.

CO409 Room 106 LONG RUN CO-MOVEMENT Chair: J Isaac Miller

CO0246: **Time varying cointegration and Kalman filter**
Presenter: **Taner Yigit**, Bilkent University, Turkey
Co-authors: Burak Eroglu, J Isaac Miller
Despite its popularity in analyzing different kinds of time varying parameter models, the Kalman Filter (KF) technique has been overlooked so far in the literature on cointegration vector (CV) instability for two reasons: i) because it suggests a very specific (stochastic) form of time variation in the estimated parameters and may fail to capture more general models such as structural breaks, regime switches, etc., and ii) it is vulnerable to the problem of spurious regression because of its Gaussian error assumption and the imposition of a cointegrating relation without the test for it. We propose a method that addresses the latter issue and offers a reliable recovery of the instability in CVs. The results show that KF provides an encompassing estimation strategy to the estimation of CV whether it is time varying or not. Thereby, our methodology proposes a "universal"

method of estimation for cointegrating vectors, suggests a method of testing for cointegration when the relation is time varying, and avoids the risk of spurious regression.

CO0327: **Unit root and mean reversion in financial time series generated from diffusion models**
Presenter: **Jihyun Kim**, Toulouse School of Economics, France
Co-authors: Joon Park
The unit root and mean reversion properties of general diffusion processes and their samples in discrete time are analyzed. In particular, we show that the Dickey-Fuller unit root test has a well-defined limit distribution if, and only if, the underlying diffusion has no mean reversion, while it diverges to minus infinity in probability if, and only if, the underlying diffusion has mean reversion. If applied to the sample from a diffusion model, the test therefore becomes a test for no mean reversion rather than a test for nonstationarity of the underlying diffusion. Diffusion processes are mean-reverting as long as their drift terms play a dominant role, and nonstationary diffusions may well have mean reversion.

CO0666: **Robust inference in structural VARs with long-run restrictions**
Presenter: **Guillaume Chevillon**, ESSEC Business School, France
Co-authors: Sophocles Mavroeidis, Zhaoguo Zhan
Long-run restrictions are a very popular method for identifying structural vector autoregressions (SVARs). A prominent example is the debate on the effect of technology shocks on employment, which has been used to test real business cycle theory. The long-run identifying restriction is that non-technology shocks have no permanent effect on productivity. This can be used to identify the technology shock and the impulse responses to it. It is well known that long-run restrictions can be expressed as exclusion restrictions in the SVAR and that they may suffer from weak identification when the degree of persistence of the instruments is high. This introduces additional nuisance parameters and entails nonstandard distributions, so standard weak-instrument-robust methods of inference are inapplicable. We develop a method of inference that is robust to this problem. The method is based on a combination of the Anderson and Rubin test with instruments derived by filtering potentially non-stationary variable to make them near stationary. We find that long-run restrictions yield very weak identification is some cases. On the hours debate, we find that a previous difference specification is very well identified, while another level specification is weakly identified.

CO0232: **Decomposing climate sensitivity: A statistical approach for a spatially heterogeneous planet**
Presenter: **J Isaac Miller**, University of Missouri, United States
Climate sensitivity, which relates anthropogenic forcings to global mean temperature, is defined by a constant parameter from an energy balance model for the Earth and identified by the requirement that net heat flux must be zero in equilibrium. Although net heat flux must sum to zero for the planet, the summands may consist of spatially heterogeneous non-zero changes in net heat flux and changes in the spatial distribution of temperature anomalies. A statistical methodology is proposed to estimate the heterogeneity, and a statistical test is developed that supports its existence using HadCRUT4 and GISS data.

CO323 Room 105 TIME-SERIES ECONOMETRICS Chair: Robert Kunst

CO0290: **Forecasting tourist arrivals with the help of web sentiment: A mixed-frequency modeling approach for big data**
Presenter: **Ulrich Gunter**, MODUL University Vienna, Austria
Co-authors: Irem Onder, Arno Scharl
Online news media coverage of a destination can affect destination image and, in turn, can influence the number of tourist arrivals. It is a form of Big Data, which can be crawled and collected in various ways. Moreover, destination image can change during and after an online browsing session when reading news related to a destination. Sentiment analysis extracts web sentiment by rating a segment of text as either positive (favorable) or negative (unfavorable), which shows the perception of the news author about a topic in discussion. The goal is to investigate whether web sentiment data, which are based on online news media coverage of four European cities (Berlin, Brussels, Paris, and Vienna), possess informative content able to predict actual tourist arrivals. To achieve this goal, sentiment analysis of online news media coverage was conducted using automated semantic routines. Due to different data frequencies of tourist arrivals (monthly) and web sentiment indicators (daily), the mixed-data sampling (MIDAS) modeling approach was applied. Results show that MIDAS models employing various types of web sentiment indicators are able to outperform time-series and naive benchmarks in terms of typical accuracy measures.

CO0707: **Distance measures implied by forecast evaluation criteria**
Presenter: **Robert Kunst**, Institute for Advanced Studies, Austria
Traditional moment-based measures of predictive accuracy, such as the mean squared error (MSE) and the mean absolute error (MAE), assess the precision of forecasts in the framework of widely accepted metric spaces. Many researchers, however, pursue more complex targets, such as the mean absolute percentage error (MAPE), often motivated by an attempt to reduce the influence of scaling. We argue that most of these measures are characterized by asymmetry in the sense that moving the actual closer to the forecast has a quite different effect from moving the forecast, and also by non-convexity of the implied environments. For some of them, even paradox effects can be generated, such as a deterioration of accuracy as the actual approaches the forecast. We illustrate all effects using contour plots and other visualization tools. Our warning against the careless usage of relative asymmetric criteria adds to the recent argument that these criteria may be hampered by the non-existence of moments.

CO1132: **Using profit persistence to predict stock returns: An alternative model**
Presenter: **Michael Hauser**, Vienna University of Economics and Business, Austria
Co-authors: Adelina Gschwandtner
The fact that past profit rates are a successful predictor for future stock returns has been previously established. The aim is to show how deviations from a firm specific long run profit rate, and a firm specific deviation from the cross-section profit rate can be used to forecast stock returns. Thereby also nonlinear but symmetric effects are found. We implement a 2-step estimation procedure. In the 1st step the long run profit rate is estimated, which is introduced as an explanatory variable measured with errors in the return equation. Date are a merged set of the Compustat and CRSP data base for 1963-2006. The persistence of profits literature in industrial economics helps to justify these effects.

CO1009: **Linear trends, fractional trends and initial conditions**
Presenter: **Heiko Rachinger**, University of Vienna, Austria
Efficient estimation is analyzed for linear trends in long memory time series. First, we consider the case of a long memory Type II error process and show that a generalized least squares (GLS) estimator that corrects the serial correlation of the error term is efficient. Second, we take into account an initial condition which bridges the two alternative definitions of long memory, Type I and Type II. In this case, a weighted least estimator (WLS), which is the efficient estimator for Type I, outperforms the GLS even for short initial conditions. It reaches efficiency when the initial condition becomes more remote. Consequently, the choice between the two estimators depends on presence and length of the initial condition. In order to illustrate the methodology, we estimate the GDP growth rates of three countries and test whether these rates are positive.

CO197 Room 111 ADVANCES IN FINANCIAL ECONOMETRICS Chair: Ostap Okhrin

CO0345: **Modeling and forecasting (un)reliable realized covariances for more reliable financial decisions**
Presenter: **Rogier Quaedvlieg**, Maastricht University, Netherlands
Co-authors: Tim Bollerslev, Andrew J Patton
A new framework is proposed for modeling and forecasting common financial risks based on (un)reliable realized covariance measures constructed from high-frequency intraday data. Our new approach explicitly incorporates the effect of measurement errors and time-varying attenuation biases into the covariance forecasts, by allowing the ex-ante predictions to respond more (less) aggressively to changes in the ex-post realized covariance measures when they are more (less) reliable. Applying the new procedures in the construction of minimum variance and minimum tracking error portfolios results in reduced turnover and statistically superior positions compared to existing procedures. Translating these statistical improvements into economic gains, we find that under empirically realistic assumptions a risk-averse investor would be willing to pay up to 170 basis points per year to shift to using the new class of forecasting models.

CO0535: **Clustering estimator of the HAC for high-frequency data**
Presenter: **Anastasija Tetereva**, University St Gallen, Switzerland
Co-authors: Ostap Okhrin
A computationally simple estimator is introduced for the multivariate hierarchical Archimedean copulae. It is proposed to estimate the structure and the parameters of a copula simultaneously based on the correlation matrix only. The advantage of the average correlation estimator is the significant reduction of the computational costs and that it can be used in cases when the maximum likelihood type estimation can not be performed. Extensive simulation studies show the superior performance and the low computational costs of the proposed estimator in comparison to the benchmark models.In the case of high-frequency data, the proposed algorithm enables the estimation based on the realized covariance matrix. The application of the estimator to the one-day-ahead Value at Risk prediction using high-frequency data gives rise to the hierarchical realized copula (RHAC). The RHAC exhibits good forecasting properties for a multivariate portfolio in comparison to the dynamic copula and realized covariance models and does not suffer under the curse of dimensionality.

CO0742: **Estimating long memory in volatility by means of factor models**
Presenter: **Roxana Halbleib**, University of Konstanz, Germany
A method is provided for modeling the long-memory of realized volatilities by means of factor models. Applying the model on large panels of realized volatilities increases the precision of the estimates and shows that the long memory of volatilities can be captured by the aggregation of short memory dynamic factors with different degrees of persistence. The model outperforms standard approaches, such as ARFIMA and HAR model both on simulated and real data.

CO0761: **On LASSO-GARCH approach**
Presenter: **Iryna Okhrin**, Dresden University for Technology, Germany
Co-authors: Sasa Zikovic
It is a stylized fact that GARCH type models are good at describing the characteristics of financial data (leverage effects, volatility clustering, etc.). List of various XGARCH models has increased tremendously over the last decade, creating a serious model selection problem. Estimation of the most general model is a challenging task and the results are not easy to explain; it also leads to overestimation. We propose a LASSO-GARCH approach which allows for estimation and simultaneous model simplification. The question estimation and simplification relates not only to the number of lags to be taken into account, but also to the specific characteristics of the data. For example, if no leverage effect is present, there is no need to estimate an asymmetric GARCH model and it is exactly what a LASSO-GARCH procedure avoids. In our LASSO-GARCH study we consider a family of GARCH models and estimate them via a penalized ML approach. Simulation study results show consistency and efficiency of the estimator under different setups. An empirical study confirms the simulation results and shows the superior performance of the approach.

CO597 Room 101 BIG DATA, HIGH-FREQUENCY TRADING AND MARKET STRUCTURE Chair: Michael Stein

CO0545: **Low-latency trading and price discovery: Evidence from the Tokyo stock exchange in the pre-opening and opening periods**
Presenter: **Darya Yuferova**, Norwegian School of Economics, Norway
Co-authors: Mario Bellia, Jun Uno, Loriana Pelizzon, Marti Subrahmanyam
The aim is to study whether the presence of low-latency traders (including high-frequency traders (HFTs)) in the pre-opening period contributes to price discovery and liquidity provision in the subsequent opening call auction. We empirically investigate these questions using a unique dataset based on server IDs provided by the Tokyo Stock Exchange (TSE), one of the largest stock markets in the world. Our data allow us to develop a more comprehensive classification of traders than in the prior literature, and to investigate the behavior of the different categories of traders, based on their speed of trading and inventory holdings. We find that HFTs dynamically alter their presence in different stocks and on different days; therefore, we focus on HFT activity only when traders utilize their low latency capacity. We find that, in spite of the lack of immediate execution, about one quarter of HFTs participate in the pre-opening period, and significantly contribute to price discovery. They also contribute to liquidity provision in the opening call auction. We also document that HFTs contribute to price discovery and are liquidity consumers during the continuous period. However, this result is driven by the three quarters of HFTs that were inactive in the pre-opening period. In contrast, those that were active in the pre-opening period contribute to liquidity provision in the subsequent continuous session.

CO0828: **The effects of artificial fragmentation of trading in equity markets**
Presenter: **Satchit Sagade**, Goethe University Frankfurt, Germany
Co-authors: Alejandro Bernales, Italo Riarte, Marcela Valenzuela, Christian Westheide
Equity markets in the US, EU and elsewhere have moved from monopolist/dominant exchanges to a competitive environment where multiple trading venues coexist. Studies - while highlighting the trade-off between competition-induced innovation and/or cost reduction, and the detrimental effects of fragmentation due to network externalities - have provided conflicting results. We argue that this can be attributed to the difficulty in disentangling the effects of fragmentation and competition. We develop a dynamic model of limit order markets, where heterogeneous agents make endogenously sequential optimal decisions in continuous time to maximize expected payoffs, taking into account markets conditions, potential future trading decisions, and other agents strategies. We compare two scenarios: a market organized as a single limit order book versus one where multiple books coexist. We find that fragmented markets benefit arbitrageurs and market-makers at the expense of liquidity-motivated traders. Consolidated markets, on the other hand, result in positive network externalities leading to improvements in market quality and welfare. We confirm the models predictions by examining a unique event at NYSE Euronext which led to a reduction in market fragmentation without affecting overall competition. Our results suggest market competition, not market fragmentation, in market design drives market quality improvements when new trading venues emerge.

CC1624: **Analysis of high-frequency order flow dynamics with machine learning techniques**
Presenter: **Adamantios Ntakaris**, Tampere University of Technology, Finland
Co-authors: Martin Magris, Juho Kanniainen, Alexandros Iosifidis, Moncef Gabbouj
We use machine learning techniques to train a predictive model based on ultra-high frequency limit order flow data to capture and predict the short term dynamics of the order book. Our ITCH feed data allows to precisely calibrate multi-class learning models to forecast several quantities such as the mid-price movement and spread crossing. Both training and prediction of our classifiers rely on a number of features designed to characterize the order book. In our application, we address the issues of effective features selection, transformation and model calibration. Based on results for several equities traded in Nasdaq Nordic market, we discuss the effectiveness of this approach in predicting books dynamics.

CO1247: **Limit-order-book models and high-frequency-trading**
Presenter: **Lars Wendt**, University of Duisburg-Essen, Germany
Order book data from the London Stock Exchange is evaluated regarding order execution times and the effect of high-frequency trading on those. From raw order data the order books of a diversity of securities from different market segments are reconstructed. Based on the order books the time to first fill and completion of the orders are measured and compared to a previous work. It is found that order execution times have declined drastically. The idea of strategic runs has been used to identify high-frequency trading. This measure is applied to compare order execution times of non-high-frequency trading to high-frequency trading. The mean order execution time of high-frequency trading orders is lower than their normal trading counterpart. A possible reason for the decline of order execution times may be a fast changing spread. From the order book the best prices for every point in time are reconstructed and the order code of the order that sets this price is recorded. With this data at hand it is evaluated how many best-price-setting orders lead to a first fill or complete execution and if this differs for high-frequency trading and normal trading. It cannot be found that best-price-setting high-frequency trading orders are executed more often than their usual trading counterparts.

CO531 Room Board meeting room I NEW DEVELOPMENTS IN DSGE MODELLING AND ESTIMATION Chair: Timo Baas

CO0762: **Limitations of macroprudential policy and implications for monetary policy**
Presenter: **Jeyakrishna Velauthapillai**, University of Hagen, Germany
Co-authors: Timo Baas, Helmut Wagner
One of the main findings of the last financial crisis is that the stability of the financial system is a necessary condition for stabilising key macroeconomic variables. Since then there is an ongoing debate about the role of monetary policy in mitigating financial imbalances. Based on a two-sector DSGE model with collateralised debt, we investigate whether interest rate policy should be used to counter excessive credit and house price growth, which can lead to substantial macroeconomic costs. As a novel feature we assume that the reaction speeds of macroprudential policy and monetary policy to shocks differ, i.e. macroprudential policy is only able to respond with a time lag. We show that integrating additional financial indicators in the reaction function of the monetary policy is welfare improving, if the macroprudential policy faces the aforementioned constraint.

CO1074: **Migration dynamics and the enlarged European labor market**
Presenter: **Timo Baas**, University of Duisburg-Essen, Germany
Co-authors: Marjan Aikimbaeva
The fifth wave of Enlargement in 2004 increased the population in the European Union by 75 million people or roughly one fifth. Old and new members agreed on a period of seven years till universal labor market access is granted for these citizens. After the end of transitional periods, labor mobility in the European Union is higher than in any previous time. The Enlarged European Labor Market is also more disperse in terms of wages and labor quality. A two-country stochastic general equilibrium model that incorporates a production function framework with migrant and native labor, rather than full substitutability, accounts for this higher degree of heterogeneity. The aggregate shocks estimated within this framework can rationalize migration dynamics between Poland and Germany, while providing a good fit to the macroeconomic data. Migration, indeed, follows cyclical patterns and fosters economic stability. This, however, holds especially true for the country of origin, as macroeconomic shocks of the destination country have a minor impact on migration flows.

CO1020: **A search and matching approach to business-cycle migration in the Euro area**
Presenter: **Janine Hart**, University of Potsdam, Berlin Doctoral Program in Economics and Management Sciences, Germany
Co-authors: Marius Clemens
Recently migration patterns in the Euro area changed markedly and reinforced the interest in labor mobility as stabilization tool against the background of heterogeneous labor market conditions. In a data set of 55 bilateral migration corridors in the Euro area over the period 1980-2010 evidence is provided for business-cycle related fluctuations in net migration flows and the crucial role of wages, unemployment and vacancies in shaping migration patterns. The empirical facts on business-cycle migration are replicated in a two-country DSGE model with migration and search frictions. Endogenous migration is modelled via the unemployed workers choice on which labor market to search for a job. The framework allows us to account for wage and unemployment gaps between natives and immigrants over the cycle as well as for factors such as language barriers that hinder the labor market integration of foreigners. In the model set up the impact of migration on country-specific average wages and unemployment depends crucially on the characteristics of immigrants and natives as well as the institutional characteristics of the total corridor, i.e. search efficiency. The model will be used to analyze the effects of different immigration and labor market policies on migration patterns and welfare.

CO1348: **Health shock and gender labor market outcome: An estimated DSGE model of the impact of Spanish flu on Swedish economy**
Presenter: **Farzaneh Shamsfakhr**, University of Duisburg-Essen, Germany
Co-authors: Timo Baas
The aim is to develop and estimate a non-linear DSGE model to address the gender labor market properties of business cycle fluctuations succeeding an unexpected health disruption in a household. For this purpose, the impact of 1918 global influenza pandemic known as Spanish Flu is examined using Swedish data during 1915-1956. The novel characteristic of our model is the presence of two kinds of labor, male and female, and the incidence of an epidemic shock affecting both labor supply and the number of households. Our results imply that after a transitory unexpected health shock, an inflow of inactive females to the labor market, in order to substitute the absence or deficiency of infected males in the household and to pay off the subsequent domestic income loss, is expected. Likewise, female labor supply is observed to respond to male labor supply fluctuations over the whole sample period. In the view of that, we see that business cycle is not gender neutral. Furthermore, it is observed that a short-lived epidemic disease can be a potential source of business cycle fluctuations through loss in labor productivity and labor disposal as transmission mechanisms, resulting in a decline in GDP and employment. However, depending on the participation rate of females in the economy, the entry of inactive women into the labor market during the shock wave, can smooth the business cycle fluctuations.

CO407 Room 112 NON-PARAMETRIC ECONOMETRICS Chair: Isabel Casas

CO0811: **Recursive estimation in large panel data models: Theory and practice**
Presenter: **Yanrong Yang**, The Australian National University, Australia
An iterative least-squares estimation method was previously for large panel data models with unobservable interactive fixed effects. The asymptotic distribution of the iterative estimators was provided under the situation of convergence. However, the impact of iteration on the asymptotic properties of the iterative least-squares estimators was not shown. We show that under the traditional assumptions of fixed interactive effects, the recursive procedure will not necessarily yield consistent estimation if the initial estimator is inconsistent. We then analyze sufficient and necessary conditions for the convergence and divergence of the iterative estimators respectively. Simulation results illustrate various examples involving both convergent and divergent iterative estimators. Empirical applications on OECD and divorce rates in United States are employed.

CO0989: **Tv-SURE or not Tv-SURE, that is the question: Time-varying coeffcient estimation in SUR models**
Presenter: **Susan Orbe**, University of the Basque Country, Spain
Co-authors: Eva Ferreira, Isabel Casas
The aim is to study the properties of a kernel estimator for a SURE system with time-varying coeffcients (tv-SURE), under general conditions and also under cross-restrictions. The asymptotical results, consistency and asymptotic normality are obtained in this framework, where the variables are locally stationary rather than stationary. We propose the use of a Kolmogorov-Smirnov statistic to test for cross-restrictions, that allows very general alternatives in a time-varying context. The theoretical results together with the simulation study support the use of tv-SURE in practice. It is important to remark that, contrary to the parametric case,the tv-SURE estimator can be different from the single equation estimator under equal regressors. Since nonparametric estimators are biased, we also need to check that a reduction in the variance does not increase the bias. As an application we estimate the infuence of the Eurostoxx and bond spreads in several domestic indexes returns. Finally we test whether this infuence can be consider the same for some European countries.

CO1062: **Nonparametric estimation and forecasting of structural time series models**
Presenter: **Joachim Schnurbus**, University of Passau, Germany
Co-authors: Harald Haupt
A non-additive nonparametric approach is proposed for estimation and forecasting of structural time series models containing trend and seasonal components as well as additional covariates. A new kernel function is proposed that allows to take into account the specific ordered structure of seasonal effects. We provide a Monte Carlo analysis of the forecasting performance against competitors such as seasonal ARIMA and nonlinear innovation state-space models.

CO0998: **High dimensional portfolio optimization by wavelet thresholding**
Presenter: **Sebastien Van Bellegem**, Universite catholique de Louvain, Belgium
The static mean-variance portfolio optimization takes the form of a quadratic programming problem under a linear constraint and uses the mean return vector and the cross-covariance matrix of an N-dimensional stationary process as inputs. Due to the fact that the process cannot be observed directly the mean return vector and the covariance matrix need to be replaced by estimates. In high-dimensional settings, e.g. when the number of assets is large relative to the sample size, the empirical covariance matrix is badly conditioned. Inversion of the covariance matrix is therefore unstable and portfolio optimization behaves poorly. We show that wavelet can achieve some decorrelation of the stationary process. We exploit this property and introduce a new thresholding rule of the empirical covariance matrix in the wavelet domain, based on a generalization of Tree Structured Wavelet (TSW) denoising. The consistency of the denoising procedure is established and we derive an optimal thresholding rule. Simulation studies show the good performance of the final optimizer compared to benchmarks and optimizers based on other regularization methods (such as e.g. Tikhonov).

CO573 Room Board meeting room II STATISTICAL METHODS FOR FINANCIAL AND ENERGY MARKETS Chair: Gernot Mueller

CO0844: **Identification of core-periphery networks in the interbank market**
Presenter: **Christian Pigorsch**, Friedrich Schiller University Jena, Germany
The recent financial crisis revealed that the interbank market structure plays an important role for the formation of systemic risk. It is thus not surprising that research on the topology of the interbank network has developed tremendously over the past few years. Most of these empirical and theoretical studies point towards a core-periphery network structure of the interbank market. In order to assess the systemic risk as well as the financial stability of the interbank network, it is therefore of interest to identify banks that belong to the core and those that belong to the periphery. The identification of the underlying network structure, however, is not trivial, as the observed interbank linkages usually do not directly coincide with a core-periphery network. Fitting a core-periphery model to an empirical network is therefore an important task. The objective is to provide a novel approach for the identification of the German interbank network from the observed network by incorporating information about the strength of the network linkages, e.g. information on the bilateral liabilities (exposures) of banks. In particular, we develop a model that attaches to each linkage a probability that signals the importance of the link.

CO0974: **Central bank interest rate policy: Bayesian analysis using a cross nested AOP model**
Presenter: **Armin Seibert**, Augsburg University, Germany
Co-authors: Gernot Mueller, Andrei Sirchenko
The decisions of central banks (e.g. ECB) to lower or raise the key interest rate have a high impact on macroeconomic conditions like other interest rates, asset prices and employment. We investigate a model for these interest rates consisting of three (cross nested) autoregressive ordered probit (CNAOP) models considering different economic covariates. The hierarchy of those probit models is intended to mimic different levels of the decision. It can handle the high number of zeros (i.e. most times the interest rate is unchanged) very well. The presence of latent variables prevents it from a maximum likelihood analysis. Therefore we develop a Bayesian estimation algorithm and test it in a simulation study. Finally, we apply the model on real world data.

CO0534: **Oscillating Ornstein-Uhlenbeck processes in electricity markets: Modelling and statistical inference**
Presenter: **Jeannette Woerner**, TU Dortmund, Germany
Co-authors: Daniel Kobe
Recently there have been proposed many models for electricity spot prices trying to capture the characteristic features of seasonalities and spikes in the prices. A common approach is to remove the seasonalities first and then divide the remaining random part in a stochastic process accounting for the normal variations of the prices and one modelling the spikes, e.g. with a combination of Ornstein-Uhlenbeck processes with different speed of mean reversion. We now propose to generalize the underlying stochastic process in such a way that we can include the seasonalities into our stochastic process and also reproduce the oscillating behaviour of the empirical autocorrelation function of the prices. We consider oscillating Ornstein-Uhlenbeck processes which belong to the class of continuous-time moving average processes. We show that a linear combination of the oscillating Ornstein-Uhlenbeck processes together with an Ornstein-Uhlenbeck process well fits the autocorrelation function of electricity spot

prices and reproduces the spikes. Furthermore, we derive an explicit formula for the forward price, which is a generalization of a previous formula. In a similar way this formula may be used for option pricing. Finally, we show, how we may infer the model parameters using empirical moments.

CO0975: **Modelling electricity prices using processes with time-varying parameters**
Presenter: **Gernot Mueller**, Augsburg University, Germany
Co-authors: Boris Buchmann, Ross Maller

Due to changing rules and regulations, changing market conditions, and a changing electricity production towards a higher proportion of renewable energies, electricity prices show a changing behaviour over time. We try to modify an existing statistical model for electricity prices by employing processes which show locally a behaviour similar to alpha-stable processes, but allow for time-varying parameters. The processes under consideration have no stationary increments, so that we look at additive (i.e. independent increment) processes instead of Levy processes. The data which motivates the analysis is taken from the data base of the European Energy Exchange EEX.

CO355 Room 102 VOLATILITY MODELLING Chair: Giuseppe Storti

CC1637: **Nonlinear conditional Value-at-Risk Granger causality**
Presenter: **Cees Diks**, University of Amsterdam, Netherlands
Co-authors: Marcin Wolski

A new methodology is proposed to assess the effects of individual companies' risk on other institutions and on the system as a whole. We build upon the Conditional Value-at-Risk (CoVaR) approach, introducing explicit Granger causal linkages while accounting for possible nonlinearities in the financial time series considered. The resulting causality measure is referred to as Nonlinear CoVaR (NCoVaR) Granger causality. The natural U-statistics estimator of NCoVaR Granger causality is shown to be asymptotically normally distributed with a long-term variance that can be consistently estimated using a HAC estimator, allowing us to construct a test for the absence of NCoVaR Granger causality. We investigate our testing approach numerically, and find it to have good size and power properties. In an empirical application we assess the feedback risk transmissions between sovereigns and banking sectors in the Euro area.

CO1269: **On the random walk assumption in high frequency stock market prices**
Presenter: **Sebastien Laurent**, AMU, France
Co-authors: Shuping Shi

The random walk hypothesis plays a central role in finance and in financial econometrics. The aim is to provide a new econometric procedure for testing this hypothesis against either the alternative of a mean reverting process or an explosive process. The new approach utilises intra-day information and tackles the specific features of high frequency financial data in an efficient and innovative way.

CO1356: **Forecasting oil price volatilities with multivariate fractionally integrated asymmetric DCC models**
Presenter: **Malvina Marchese**, University of Genoa, Italy
Co-authors: Lorenzo Trapani

The evidences of asymmetries and long-range dependence in the volatilities of oil prices is reassessed using a multivariate fractionally integrated exponential DCC model for three markets- Brent, Dubai, and West Texas Intermediate. We estimate several MGARCH models, compare their in-sample performance and their predictive ability with three approaches: the Superior Predictive Abilty test, the Model Confidence Set (MCS) method and the Value-at-Risk approach. In doing so, we extend the MCS method to include cases where the forecast error loss differential is strongly autocorrelated as arising in time series with long memory. In the overall, our results indicate significant gains from using models that include long-range dependence and asymmetries against short-memory models.

CO1371: **Flexible realized GARCH models**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Richard Gerlach, Antonio Naimoli

A new class of flexible Realized GARCH models is introduced. Our model generalizes the original specification along three different directions. First, it features a time varying volatility persistence. Namely, the shock response coefficient in the volatility equation adjusts to the time varying accuracy of the associated realized measure. Second, our framework allows us to consider, in a parsimonious way, the inclusion of multiple realized measures. Finally, it allows for heteroskedasticity of the noise component in the measurement equation. The appropriateness of the proposed class of models is appraised by means of an application to a set of stock returns data.

CG706 Room 109 CONTRIBUTIONS IN ALGORITHMS AND SOFTWARE FOR FINANCIAL ECONOMETRICS Chair: Georgi Boshnakov

CC1621: **Various versatile variances: An object-oriented implementation of clustered covariances in R**
Presenter: **Susanne Berger**, University of Innsbruck, Austria
Co-authors: Nathaniel Graham, Achim Zeileis

Clustered covariances or clustered standard errors are very widely used to account for correlated or clustered data, especially in economics, political sciences, or other social sciences. They are employed to adjust the inference following estimation of a standard least-squares regression or generalized linear model estimated by maximum likelihood. Although many publications just refer to "the" clustered standard errors, there is a surprisingly wide variation in clustered covariances, particularly due to different flavors of bias corrections. Furthermore, while the linear regression model is certainly the most important application case, the same strategies can be employed in more general models (e.g. for zero-inflated, censored, or limited responses). We discuss an object-oriented implementation based on the building blocks provided by R package "sandwich" and assess its performance in a simulation study.

CC1512: **Estimating the variance-gamma distribution: A comparison of algorithms and of estimated option prices**
Presenter: **Marco Bee**, University of Trento, Italy
Co-authors: Maria Michela Dickson, Flavio Santi

The variance-gamma (VG) process has been proposed in the theory of option pricing as an alternative to the geometric Brownian motion for modeling the underlying of financial derivatives. The VG distribution is the marginal distribution of this process and has the capability of capturing skewness and kurtosis of the underlying asset's distribution. We exploit the fact that the VG distribution is a special case of the generalized hyperbolic distribution to develop maximum likelihood estimation (MLE) via the EM algorithm. Two main conclusions are reached. First, according to extensive simulation experiments, the EM-based estimators are preferable to the MLEs obtained via standard numerical optimization routines implemented in the `VarianceGamma` and `ghyp R` packages. Second, given the availability of closed-form formulas for computing the prices of the European-style options with VG-distributed underlying, the simulated distributions of the estimators are used to study the precision of estimated option prices. Besides shedding some light on the controversy about the estimated values of the parameters of the VG distribution, these outcomes permit to assess the impactof the errors associated to the different estimation procedures on the prices of options.

CC0775: **Real-time analysis of the intraday financial volatility: Big data, simulations and stochastic volatility using R**
Presenter: **Antonio Santos**, University of Coimbra, Portugal

The financial volatility is a key element for economic agents that make decisions in financial markets. To define the measures of volatility through the financial models, data need to be collected, models need to be estimated, and the relevant results need to be presented in an integrated way. Using the capabilities of R these tasks can be performed in an integrated form, allowing a more efficient use of the data, models and measures to characterize the volatility evolution in the financial markets. A package in R that integrates the three tasks of collecting and treating big financial datasets, estimation of the models and definition relevant measures of volatility, and also the presentation the results in an intuitive and iterative form is certainly useful. The capabilities of R to retrieve public available data from different sites and to organize the data conveniently are used to deal with big data sets. Estimation of the parameters within the stochastic volatility model, and the forecasting of the volatility are usually made through the Bayesian statistics and Markov chain Monte Carlo methods. A mix of code in R and C is used to accomplish these tasks. The presentation of the measures of volatility forecasts can take advantage of the resources available in R. This will be done by an R Shiny web application. A package in R was developed to perform the three aforementioned tasks, and some of the main functions will be described.

CC0474: **Nowcasting economic turning points with a simple machine-learning algorithm**
Presenter: **Thomas Raffinot**, Paris Dauphine, France

To nowcast economic turning points, probabilistic indicators are created from a simple and transparent machine-learning algorithm known as Learning Vector Quantization (LVQ). The real-time ability of the indicators to quickly and accurately detect economic turning points in the United States and in the euro area is gauged. To assess the value of the indicators, profit maximization measures based on trading strategies are employed in addition to more standard criteria. When comparing predictive accuracy and profit measures, the model confidence set procedure is applied to avoid data snooping. A substantial improvement in profit measures over the benchmark is found: macroeconomists can get rich nowcasting economic turning points.

EO257 Room 204 INCOMPLETE DATA IN EXTREME VALUE THEORY Chair: Laurent Gardes

EO0175: **Tail product-limit process for truncated data with application to extreme value index estimation**
Presenter: **Abdelhakim Necir**, Mohamed Khider University of Biskra, Algeria

A weighted Gaussian approximation to tail product-limit processes for Pareto-like distributions of randomly right-truncated data is provided and a new consistent and asymptotically normal estimator of the extreme value index is introduced. Moreover, we propose an estimator of the second-order parameter and an asymptotically unbiased estimator for the tail index and establish their asymptotic normality. A simulation study is carried out to evaluate the finite sample behavior of the proposed estimators and compare it to one recently proposed. Also, a new approach of estimating extreme quantiles, under random right truncation, is derived and applied to a real dataset of lifetimes of automobile brake pads.

EO0492: **Assessing the validity of certain hypotheses in randomly right-censored extremes**
Presenter: **Gilles Stupfler**, The University of Nottingham, United Kingdom
Co-authors: Claudia Neves

Studying the extremes of a randomly right-censored variable is a recent developing field. To the best of our knowledge, the existing studies on this topic stem from a couple of hypotheses of which a consequence is that the problem studied is nontrivial and has a solution, that is, the extremes of the censored variable may be recovered. Our goal is to give some elements so as to make it possible to check the validity of such hypotheses in practice. Our techniques are showcased on a simulation study and their practical applicability is illustrated on a real data set.

EO0518: **Global fits using splicing for censored data: mixed Erlang and extreme value distributions**
Presenter: **Jan Beirlant**, KULeuven, Belgium
Co-authors: Tom Reynkens, Roel Verbelen, Katrien Antonio

In risk analysis, a global fit that appropriately captures the body and the tail of the distribution is essential. Modelling the whole range of the loss distribution using a standard distribution does not seem to be suitable. A possible solution is to combine two distributions in a splicing model: a light-tailed distribution for the body, i.e. concerning light and moderate losses, and a heavy-tailed distribution for the tail, i.e. concerning large losses. We propose a splicing model with a mixed Erlang (ME) distribution for the body and a Pareto distribution for the tail. This combines the flexibility of the ME distribution with the ability of a Pareto distribution to model extreme values. We extend our splicing approach for censored and/or truncated data. Relevant examples of such data can be found in insurance loss data. We illustrate the flexibility of this splicing model using practical examples coming from actuarial science.

EO0523: **Extreme value index estimation for randomly censored data with competing risks**
Presenter: **Julien Worms**, University of Versailles-Saint-Quentin / University Paris-Saclay, France
Co-authors: Rym Worms

The study of the tails of (randomly and independently) right-censored data has attracted some attention in the recent years. Concerning the estimation of the extreme value index (e.v.i.) of such data, a strategy consists in using a trick which relates the e.v.i. of the censored sample to the e.v.i. of the complete sample and the probability of not being censored in the tail. We recently proposed an alternative method which, in the context of heavy tail data, led to an adaptation of the famous Hill estimator taking the form of a Kaplan-Meier functional : it weights the data in the tail in a natural way in this survival analysis framework. We extend this method to the case where competing risks are present: in order to estimate useful extreme quantiles, the target is now the e.v.i. of one of the cause-specific survival functions (also called cumulative incidence functions, assumed here to have heavy tails). The proposed estimator is now a functional of the Aalen-Johansen estimator. Note that in this context, it does not seem possible to rely on the trick previously referred to (the first strategy, when there is only one cause). Finite sample behavior and asymptotic properties of the estimator (consistency and asymptotic normality) will be discussed.

EO593 Room 213 SURVIVAL ANALYSIS Chair: Thomas Scheike

EO0194: **Instrumental variables estimation with competing risk data**
Presenter: **Torben Martinussen**, University of Copenhagen, Denmark

Semiparametric structural cumulative failure time model and instrumental variables (IV) has been previously used to estimate causal exposure effects for survival data. No restrictions on the type of the instrument nor on the exposure variable are imposed. Furthermore the method allows for nonparametric estimation of possible time changing exposure effect. We extend that method to handle competing risk data. Such data are very common in practice when studying the timing of initiation of a specific disease since death will often be a competing event. Also when studying death due to a specific cause, such as death from breast cancer as was of interest in the HIP-study, death from any other cause is a competing event. The HIP-study comprises approximately 60000 women and in the first 10 years of follow-up there are 4221 deaths, but only 340 were deemed due to breast cancer. Hence, competing risks is a major issue in these data. Due to non-compliance it is not straightforward to estimate the screening effect. Randomization can, however, be used as an IV and, hence, for these data it is pertinent to have IV-methods for competing risk data to learn about the causal effect of breast cancer screening on the risk of dying from breast cancer.

EO1052: **A general class of nonparametric test statistics for interval-censored data**
Presenter: **Ramon Oller**, Fundacio Universitaria Balmes, Spain

An extension of the Jones and Crowley's class of nonparametric test statistics is proposed for interval-censored data. The test statistics are formulated in terms of the estimated distribution function $\hat{F}_i(t)$ of each individual in the sample. They measure the relationship between survival times and a quantitative vector label $\mathbf{z}(t)$ attached to each individual. Under the null hypothesis of no relationship and the assumption that the underlying censoring process is identical for all individuals, the permutation distribution of the test statistics is derived. In this general class, we obtain and investigate several particular statistics. Some of them are well known, for instance a set of weighted log-rank test statistics. Others are new and give new insights for analyzing interval-censored data, for instance a set of Kendall-type test statistics.

EO1284: **Penalised competing risks regression**
Presenter: **Federico Ambrogi**, University of Milan, Italy
Co-authors: Thomas Scheike

High dimensional data analysis is an important topic in many research fields. For example, biomedical research generates increasing amount of data to characterise patients bio-profiles (e.g. from genomic high-throughput assay, imaging, physiological measurements, laboratory tests, patient monitoring, etc.). Variable selection is a long-established problem in statistical research and is every day more and more important. In the last decades many forms of penalised regression have been developed, as a modern form of variable selection, to cope with high and ultra-high dimensional settings. Although the number of contributions in standard survival analysis have increased, the research regarding competing risks is less developed. The aim is to consider how to do penalized regression when considering the crude cumulative incidence. The direct binomial regression model is reformulated in a penalised framework to possibly fit a sparse regression model. The proposed approach is easily implementable using existing high performance software to do either ridge, or lasso or elastic net penalization. Results from simulation studies are presented together with an application to genomic data when the endpoint is progression free survival.

EO1170: **Permutation tests for general dependent truncation**
Presenter: **Rebecca Betensky**, Harvard School of Public Health, United States
Co-authors: Sy Han Chiou, Jing Qian

Quasi-independence is a common assumption for analyzing truncated survival data that are frequently encountered in biomedical science, astronomy, and social science. While the concept of censoring has been rigorously studied, many are not aware of the analytic issues that arise with delayed entry, or general truncation. Ignoring dependent truncation can lead to severely biased estimation and inference. Current methods for testing quasi-independent truncation are powerful for monotone alternatives, but not otherwise. We extend methods for detecting highly non-monotone and even non-functional dependencies and develop nonparametric tests for dependent truncation that are powerful against non-monotone alternatives. We compare computation time, size and power of both conditional and unconditional permutation procedures. We apply our results to a study on cognitive and functional decline that had delayed entry due to post-baseline imaging.

EO712 Room 007 BAYESIAN SEMI-AND NONPARAMETRIC MODELLING II Chair: Matteo Ruggiero

EO0206: **Bayesian nonparametric approaches to quantifying dependence between random variables**
Presenter: **Sarah Filippi**, University of Oxford, United Kingdom
Co-authors: Chris Holmes, Luis E Nieto-Barajas

Nonparametric and nonlinear measures of statistical dependence between pairs of random variables have proved themselves important tools in modern data analysis, where the emergence of large data sets can support the relaxation of linearity assumptions implicit in traditional association scores such as correlation. We will present two Bayesian nonparametric procedures to test for dependences. In the first one a tractable, explicit and analytic quantification of the relative evidence of dependence vrs independence, using Polya tree priors on the space of probability measures which can then be embedded within a decision theoretic test for dependence. In the second procedure the unknown sampling distribution is specified via Dirichlet Process Mixtures (DPM) of Gaussians, which provide great flexibility while also encompassing smoothness assumptions. These procedures can accommodate known uncertainty in the form of the underlying sampling distribution and provides an explicit posterior probability measure of both dependence and independence. Well-known advantages of having an explicit probability measure include the easy comparison of evidence across different studies, the inclusion of prior information, and the integration of results within formal decision analysis.

EO1123: **Bayesian nonparametric regression on manifolds**
Presenter: **Vinayak Rao**, Purdue University, United States

The problem of constructing flexible nonparametric surfaces indexed by manifold-valued covariates is addressed. Such objects are useful in regression and classification, when covariates lie on a manifold, and in density modeling, when i.i.d. observations lie on a manifold. Such problems arise in applications like brain imaging data, diffusion-tensor imaging, shape modeling, etc. The standard nonparametric Bayesian approach involving Gaussian processes requires constructing positive-definite covariance kernels, something that is not straightforward for general manifolds. Instead, we propose defining a Gaussian process on a tangent plane to a manifold, which only requires standard machinery from Euclidean geometry. To evaluate the function at any point on the manifold, we project that point up to the tangent place, and evaluate the GP there. More flexible surfaces can be obtained by mixing over different GPs defined on different tangent planes. We demonstrate our methodology on different manifolds, and on synthetic and real datasets.

EO0199: **Bayesian nonparametric conditional copula estimation of twin data**
Presenter: **Fabrizio Leisen**, University of Kent, United Kingdom
Co-authors: Luca Rossini, Luciana Dalla Valle

Several studies on heritability in twins aim at understanding the different contribution of environmental and genetic factors to specific traits. Considering the National Merit Twin Study, our purpose is to correctly analyse the influence of the socioeconomic status on the relationship between twins' cognitive abilities. Our methodology is based on conditional copulas, which allow us to model the effect of a covariate driving the strength of dependence between the main variables. We propose a flexible Bayesian nonparametric approach for the estimation of conditional copulas, which can model any conditional copula density. Our methodology extends previous work by introducing dependence from a covariate in an infinite mixture model. Our results suggest that environmental factors are more influential in families with lower socio-economic position.

EO0910: **A Bayesian nonparametric approach to ecological risk assessment**
Presenter: **Guillaume Kon Kam King**, University of Torino, Italy
Co-authors: Julyan Arbel, Igor Pruenster

A classical method for ecological risk assessment, the Species Sensitivity Distribution (SSD) approach, is revisited in a Bayesian nonparametric framework. SSD is a mandatory diagnostic required by environmental regulatory bodies from the European Union, the United States, Australia, China etc. Yet, it is subject to much scientific criticism, notably concerning a historically debated parametric assumption for modelling species variability. Tackling the problem using nonparametric mixture models, it is possible to shed this parametric assumption and build a statistically sounder basis for SSD. The Bayesian nonparametric approach offers another advantage: the ability to deal with small datasets, typical in the field

of ecological risk assessment. We use Normalised Random Measures with Independent Increments (NRMI) as the mixing measure because they offer a greater flexibility than the Dirichlet process. Indeed, NRMI can induce a prior on the number of components in the mixture model that is less informative than the Dirichlet process. This feature is consistent with the fact that SSD practitioners do not usually have a strong prior belief on the number of components. We illustrate the advantage of the nonparametric SSD over the classical normal SSD and a kernel density estimate SSD on several real datasets. We then perform a systematic comparison on simulated data, and finish by studying the clustering induced by the mixture model to examine patterns in species sensitivity.

EO263 Room 217 STATISTICAL SIZE DISTRIBUTIONS Chair: Yves Dominicy

EO0230: **An interpolating family of size distributions**
Presenter: **Corinne Sinner**, ULB, Belgium
Co-authors: Yves Dominicy, Christophe Ley, Julien Trufin, Patrick Weber
Size distributions are important tools in describing phenomena like the behavior of claim sizes in actuarial sciences or wind speed in environmetrics. Despite a large number of available size distributions in the literature, the counterplay between tractability and flexibility makes the practical use of most of these distributions quite challenging. We introduce a new five-parameter family of distributions on the semi-infinite interval $[x_0, +\infty)$. Its genesis is based on power laws and power laws with exponential cutoff. To be concise, this new family interpolates between these power laws and power laws with exponential cutoff. It includes some very well known special cases, as for example the Pareto distributions and extreme value distributions. As competitor to the generalized Beta distribution, we show that in terms of tractability it has some advantages over the latter. Indeed, its normalizing constant does not involve any special functions such as the Beta function, and consequently its cumulative distribution function can be written under closed form. Moreover, its moments are expressed as combinations of Beta and/or Gamma functions. To conclude, we will apply it to three real data sets.

EO0299: **A simple cylindrical model as a combination of the Weibull and sine-skewed von Mises distributions**
Presenter: **Toshihiro Abe**, Nanzan University, Japan
A cylindrical model is considered obtained by combining the sine-skewed von Mises distribution with the Weibull distribution. The WeiSSVM model has numerous good properties such as simple normalizing constant, and hence very tractable density, parameter-parsimony and interpretability, maximum entropy characterization, good circular-linear dependence structure, easy random number generation thanks to known marginal/conditional distributions, and flexibility illustrated via excellent fitting abilities. As an illustrative example, our model is applied in analyses of periwinkle data set.

EO0657: **(non-)Informative priors for the shape parameter of log-two-piece and log-skew-symmetric distributions**
Presenter: **Francisco Javier Rubio**, London School of Hygiene & Tropical Medicine, United Kingdom
An innovative method will be presented for constructing proper priors for the shape parameter in the log-skew-symmetric and log-two-piece family of distributions. These families of distributions have been used in several contexts such as particle size modelling, biometry, finance, medicine, biology, among others. The proposed method is based on assigning a prior distribution on the perturbation effect of the shape parameter, which is quantified in terms of the Total Variation distance, and leads to closed-form expressions in many cases. We will discuss strategies to construct informative and non-informative priors. Illustrative examples are presented using real data.

EO0825: **Efficient and simple testing procedures for the (extended) generalized inverse Gaussian models**
Presenter: **Efoevi Angelo Koudou**, Universite de Lorraine, France
Co-authors: Christophe Ley
The standard efficient testing procedures in the Generalized Inverse Gaussian (GIG) family are likelihood ratio tests, hence rely on Maximum Likelihood (ML) estimation of the three parameters of the GIG. The particular form of GIG densities, involving modified Bessel functions, prevents in general from a closed-form expression for ML estimators, which are obtained at the expense of complex numerical approximation methods. On the contrary, Method of Moments (MM) estimators allow for concise expressions, but tests based on these estimators suffer from a lack of efficiency compared to likelihood ratio tests. This is why, in recent years, trade-offs between ML and MM estimators have been proposed, resulting in simpler yet not completely efficient estimators and tests. The presented method does not propose such a trade-off but rather an optimal combination of both methods. The proposed tests inherit efficiency from an ML-like construction and simplicity from the MM estimators of the nuisance parameters. They rely on the Le Cam methodology. Besides providing simple efficient testing methods, the theoretical background of this methodology further allows to write out explicitly power expressions for the proposed tests. A Monte Carlo simulation study shows that, also at small sample sizes, these simpler procedures do at least as good as the complex likelihood ratio tests. Steps towards an application of the same method to the Extended GIG models will be discussed.

EO051 Room 203 ADVANCES AND APPLICATIONS OF FUNCTIONAL DATA ANALYSIS Chair: Ying Zhu

EC0244: **Sensible functional linear discriminant analysis**
Presenter: **Lu-Hung Chen**, National Chung-Hsing University, Taiwan
Co-authors: Ci-Ren Jiang
The focus is to extend Fisher's linear discriminant analysis (LDA) to both densely recorded functional data and sparsely observed longitudinal data for general c-category classification problems. We propose an efficient approach to identify the optimal LDA projections in addition to managing the noninvertibility issue of the covariance operator emerging from this extension. A conditional expectation technique is employed to tackle the challenge of projecting sparse data to the LDA directions. We study the asymptotic properties of the proposed estimators and show that asymptotically perfect classification can be achieved in certain circumstances. The performance of this new approach is further demonstrated with numerical examples.

EO1039: **Functional data analysis with its application to discrimination of Chinese herbal medicine**
Presenter: **Ying Zhu**, Nanyang Technological University, Singapore
Fourier transform infrared (FTIR) spectra of herbal medicine consist of many overlapping absorption bands sensitive to the physical and chemical states of compounds and thus show highly correlated structure due to the complex system of herbal medicine. It could be more informative to describe the FTIR spectrum as a function rather than as a sequence of individual variables. A model based on functional data analysis (FDA) was developed by taking advantage of the characteristic of FTIR spectroscopic data, being a sum of absorption peaks caused by the different chemical constituents present in the sample under study. Different basis functions were explored while taking into account the spatial correlation between the variables. This FDA model has been applied to distinguish between different species of Ganoderma lucidum. The model performance has been compared with conventional multivariate statistical techniques in terms of discrimination and interpretation. The informative spectral absorption bands for discrimination have provided better interpretations of the major chemical compounds of Ganoderma lucidum in relation to its medicinal efficacy.

EO1714: **On the extension of Lo's modified R/S statistics against to locally stationary short-range dependence**
Presenter: **Junichi Hirukawa**, Niigata University, Japan

A test for long-run memory has been developed that is robust to short-range dependence. It was an extension of the range over standard deviation or R/S statistic, for which the relevant asymptotic sampling theory was derived via functional central limit theory. We consider an extension of thes modified R/S statistics which is aimed to be robust against locally stationary short-range dependence.

EC1513: **Functional data analysis by matrix completion**
Presenter: **Marie-Helene Descary**, EPFL, Switzerland
Co-authors: Victor Panaretos

A new approach is put forward for analysing functional data, where one replaces the traditional smoothing and PCA steps by a variant of PCA based on matrix completion. More specifically, our method is a band-deleted PCA, where we seeks low rank matrices that best complete a partial version of the raw empirical covariance where a non-trivial band around the diagonal is removed. Since the band near the diagonal carries almost entirely local variations and noise, the resulting low rank covariances encapsulate the smooth global features that are the main focus of FDA. Moreover, separate analysis of the remaining banded covariance residual allows us to recover localised features of functional variation that would have been annihilated or confounded with smooth variation by a traditional smoothing plus PCA analysis. We illustrate the advantages of our method by simulation and analysis of real data, and provide theory to show the validity of the method, including consistency and rates of convergence.

EO171 Room 212 ORDINAL PATTERN DEPENDENCE IN TIME SERIES ANALYSIS Chair: Alexander Schnurr

EO0253: **Detecting climatic coherences in hydrological time series using ordinal patterns**
Presenter: **Svenja Fischer**, Ruhr-University Bochum, Germany
Co-authors: Andreas Schumann, Alexander Schnurr

In hydrology it is not only important to consider single flood series at specific gauges but one also has to take into account the processes influencing the catchment area, because the simultaneous appearance of certain phenomena can even lead to a higher risk than a single extreme flood. We therefore want to detect certain clusters of gauges that show a coherence in their behaviour. An intuitive measure for dependence between two time series are ordinal patterns. This rank-based method is not only robust but independent of the distribution and scale, which is very important when considering hydrological time series. The significance of coherence between two time series can then be evaluated by comparison with the relative frequency. Nevertheless, hydrological time series underlie particular uncertainties, especially in the left tail. These uncertainties are taken into account by a pre-processing of the data as well as a modification of the ordinal patterns to a weighting approach. This ensures that coherences are not influenced by erroneous data. In a German case study we want to show the detected clusters and interrelate these with certain climatic events.

EO0288: **Ordinal pattern dependence: Background, origins and areas of application**
Presenter: **Alexander Schnurr**, University Siegen, Germany
Co-authors: Herold Dehling

Ordinal pattern dependence is a new way of measuring the degree of dependence between time series. Since it only relies on the ordinal structure of the data it is robust against monotone transformations and measurement errors. We emphasize the applicability of the method by analyzing the interplay between the S& P 500 and its volatility index (VIX). In the short range dependent setting we present limit theorems and a test for structural breaks.

EO0348: **Entropies based on ordinal pattern distributions**
Presenter: **Karsten Keller**, Luebeck, Germany

The quantification of complexity of a time-developing system on the base of time series obtained from it is a challenging problem. One reason for this is that many well-motivated theoretical complexity concepts are 'infinite', e.g. are given by a 'limit procedure', but real data are finite. The purpose is to discuss the relatively new concept of permutation entropy in a wider context. We consider the concept of ordinal patterns which is central for the definition of permutation entropy and variants of this entropy and show that the distribution of ordinal patterns contains substantial information on a given system and on dependencies between system components. This is our theoretical background for describing time series analysis methods aimed to quantify complexity, similarity and dependence.

EO0739: **Ordinal patterns under long-range dependence**
Presenter: **Jannis Buchsteiner**, Ruhr-University Bochum, Germany
Co-authors: Herold Dehling, Alexander Schnurr, Jeannette Woerner

While in recent years many authors studied ordinal patterns of weakly dependent time series, we consider long-range dependent data. More precisely, we investigate the ordinal information of a subordinated Gaussian process with a non-summable autocovariance function. We establish the asymptotic behavior of different estimators for ordinal pattern probabilities by using a multivariate Hermite decomposition.

EO183 Room 211 DATA-DRIVEN ESTIMATION PROCEDURES IN THE PRESENCE OF DEPENDENCE Chair: Jan Johannes

EO0303: **Adaptive functional linear regression in the presence of dependence**
Presenter: **Nicolas Asin**, Universite Catholique de Louvain, Belgium
Co-authors: Jan Johannes

In presence of dependence in the data generating process we consider the estimation of a slope function in functional linear regression which models a relationship between a scalar response and a random function. Assuming an independent and identically distributed (iid) sample it has been shown in the literature that a least squares estimator based on dimension reduction and thresholding can attain minimax-optimal rates of convergence up to a constant. Assuming sufficiently weak dependent observations characterized by fast decreasing mixing coefficients it is shown that the estimator attains minimax optimal rate considering smoothness assumptions. In particular the derived risk bound coincides up to a constant with those obtained in the independent case. However as for the iid case, this estimation procedure requires an optimal choice of a dimension parameter with regard amongst others to certain characteristics of the unknown slope function. Borrowing ideas from the iid case we investigate its fully data-driven choice which combines model selection and Lepskis method. We show that this data-driven estimator can attain the iid lower risk bound up to a constant provided a fast decay of the mixing coefficients.

EO0312: **Nonparametric estimation for Poisson inverse problems in a circular model**
Presenter: **Martin Kroll**, Universitaet Mannheim, Germany

Nonparametric estimation of the intensity function of a Poisson point process in a circular model is considered when only observations contaminated with error are available. In general, the observed processes are not Poisson again, but contain inherent dependencies. Firstly, we approach this estimation problem from a minimax point of view. We propose an orthogonal series estimator whose performance depends on the appropriate choice of a tuning parameter. Choosing this parameter optimally, we obtain an upper bound for the maximum risk of the estimator over classes of potential intensities and error densities. Afterwards, we discuss conditions under which this upper bound is also a lower bound. Since our

optimal estimator in the minimax framework depends crucially on some prior knowledge concerning the intensity and the error density, it is not completely data-driven. This drawback is tackled and an adaptive estimation procedure is proposed. Following a model selection paradigm, the tuning parameter of the series estimator is chosen as the minimizer of a penalized contrast criterion. The resulting adaptive estimator attains optimal rates in a variety of scenarios.

EO1288: **Nonparametric estimation of the interarrival distribution of a renewal process**
Presenter: **Celine Duval**, Universite Paris Descartes, France
Co-authors: Fabienne Comte

Nonparametric estimation of interarrival times density of a renewal process is considered. First, we assume continuous observation of the process and build a projection estimator in the Laguerre basis. We study its mean integrated squared error and compute rates of convergence on Sobolev-Laguerre spaces when the length of the observation interval gets large. Second, we consider a discrete time observation. A first strategy consists in neglecting the discretization error, under suitable conditions, an analogous MISE is obtained. Then, taking into account the structure of the data, a deconvolution estimator is defined and studied. In that case, we work under a simplifying "dead-zone" condition. In the three cases, an automatic model selection procedure is described and gives the best MISE, up to a logarithmic term. The results are illustrated through a simulation study.

EO1612: **Adaptive aggregation in circular deconvolution in the presence of dependence**
Presenter: **Jan Johannes**, Ruprecht-Karls-Universitat Heidelberg, Germany
Co-authors: Xavier Loizeau, Fabienne Comte

A circular deconvolution problem is considered, where the density of a circular random variable X has to be estimated nonparametrically from a noisy observation Y of X. The additive measurement error is supposed to be independent of X and its density is known. The objective is the construction of a fully data-driven estimation procedure based on a dependent sample from Y. Assuming an independent and identically distributed (iid) sample it has been previously shown that an orthogonal series estimator with adaptive choice of the dimension parameter using a model selection approach can attain minimax-optimal rates of convergence up to a constant. We propose a fully data-driven shrinkage estimator which is inspired by a Bayes estimator in an indirect sequence space model with hierarchical prior and can be interpreted as a fully data-driven aggregation of the orthogonal series estimators. Considering first an iid sample we show that the fully data-driven shrinkage estimator can attain minimax-optimal rates of convergence over a wide range of density classes covering in particular ordinary and super smooth densities. Dismissing then the independence condition and assuming sufficiently weak dependence characterised by a fast decay of the mixing coefficients we derive an upper risk bound for the shrinkage estimator which coincides up to constant with the minimax-optimal risk bound in the iid case.

EO437 Room 206 ADVANCES IN ROBUST DATA ANALYSIS Chair: Agustin Mayo-Iscar

EO0316: **Robust morphometrics with geographical information systems**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain
Co-authors: Yolanda Cabrero-Ortega

An application of Morphometrics is the classification of objects based on their Shapes. To do this, we have first to remove the Size effect. Modern morphometric methods propose to fix some reference points called Landmarks that must be common to all objects. For instance, if the objects to be classified are arrows, they become polygons from a digital image with the help of a Geographical Information System and the landmarks should be the vertices. The collection of all the landmarks coordinates constitutes a Configuration that has a centroid. The Size of the configuration is defined as the Euclidean distance from the landmarks to the centroid of the configuration. The standardize data matrix is the matrix of the landmarks coordinates divided by the size. We cannot use now traditional multivariate techniques to classify the objects because the landmarks are not independent. First we have to transform the rows of this matrix in shape variables using one of the superimposition method; namely, the Procrustes method, projecting the rows of the standardized data matrix on the tangent shape space. The Kendall's method estimates the superimposition parameters using least squares. Here we shall consider robust methods instead of least squares and a Geographical Information System, QGIS, to classify objects (polygons) incorporated into QGIS as raster data. We conclude using these techniques in the classification of real Solutrean Projectile Points (lithics hunting arrows).

EO1227: **Highly robust and efficient estimates for the generalized linear model with a dispersion parameter**
Presenter: **Alfio Marazzi**, University of Lausanne, Switzerland
Co-authors: Michael Amiguet, Marina Valdora, Victor Yohai

Let $Y|x \sim F_{\mu(x),\alpha}$, where $F_{\mu,\alpha}$ is a continuous or a discrete distribution with density $f_{\mu,\alpha}$, $\mu(x) = E(Y|x) = h(\beta^{\mathrm{T}}x)$, α is nuisance parameter, x a covariate vector, β a coefficient vector, h a link function. Assume that for any α, if $\mu_2 > \mu_1$, $Y_1 \sim F_{\mu_1,\alpha}$, $Y_2 \sim F_{\mu_2,\alpha}$, then Y_2 is stochastically larger than Y_1. We propose a robust estimator of (β,α) based on four steps. In the first step the maximum rank correlation estimator is used to consistently estimate the slopes up to a scale factor. In the second step, the scale factor, the intercept, and the dispersion parameter are estimated using M-estimates. In the third step, outliers are identified using randomized quantile residuals (rqr) defined as $z_i = F_{\mu(x_i),\alpha}(y_i)$ in the continuous case or $z_i = F_{\mu(x_i),\alpha}(y_i) - u_i f_{\mu(x_i),\alpha}(y_i)$ in the discrete case, where u_i is uniformly distributed. Under the model, the rqr-s are uniformly distributed. A lower and an upper cutoff values for the z_i-s can then be determined by comparing their empirical and theoretical distributions. Outliers are observations for which z_i is not within the cutoff values and are rejected. The final estimator is a conditional maximum likelihood estimator given the retained observation. We discuss the special cases of negative binomial and Beta regression.

EO1244: **A user approach to robust multivariate analysis in R**
Presenter: **Valentin Todorov**, UNIDO, Austria
Co-authors: Peter Filzmoser

In the recent decades a lot of effort was invested in providing robust alternatives to traditional multivariate methods. The theory has been further consolidated and many computational challenges have been solved. We might have expected that by the time the robust methods would have overtaken the classical ones and would have been widely used in practice. However this did not happen and one of the main reasons, often cited is the availability of too many methods with too many parameters to tune which by itself poses a serious challenge to the practitioner. A second reason might be the lack of adequate, easy to use software, competitive to the well-developed procedures available in the standard statistical packages. Currently the packages rrcov and rrcovHD cover a large number of robustified multivariate analysis methods, both for low and high dimensional data. We will assess critically the user friendliness and applicability of these methods based on use cases from the practice and will present the new features added in the recent years. Of particular interest are the methods for dimension reduction and classification.

EC1761: **FSDA4R: Porting the FSDA toolbox to R**
Presenter: **Emmanuele Sordini**, Joint Research Centre (JRC), Italy
Co-authors: Valentin Todorov, Aldo Corbellini

The FSDA (Flexible Statistics for Data Analysis) toolbox is a software library that extends MATLAB and its Statistics Toolbox to support a robust and efficient analysis of complex datasets, affected by different sources of heterogeneity. FSDA was born around the forward search approach, and has since developed to become a comprehensive and computationally efficient software package for robust statistics, thanks to its tools for

regression and multivariate analysis, robust transformations, cluster analysis and model selection. A distinctive feature motivating the choice of the MATLAB platform is its high level of interactivity, which is particularly useful for exploratory data analysis and for a deep understanding of the influence of each individual observation on statistical estimates. One downside of FSDA, however, is that it is based on commercial software (MATLAB), and therefore it is not so appealing to the bulk of the statistical community, where the use of R is predominant. For this reason, we started a new project of porting selected FSDA tools to R (FSDA4R) which we hope to make available to the R community as an R package in CRAN. In particular, we plan to implement several general-purpose R methods, along the lines of the approach adopted in rrcov and robustbase. This presentation will focus on: i) The code porting approach; ii) The pre-requisites to be met in order to use FSDA functions in R; iii) The features of FSDA currently available in R and iv) A live demo.

EO097 Room 209 RECENT ADVANCES IN QUANTILE REGRESSION Chair: Carlos Lamarche

EO0320: **Quantile-regression inference with adaptive control of size**
Presenter: **Juan Carlos Escanciano**, Indiana University, United States
Wald-type inference in the context of regression-quantile analysis generally requires the preliminary estimation of the conditional densities of the response variables at the quantile of interest. A new approach is developed to estimating the asymptotic variance of regression quantiles that leads the resulting Wald-type tests or confidence regions to behave as well in large samples as their infeasible counterparts in which the true conditional response densities are embedded. The new estimators of the regression quantile asymptotic variances are based on a new kernel-based approach to estimating the conditional response densities. Explicit guidance on implementing these density estimators is given, including a procedure for estimation of the optimal bandwidth from the point of view of adaptively controlling the size of any resulting Wald-type test. Monte Carlo evidence indicates the potential of our approach to deliver scalar confidence intervals for median regression parameters with excellent coverage accuracy. An empirical application is also included.

EO0728: **Quantile difference in differences**
Presenter: **Brantly Callaway**, Temple University, United States
Co-authors: Tong Li, Tatsushi Oka
The effect of a binary treatment on the quantiles of some outcome is identified and estimated under a Conditional Quantile Difference in Differences assumption. This assumption says that, for individuals with similar covariates, the path of the p-th quantile ($p \in (0,1)$) is the same as the observed path for untreated observations. We show that this assumption is valid in common panel data-type models with unobserved heterogeneity. Our estimates of unconditional quantile treatment effects converge at the parametric rate. Moreover, we provide simple estimators for quantile treatment effects that are easy to implement in practice, provide results on the validity of the bootstrap for computing standard errors, and consider the empirically relevant case with more than two time periods and where treatment can occur in different periods for different individuals. Finally, we consider an application on the effect of union membership at different points in the skill distribution.

EO0936: **Common correlated effects estimation of heterogeneous dynamic panel quantile regression models**
Presenter: **Carlos Lamarche**, University of Kentucky, United States
Co-authors: Matthew Harding, Hashem Pesaran
A quantile regression estimator is proposed for a heterogeneous panel model with lagged dependent variables and interactive effects. We adopt the Common Correlated Effects (CCE) approach and demonstrate that the extension to the estimation of dynamic quantile regression models is feasible under similar conditions to the ones used in the literature. We establish consistency and derive the asymptotic distribution of the new quantile estimator. Monte Carlo studies are carried out to study the small sample behavior of the proposed approach. The evidence shows that the estimator can significantly improve the performance of existing estimators as long as the time series dimension of the panel is large. We present an application to the evaluation of Time-of-Use pricing using a large randomized control trial.

EO1184: **Regression rankscores for inference in semiparametric quantile regression models**
Presenter: **Thomas Parker**, University of Waterloo, Canada
Quantile regressions usually possess a set of associated regression rankscores, which are the dual solution in the quantile regression optimization program. Although simply a computational byproduct of the quantile regression optimization routine, they are also very useful in testing linear hypotheses in parametric quantile regression models and correspond to score tests commonly used in mean regression models. However, the use of regression rankscores as an inferential tool in semi- and nonparametric quantile regression models has not yet been investigated. Methods are developed to investigate hypotheses on quantile treatment effects in flexible quantile regression models using regression rankscores and compared to corresponding Wald-type tests.

EO457 Room 002 RECENT ADVANCES IN HIGH-DIMENSIONAL INFERENCE AND TESTING Chair: Zuofeng Shang

EO0383: **Two-Sample tests for high-dimensional linear regression with an application to detecting interactions**
Presenter: **Yin Xia**, Fudan University, China
Motivated by important applications in genomics, we consider global and coordinatewise tests for the comparisons of two high-dimensional linear regression models. A procedure for testing the equality of the two regression vectors globally is proposed and shown to be particularly powerful against sparse alternatives. In addition, we introduce a multiple testing procedure for identifying unequal coordinates while controlling the false discovery rate (FDR) and false discovery proportion (FDP). Theoretical justifications are provided to guarantee the validity of the proposed tests and optimality results are established under sparsity assumptions on the regression coefficients. Simulation results show that the proposed tests maintain the desired error rates under the null and have good power under the alternative at moderate sample sizes. The procedures are applied to the Framingham Offspring study to investigate the interactions between smoking and cardiovascular related genetic mutations important for an inflammation marker.

EO0396: **Temporal homogeneity detection of high-dimensional means**
Presenter: **Jun Li**, Kent State University, United States
Co-authors: Ping-Shou Zhong
Motivated by the objective of localizing the brain regions activated by certain stimuli in the fMRI data analysis, we consider the problem of testing temporal homogeneity of p-dimensional population mean vectors from the repeated measurements of n subjects over T times. To cope with the unique features of fMRI data, we propose a test statistic that takes into account not only the "large p, large T and small n" situation, but also the complex temporospatial dependence of the fMRI data. The asymptotic distribution of the proposed test statistic is established under mild conditions. When the null hypothesis of temporal homogeneity is rejected, we further propose a binary segmentation method shown to be consistent for multiple change-points identification. Simulation studies and an application to fMRI data are provided to demonstrate the performance of the proposed methods.

EO1293: **Exploring complex systems using semi-parametric graphical models**
Presenter: **Mladen Kolar**, University of Chicago, United States

Extracting knowledge and providing insights into the complex mechanisms underlying noisy high-dimensional data sets is of utmost importance in many scientific domains. Networks are an example of simple, yet powerful tools for capturing relationships among entities over time. For example, in social media, networks represent connections between different individuals and the type of interaction that two individuals have. In systems biology, networks can represent the complex regulatory circuitry that controls cell behavior. Unfortunately the relationships between entities are not always observable and need to be inferred from nodal measurements. A line of work will be presented that deals with the estimation and inference in high-dimensional semi-parametric elliptical copula models. We will explain why these models are useful in exploring complex systems, how to efficiently estimate parameters of the model, and also provide theoretical guarantees that justify usage of the models in scenarios where more rigid Gaussian graphical models are commonly used.

EO1665: **Inference in high-dimensional linear models with the bootstrap**
Presenter: **Karl Gregory**, University of South Carolina, United States
Co-authors: Soumendra Lahiri

Bootstrap is studied in the context of the high-dimensional linear regression model, in which the number of predictors is very large compared to the sample size. We concern ourselves with the sparse setting, in which many predictors do not influence the response, but the number of predictors which do influence the response is regarded as growing with the sample size. In this setting we consider applying the bootstrap to variants of Lasso-type estimators. We consider the de-sparsified Lasso estimator, which is asymptotically Normal under mild conditions and allows for the construction of confidence intervals for individual regression coefficients in the high-dimensional linear regression model. This estimator is ground-breaking in that it allows one to make inference on the regression coefficients of the active as well as of the inactive predictors without necessitating strong conditions. We explore ways to improve the finite-sample coverage accuracy of confidence intervals in the high-dimensional linear model setting via application of the bootstrap.

EO618 Room 201 MODERN METHODS IN THE ANALYSIS OF BRAIN IMAGING DATA Chair: Jeff Goldsmith

EO0388: **Topological data analysis for functional neuroimaging**
Presenter: **Nicole Lazar**, University of Georgia, United States

Topological data analysis (TDA) is a relatively new development at the intersection of mathematics, statistics, and computer science. The approach, which focuses on topological features of a data set, is particularly suited for "Big Data", often characterized by complicated structure such as images. We will discuss the basics of TDA, emphasizing persistent homology and its uses for functional neuroimaging data.

EO1034: **A Bayesian approach for estimating dynamic functional connectivity networks in fMRI data**
Presenter: **Erik Erhardt**, University of New Mexico, United States
Co-authors: Ryan Warnick, Michele Guindani, Elena Allen, Vince Calhoun, Marina Vannucci

Dynamic functional connectivity, i.e. how the interactions of brain regions change during an fMRI experiment, has recently received wide interest in the neuroimaging literature. We propose a principled Bayesian approach for determining time varying functional connectivity in a single subject fMRI experiment. Our method utilizes a hidden Markov model for classification of latent neurological states, achieving estimation of the connectivity networks in an integrated framework by borrowing strength over the entire time course of the experiment. Our modeling framework further assumes that the graph structures, which define the connectivity states at each time point, are related within a super-graph, so to allow immediate assessment of the differences between networks by computing posterior probabilities of both shared and differential edges in the time-varying functional networks. We assess the performance of our proposed approach in a simulation study, and also show results for an fMRI sensorimotor task experiment.

EO1234: **A spatio-temporal nonparametric Bayesian model of multi-subject fMRI data**
Presenter: **Michele Guindani**, University of California, Irvine, United States
Co-authors: Marina Vannucci

A unified, probabilistically coherent framework is proposed for the analysis of task-related brain activity in multi-subject fMRI experiments. This is distinct from two-stage group analysis approaches traditionally considered in the fMRI literature, which separate the inference on the individual fMRI time courses from the inference at the population level. In our modeling approach we consider a spatio-temporal linear regression model and specifically account for the between-subjects heterogeneity in neuronal activity via a spatially informed multi-subject nonparametric variable selection prior. For posterior inference, in addition to Markov chain Monte Carlo sampling algorithms, we develop suitable Variational Bayes algorithms. We show on simulated data that variational Bayes inference achieves satisfactory results at a much reduced computational costs than using MCMC, allowing scalability of our methods. In an application to data collected to assess brain responses to emotional stimuli our method correctly detects activation in visual areas when visual stimuli are presented.

EO1402: **Inferring brain signals synchronicity from a sample of EEG readings**
Presenter: **Donatello Telesca**, UCLA, United States

Inferring patterns of synchronous brain activity from a heterogeneous sample of electroencephalograms (EEG) is scientifically and methodologically challenging. While it is intuitively and statistically appealing to rely on readings from more than one individual in order to highlight recurrent patterns of brain activation, pooling information across subjects presents with non-trivial methodological problems. We discuss some of the scientific issues associated with the understanding of synchronized neuronal activity and propose a methodological framework for statistical inference from a sample of EEG readings. Our work builds on classical contributions in time-series, clustering and functional data analysis, in an effort to reframe a challenging inferential problem in the context of familiar analytical techniques. Some attention is paid to computational issues, with a proposal based on the hybrid combination of machine learning and Bayesian techniques.

EO453 Room 208 HIGH DIMENSIONAL STATISTICS AND FINANCE Chair: Carlos Tolmasky

EO0411: **Cleaning large correlation matrices: Eigenvector overlaps, rotationally invariant estimators and financial applications**
Presenter: **Marc Potters**, Capital Fund Management, France
Co-authors: Jean-Philippe Bouchaud, Joel Bun

A large symmetric matrix corrupted by additive or multiplicative rotational invariant noise is considered. For example, the sample covariance matrix can be written as the true covariance matrix multiplied by a random Wishart matrix. The optimal estimator (RIE) of the true matrix requires the knowledge of eigenvector overlaps between the true matrix and its noisy version. Extending a previous work, we compute these overlaps and find the optimal estimator for a large class of additive and multiplicative processes. We also show how to compute the eigenvector overlaps between two noisy matrices with different realization of the noise. In particular, we find explicit formulas for the eigenvector overlaps of two large correlated GOE matrices. Finally, we propose a simple implementation of the RIE estimator of the sample covariance matrix. Considering the Markowitz problem on real financial data, we compare the performance of the RIE estimator with that of other proposed cleaning methods.

EO0704: **Likelihood ratio test for partial sphericity in high and ultra-high dimensions**
Presenter: **Maria Antonella Gieco**, CONICET - Facultad de Ingenieria Quimica, UNL, Argentina
Co-authors: Liliana Forzani, Carlos Tolmasky
The spiked population model for covariance matrices is revisited in the setting of p and n large. This model assumes that all the population eigenvalues, with the exception of the first (largest) few, are all equal. The asymptotic distribution of the partial likelihood ratio statistic is studied and used to test for the dimension of the spike subspace. The analysis is extended to the ultra-high dimensional case, i.e. $p > n$. A thorough study of the power of the test gives a correction that allows to test for the dimension of the spike subspace even for values of p/n close to 1, a setting where other approaches proved to be deficient.

EO1302: **Spatial weight matrix estimation and financial applications**
Presenter: **Cheng Qian**, London School of Economics and Political Science, United Kingdom
Spatial econometrics focus on cross sectional interaction between physical or economic units. However, most of studies apply a prior knowledge about spatial weight matrix in spatial econometrics model. Therefore misspecification on spatial weight matrix could affect significantly accuracy of model estimation. An error upper bound for the spatial regression parameter estimators in a spatial autoregressive model has been recently provided, showing that misspecification can indeed introduce large bias in the final estimates. Meanwhile, new researches on spatial weight matrix estimation only consider static effects but not include dynamic effects between spatial units. Our model firstly use the different linear combinations of same spatial weight matrix specifications for different time-lag responds in proposed spatial econometrics model. Moveover, by introducing penalization on the coefficients of the linear combination of spatial weight matrix specifications, the best specification or the best linear combination of specifications for different lag spatial effect can be selected. To overcome endogeneity from autoregression, instrumental variables are introduced. The model we use can also find fixed effects and spillover effects. Finally, we also develop asymptotic normality for our estimation under the framwork of a previous functional dependence measure. The proposed methods are illustrated using both simulated and financial data.

EC1482: **Asymptotic theory of the sparse group lasso**
Presenter: **Benjamin Poignard**, ENSAE - CREST, France
A general framework is proposed for penalized convex empirical criteria and a new version of the Sparse-Group LASSO (SGL) is proposed, which is called the adaptive SGL, where both penalties of the SGL are weighted by preliminary random coefficients. We explore extensively its asymptotic properties and prove that this estimator satisfies the so-called oracle property, that is the sparsity based estimator recovers the true underlying sparse model and is asymptotically normally distributed. Then we study its asymptotic properties in a double-asymptotic framework, where the number of parameters diverges with the sample size. We compare our approach with some other oracle-like methods through some simulated experiments.

EO465 Room 205 ADVANCES IN CAUSAL INFERENCE Chair: Jason Roy

EO0528: **Confounding, mediation and colliding in sibling comparison designs**
Presenter: **Arvid Sjolander**, Karolinska Institute, Sweden
The sibling comparison design is an important epidemiological tool to control for unmeasured confounding, in studies of the causal effect of an exposure on an outcome. It is routinely argued that within-sibling associations are automatically controlled for all measured and unmeasured covariates that are shared (constant) within sets of siblings, such as early childhood environment and parental genetic make-up. However, an important lesson from modern causal inference theory is that not all covariate adjustments are desirable. In particular, it has been argued that collider adjustment always lead to bias, and that mediator adjustment may or may not lead to bias, depending on the research question. We use Directed Acyclic Graphs (DAGs) to distinguish between shared confounders, shared mediators and shared colliders, and we discuss which of these shared covariates that the sibling comparison design really controls for.

EO0810: **A Bayesian approach for causal inference using electronic health records**
Presenter: **Michael Daniels**, University of Texas at Austin, United States
Co-authors: Dandan Xu
To conduct comparative effectiveness research using electronic health records (EHR), we propose a causal inference approach using Bayesian nonparametric (BNP) methods. Our approach is quite general and allows causal inference beyond average causal effect (e.g. quantile causal effects). We define relevant causal quantities and specify BNP models to avoid bias from restrictive parametric assumptions. We explore the operating characteristics of our approach via simulations and use it to answer a causal question using EHRs from the University of Florida hospital.

EO0774: **Recovering experimental benchmarks with multilevel matching algorithms**
Presenter: **Luke Keele**, Georgetown University, United States
Many observational studies of causal effects occur in settings with clustered treatment assignment. In studies of this type, treatment is applied to entire clusters of units. For example, an educational intervention might be administered to all the students in a school. We develop a matching algorithm for multilevel data based on a network flow algorithm. Earlier work on multilevel matching relied on integer programming, which allows for balance targeting on specific covariates, but can be slow with larger data sets. While we cannot target balance on specific covariates, our algorithm is quite fast and scales easily to larger data sets. We also consider complications that arise from the common support assumption. In one variant of the algorithm, where we match both schools and students, we change the causal estimand to better maintain common support. In a second variant, we relax the common support assumption to preserve the causal estimand by only matching on schools. We apply the algorithm to data from a clustered RCT. We show that we can recover an experimental estimate from observational data.

EO1112: **Causal inference under selection**
Presenter: **Vanessa Didelez**, Leibniz Institute for Prevention Research and Epidemiology BIPS GmbH, Germany
The notion of "selection" we use here refers to situations where the sample is drawn, directly or indirectly, by design or by accident, in a way that depends on (i.e. conditional on) the outcome of interest, e.g. case-control sampling. Causal inference methods mainly deal with confounding; instead, we consider and suggest new approaches to causal modelling under selection. When it comes to other causal parameters of interest, such as marginal parameters, most methods proposed to date require additional external information, such as the population disease prevalence. Our approach exploits structural assumptions to obtain generic identification, e.g. knowledge that the exposure of interest and a covariate are independent, in which case the full joint distribution can be reconstructed. Such knowledge may for instance be available if it is known that the exposure and a measured genetic variant are independent. We particularly consider conditional independence constraints, which arise from graphical causal models, although other constraints can also be used. This allows us to obtain graphical rules for the generic identification of general causal parameters under selection. We show that inference on causal effects under selection is possible in a much wider class of causal models than had previously been known, as well as in more general situations other than case-control sampling.

EO169 Room 202 ECONOMETRICS AND STATISTICS FOR FINANCIAL RISK MANAGEMENT Chair: Katrien Antonio

EO0718: **A methodological overview for quantifying the risk of an accident in usage-based insurance**
Presenter: **Montserrat Guillen**, University of Barcelona, Spain
Co-authors: Ana M Perez-Marin, Mercedes Ayvso, Jens Perch Nielsen

Methods are presented to quantify risk with application to usage-based motor insurance. We show illustrations derived from a study of young drivers with PAYD Insurance where the risk of being involved in a crash by estimating the time and distance travelled to first accident is analyzed. The comparison between novice vs. experienced young divers shows that vehicle usage differs between these two groups and that the time to first crash is shorter for those drivers with less experience. Men show riskier driving patterns than women. The risk of accident increases with excessive speed, but the effect is higher for men than for women among the more experienced drivers. Additionally, nighttime driving reduces the time to first accident for women but not for men. Gender differences in the risk of accident are mainly attributable to the fact that men drive more often than women. So, we explore alternative methods to include mileage in the quantification of risk, as well as the way exposure to risk is contemplated in generalized linear models. Finally, we have also investigated changes in driving patterns after having an accident, and conclude that those who speed more and have accidents with bodily injuries reduce their proportion of speed violations after the accident.

EO0794: **Unraveling the predictive power of telematics data in car insurance pricing**
Presenter: **Roel Verbelen**, KU Leuven, Belgium
Co-authors: Katrien Antonio, Gerda Claeskens

Telematics data are analyzed from a Belgian portfolio of young drivers who underwrote a pay-as-you-drive insurance product in between 2010 and 2014. Using installed black box devices, telematics data are collected on how many kilometers are driven, during which time slots and on which type of roads. Car insurance is traditionally priced based on self-reported information from the policyholder, most importantly: age, license age, postal code, engine power, use of the vehicle, and claims history. However, these observable risk factors are only indirect indicators of the accident risk and don't reflect the real driving behavior. By constructing predictive models for the claim frequency, we compare the performance of different sets of predictor variables (e.g. traditional vs purely telematics) and discover the relevance and impact of adding telematics insights. In particular, we contrast the use of time and distance as exposure-to-risk, the basic rating unit underlying the insurance premium. We show how to incorporate the divisions of the distance driven by road type and time slots as compositional predictors in the model and how to interpret their effect on the average claim frequency. We find that the telematics variables increase the predictive power and render the use of gender as a discriminating rating variable redundant.

EO0915: **Aggregating and disaggregating risks**
Presenter: **Miguel Santolino**, University of Barcelona, Spain
Co-authors: Catalina Bolance, Montserrat Guillen

Risk quantification is especially complex when risks are aggregated and/or disaggregated. In both situations the interconnection between individual risks and the aggregate risk has to be addressed. In the case of risk aggregation, risk measures are frequently asked to be subadditive. A new class of distortion risk measures is presented and the notions of tail-subadditivity discussed. Non-parametric methods are not popular in risk quantification due to: (1) estimating the risk implies evaluating the distribution in the tail; (2) sample information associated with tails is scarce; (3) the difficulty increases with heavy tail distributions. We analyze modified non-parametric methods, also based on distorting the distribution, that improve the efficiency of non-parametric estimations of the risk. When an aggregate risk is disaggregated, managers have to decide how the aggregate amount is distributed among individual risks and there are alternative criteria to do it. A particular allocation criterion is a capital allocation principle. Many allocation principles may be found, but the literature dealing how to compare them is scarce. The relationship between capital allocation principles and compositional data is provided, where compositions are quantitative descriptions of the components of a whole. Compositional methods provide quantitative criteria to compare capital allocation principles. An immediate application is that they can be ranked based on their distances.

EO1273: **Sparse modeling of risk factors for insurance analytics**
Presenter: **Sander Devriendt**, KU Leuven, Belgium
Co-authors: Katrien Antonio, Edward Frees, Roel Verbelen

Insurance companies use predictive models for a variety of analytic tasks, including pricing, marketing campaigns, claims handling, fraud detection and reserving. Typically, these predictive models use a selection of categorical, continuous, spatial and multi-level (e.g. car brand and model) covariates to differentiate risks. Insurance companies compete by setting their prices through risk classification or segmentation. The pricing model should not only be competitive, but also interpretable by stakeholders (including the policyholder and the regulator) and easy to implement and maintain in a production environment. That is why current literature on actuarial pricing puts focus on generalized linear models where risk factors are binned (or: categorized) up front, using ad hoc techniques or professional expertise. Relying on the statistical literature on sparse modeling with penalization techniques, we present a strategy which includes variable selection and the binning or grouping of risk factors within the model estimation process. As such, we are able to simultaneously select, estimate and group, in a statistically sound way, any combination of categorical, continuous, spatial and multi-level risk factors. We explain the general framework and show how this method incorporates different adjustable penalties to handle categorical, ordinal, nominal and spatial information. We illustrate the approach with a case-study on a motor third party liability dataset.

EO652 Room 207 ADVERSARIAL RISK ANALYSIS Chair: Jacinto Martin Jimenez

EO0751: **An artificial bee colony-based optimization approach to design a water quality monitoring network in a river system**
Presenter: **Carlos Javier Perez Sanchez**, University of Extremadura, Spain
Co-authors: Klara Reder, Miguel A Vega-Rodriguez, Martina Florke, Joseph Alcamo

The risk of diseases or even death increase when drinking or coming into contact with polluted water. Water quality monitoring is determinant in the management of freshwater resources. Water quality monitoring network design has been traditionally based on intuition, experience and subjective judgement in locating monitoring stations. In the last years some approaches have tried to objectivize station location based on different criteria. An artificial bee colony-based optimization approach has been implemented to minimize the number of stations in a river at the same time that maximizing the probability of detecting threshold violations of three water pollution parameters: biochemical oxygen demand (BOD5), faecal coliform bacteria (FC) and total dissolved solids (TDS). The approach has been applied to the Great Fish River (South Africa) based on WorldQual data, where the whole watershed is structured as a grid. The results show that the approach succeeds in finding the locations where more pollution is concentrated whereas the number of stations remains at a minimum. A great advantage of the proposed approach is that it can be used to initially locate monitoring stations based only on WorldQual simulated data in rivers where there is no previous monitoring network. The approach can be extended by including other criteria (e.g. those involving distances or data variability), which may enhance the proposal.

EO0820: **Social network analysis for fraud detection using complicity functions**
Presenter: **Alfonso Mateos Caballero**, Universidad Politecnica de Madrid, Spain

In social networks the actors are people o groups of people, whose connections are based on any form of social interaction between them, such as economic transactions. Social network analysis has important implications in the fight against organized crime, business associations with fraudulent purposes or terrorism. We propose several complicity functions to measure the degree of complicity between the actors in a social network with a set of previously identified fraudster actors, with the aim of detecting the partners of an organized crime plot. We also propose a procedure for ring detection consisting on five steps. The first and second steps compute the complicity of each actor with each fraudster actor and the strength of attraction between fraudster actors, respectively. The third step projects the set of toxic actors using multidimensional scaling in a plane. In the fourth step the points of the projection are grouped according to the DBSCAN algorithm. Finally, the fifth step adds each non-fraudster actor that maximizes its complicity. The ring detection procedure is illustrated with a real example including 835 linked companies, of which eight are fraudulent.

EO1209: **Model for management risk of burglaries**
Presenter: **Isabel Serra**, Universitat Autonoma de Barcelona, Spain

The DefendAttackDefend model in adversarial risk analysis proposes a process of take-decission that can be measured by the losses in each step. This learning process lets to characterize the evolution of shape of the extremes losses in a geo-localization models that can be applied to manage risk of burglaries.

EO0912: **A Bayesian hierarchical model for extreme rainfall in Extremadura**
Presenter: **Jacinto Martin Jimenez**, Universidad de Extremadura, Spain
Co-authors: Lizbeth Naranjo Albarran, Francisco Javier Acero Diaz, Jose Agustin Garcia Garcia

Motivated by extreme rainfall data in Extremadura (Spain), a Bayesian hierarchical spatio-temporal model is proposed to address environmental data distributed in space and time in a more flexible way. There exists several ways to address the statistical analysis of extreme events, one of them considers the extreme value theory. In addition to the extreme values, due to the spatial character of the rainfall, it is needed to address the problem of spatial extremes. Moreover, considering the consequences of the climate change, the time has also an important effect in the meteorological events. One of the benefits of using spatio-temporal extreme models, is to improve the precision in the parameters estimation sharing information from similar sites and considering jointly the temporal change effect. A Bayesian hierarchical model is developed. It is assumed that the observations follow a generalized extreme value distribution for which the location, scale and shape parameters have space-time structures, depending of localization coordinates of the sites in Extremadura and the rainfall year.

EO515 Room 006 TIME SERIES OF NETWORKS Chair: Carsten Jentsch

EO0994: **Markov process models for time-varying networks**
Presenter: **Harry Crane**, Rutgers University, United States

Many models for dynamic networks, such as the preferential attachment model, describe evolution by sequential addition of vertices and/or edges. Such models are not suited to networks whose connectivity varies over time, as in social relationships and other kinds of temporally varying interactions. For modeling in this latter setting, I develop the general theory of exchangeable Markov processes for time-varying networks and discuss relevant consequences.

EO1193: **Measuring dynamic connectedness with large TVP-VAR models**
Presenter: **Kamil Yilmaz**, Koc University, Turkey

The connectedness index methodology recently proposed relies on rolling sample windows estimation of VAR model. The resulting dynamic total connectedness measure possesses excessive persistence which potentially renders the identification of the beginning and end of the crisis episodes difficult. In order to provide a solution to the excessive persistence problem, we estimate a large TVP-VAR model of daily stock return volatilities for 28 U.S. and European financial institutions. Our results clearly show that TVP-VAR dynamic connectedness measures from the TVP-VAR model do not suffer from the excessive persistence problem. As a direct consequence, the dynamic total connectedness index obtained from the TVP-VAR model shows more pronounced jumps during important crisis moments than the DY index obtained from rolling-windows estimation.

EO1067: **Network time series modelling**
Presenter: **Marina Knight**, University of York, United Kingdom
Co-authors: Matthew Nunes, Guy Nason

A network time series is a multivariate time series augmented by a graph that describes how variables (or nodes) are connected. We introduce the network autoregressive (integrated) moving average (NARIMA) processes: a set of flexible models for network time series. For fixed networks the NARIMA models are essentially equivalent to vector autoregressive moving average-type models. However, NARIMA models are especially useful when the structure of the graph, associated with the multivariate time series, changes over time. Such network topology changes are invisible to standard VARMA-like models. For integrated NARIMA models we introduce network differencing, based on the network lifting (wavelet) transform, which removes trend. We exhibit our techniques on a network time series describing the evolution of mumps throughout counties of England and Wales weekly during 2005.

EO1194: **Modeling dynamic networks using high-dimensional binary autoregressive time series processes**
Presenter: **Carsten Jentsch**, University of Mannheim, Germany
Co-authors: Shaikh Tanvir Hossain, Lena Reichmann

Suppose a time series of networks is identified by their adjacency matrices $A_1,\ldots,A_T$, where $A_t = (a_{ij;t})_{i,j=1,\ldots,N}$ with $a_{ij;t} \in \{0,1\}$ and $a_{ij,t} = 1$ indicating that there is a directed edge pointing from vertex i to vertex j at time t. We propose to model the joint dynamics of the edges by multivariate binary time series processes. We adopt a previous class of Discrete AutoRegressive Moving-Average (DARMA) models to model univariate categorical time series. For our purpose of modeling multivariate binary data, we extend their approach in two aspects. First, we consider vector-valued extensions. Second, we propose a simple modification that allows for negative autocorrelations. The resulting model class is flexible enough to capture very general autocorrelations driving the dynamic network structure. The model class fulfills Yule-Walker-type equations that allow in principle an explicit estimation of the model parameters. However, as the dimension of the adjacency matrices grows quadratically with the number of vertices, we have to make use of Lasso-penalization techniques to estimate sparse models. We adopt an approach which guarantees consistent estimation for high-dimensional vector autoregressive models via ℓ_1-regularization for a large class of stable processes under sparsity constraints.

EO205 Room 214 RECENT STATISTICAL ADVANCES FOR CLINICAL TRIALS Chair: Jacobo de Una-Alvarez

EO1275: **Causal mediation analysis in clinical survival trials incorporating repeated mediator measurements**
Presenter: **Susanne Strohmaier**, Brigham and Women's Hospital, Harvard Medical School, United States

Important questions within the fields of social sciences, epidemiology as well as clinical trial research involve the challenge of decomposing the effect of an intervention into direct and indirect effects working through a defined mediator, thereby aiming for a better understanding of the underlying mechanisms. For the case of a single and multiple mediators measured at a single point in time, researchers have established theoretical properties and developed practical tools for the analysis of a broad class of mediator and outcome models. However, data structures are often more complex than the described scenarios. Especially when it comes to clinical survival trials, usually collection of time-to-event data goes hand in hand with the collection of information on other internal time-dependent covariates, but this potentially useful information is hardly ever used. We present dynamic path analysis, a method that allows us to investigate mediation processes by integrating more of the routinely collected data to gain a better understanding how treatment effects evolve over time. Additionally, we suggest a definition of direct, indirect and total effects for this particular setting of a time-to-event outcome and a repeatedly measured mediator that allows a causal interpretation without making explicit use of the counterfactual framework.

EO1755: **On the selection of a randomization procedure in clinical trials**
Presenter: **Ralf-Dieter Hilgers**, RWTH University Aachen, Germany

In randomized clinical trials randomization is the most important design techniques to avoid bias. This statement is based on long run consideration and might fail in small clinical trials. Consequently, an investigation of the properties of randomization procedures with respect to the impact of bias on the study result is of interest. On the other hand, this investigation implies a comparison study of randomization procedure in the planning phase o a trial.In this talk, I will show the impact of selection bias on the study results i.e. the type 1 error probability. As a consequence a scientific argumentation to select a randomization procedure is enabled.

EO1251: **Planning clinical trials with CompARE**
Presenter: **Guadalupe Gomez Melis**, Universitat Politecnica de Catalunya, Spain
Co-authors: Moises Gomez Mateu, Marta Bofill

Randomized clinical trials provide compelling evidence that a study treatment causes an effect on human health. A primary endpoint ought to be chosen to confirm the effectiveness of the treatment and is the basis for computing the number of subjects in a Randomized clinical trial. Often a Composite Endpoint based on a combination of individual endpoints is chosen as a Primary Endpoint. As a tool for a more informed decision between using the Composite Endpoint as Primary Endpoint or one of its components, the ARE method is proposed. This method uses the Asymptotic Relative Efficiency (ARE) between the two possible logrank tests to compare the effect of the treatment. CompARE, a web-based interface tool is presented. CompARE computes the asymptotic relative efficiency in terms of interpretable parameters such as the anticipated probabilities of observing the primary and secondary endpoints and the relative treatment effects on every endpoint given by the corresponding hazard ratios. The ARE method is extended to observational studies as well as to Binary Composite Endpoints. A discussion on how to use the ARE method for the derivation of the sample size when the proportional hazards assumption does not hold, will conclude.

EO1546: **Some statistical aspects of clinical trials for personalized medicine**
Presenter: **Tim Friede**, University Medical Center Goettingen, Germany
Co-authors: Marius Placzek

In personalized medicine patient populations are stratified with a view to improve treatment outcome in terms of efficacy and tolerability. Stratification is often carried out by biomarkers. A particular case are nested subgroups that might arise from using several thresholds of a continuous marker. Efficient testing strategies for normal data are derived under homoscedasticity and heteroscedasticity assumptions across the subgroups. Furthermore, procedures for sizing a study with several nested subgroups are presented. These depend among other quantities on nuisance parameters such as the variances of the outcomes in the subgroups and the prevalences of the subgroups. Knowledge of these might be very scarce in the planning phase of such a trial resulting in a considerable risk of choosing an inappropriate sample size. To mitigate these risks an internal pilot study design is proposed and its properties including type I error rate, power and sample size distribution are explored in Monte Carlo simulations. Adaptive enrichment designs allow to restrict recruitment to certain subgroups following interim analyses. Approaches to hypothesis testing in such designs are reviewed and their properties are compared in a simulation study. Finally, an adaptive enrichment design including an internal pilot study is presented and its characteristics are discussed.

EO021 Room 210 RECENT DEVELOPMENTS IN SPATIAL STATISTICS Chair: Yulia Gel

EO0313: **An approach to hypothesis testing based on local indicators of spatio-temporal association**
Presenter: **Francisco Javier Rodriguez Cortes**, Universitat Jaume I, Spain

The detection of clustering structure in a point pattern is one of the major focus of attention in spatio-temporal data mining. For instance, statistical tools for clustering detection and identification of events belonging to clusters are welcome in epidemiology and seismology. Local second-order statistics can provide information on how an event relates to nearby events. We extend local indicators of spatial association (known as LISA functions) into the spatio-temporal context (which then will be called LISTA functions). These functions can be used for local tests in the context of case-control spatio-temporal point patterns, and are able to assess in the neighbourhood of each event if the two point patterns have a different structure. We present a simulation study and apply this methodology to earthquakes data.

EC1449: **A multivariate spatial econometrics model with an intra-location feedback effect**
Presenter: **Nima Shariati Fokalaei**, Shariati Fokalaei, Sweden
Co-authors: Hossein Asgharian

In the spatial multivariate econometrics models, the relations are typically built for the between location dependencies, which is adopted from univariate case like SEM and SLM models. However, this does not allow for intra-location dependencies to be accounted directly. The weakness of the previous models is shown analytically and using examples. A new multivariate spatial econometric model is presented that accounts for a feedback effect between variables within the same location, called intra-location feedback effect, which was not considered before in this area. Model identification and other basic properties of the spatial multivariate econometrics model, and especially for the extended version, are established. Statistical inferences are presented using a proposed analytical method based on empirical precision matrix and also the maximum likelihood. Model validation for new data sets is discussed.

EC1627: **Bootstrapping and bandwidth selection for kernel intensity estimation in point processes with covariates**
Presenter: **Maria Isabel Borrajo**, University of Santiago de Compostela, Spain
Co-authors: Wenceslao Gonzalez-Manteiga, Lola Martinez-Miranda

Point processes are used in spatial statistics to study the geometrical structure of patterns formed by objects or events, which are distributed randomly in number and space. In point processes we can have extra information through marks that characterize every event, but also, as it is the

case, through covariates that give information over the whole spatial region where the process is defined. In this situation we focus on the estimation of the intensity function that characterizes the point process. We propose a new kernel estimator for this function based on the information given by a single covariate and we provide asymptotic derivations about its mean and its variance. A new bootstrap method is also defined and used for data-driven bandwidth selection. Finally we show the finite-sample performance of the proposals through an extensive simulation study.

EO131 Room 215 MODERN MULTIVARIATE METHODS WITH R Chair: Anne Ruiz-Gazen

EO0809: **Multivariate analysis of mixed data: The PCAmixdata R package**
Presenter: **Marie Chavent**, University of Bordeaux, France
Mixed data type arise when observations are described by a mixture of numerical and categorical variables. The R package PCAmixdata extends standard multivariate analysis methods to incorporate this type of data. The key techniques included in the package are PCAmix (PCA of a mixture of numerical and categorical variables), PCArot (rotation in PCAmix) and MFAmix (multiple factor analysis with mixed data within a dataset). A synthetic presentation of the three algorithms will be provided and the three main procedures will be illustrated on real data composed of four datasets caracterizing conditions of life of cities of Gironde, a south-west region of France.

EO0789: **Blind source separation based on joint diagonalization in R: The packages JADE and BSSasymp**
Presenter: **Klaus Nordhausen**, University of Turku, Finland
Co-authors: Jari Miettinen, Sara Taskinen
Blind source separation (BSS) is a well-known signal processing tool which is used to solve practical data analysis problems in various fields of science. In BSS is assumed that the observed data consists of linear mixtures of latent variables. The mixing system and the distributions of the latent variables are unknown. The aim is to find an estimate of an unmixing matrix which then transforms the observed data back to latent sources. The package JADE offers several BSS methods which are based on joint diagonalization. Package BSSasymp contains functions for computing the asymptotic covariance matrices as well as their data-based estimates for most of the BSS estimators included in the package JADE. Both packages and their underlying methodology are introduced using simulated and real data.

EO0831: **Using SOMbrero for clustering and visualizing complex data**
Presenter: **Madalina Olteanu**, Pantheon-Sorbonne University, France
Co-authors: Nathalie Villa-Vialaneix
Over the years, the self-organizing map (SOM) algorithm was proven to be a powerful and convenient tool for clustering and visualizing data. While the original algorithm had been initially designed for numerical vectors, the available data in the applications became more and more complex, being frequently too rich to be described by a fixed set of numerical attributes only. This is the case, for example, when the data are described by relations between objects (individuals involved in a social network) or by measures of resemblance/dissemblance. The aim will be to illustrate how the SOM algorithm can be used to cluster and visualize complex data such as graphs, categorical time series or panel data. In particular, the focus is on the use of the R package SOMbrero, which implements an online version of the relational self-organizing map, able to process any dissimilarity data. The package offers many graphical outputs and diagnostic tools, and comes with a user-friendly web graphical interface based on R-Shiny. Several examples on various real-world datasets will be given for highlighting the functionalities of the package.

EO1276: **Analysis, simulation and prediction of multivariate random fields**
Presenter: **Martin Schlather**, Universitat Mannheim, Germany
Multivariate data in space and time, such as temperature and pressure, are ubiquitous in environmental sciences, physics and materials science. The analysis and the modeling of the data is challenging with respect to the search for useful models, the load of parameters, and the amount of data. The R package RandomFields has started to include multivariate geostatistics and deals with several cross-covariance models of multivariate and vector-valued models, allowing for trends and anisotropies. The package supports the simulation, the parameter estimation and the prediction. An overview over multivariate geostatistical models and their implementation in RandomFields will be given. The concept of how trend modeling can be included in general will be also be discussed.

Sunday 11.12.2016 10:50 - 12:05 Parallel Session M – CFE-CMStatistics

CO301 Room 107 BAYESIAN ECONOMETRICS Chair: Luca Rossini

CO0243: **Bayesian nonparametric aparse seemingly unrelated regression model (SUR)**
Presenter: **Luca Rossini**, Ca Foscari University of Venice, Italy
Co-authors: Monica Billio, Roberto Casarin
Seemingly unrelated regression (SUR) models are used in studying the interactions among economic variables of interest. In a high dimensional setting and when applied to large panel of time series, these models have a large number of parameters to be estimated and suffer of inferential problems. We propose a Bayesian nonparametric hierarchical model for multivariate time series in order to avoid the overparametrization and overfitting issues and to allow for shrinkage toward multiple prior means with unknown location, scale and shape parameters. We propose a two-stage hierarchical prior distribution. The first stage of the hierarchy consists in a lasso conditionally independent prior distribution of the Normal-Gamma family for the SUR coefficients. The second stage is given by a random mixture distribution for the Normal-Gamma hyperparameters, which allows for parameter parsimony through two components. The first one is a random Dirac point-mass distribution, which induces sparsity in the SUR coefficients; the second is a Dirichlet process prior, which allows for clustering of the SUR coefficients. We provide a Gibbs sampler for posterior approximations based on introduction of auxiliary variables. Some simulated examples show the efficiency of the proposed model. We study the effectiveness of our model and inference approach with an application to macroeconomics.

CO1283: **Likelihood shape and regularization priors for econometric models with reduced rank**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Nalan Basturk, Lennart Hoogerheide
'Reporting the shape of the likelihood and its properties is an important task for a Bayesian econometrician', was argued by Hildreth in 1963. This viewpoint is based on the idea that scientific evidence should be reported in such a way that the information specified in the econometric model dominates with respect to other sources of information. Shape and features of the posterior probability density function of the parameters of interest of reduced rank econometric models under regularization priors are analyzed. Given the indicated relative prior ignorance, the likelihood information strongly dominates. We obtain, to the best of our knowledge, novel results.

CO0686: **Bayesian matrix regression**
Presenter: **Matteo Iacopini**, Ca Foscari University of Venice, Italy
Co-authors: Roberto Casarin, Monica Billio
A general model is proposed for linear regression with matrix variate response, which encompasses univariate and multivariate regression as special cases. For dealing with the issue of dimensionality, we exploit a suitable decomposition which enables to achieve both parsimony and to incorporate time-varying sparsity on the coefficients with the additional aim of capturing the change over time of the relevance of covariates. Inference is carried out in the Bayesian framework via Gibbs sampler.

CO423 Room 106 INFERENCE ON NON-CAUSAL OR NON-INVERTIBLE TIME SERIES Chair: Carlos Velasco

CO0293: **Noncausality and the commodity currency hypothesis**
Presenter: **Henri Nyberg**, University of Turku and University of Helsinki, Finland
Co-authors: Matthijs Lof
New evidence is provided on the role of exchange rates in forecasting commodity prices. Consistent with previous studies, we find that commodity currencies hold out-of-sample predictive power for commodity prices when using standard linear predictive regressions. After, we reconsider the evidence using noncausal autoregressions, which provide a better fit to the data and are able to accommodate the effects of nonlinearities and omitted variables, the predictive power of exchange rates disappears.

CO0361: **Frequency domain minimum distance inference for possibly noninvertible and noncausal ARMA models**
Presenter: **Carlos Velasco**, Universidad Carlos III de Madrid, Spain
Co-authors: Ignacio Lobato
Frequency domain minimum distance procedures are introduced for performing inference in general, possibly non causal and/or noninvertible, autoregressive moving average (ARMA) models. We use information from higher order moments to achieve identification on the location of the roots of the AR and MA polynomials for non-Gaussian time series. We study minimum distance estimation that combines the information contained in second, third, and fourth moments. Contrary to existing estimators, the proposed estimator is consistent under general assumptions, and improves on the efficiency of the estimates based on only second order moments.

CC1007: **Multivariate all-pass time series models: Modelling and estimation strategies**
Presenter: **Juho Nyholm**, University of Helsinki, Finland
Co-authors: Bernd Funovits
Multivariate all-pass time series models are analyzed, i.e. rational matrix functions $Q(z)$ for which $Q(z)Q(z^{-1})' = I$ holds. In particular, the poles of the determinant of all-pass matrix functions are the reciprocals of its roots. Multivariate all-pass models are important for analysing non-causal and non-invertible time series models as recently proposed. For i.i.d white noise input, all-pass models generate uncorrelated (white noise) processes; however, these processes are not independent in the non-Gaussian case. We use the theory of rational matrix factorization to gain insight into the structure of all-pass models. Furthermore, we propose estimation and identification strategies.

CO445 Room 104 LARGE-SCALE MULTIVARIATE MODELING OF ASSET RETURNS AND PORTFOLIO ALLOCATION Chair: Marc Paolella

CO0318: **Robust normal mixtures for financial portfolio allocation**
Presenter: **Marco Gambacciani**, University of Zurich and Swiss Financial Institute, Switzerland
Co-authors: Marc Paolella
A new approach for multivariate modeling and prediction of asset returns is proposed. It is based on a two-component normal mixture, estimated using a fast new variation of the minimum covariance determinant (MCD) method made suitable for time series. It outperforms the (shrinkage-augmented) MLE in terms of out-of-sample density forecasts and portfolio performance. In addition to the usual stylized facts of skewness and leptokurtosis, the model also accommodates leverage and contagion effects, but is i.i.d., and thus does not embody, for example, a GARCH-type structure. Owing to analytic tractability of the moments and the expected shortfall, portfolio optimization is straightforward, and, for daily equity returns data, is shown to substantially outperform the equally weighted and classical long-only Markowitz framework, as well as DCC-GARCH (despite not using any kind of GARCH-type filter).

CO0515: **A flexible regime-switching model for asset returns**
Presenter: **Patrick Walker**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Pawel Polak

A Markov regime-switching correlation model for a multivariate set of asset returns is proposed. The univariate series are endowed with the usual GARCH structure, and the underlying innovations are multivariate generalized hyperbolic distribution (MGHyp). The multivariate conditional predictive distribution is MGHyp, hence weighted sums of marginals are themselves GHyp and thus tractable, enabling, e.g., portfolio optimization. To accomplish joint likelihood estimation of all the model parameters, a new, fast and efficient two-stage EM-algorithm is developed for estimation. This is coupled with shrinkage estimation of the correlation matrices via a quasi-Bayesian prior, enhancing both estimation ease and forecast quality. Based on Dow Jones 30 data from 1999 to 2014, the new model is demonstrated to outperform all special cases in terms of in-sample fit and out-of-sample density forecasts. An application to portfolio optimization shows the importance of dynamical correlations for optimal asset allocation by providing consistently higher Sharpe ratios for all RSDC models, compared to their CCC counterparts.

CO1299: **Portfolio selection with active risk monitoring**
Presenter: **Pawel Polak**, Columbia University, United States
Co-authors: Marc Paolella

A framework is proposed for large-scale portfolio optimization which accounts for all the major stylized facts of multivariate financial returns, including volatility clustering, dynamics in the dependency structure, asymmetry, heavy tails, and non-ellipticity. It introduces a so-called risk fear portfolio strategy which combines portfolio optimization with active risk monitoring. The former selects optimal portfolio weights. The latter, independently, initiates market exit in case of excessive risks. The strategy agrees with the stylized fact of stock market major sell-offs during the initial stage of market downturns. The advantages of the new framework are illustrated with an extensive empirical study. It leads to superior multivariate density and Value-at-Risk forecasting, and better portfolio performance. The proposed risk fear portfolio strategy outperforms various competing types of optimal portfolios, even in the presence of conservative transaction costs and frequent rebalancing. The risk monitoring of the optimal portfolio can serve as an early warning system against large market risks. In particular, the new strategy avoids all the losses during the 2008 financial crisis, and it profits from the subsequent market recovery.

CO585 Room 112 COMMODITY PRICES: FORECASTING AND POLICY DESIGN Chair: Anthony Garratt

CO0389: **Commodity prices and fiscal policy design: Procyclical despite a rule**
Presenter: **Hilde Bjornland**, BI Norwegian Business School, Norway
Co-authors: Leif Anders Thorsrud

The aim is to analyse if the adoption of a fiscal rule insulates the domestic economy from commodity price fluctuations in a resource-rich economy. To do so, we develop a time-varying Dynamic Factor Model, in which both the volatility of structural shocks and the systematic fiscal policy responses are allowed to change over time. We focus on a particular country, Norway, that is put forward as exemplary with its handling of resource wealth; income from the sale of petroleum is first saved in a sovereign wealth fund for then to be spent following a fiscal rule. We find that, contrary to common perception, fiscal policy has been more (not less) procyclical with commodity prices since the adoption of the rule. Fiscal policy has thereby exacerbated the commodity price fluctuations on the domestic economy. Still, compared to many other resource-rich economies practicing a more spend-as-you-go strategy the responses are modest, as also documented in our counterfactual analysis. From a policy point of view, the implications of our findings are therefore of general interest since they highlight strengths and weaknesses of fiscal rules adopted in resource rich countries.

CO0432: **Risk premia and seasonality in commodity futures**
Presenter: **Ivan Petrella**, Warwick Business School, United Kingdom

A multifactor affine model of commodity futures is developed and estimated that allows for stochastic variations in seasonality. We show conditions under which the yield curve and the cost-of-carry curve adopt augmented Nelson and Siegel functional forms. This restricted version of the model is parsimonious, does not suffer from identification problems, and matches well the yield curve and futures curve over time. We estimate the model using heating oil futures prices over the period 1984-2012. We find strong evidence of stochastic seasonality in the data. We analyse risk premia in futures markets and discuss two traditional theories of commodity futures: the theory of storage and the theory of normal backwardation. The data strongly supports the theory of storage.

CO0394: **Disaggregate commodity prices, convenience yields and inflation densities**
Presenter: **Anthony Garratt**, University of Warwick, United Kingdom
Co-authors: Ivan Petrella

The ability of different commodity groups to generate well calibrated h-step ahead US consumer price inflation predictive densities is examined. Using a model averaging approach, in contrast to the existing literature which often adopts a factor based approach, we combine a large number of simple linear models which explain h-step ahead inflation. The models use either a single commodity spot price or a single commodity convenience yield, defined over different lag lengths of the commodity variable and inflation. Two averaging methods are examined, the Linear Opinion Pool and the Log Linear opinion pool, where time-varying) weighted averages are constructed using both log score (LS) and continuous rank probability score (CRPS) weights. Averages which allows the contribution of each commodity group, and the second for the effects of spot prices versus convenience yields, are examined. The general features of the results are (i) there are significant roles for most commodity group considered and for models containing both spot prices or convenience yields (ii) where a single commodity group does take a more dominant role, this is most often Grain group, but the degree of dominance is modest and varies over the forecast horizon being considered (iii) time variation is a feature of the weightings.

CO587 Room Board meeting room II HIGH FREQUENCY EMPIRICAL FINANCE Chair: Wenying Yao

CO0558: **Asymmetric jump beta estimation with implications for portfolio risk management**
Presenter: **Wenying Yao**, Deakin University, Australia
Co-authors: Vitali Alexeev, Giovanni Urga

The aim is to evaluate the impact of extreme market shifts on equity portfolios. Assuming that investors care differently about downside losses as opposed to upside gains, we estimate jump sensitivities for the negative and positive market shifts. We investigate the implications of the difference in negative and positive sensitivities to market jumps for portfolio risk management by contrasting the results for individual stocks with the results for portfolios with varying number of holdings. In the context of a portfolio, we investigate to what extent the downside and upside jump risks can be diversified away. This can have a direct impact on the pricing of jump risks and subsequently, investors' decision-making. Varying the jump identification threshold, we show that the asymmetry is more prominent for more extreme events and that the number of holdings required to diversify portfolios' sensitivities to negative jumps is higher than that required for positive jump diversification. We found that ignoring the

asymmetry in sensitivities to negative versus positive market jumps may result in under-diversification of portfolios and increased exposure to extreme negative market shifts.

CO1410: **Estimation of the discontinuous leverage effect: Evidence from the NASDAQ order book**
Presenter: **Lars Winkelmann**, Freie Universitaet Berlin, Germany

A new nonparametric estimation approach is provided for studying the correlation of price and volatility jump sizes in high-frequency data. A spectral estimator is proposed which allows to asymptotically remove market microstructure noise from observable prices and to locate and estimate the size of price and volatility jumps in the efficient price process. The leverage effect is estimated by the scaled discontinuous part of the quadratic covariation. Based on five years of NASDAQ transaction data of about 300 firms, we are able to confirm evidence about an elusive relation of price and volatility jump sizes in terms of sign, magnitude and significance. We show that this inconsistency is due to different sources of jumps and a time-varying correlation. The filtering of market-wide events and important macroeconomic news releases gives uniform and highly significant firm specific correlations. Idiosyncratic price volatility jumps are identified as one disrupting component of the discontinuous leverage effect.

CC1501: **Co-jumps asymmetry**
Presenter: **Haiyue Yu**, University of Hong Kong, China
Co-authors: Qi Lin

The co-jumps between individual assets and market index are asymmetric and co-downside jumps are significantly different from co-upside jumps. We make use of high-frequency stock return data to estimate downside jump beta and upside jump beta and find the difference between downside jump beta and upside jump beta is statistically significant for stocks included in Dow Jones 30 Index. We propose a measure for co-jumps asymmetry (CJA) and find that stocks with higher CJA (more left-skewed co-jumps) outperform their counterpart by a monthly return of 0.48%, associated with t-statistics 2.94. The prediction effects of CJA are stronger for firms with smaller idiosyncratic volatility, higher liquidity and larger firm size, which implies that the relationship between CJA and cross-sectional stock return is driven by risk-return trade off rather than behavioral bias of the investors.

CO335 Room 110 OIL PRICE DYNAMICS Chair: Ivan Paya

CO0561: **On speculative bubbles in oil markets: An analysis based on market expectations**
Presenter: **Ivan Paya**, Lancaster University, United Kingdom
Co-authors: Efthymios Pavlidis, David Peel

Competing arguments have been put forward to explain the huge fluctuations in oil prices from 1990 to 2012, in particular, to the run-up of prices in 2007-08 and its subsequent collapse. The disconnect between oil prices and fundamental factors has been analysed by testing for speculative bubbles. Econometric tests are typically not conclusive because they examine a joint null hypothesis of absence of bubbles, and that the econometrician uses the true model for fundamentals. We contribute to this debate by adopting an approach similar to a recent one proposed to deal with the joint hypothesis problem. Instead of utilising derivative prices to avoid specifying an asset price determination model, we advocate the use of market expectations about future oil prices. A major advantage of market expectations is that the former is not contaminated by a risk premium. By analysing survey data obtained from Consensus Economics, we show that both the spot real oil price and the expected future oil price do individually appear to display periods of explosive dynamics -in line with the presence speculative bubbles in the market. However, recursive right-tail unit root tests on the difference between the expected future price of oil and the actual future spot price does not reject the null of no bubbles for any time period between 1990 and 2012. We therefore conclude against the presence of speculative bubbles in real oil prices during this sample period.

CO1316: **Oil price shocks and the UK fiscal regime: 1990-2005**
Presenter: **Marco Lorusso**, Heriot-Watt University, United Kingdom
Co-authors: Charles Nolan

The aim is to study the main transmission channels of oil price fluctuations for the UK economy and the consequences of oil price changes on its public finances. Our results show that the underlying source of oil price shocks matters. Accordingly, UK public finances deteriorate in response to positive shocks to foreign oil intensity and supply that induce a fall in the oil price. In such cases, the reductions in the VAT, fuel duty and PRT tax receipts cause the rise in the government debt. Conversely, the real oil price remains almost unchanged following positive shocks to domestic oil production but, in this case, the UK public finances are positively affected because the main tax revenues increase. In contrast to existing studies we find that the variation of the UK government debt is significantly affected by changes in the PRT revenues following domestic oil production shocks. We also extend the previous literature by estimating the parameters of several fiscal policy rules related to the major UK tax rates.

CO1460: **Mild explosivity in recent crude oil prices**
Presenter: **Isabel Figuerola-Ferretti**, ICADE, Spain
Co-authors: Roderick McCrorie, Ioannis Paraskevopoulos

The new, mildly explosive/multiple bubbles technology is used to assess whether crude oil prices over the last decade have exhibited departures from martingale trend behavior and to explore whether any such departures indicate divergence from fundamental value. The test dates two significant time periods in both Brent and WTI front-month futures: a mildly explosive episode within the 2007-08 oil price spike, immediately prior to the peak of the Global Financial Crisis; and a negative such episode during the recent oil price decline, whose commencement is dated around a key OPEC meeting in November 2014. Evidence using other commodity prices and indices, themselves declining recently, points to factors beyond commodity markets. We find the CBOE Volatility Index (VIX), a financial variable which acts as an omnibus gauge of market-based expectations, echoes both departures in trend; however there is no one factor among fundamental proxy variables such as global economic activity or inventories, or other financial variables, which is itself decisive in explaining the recent price decline.

CC1767: **Potential output of net oil exporters after the boom**
Presenter: **Samya Beidas-Strom**, International Monetary Fund, United States

Oil prices have declined sharply over the past three years, and output growth has slowed considerably among countries that are net oil exporters. A critical question for policy makers in these economies is whether oil windfalls influence potential output. Our analysis suggests that both actual and potential output move together with oil terms of trade, but that actual output co-moves twice as strongly as potential output. The weak oil price outlook is estimated to have subtracted 1 to 2 percentage points from actual output growth annually on average during 2015 to 2017. The forecast drag on potential output is about one-third of that for actual output.

CO579 Room 101 NOWCASTING AND FORECASTING AT CENTRAL BANKS Chair: Knut Are Aastveit

CO0584: **Disaggregated model-based inflation forecasts: A univariate approach**
Presenter: **Jan-Oliver Menz**, Deutsche Bundesbank, Germany
Co-authors: Thomas Goetz
In the context of the Narrow Inflation Projection Exercise the national central banks of the Euro area are required to deliver various short-term inflation forecasts to the European Central Bank. Apart from the Harmonized Index of Consumer Prices these are particularly five ECB special aggregates (unprocessed food, processed food, energy, industrial goods ex energy and services), which are derived from the roughly 90 COICOP 4-digit-components, the currently deepest available disaggregation level. Naturally, several disaggregation levels and schemes could be considered in-between predicting the target series directly and aggregating forecasts obtained on the deepest disaggregation level. One example is a set of so-called NIPE-aggregates being used in the Deutsche Bundesbank. We aim at systematically assessing which level of disaggregation is optimal for predicting the special aggregates. Therefore we first analyse which of several univariate models (among which are a UCM-SV- and a time-varying AR model) is optimal for predicting which target series for each available aggregation level. We do so by conducting an extensive out-of-sample pseudo-real-time forecast exercise involving over 120 price series for Germany. Given the insights gained in the first step, we then propose a data-driven selection of disaggregate price series in order to obtain a modified set of NIPE-aggregates, which to base our disaggregate inflation forecasts on.

CO0760: **How the central banks reaction function in small open economies evolved during the crisis**
Presenter: **Aleksandra Halka**, Narodowy Bank Polski, Poland
The experience of the global financial crisis caused central banks to change the way the monetary policy is conducted, in particular it changed their reaction function. We investigate how four selected European central banks in small open economies have changed their reaction function in response to the GFC. To address this problem a logit model is used to see, firstly, how the relative importance of GDP growth forecasts in the process of setting interest rates evolved over time, secondly, how the CPI forecast horizon which central banks take into consideration has changed and, finally, how the monetary policy stance has changed. The outcomes indicate that all banks in the course of the crisis have become more flexible in the way they conduct monetary policy - Polish and Hungarian central banks increased the relative importance of GDP growth as compared to inflation development, Czech and Swedish central banks increased the forecast horizon for inflation and all but the Swedish central bank started to conduct a more accommodative monetary policy.

CO1093: **Exchange rate predictability and model incompleteness**
Presenter: **Knut Are Aastveit**, Norges Bank, Norway
Co-authors: Herman van Dijk, Francesco Ravazzolo
It is well known that exchange rate fluctuations are very difficult to predict using economic models, and that a random walk forecasts exchange rates better than any economic model (the Meese and Rogoff puzzle). The recent literature has identified a series of fundamentals/ methodologies that claim to have resolved the puzzle. However, although these studies find predictors that provide exchange rate predictability for some countries at specific time periods and specific forecast horizons, typically the relationships, and thus the predictability, are unstable and break down over time. We employ a combined density forecasting framework that allows us to pin down several sources of instability that might affect the out-of-sample forecasting performance of exchange rate models. The latent weights of the combination scheme depend on past forecasting performance and other learning mechanisms and allows for model incompleteness. We show that allowing for these features systematically improve upon the random walk benchmark in an out-of-sample forecasting exercise.

CO581 Room Board meeting room I FINANCIAL NETWORKS Chair: Marco Petracco

CO0588: **Risk and the credit default swap market**
Presenter: **Tuomas Peltonen**, European Systemic Risk Board, Germany
Co-authors: Marco Derrico, Stefano Battiston, Martin Scheicher
A Credit Default Swap (CDS) is a contract built upon the transfer of underlying risk and the creation of counterparty risk. We provide theoretical and empirical results about the complex network arising from the set of risk transfers in the CDS network. First, we introduce the notion of flow-of-risk and provide sufficient conditions fora bow-tie network architecture to endogenously emerge as a result of intermediation. We show that the probability of a widespread systemic loss due to counterparty risk is higher in a bow-tie architecture than in more fragmented network structures. Empirically, we use a unique global dataset of CDS bilateral exposures on major sovereign and financial reference entities in the years 2011 - 2014. We find that the CDS network shows precisely a bow-tie network architecture for the vast majority of reference entities. The flow-of-risk originates from a large number of Ultimate Risk Sellers (URSs) and ends up in a very few leading Ultimate Risk Buyers (URBs), the largest majority of which are non-banks. The analysis of the CDS portfolio composition of the URBs shows a high level of concentration in the market; in particular, the top URBs often show large exposures to potentially correlated reference entities.

CO0818: **Identifying the systemically important financial communities through the weighted stochastic block model**
Presenter: **Michele Costola**, SAFE, Goethe University Frankfurt, Germany
Co-authors: Roberto Casarin, Erdem Yenerdag
Systemic risk and contagion channels of financial markets are analyzed proposing the stochastic block models (SBM) as generative models for the financial networks and exploiting the topological features of its communities. Furthermore, they propose a ranking method to identify the systemically important communities (SIC), which naturally represent the groups of financial institutions playing a central role in the vulnerability of a financial system. The empirical SBM analysis is performed on the European financial firms (active and dead) selected by the World Scope lists at daily frequency from December 1995 to January 2013. Finding shows that the number of communities tends to increase during the financial crisis which suggests the potential role of SBM communities as a new tool to be monitored by central authorities for early warning purposes.

CO1146: **Multiplex interbank networks and systemic importance: An application to European data**
Presenter: **Ivan Alves**, European Central Bank, Germany
Co-authors: Ignacio Aldasoro
Research on interbank networks and systemic importance is starting to recognise that the web of exposures linking banks balance sheets is more complex than the single-layer-of-exposure approach. We use data on exposures between large European banks broken down by both maturity and instrument type to characterise the main features of the multiplex structure of the network of large European banks. This multiplex network presents positive correlated multiplexity and a high similarity between layers, stemming both from standard similarity analyses as well as a core-periphery analyses of the different layers. We propose measures of systemic importance that fit the case in which banks are connected through an arbitrary number of layers (be it by instrument, maturity or a combination of both). Such measures allow for a decomposition of the global systemic importance index for any bank into the contributions of each of the sub-networks, providing a useful tool for banking regulators and supervisors in identifying tailored policy instruments. We use the dataset of exposures between large European banks to illustrate that both the methodology and the specific level of network aggregation matter in the determination of interconnectedness and thus in the policy making process.

CO449 Room 105 DATA ANALYTICS FOR FINANCIAL AND INSURANCE RISK MANAGEMENT Chair: Gareth Peters

CO1391: **Multi-curve interest rate modelling and inflation-linked pricing**
Presenter: **Andrea Macrina**, University College London, United Kingdom
Co-authors: Henrik Dam, Obeid Mahomed, David Skovmand
The valuation and management of risk exposures often leads to the question of how to model spreads. Multi-curve discounting, for instance, relies on modelling the risk exposures manifested in the spreads between the curves linked to the tenors of interbank offer rates and an overnight index swap rate. Another example is inflation-linked pricing and hedging where one of the main ingredients is a model for the consumer price index, which we view as a spread between the nominal and the real value of goods and services. The list of examples may also include credit and foreign exchange, and certain aspects relevant to asset valuation in emerging markets also appear to pose questions revolving around spread modelling. We consider stochastic models for interest rates and inflation-linked pricing, which share in common an approach that offers good malleability and facilitates financial understanding while keeping mathematical complexity low.

CO1708: **On risk factors which drive oil futures price curves: Speculation and hedging in the short-term and long-term**
Presenter: **Guillaume Bagnarosa**, ESC Rennes, France
Co-authors: Gareth Peters, Matthew Ames, Pavel Shevchenko, Tomoko Matsui
A consistent estimation framework is developed, which builds on the familiar two-factor model, to allow for an investigation of the influence of observable covariates on the term structure of WTI crude oil futures prices. The proposed framework incorporates observable covariates, such as inventories, production or hedging pressure, into the spot price dynamics. This novel approach builds on recent literature exploring post model fit regressions of convenience yield on covariates to provide a model with a number of key advantages. In particular, we are able to assess the influence of the covariates at any point along the term structure. From a risk management perspective, it is straightforward with our model to conduct stress testing of the futures curve to shocks in the covariates. From a speculative trading perspective, if one is able to forecast the covariates with some degree of accuracy then our model could be very useful in forecasting and profiting from changes in the futures curve.

CO1749: **A rational model for inflation**
Presenter: **David Glavind Skovmand**, University of Copenhagen, Denmark
Co-authors: Henrik Dam, David Sloth Pedersen
Stochastic models for interest rates and inflation-linked pricing are considered, which share in common an approach that offers good malleability and facilitates financial understanding while keeping mathematical complexity low. The approach is built on using the so-called rational framework for constructing both the nominal and real term structures. The approach enables analytic expressions for linear and some nonlinear inflation linked derivatives.

CO417 Room 111 ECONOMETRIC ANALYSIS OF COMMODITY AND ENERGY MARKETS Chair: Helena Veiga

CO0189: **On the volatility of commodity prices and the macroeconomic uncertainty**
Presenter: **Marc Joets**, Banque de France and University of Paris, France
While there exists numerous studies on the macroeconomic effects of oil and commodity shocks, the literature is quite silent on the impact of macroeconomic uncertainty on oil and commodity prices and, especially, on their volatility. The aim is to tackle this issue through the estimation of a structural threshold vector autoregressive (TVAR) model on a sample of 19 commodity markets. We aim at (i) assessing whether the effect of macroeconomic uncertainty shocks on commodity price returns depends on the degree of uncertainty, and (ii) investigating the transfer from macroeconomic uncertainty to price uncertainty using a newly developed measure of commodity price uncertainty. Our findings show that both agricultural and industrial markets are highly sensitive to the variability and the level of macroeconomic uncertainty, while the impact on precious metals is more parsimonious given their well-identified safe-haven role in time of economic turmoil. In addition, we find evidence that the recent 2007-09 recession has generated an unprecedented episode of high uncertainty in numerous commodity prices. Interestingly, our analysis further reveals that volatility and uncertainty in prices can be disconnected.

CO0738: **Commodities common factor: An empirical assessment of the markets drivers**
Presenter: **Johannes Luebbers**, TU Dortmund, Germany
Co-authors: Peter N Posch
Applying a generalized dynamic factor model to the energy and raw material market a latent common factor in commodity futures returns is identified. This factor is driven by five market shocks and accounts for a major part of the total return variation. Before the global financial crisis the authors find commodity futures returns to be significantly exposed to an agricultural specific component and to an energy specific component afterwards which emphasizing the rising importance of the energy sector. This provides an indication of how a global change in the energy market affects the co-movement of commodities. The findings suggest to considering commodities as a heterogeneous asset class which is exposed to fundamental driving forces.

CO1072: **Energy industry's market value and oil price**
Presenter: **Sofia Ramos**, ESSEC Business School, France
Co-authors: Helena Veiga
The long-run equilibrium between the market value of energy firms and oil prices is analyzed. Using a sample of industry indexes from Canada, France, Japan, the UK and the US, we find that UK oil producers and US oil integrated industry firms are jointly cointegrated with oil and stock market values. We find cointegration with the stock market value for French integrated firms and UK and US renewable energy firms. Further analysis support causality relations between oil and stock market changes and the value of energy firms. For the others industry indexes of our sample, we only find short term effects from shocks of oil, stock market prices and exchange rates. For renewable energy firms the effects of oil shocks are mixed that nevertheless are not statistically significant. The model provides us a long-term relationship that is then tested outsample, presenting a good fit, in particular for US integrated firms. Overall, the results confirm that oil and stock markets are drivers of the market value of the UK and the US energy industry; for the other country-industries we find a puzzingly absence of relation with oil and stock markets. The results have implications for investors and managers of firms in energy sector.

CO583 Room 102 THE ECONOMETRICS OF INTERGENERATIONAL MOBILITY Chair: Emma Tominey

CO0749: **Measurement error and rank correlation**
Presenter: **Toru Kitagawa**, University College London, United Kingdom
Co-authors: Jan Stuhler, Martin Nybom
A method is proposed to correct measurement error biases in the estimation of rank correlation coefficients (Spearman's rho and Kendall's tau). We approximate the measurement error biases using the technique of small error variance approximation developed and construct a bias-corrected estimator by subtracting the bias estimate from the sample rank correlations. Using the rich Swedish data in which the underlying precise measurements are available, we empirically assess approximation performances of our method in the two empirical applications; rank correlations of father and son's permanent income and father and son's cognitive abilities.

CO0822: **Estimation of dynastic life-cycle discrete choice models**
Presenter: **Mehmet Soytas**, Ozyegin University, Turkey
Co-authors: George-Levi Gayle, Limor Golan
The estimation of a class of life-cycle discrete choice inter-generational models is explored. A new semi-parametric estimator is proposed. It is shown that it is root N-consistent and asymptotically normally distributed. We compare our estimator with a modified version of the full solution maximum likelihood estimator (MLE) in a Monte Carlo study. Our estimator performs comparably to the MLE in a finite sample but greatly reduces the computational cost. The quantity-quality trade-offs is shown to depend on the household composition and specialization in the household. Using the proposed estimator, we estimate a dynastic model that rationalizes these observed patterns.

CO0977: **Family income shocks and adolescent human capital**
Presenter: **Emma Tominey**, University of York, United Kingdom
Co-authors: Pedro Carneiro, Kjell Salvanes
How do shocks to parental income drive adolescent human capital, including years ofschooling, high school dropout, university attendance, IQ and health? A structural model decomposes household shocks into permanent and transitory components, then the effect of shocks at age 1-16 is estimated for 600,000 Norwegian children. The effect of permanent shocks declines - and of transitory shocks is small and constant across child age, suggesting parents optimise similarly to consumption. We explore possible insurance mechanisms available to households and find direct evidence of spousal income and indirect evidence of government transfers partially smoothing the effect of income shocks on child human capital.

CG532 Room 109 CONTRIBUTIONS IN DSGE MODELLING AND ESTIMATION Chair: Stephen Pollock

CC1420: **Housing markets - A DSGE analysis of the German case**
Presenter: **Chong Dae Kim**, Technical University of Cologne and Universitat Witten/Herdecke, Germany
Dynamic stochastic general equilibrium (DSGE) models have been a part of the policy making tool kits of central banks for many years. We show that the Iacoviello and Neri model, a closed economy DSGE model with a housing market, can be used as is for the German economy by comparing the results with independent historical observations and use this model to show the interaction between the German housing market and its economy. The data - the real price of housing in Germany between 1991 Q2 and 2014 Q4 - used show that housing prices have started rising significantly in 2010, as in Arhelger and Kim, the results however cannot confirm if this is due to the creation of a bubble in the German housing market. The results show that a monetary policy shock has only a short term effect on housing prices. Effects of the shock dissipate within two years of impact. Housing demand however, due to overheating in the construction sector as a result of a large amount of subsidies in the post reunification period, plays a major role in housing prices and investment, with a strong negative influence on both.

CC1014: **The dimension of the set of causal solutions of linear multivariate rational expectations models**
Presenter: **Bernd Funovits**, University of Helsinki, Finland
The aim is to analyze the number of free parameters and solutions of the structural difference equation obtained from a linear multivariate rational expectations model. The theory of polynomial matrix factorization is used to amend errors previous theorems. Stochastic singularity is analyzed and certain restrictions on the solution set are introduced in order to compare the approach with the most popular solution methods. This model is the basis for identifiability analysis of determinate as well as indeterminate equilibria, allowing for singular spectral densities and without assuming minimality of the associated state space system.

CC1660: **Exploiting special features of a new series representation for non-linear stochastic dynamic models**
Presenter: **Gary Anderson**, Federal Reserve Board, United States
It has been developed a new series representation useful for solving a wide class of nonlinear dynamic stochastic models. The solutions computed by the technique accommodate the possibility that model trajectories can depart from and re-engage occasionally binding constraints as well as transition between various regimes. Remarkably, this series representation makes it possible to characterize model solutions as a linear sum of orthogonal functions. We report on the efficiency gains associated with exploiting this linearity and with parallelizing the computation of the function approximations.

EO614 Room 208 RECENT ADVANCES IN STATISTICAL ANALYSIS OF HUMAN BRAIN DATA Chair: Tingting Zhang

EO0154: **Bayesian inference for directional brain networks modeled by damped harmonic oscillators using ECoG data**
Presenter: **Tingting Zhang**, University of Virginia, United States
A high-dimensional ordinary differential equation (ODE) model is proposed to explore connectivity among multiple small brain regions. The new model, called modular oscillator dynamic directional model (MODDM), captures the oscillatory behavior of the temporal activity of the brain while featuring a cluster structure consisting of modules of densely connected brain regions. We develop a unified Bayesian framework for quantifying uncertainty or the inadequacy in the assumed ODE model from the underlying truth, identifying clusters, selecting strongly connected brain regions, and making inferences about the MODDM. The prior distributions in the Bayesian model for MODDM parameters are carefully designed such that the ensuing joint posterior distributions for ODE state functions and the MODDM parameters have well-defined and easy-to-simulate posterior conditional distributions. To further speed up the posterior simulation, we employ parallel computing schemes in two Markov Chain Monte Carlo steps. We show that the proposed Bayesian approach outperforms an optimization-based method for ODE estimation through simulation studies. We use the MODDM to study the auditory function of the brain using an electrocorticography dataset.

EO0264: **Multiscale information extraction of structure connectivity of the brain**
Presenter: **Jie Peng**, University of California Davis, United States
A major limitation of most diffusion MRI data analysis is that the spatial scale is determined a priori and is fixed throughout the analysis. We will introduce a computationally efficient, bottom-up representation of structural connectivity of the brain based on D-MRI data such that coherent

neuronal fiber directionality is modeled over arbitrary-sized spatial neighborhoods. Particularly, we will derive new features that capture structural connectivity at different spatial scales which are easy to be related to external variables (such as age, cognitive measures) in statistical models.

EO0387: **Estimating information flow in large brain networks via pathway lasso**
Presenter: **Xi Luo**, Brown University, United States
Co-authors: Yi Zhao
The brain can be conceptualized as a dynamic network of connected nodes, and information, such as external stimuli, is processed while passing through series of nodes that form pathways. This provides a foundational model for the spatial-temporal processes in the brain. Task fMRI data are used to study the problem of estimating information flow in large brain networks. We model such an information flow as dynamic weights on each edge of the network. One challenge is that the number of pathways between the source and target nodes grows exponentially with the number of the nodes in the network. To this challenge, we develop a large-scale structural equation model, and we propose a convex optimization approach to infer the model parameters. Our approach enjoys the following advantages. It relaxes the original non-convex pathway-search problem to a convex one with an innovation on a new penalty formulation that selects the major pathways, which yields improved pathway selection and estimation accuracies. The numerical merits are illustrated using simulated data and a real fMRI dataset.

EO563 Room 207 RECENT ADVANCES IN NONPARAMETRIC AND SEMIPARAMETRIC INFERENCE Chair: Zheyu Wang

EO0222: **A semiparametrically efficient estimator of the time-varying effects for survival data with time-dependent treatment**
Presenter: **Huazhen Lin**, Southwestern University of Finance and Economics, China
The timing of time-dependent treatment (e.g., when to perform kidney transplantation) is an important factor for evaluating treatment efficacy. A naive comparison between the treatment and nontreatment groups, while ignoring the timing of treatment, typically yields results that might biasedly favor the treatment group, as only patients who survive long enough will get treated. On the other hand, studying the effect of time-dependent treatment is often complex, as it involves modeling treatment history and accounting for the possible time-varying nature of the treatment effect. We propose a varying-coefficient Cox model that investigates the efficacy of time-dependent treatment by utilizing a global partial likelihood, which renders appealing statistical properties, including consistency, asymptotic normality and semiparametric efficiency. Extensive simulations verify the finite sample performance, and we apply the proposed method to study the efficacy of kidney transplantation for end-stage renal disease patients in the U.S. Scientific Registry of Transplant Recipients (SRTR).

EO0342: **Nonparametric Bayesian learning of heterogeneous dynamic transcription factor networks**
Presenter: **Xiangyu Luo**, The Chinese University of Hong Kong, China
Co-authors: Yingying Wei
Gene expression is largely controlled by transcription factors (TFs) in a collaborative manner. Therefore, understanding TF collaboration is crucial for elucidating gene regulation. The co-activation among TFs can be represented by networks. These networks are dynamic over diverse biological conditions and heterogeneous across the genome within each biological condition. Existing methods for constructing TF networks lack solid statistical models, analyze each biological condition separately, and enforce a single network for all the genome positions within one biological condition, which suffer from low statistical power and result in misleading spurious association. We present a novel Bayesian nonparametric dynamic Poisson graphical model for inference on heterogeneous dynamic TF networks. Our approach automatically teases out the genome heterogeneity and borrows information across conditions to improve signal detection, thus offering a valid and efficient measure of TF co-activations. An efficient parallel Markov Chain Monte Carlo algorithm is developed for posterior computation. The proposed approach is applied to study TF associations in ENCODE cell lines.

EO1353: **Structural equation models with mixed continuous and partially ordered data**
Presenter: **Xiaoqing Wang**, The Chinese University of Hong Kong, China
Co-authors: Xiangnan Feng, Xinyuan Song
Structural Equation modeling (SEM) is well recognized in many disciplines as the most important multivariate technique for assessing the interrelationships among latent variables that are grouped based on correlated observable variables. Owing to the questionnaire design and problem nature, discrete categorical data, such as ordered data, unordered data, and partially ordered data, are routinely collected in social, medical and behavioural research. Among literature, substantial efforts have been devoted on investigating SEM with ordered and unordered categorical variables, whereas litter attention has been focused on the analysis of correlated partially ordered data under the SEM framework. To fill the gap, we propose a general class of SEM that is capable of operationalizing latent variables through mixed continuous and partially ordered variables. Partially ordered set theory and the specification of the model are discussed. Bayesian procedures implemented through Markov Chain Monte Carlo algorithms are developed for statistical inference. Extensive simulation studies demonstrate that the developed methodology enjoys satisfactory performance.

EO027 Room 211 RECENT ADVANCES IN FAILURE TIME DATA ANALYSIS Chair: Jianguo Sun

EO0291: **A modified log rank test for interval censored competing risk data**
Presenter: **Yang-Jin Kim**, Sookmyung Women University, Korea, South
A competing risk analysis has been applied when a patient fails by one of several causes. Then available data include information about a failure time and a failure cause. Gray suggested a log rank test with a weighted risk process for right censored competing risk data. Our interest is to suggest a test statistics with an interval censored competing risk data. The available data is two sequential inspection times embracing a failure time and a cause indicator. We proposed a modified log rank test by combining Sun's test for interval censored data and Gray's test for competing risk. Therefore, the suggested test reduces to Sun's test statistic under one cause failure. The suggested method is applied to AIDS dataset and some simulation results are presented to evaluate the performance.

EO0292: **Regression analysis with K interval-censored failure time data in presence of informative censoring**
Presenter: **Peijie Wang**, Jilin University, China
Co-authors: Hui Zhao, Jianguo Sun
Interval-censored failure time data occur in many fields such as demography, economics, medical research and reliability and many inference procedures on them have been developed. However, most of the existing approaches assume that the mechanism that yields interval censoring is independent of the failure time of interest and it is clear that this may not be true in practice. We consider regression analysis with K interval-censored failure time data when the censoring mechanism may be related to the failure time of interest. An estimated sieve maximum likelihood approach is proposed for the data arising from the proportional hazards frailty model and for estimation, a two-step procedure is presented. In the addition, the asymptotic properties of the proposed estimators of regression parameters are established and an extensive simulation study suggests that the method works well. Finally, we apply the method to a set of real interval-censored data that motivated this study.

EO0400: **Joint modeling on multivariate longitudinal and recurrent event data and its application on urea cycle disorder**
Presenter: **DoHwan Park**, Univ. ov Maryland – Baltimore County, United States

Semiparametric joint models are proposed to analyze multivariate longitudinal data with recurrent events. An proportional hazard with frailty model for a cumulative intensity is used for the recurrent events. Multivariate linear mixed model with random effect is used to accommodate association for longitudinal outcomes. We proposed to estimate all the parameters using the nonparametric maximum likelihood estimators. Properties of the parameter estimates are studied. We examine the performance of our proposed method through simulation studies. The method are then applied to analyze an urea cycle disorder patients' hyperammonemia episodes.

EO109 Room 203 SOME NEW DEVELOPMENT IN FUNCTIONAL DATA ANALYSIS Chair: Kehui Chen

EO0405: **Testing equality of surface mean functions in longitudinal functional data**
Presenter: **Jin Yang**, The Hong Kong polytechnic University, China
Co-authors: Catherine Liu, Tao Zhang

Two method of two sample testing are proposed for dependent functional data. Specifically, we consider the setting of functional data with repeatedly observed where curves are recorded repeatedly for each subject. The asymptotic null distribution of the test statistic is developed under mild regularity conditions. It indicates that the proposed method has a nice performance both in terms of size and power. The approaches are illustrated through the analysis of longitudinal mortality data from period lifetables that are repeatedly observed for a sample of countries over many years and environmental pollution data from different environmental control station.

EO0362: **Functional data analysis using factor model with Toeplitz error structure**
Presenter: **Chi Tim Ng**, Chonnam National University, Korea, South

A factor model with errors generated from independent Gaussian stochastic processes is proposed for the functional data analysis. In the special cases where the observations are evenly-spaced, the covariance matrix is a Toeplitz matrix. The applications of the proposed model in the climate data analysis are discussed.

EO1527: **Nonlinear mixed-effects scalar-on-function models and variable selection for kinematic upper limb movement data**
Presenter: **Yafeng Cheng**, MRC Biostatistics Unit, United Kingdom
Co-authors: Jian Qing Shi, Janet Eyre

The motivation was to model clinical assessments of upper limb function after stroke using 3D kinematic data. We present a new nonlinear mixed-effects scalar-on-function regression model with a Gaussian process prior focusing on variable selection from large number of candidates including both scalar and function variables. A novel variable selection algorithm has been developed, namely functional least angle regression (fLARS). As they are essential for this algorithm, we studied the representation of functional variables with different methods and the correlation between a scalar and a group of mixed scalar and functional variables. We also propose two new stopping rules for practical usage. This algorithm is able to do variable selection when the number of variables is larger than the sample size. It is efficient and accurate for both variable selection and parameter estimation. Our comprehensive simulation study showed that the method is superior to other existing variable selection methods. When the algorithm was applied to the analysis of the 3D kinetic movement data the use of the non linear random-effects model and the function variables significantly improved the prediction accuracy for the clinical assessment.

EO461 Room 215 MATRIX METHODS IN MULTIVARIATE MODELS Chair: Katarzyna Filipiak

EO0722: **Inference for the growth curve model with orthogonal covariance structure**
Presenter: **Miguel Fonseca**, NOVA.id.FCT, Portugal
Co-authors: Martin Singull

The growth curve model is a well-documented multivariate model, with a well-established methodology of maximum likelihood estimation. We propose a growth curve model family with an orthogonal covariance structure for lines and columns, proceeding with the derivation of maximum likelihood statistics. Many familiar models fall within this model family, as it will be shown.

EO1222: **Mean testing under orthogonal covariance structure**
Presenter: **Daniel Klein**, P.J. Safarik University, Slovakia
Co-authors: Martin Singull

Covariance matrices with different known structures arise in a variety of contexts and have been studied by many authors. Statistical test criteria for testing equality in means assuming compound symmetry or block compound symmetry covariance structure has been developed using the likelihood ratio principle as well as the modification of Hotelling's T^2. The compound symmetry structure belongs to the class of orthogonal structures. It will be discussed and compared several test statistics for testing the mean assuming orthogonal structure for the covariance matrix.

EO0783: **Using matrix derivatives in influential analysis**
Presenter: **Dietrich von Rosen**, Swedish University of Agricultural Sciences, Sweden

A general approach to identify influential observations in multivariate linear models is presented. The main idea is to perturb models which are evaluated via Taylor expansions including matrix derivatives. These derivatives can be used as measures of influence. In particular bilinear models such as the growth curve model and its extensions are discussed.

EO141 Room 204 ROBUST HIGH-DIMENSIONAL/FUNCTIONAL DATA ANALYSIS Chair: Stefan Van Aelst

EO0741: **Robust PCA in the presence of outlying cells**
Presenter: **Wannes Van den Bossche**, KU Leuven, Belgium
Co-authors: Mia Hubert, Peter Rousseeuw

Principal component analysis (PCA) is a popular dimension reduction technique that is typically used as a first step when exploring high-dimensional data. It is known that classical PCA is highly sensitive to outliers. Several robust alternatives have been developed which yield accurate loadings in the presence of outliers. However, these methods consider outliers to be entire rows of the data matrix while it often happens that only a few cells in a row are outlying. Downweighting an entire row then leads to an unnecessary loss of information. Furthermore, in high-dimensional data it could easily happen that more than half of the rows contain such cellwise outliers which causes the current rowwise robust PCA methods to break down. We introduce a new method for robust principal component analysis which can handle cellwise outliers. In addition it provides estimates for deviating data cells, as well as for missing values. The code will be made available in the Matlab toolbox LIBRA and in R. Some real data examples will be shown.

EO0651: **Finding outliers in image data and video**
Presenter: **Jakob Raymaekers**, KULeuven, Belgium
Co-authors: Mia Hubert, Peter Rousseeuw
Images and video can be considered as functional data with a bivariate domain, where the data per grid point can be univariate (e.g. grayscale values) or multivariate (e.g. red, green, and blue intensities). This often yields large datasets, in which outliers may occur that can distort the analysis. At each grid point we propose to compute a fast measure of outlyingness which accounts for skewness in the data. It can be used for univariate data and, by means of projection pursuit, for multivariate data. The influence function of this outlyingness measure is computed. We also construct a cutoff value for the outlyingness. Heatmaps of the outlyingness indicate the regions in which an image deviates most from the majority of images. To illustrate the performance of the method, it is applied to real multivariate functional data. One example consists of MRI images which are augmented with their gradients. We also show an example of video surveillance data, where we compare the exact method with faster approximations.

EO1445: **Robust and sparse classification by the optimal scoring approach**
Presenter: **Irene Hoffmann**, Vienna University of Technology, Austria
Co-authors: Peter Filzmoser, Christophe Croux
In supervised classification problems observations belong to one of K different classes. The optimal scoring approach transforms these K class labels into continuous response variables and thereby enables a propagation of regression methods to the classification framework. Resent developments in regression analysis are used to propose a robust and spares classifier based on optimal scoring, which is also applicable if the number of variables exceed the number of observations. The robust sparse optimal scoring (RSOS) algorithm transforms the data into a low dimensional subspace, which is obtained by linear combinations of a subset of the original variables, aims for optimal group separation and is not distorted by outliers. An L1 norm penalty leads to intrinsic variable selection which excludes uninformative variables and thereby improves the model precision and its interpretation. Then the classification rule is defined by robust Mahalanobis distances to the group centers in the low dimensional subspace. Simulation studies and real data examples illustrate the properties of the proposed method.

EO549 Room 209 HIGH DIMENSIONAL DATA ANALYSIS Chair: Zuofeng Shang

EO0772: **Projection under pairwise distance control**
Presenter: **Hiba Alawieh**, University of Lille 1, France
Co-authors: Nicolas Wicker, Christophe Biernacki
Visualization of high-dimensional and possibly complex (non continuous for instance) data onto a low-dimensional space may be difficult. Several projection methods have been already proposed for displaying such high-dimensional structures on a lower-dimensional space, but information lost on initial data is not always easy to use. A new projection paradigm is presented to describe a non-linear projection method that takes into account the projection quality of each projected point in the reduced space, this quality being directly available in the same scale as this reduced space. More specifically, this novel method allows a straightforward visualization data in $\mathbb{R}^2$ with a simple reading of the approximation quality, and provides then a novel variant of dimensionality reduction.

EO0885: **Difference-based estimators of (autoco)variance in nonparametric models with a discontinuous signal**
Presenter: **Inder Tecuapetla-Gomez**, University of Goettingen, Germany
A class of difference-based estimators for the (autoco)variance in nonparametric regression is discussed for the case when the signal is discontinuous (change-point regression), possibly highly fluctuating, and the errors form a stationary m-dependent process. These estimators circumvent the explicit pre-estimation of the unknown regression function, a task which is particularly challenging for these signals. We derive their finite sample mean squared errors when the signal function is piecewise constant (segment regression) and the errors are Gaussian; based on this we derive biased-optimized estimates which do not depend on the particular (unknown) autocovariance structure. Notably, for positively correlated errors, that part of the variance of our estimators which depends on the signal is minimal as well. Asymptotic properties of these estimators in this and more general change-point models will be discussed as well. Simulation studies and an application to biophysical measurements will be shown.

EO1712: **High-dimensional consistency in score-based and hybrid structure learning**
Presenter: **Marloes Maathuis**, ETH Zurich, Switzerland
The main approaches for learning Bayesian networks can be classified as constraint-based, score-based or hybrid methods. Although high-dimensional consistency results are available for the constraint-based PC algorithm, such results have been lacking for score-based and hybrid methods, and most hybrid methods are not even proved to be consistent in the classical setting where the number of variables remains fixed. We study the score-based Greedy Equivalence Search (GES) algorithm, as well as hybrid algorithms that are based on GES. We show that such hybrid algorithms can be made consistent in the classical setting by using an adaptive restriction on the search space. Moreover, we prove consistency of GES and adaptively restricted GES (ARGES) for certain sparse high-dimensional scenarios. ARGES scales well to large graphs with thousands of variables, and our simulation studies indicate that both ARGES and GES generally outperform the PC algorithm.

EO193 Room 206 STATISTICS IN ECOLOGY Chair: Eva Cantoni

EO0803: **Efficient model-fitting via a semi-complete data likelihood approach**
Presenter: **Ruth King**, University of Edinburgh, United Kingdom
Many ecological models lead to analytically intractable likelihoods, where the observed data likelihood is expressible only as an integral. In these circumstances a common approach has been to apply a Bayesian complete data likelihood approach, using a data augmentation technique. This often permits the models to be fitted within standard Bayesian packages, such as BUGS/JAGS. However, in many applications, this approach leads to poor mixing, and hence performance, of the Markov chain Monte Carlo algorithm. We propose an alternative semi-complete data likelihood approach. The idea is to combine a complete data likelihood component with associated auxiliary variables; with an observed data likelihood component expressible as a low-dimensional integral that can be easily approximated via standard numerical integration techniques. We demonstrate the improved computational improvement using this semi-complete data likelihood applied to real applications, and fitted within the standard Bayesian package JAGS.

EO1161: **Robust fitting of state-space models for reliable fish stock assessment**
Presenter: **William Aeberhard**, Dalhousie University, Canada
Co-authors: Joanna Mills Flemming, Chris Field, Eva Cantoni, Ximing Xu
The sustainable management of fisheries strongly relies on the output of fish stock assessment models fitted to scarce and noisy data. State-space models represent a relevant general framework for accounting for both measurement error and a complex dependence structure of latent (unobserved) random variables. Classical estimation of fixed parameters in such models, for instance by maximizing an approximated marginal likelihood, is known to be highly sensitive to the correct specification of the model. This sensitivity is all the more so problematic since assumptions about latent variables cannot be verified by the data analyst. We introduce robust and consistent estimators for general state-space models which

remain stable under deviations from the assumed model. These estimators are shown to yield reliable inference for fish stock assessment in various scenarios.

EO1125: **Modelling animal movement with state-space models**
Presenter: **Joanna Mills Flemming**, Dalhousie University, Canada
Animals move in order to maximize their probability of survival and reproduction. The movement of an animal therefore reflects its response to its current physical needs and available environment. In the marine realm, where direct observation of animal movements is often impossible, researchers typically employ satellite telemetry positioning systems to obtain series of estimates of locations of animals in space through time. Each series resembles an animal path or track. State-space models represent an ideal framework for accounting for both measurement error and process error in these tracks. We discuss novel formulations that allow us to model these tracks robustly and at the same time infer animal behavioral states. Information gained from these efforts is currently being used for proper management of both species and ecosystems as will be evident from the applications presented.

EO165 Room 205 TOPOLOGICAL DATA ANALYSIS Chair: Pedro Teran

EO1103: **Hypothesis testing with persistent homology of the large-scale structure of the universe**
Presenter: **Jessi Cisewski**, Yale University, United States
The large-scale structure (LSS) of the Universe is an intricate and spatially complex web. In order to understand the physics of the Universe, theoretical and computational cosmologists develop large-scale simulations that allow for visualizing and analyzing the LSS under varying physical assumptions. However, rigorous comparisons and inference on such complicated structures can be problematic. We present a framework for hypothesis testing of LSS using persistent homology. The randomness in the data (due to measurement error or topological noise) is transferred to randomness in the topological summaries, which provides an infrastructure for inference. These tests allow for statistical comparisons between complicated spatial data such as LSS in cosmology, but are also present in other areas of science. We present several possible test statistics using persistence diagrams, carryout a simulation study to investigate the suitableness of the proposed test statistics, and finally we apply the proposed inference framework to WDM vs. CDM cosmological simulation data.

EO1547: **Subsampling methods for persistent homology**
Presenter: **Frederic Chazal**, INRIA, France
Computational topology has recently seen an important development toward data analysis, giving birth to Topological Data Analysis. Persistent homology appears as a fundamental tool in this field. It is usually computed from filtrations built on top of data sets sampled from some unknown (metric) space, providing "topological signatures" revealing the structure of the underlying space. When the size of the sample is large, direct computation of persistent homology often suffers two issues. First, it becomes prohibitive due to the combinatorial size of the considered filtrations and, second, it appears to be very sensitive to noise and outliers. We present a method to overcome these issues by computing persistent diagrams from several subsamples and combining them in order to efficiently infer robust and relevant topological information from data.

EO1607: **Topological consistency via kernel estimation**
Presenter: **Sayan Mukherjee**, Duke University, United States
Co-authors: Omer Bobrowski, John Taylor
A consistent estimator is introduced for the homology (an algebraic structure representing connected components and cycles) of level sets of both density and regression functions. Our method is based on kernel estimation. We apply this procedure to two problems: 1) inferring the homology structure of manifolds from noisy observations, 2) inferring the persistent homology (a multi-scale extension of homology) of either density or regression functions. We prove consistency for both of these problems. In addition to the theoretical results we demonstrate these methods on simulated data for binary regression and clustering applications.

EO698 Room 007 RECENT ADVANCES IN SEQUENTIAL MONTE CARLO AND RELATED METHODS Chair: Carles Breto

EO1233: **Anytime Monte Carlo**
Presenter: **Lawrence Murray**, University of Oxford, United Kingdom
Monte Carlo algorithms typically simulate some fixed number of samples, n, with the real time taken to do so a random variable, $T(n)$. For the purposes of real-time deadlines, particularly in a distributed computing context, an alternative is to fix the real time, t, and allow the number of samples drawn in this time to be a random variable, $N(t)$. Naive estimators constructed from these $N(t)$ samples are not necessarily consistent, however, and in general exhibit length bias with respect to compute time. A framework will be introduced for dealing with the length bias for both iid and Markov chain Monte Carlo samplers, and demonstrate the utility of the approach on a large scale sequential Monte Carlo deployment using the LibBi software on the Amazon EC2 cloud computing infrastructure.

EO1160: **Multilevel sequential Monte Carlo samplers**
Presenter: **Alexandros Beskos**, University College London, United Kingdom
Multilevel Monte Carlo methods provide a powerful computational technique for reducing the computational cost of estimating expectations for a given computational effort. They are particularly relevant for computational problems when approximate distributions are determined via a resolution parameter h, with $h = 0$ giving the theoretical exact distribution (e.g. SDEs or inverse problems with PDEs). The method provides a benefit by coupling samples from successive resolutions, and estimating differences of successive expectations. We develop a methodology that brings Sequential Monte Carlo (SMC) algorithms within the framework of the Multilevel idea, as SMC provides a natural set-up for coupling samples over different resolutions. We prove that the new algorithm indeed preserves the benefits of the multilevel principle, even if samples at all resolutions are now correlated.

EO0548: **Stability of the optimal filter in continuous time: Beyond the Benes filter**
Presenter: **Sylvain Rubenthaler**, Universite Nice Sophia Antipolis, France
Co-authors: Van Bien Bui
A case is exhibited in which the optimal filter is stable with respect to its initial condition. The computations are such that this result will allow us to prove the stability of an associated particle filter. The set of assumptions we propose is a neighborhood of the Bene filter. Observations in continuous time could be understood as "arbitrarily frequent observations". The proof is based on a decomposition which takes us back to optimal filters with observations which are not continuous in time.

EO650 Room 202 UNCERTAINTY, ERROR PROPAGATION AND QUALITY CONTROL OF SPATIAL DATA Chair: Jose Rodriguez-Avi

EO1243: **An alternative approach to spatial regression for count data: An application to Sierra Magina**
Presenter: **Jose Rodriguez-Avi**, University of Jaen, Spain
Co-authors: Maria Jose Olmo-Jimenez, Francisco Javier Ariza-Lopez
Spatial regression has been taken up by different approaches. When the response variable is a count data one, Poisson and Negative binomial have usually been proposed as underlying distribution of the regression model. Furthermore, different ways of introducing spatial effects have been considered. We propose the use of the generalized Waring regression model in order to explain count data variables in function of spatial and non-spatial covariates simultaneously. So, the response variable given the covariates is assumed to follow a univariate Waring distribution. Moreover, since data are obtained by counting items in a spatial domain which is composed by tesserae of the same size, the spatial covariates are introduced in the model as dummy variables, taking into account the neighbourhood relationships. This new approach allows us to consider other non-spatial covariates at the same time as spatial covariates and, in consequence, to analyse the possible interactions. To illustrate this methodology, we have studied some count data variables measured in the Natural Park of Sierra Magina and its surrounding area in the province of Jaen (Andalusia).

EO1278: **On the use of ϕ-divergences for controlling spatial point patterns**
Presenter: **Virtudes Alba-Fernandez**, University of Jaen, Spain
Co-authors: Francisco Javier Ariza-Lopez
The understanding of spatial point patterns is one of the major challenges of geographical analysis and has interest in many science. Spatial patterns and spatial statistical sampling is also a major issue in spatial data quality assessment because sampling is a very common procedure in order to derive estimates or perform tests. For these reasons the evaluation of the spatial similarity of two observed samples can be of interest in many cases. We develop a procedure to determine whether two spatial point patterns come from the same spatial point distribution by means of an area-based test centered on the counting of positional events. This method follows two steps; the first one is to make use of the space-filling curves as a tool to order the space, with independence of its dimension. The second one is to model the count of points in the resultant grid by means of the multinomial law. As a consequence, to test the similarity between two spatial point patterns is equivalent to test the homogeneity of two multinomial distributions. The homogeneity test of the resultant multinomial populations is carried out by using *phi*-divergence.

EC1340: **Uncertainty assessment in soil erosion risk maps**
Presenter: **Francisco Rueda-Valdivia**, University of Granada, Spain
Co-authors: Juan Francisco Reinoso-Gordo, Ana Diaz-Castro
Soil erosion is one of the most important of today's environmental problems world-wide. Not only it compromises the long-therm sustainability of agriculture, but it also poses serious problems related to sedimentation of waterways and eutrophication of water bodies. It is esteemed that reservoirs are losing one percent of their storage capacity annually. There is an increasing interest among policy-makers and environmental managers to quantify the risk of soil erosion, and to develop soil erosion risk maps, as decision support for the implementation of control measures. The Universal Soil Loss Erosion Equation USLE, is the base of most regional soil erosion maps. These methods are subject to a large degree of uncertainty, associated to (1) the equation does not have the level of complexity needed to represent all processes at play; and (2) the spatial variabilty of the factors controlling erosion rates. The values of these factors may not known but in a few points, hence, one needs to interpolate in space, an exercise which is not error-free. The predictions contained in soil erosion risk maps are uncertain, which may result in biased estimates and consequently in the adoption of a non-efficient erosion management strategies. Our goal is to propose some Generalized Likelihood Uncertainty Estimation procedures, to construct soil erosion risk maps in which spatially distributed estimates of erosion rates are provided together with the likelihood of those estimates.

EO137 Room 006 HEALTH ECONOMICS Chair: Michael Talias

EO1347: **Bayesian networks for the analysis of inpatient admissions**
Presenter: **Vincenzina Vitale**, Roma Tre, Italy
Co-authors: Caterina Conigliani, Tommaso Petitti
Inpatient admissions in different wards/clinics between 1998 and 2014 are considered. A Bayesian network (BN) structure is estimated, directly, from data in order to compute the joint probabilities of the different patient profiles as one of the main objective is to identify the most probable configurations of the wards/clinics. Knowing which wards/clinics are more interrelated could be useful for a better organization of the hospital. Once the Bayesian network is estimated, evidence for some nodes (in our case the history of a patient up to a certain stage) can be propagated through the graph, and the BN shows how such evidence changes the marginal distributions of the remaining nodes. Therefore, it is possible to predict in which ward/clinic there will be a next admission, assuming that there is one. Cross-validation are also performed to test the predictive ability of the BN.

EO1603: **Modelling unmeasured heterogeneity on different health systems using a Bayesian metafrontier framework**
Presenter: **Michael Talias**, Open University of Cyprus, Cyprus
Co-authors: Kostas Kounetas
The pursuit of efficiency has become a central objective of policy makers within health systems. Adopting a metafrontier framework on different health systems we account for unmeasured heterogeneity and technology gaps. Moreover, we consider the case of a directional distance function model for handling asymmetrically both desirable and undesirable outputs in the health care process. The above-mentioned technique has been implemented using a Bayesian procedure for heterogeneous set of world countries for the 1995-2013 period.

EO1359: **Efficiency measurement of healthcare units: A Bayesian DEA approach**
Presenter: **Panagiotis Zervopoulos**, University of Sharjah, United Arab Emirates
It is proven that only asymptotically Data Envelopment Analysis (DEA) yields unbiased efficiency estimators and that DEA estimators are upward biased for finite samples. DEA bootstrap methods are widely used to correct biases of DEA efficiency estimators. The combination of DEA and a modified bootstrap expression enhances the statistical properties of DEA estimators. However, major limitations, such as the asymptotic justification of the bootstrap technique and the overlapping confidence intervals of greatly diverse bootstrapped efficiency estimators, are still present. A Bayesian DEA method yields unbiased efficiency estimators for every sample size. Unlike extant Bayesian DEA studies, the new method does not use simulation techniques to estimate the posterior distribution but assumes a posterior beta distribution, which achieves the best fit to DEA efficiency estimators. The new Bayesian DEA method is applied to healthcare units operating in Greece and Cyprus.

EO616 Room 201 CHALLENGES IN ANALYSIS OF COMPLEX BIOMEDICAL DATA Chair: Ronglai Shen

EO1717: **A hierarchical hidden Markov random field model for peak calling across multiple Hi-C datasets**
Presenter: **Yun Li**, University of North Carolina, Chapel Hill, United States
The constantly accumulating Hi-C data provide rich information for calling peaks across multiple tissue/cell types, experimental conditions, and/or cell differentiation stages. However, statistical models and computational tools are still in their infancy. Multiple factors, including sequencing depth and heterogeneity across Hi-C experiments, pose great challenges for the development of proper and efficient methods. We propose a peak caller based on a hierarchical hidden Markov random field (HHMRF) model to detect long range chromatin interactions from multiple Hi-C datasets. In addition to model the spatial dependency in the local neighborhood, HHMRF is able to model dependency across multiple Hi-C datasets, leading to further improved statistical power. We conducted comprehensive simulation studies, and showed that HHMRF model outperforms competing methods that ignore the dependency structure and call peaks separately in each individual Hi-C dataset. Next, we analyzed a real Hi-C dataset on human H1 embryonic stem cells and four H1 derived cells, and found that the cell-type-specific peaks identified by HHMRF show higher overlap with cell-type-specific epigenetic features and cell-type-specific gene expression, compared to those identified by competing methods. HHMRF model has the potential to unveil the structural basis of cell-type-specific transcription regulation mechanism.

EO1718: **Integrative omics analyses using tensor decomposition and regularization**
Presenter: **Qi Long**, University of Pennsylvania, United States
Co-authors: Eun Jeong Min, Yijuan Hu
Advances in technology have led to generation of -omics data (such as genetics, transcriptomics, metabolomics data), which in some studies may be collected from multiple conditions (or cell types, tissue samples, disease states). Such data can be used to explore expression quantitative trait loci (eQTL) or metabolomic quantitative trait loci (mQTL) networks across multiple conditions. However, integrative analysis of omics data across multiple conditions presents a number of analytical and numerical challenges. To tackle these challenges, we propose a novel, computationally efficient approach based on tensor decomposition and regularization techniques. The proposed approach enables us to identify common network structures across multiple conditions as well as condition-specific network structures, while encouraging sparsity in latent factors used in the tensor decomposition. Synthetic data and real data are used to demonstrate the performance of the proposed approach.

EO1720: **Integrating omics data for cancer prognostic assessment**
Presenter: **Ronglai Shen**, Memorial Sloan-Kettering Cancer Center, United States
The aim is to present a supervised clustering approach to link cancer patient survival outcome with genomic data. This is accomplished by simultaneously optimizing integrated distance metric between the classes based on the multiple data types as well as the logrank statistic comparing the survival distributions of the classes to derive prognostically distinct subclasses. Cross-validation is used to prevent overfitting and obtain unbiased estimates of prediction accuracy. Application to the Cancer Genome Atlas (TCGA) data sets will be demonstrated.

EO537 Room 212 MULTIPLE TESTING AND SIMULTANEOUS INFERENCE PROCEDURES Chair: Frank Konietschke

EO0509: **Wild bootstraping rank-based procedures: Factorial designs and multiple testing**
Presenter: **Maria Umlauft**, Ulm University, Germany
Co-authors: Frank Konietschke, Markus Pauly
Scientific experiments comparing more than two groups are usually inferred by a classical ANOVA model. If the global null hypothesis of no treatment effect is rejected, multiple comparisons between the groups are usually performed, e.g. testing all pairwise hypothesis by means of unpaired t-test type Statistics and/or adequate simultaneous confidence intervals for the corresponding treatment effect. However, the underlying distributional assumptions (such as normality and variance homogeneity) are often not met in real data. Furthermore, the used effect sizes (mean differences or more general mean contrasts) may not be appropriate measures, especially if ordinal or ordered categorical data are present. To this end, several nonparametric procedures for simultaneous inference in general factorial designs have been studied. Here, the current approaches do not lead to simultaneous confidence intervals for contrasts in adequate effect measures. Thus, global inference and multiple testing procedures for an adequate nonparametric effect measure are required. We discuss rank-based multiple comparison procedures which can be used to test hypotheses formulated in terms of purely non-parametric treatment effects. In particular, different resampling methods as small sample size approximations will be discussed.

EO0791: **Visualization of simultaneous statistical inference through the non-overlapping confidence intervals**
Presenter: **Kimihiro Noguchi**, Western Washington University, United States
Visualization of the statistical significance through the appropriately adjusted non-overlapping confidence intervals facilitates the understanding of statistical results. However, previously proposed confidence level adjustment procedures for the multi-sample cases, such as the Bonferroni adjustment, are often very conservative. Thus, the purpose of the study is to discuss appropriate adjustment procedures that approximately control common error rates such as the familywise error rate and false discovery rate. In addition, an application of the proposed procedure to genomewide studies is demonstrated.

EO1198: **Detecting global and local signals using multivariate and multiple inference**
Presenter: **Arne Bathke**, University of Salzburg, Austria
When there are several endpoints and different predictors, researchers typically want to find out which predictors are relevant, and for which endpoints. We present two rather general approaches trying to accomplish these goals, accommodating binary, ordinal, and metric endpoints, and different nominal factors. We also try to address the question of how well the proposed methods actually accomplish their goals.

EO127 Room 002 COMBINING STATISTICS AND IMPRECISION Chair: Ana Belen Ramos-Guajardo

EO1268: **Nonparametric inference about the mean for big-but-biased data**
Presenter: **Ricardo Cao**, University of Coruna, Spain
Co-authors: Laura Borrajo
The risks of the sentence 'with enough data, the numbers speak for themselves' have been recently highlighted. Some of the problems coming from ignoring sampling bias in big data statistical analysis has also been recently reported. The problem of nonparametric statistical inference in big data under the presence of sampling bias is considered. The mean estimation problem is studied in this setup, in a nonparametric framework, when the biasing weight function is known (unrealistic) as well as for unknown weight functions (realistic). In the latter setup the problem is related to nonparametric density estimation. Asymptotic expressions for the mean squared error of the estimators proposed are considered. This leads to some asymptotic formula for the optimal smoothing parameter and some proposal for bandwidth selection. The question of how big the sample size has to be to compensate the sampling bias in big data is considered. Some simulations illustrate the performance of the proposed nonparametric methods.

EO1086: **Nonparametric tests for interval-valued data**
Presenter: **Przemyslaw Grzegorzewski**, Warsaw University of Technology, Poland
Co-authors: Martyna Spiewak
Interval-valued data have drawn an increasing interest recently. However, closed intervals applied for modeling data may deliver two different types of information: the imprecise description of a point-valued quantity (epistemic view) or the precise description of a set-valued entity (ontic view). Each view yields its own approach to data analysis and the way of carrying on the statistical inference. In the case of hypothesis testing based on interval-valued data it results in different formulation of hypotheses, test construction, the way of making final decisions and their interpretation. We illustrate these differences by suggesting the generalizations of some well-known nonparametric test for the one-sample and two-sample interval-valued data.

EC1094: **Computing maximum variance for interval uncertainty**
Presenter: **Adam Kolacz**, Warsaw University of Technology, Poland
Co-authors: Przemyslaw Grzegorzewski
The specificity of interval-valued data may cause some considerable problems even at the initial step of data analysis and statistical inference. For example, it is known that in general the problem of computing sample variance under interval uncertainty (perceived from the epistemic view) is NP hard. Therefore, for practical applications one has to consider only such cases when efficient computation may be possible. Some particular classes of interval-valued data for which efficient algorithms reaching the goal in an acceptable time have been found. Unfortunately, these classes require limitations on the intervals that sometimes appear too strong for the practical use. We propose an asymptotic approach leading to a novel class of intervals for which an efficient algorithm for computing the upper endpoint of the sample variance exists (the lower endpoint can be always computed in a feasible time). Our result shows not only a broad class of intervals of interest but also characterizes this class by conditions that could be easily verified.

EO105 Room 214 RECENT DEVELOPMENT IN DESIGN OF EXPERIMENTS Chair: Chang-Yun Lin

EO1164: **Aberration in qualitative multilevel designs: Computational aspects**
Presenter: **Roberto Fontana**, Politecnico di Torino, Italy
Generalized Word Length Pattern (GWLP) is an important and widely-used tool for comparing fractional factorial designs. We consider qualitative factors and we code their levels using the roots of the unity. We write the GWLP of a fraction $\mathcal{F}$ using the polynomial indicator function, whose coefficients encode many properties of the fraction. We show that the coefficient of a simple or interaction term can be written using the counts of its levels. This apparently simple remark leads to major consequence, including a convolution formula for the counts. We also show that the *mean aberration* of a term over the permutation of its levels provides a connection with the variance of the level counts. Moreover, using *mean aberrations* for symmetric s^m designs with s prime, we derive a new formula for computing the GWLP of $\mathcal{F}$. It is computationally easy, does not use complex numbers and also provides a clear way to interpret the GWLP. As case studies, we consider non-isomorphic orthogonal arrays that have the same GWLP. The different distributions of the *mean aberrations* suggest that they could be used as a further tool to discriminate between fractions.

EO1521: **Some properties of optimal foldover designs with column permutations**
Presenter: **Po Yang**, University of Manitoba, Canada
Foldover is a follow-up technique used in design of experiments. Traditional foldover designs are obtained by changing the sign of some columns of an initial design. We further consider to perform a column permutation. We investigate when a column permutation results a combined foldover design with better G- and G2-aberrations. Properties of such foldover designs are studied. Optimal foldover designs with column permutations are searched and tabulated for practical use.

EO1606: **Optimal experimental designs for mixed categorical and continuous responses**
Presenter: **MingHung Kao**, Arizona State University, United States
Experiments with mixed continuous and categorical responses that are possibly correlated are not uncommon in engineering, medical studies and many other fields. We develop optimal experimental designs for such experiments, which a joint model for both continuous and categorical responses is considered. We develop useful results that help to significantly reduce the number of candidate designs, and implement a convex optimization algorithm to search for optimal designs. The optimality of our designs is verified with the generalized equivalence theorem.

Sunday 11.12.2016 14:30 - 15:50 Parallel Session O – CFE-CMStatistics

CG298 Room 101 CONTRIBUTIONS IN MACROECONOMETRICS AND TIME SERIES II Chair: Guillaume Chevillon

CC0225: **The linear systems approach to linear rational expectations models**
Presenter: **Majid Al Sadoon**, Universitat Pompeu Fabra, Spain
Linear rational expectations models are considered from the linear systems point of view. Using a generalization of the Wiener-Hopf factorization, the linear systems approach is able to furnish very simple conditions for existence and uniqueness of both particular and generic linear rational expectations models. Two applications of this approach are provided; the first describes necessary and sufficient condition for exogeneity in linear rational expectations models and the second provides an exhaustive description of stationary and cointegrated solutions, including a generalization of Granger's representation theorem. Finally, an innovative numerical solution to the Wiener-Hopf factorization and its generalization is provided.

CC0961: **Mixed time aggregation of dynamic multivariate linear processes**
Presenter: **Michael Thornton**, University of York, United Kingdom
The time aggregation of vector linear processes: (i) containing mixed stock-flow data; and, (ii) aggregated at mixed frequencies is explored, showing how the parameters of the underlying model translate into those of the equivalent model of the aggregate. Based on manipulations of a general state-space form, the results may be applied to a wide range of linear ARMAX processes, including the discrete representation of a continuous time process, and may be iterated to model multiple frequencies or aggregation schemes. Estimation via the Kalman-Bucy filter and via the ARMAX representation of the observable data is discussed.

CC1343: **Output gap dispersion persistence and convergence: A stochastic volatility approach**
Presenter: **Thanasis Stengos**, University of Guelph, Canada
Co-authors: Serda Selin Ozturk
In empirical growth, the growth convergence hypothesis has been one of the main focal points of the active research. The main finding of the literature based on an analysis of (the mean function) of output gaps is that convergence in output gaps is hard to obtain and that most countries seem to follow divergent paths from each other. We use a stochastic volatility methodology based on sequential Efficient Importance Sampling (EIS) to examine the persistence of the unobserved volatility of output gaps and offer additional evidence that complements the findings of the mean based convergence approach. This is the first attempt in the literature to employ a stochastic volatility model to examine convergence. Our results confirm the results of weak convergence or lack of it altogether found in the literature using the mean function of output gaps.

CC1566: **Heteroskedasticity-robust unit root testing for panels with linear trends**
Presenter: **Yabibal Walle**, Georg-August-University Goettingen, Germany
Co-authors: Helmut Herwartz
Volatilities of most economic time series display significant fluctuations over time. Standard (panel) unit root tests are not robust in such cases and, hence, could lead to misleading conclusions. Consequently, recent studies have proposed heteroskedasticity-robust panel unit root tests. A major limitation of these tests, however, is that they work well only as long as the data is not trending. We propose a modification of the detrending procedure previously suggested that enables heteroskedasticity-robust panel unit root tests to work well in the presence of linear trends and innovation variance breaks. Evidence from Monte Carlo simulations shows satisfactory small sample performance of the tests under the new detrending strategy. As an empirical illustration, the evidence on the stationarity of GDP per capita is re-examined. Results show that GDP per capita is best characterized as a unit root process.

CG296 Room 106 CONTRIBUTIONS IN ECONOMETRICS OF BOND PRICES Chair: Ranko Jelic

CC0459: **Correcting estimation bias in regime switching dynamic term structure models**
Presenter: **Liu Liu**, The University of Manchester, United Kingdom
Co-authors: Sungjun Cho
The small sample bias in the estimation of a regime switching dynamic term structure model is assessed and corrected for. Using the dataset from 1971 to 2009, there are two regimes driven by the conditional volatility of bond yields and risk factors. In both regimes, the process of bond yields is highly persistent, which is the source of estimation bias when the sample size is small. After bias correction, the inference about expectations of future policy rates and long-maturity term premia changed dramatically in two high-volatility periods: the famous 1979–1982 episode and the recent financial crisis. Empirical findings are supported by Monte Carlo simulation, which shows that correcting small sample bias leads to more accurate inference about expectations of future policy rates and term premia compared to before bias correction.

CC1633: **TIPS liquidity premium and quantitative easing**
Presenter: **Laura Coroneo**, University of York, United Kingdom
In the context of a state-space model for nominal and TIPS yields, we identify the liquidity premium in the TIPS market as the common component in TIPS yields that is unspanned by nominal yields. This identifying assumption allows us to obtain a measure of the liquidity premium in the TIPS market that does not require selecting a market liquidity proxy, specifying a term structure model, or using inflation swaps. We estimate a joint state-space model for nominal and TIPS yields that treats the liquidity premium in the TIPS market as an unobservable component that we extract simultaneously with the yield curve factors. Using daily US yields, we find that the TIPS liquidity premium explains up to 22% of the variation in TIPS yields and that it sharply spiked during the 2008 financial crisis. We also find evidence of a flight to liquidity effect in the US Treasury market, as yields on nominal bonds decrease following a shock to the liquidity premium in the TIPS market. In addition, a counterfactual exercise shows that the QE2 programme had only limited effect on the liquidity premium in the TIPS market.

CC1654: **Forecasting the yield curve dynamics: A just-identified no-arbitrage FAVAR approach with tractable economic factors**
Presenter: **Alena Audzeyeva**, Keele University, United Kingdom
Co-authors: Robin Bladen-Hovell, Sadeeptha Jayathilaka
A term structure model formulation is proposed that links the term structure dynamics to evolution in underlying economic factors within a data-rich setting. Using a factor-augmented vector autoregression, we examine the influence of dynamic factors that exploit large information sets tracking real activity, inflation, money and market asset pricing and also a latent factor during different economic policy regimes. In contrast to many popular canonical formulations of Gaussian affine term-structure models, our model is just-identified which ensures an improved robustness of the parameter estimation, based on no-arbitrage restrictions, by means of minimum-chi-square estimation. While economically and practically appealing, our term-structure formulation affords more accurate out-of-sample forecasts than a number of previously suggested affine term-structure modeling approaches during both the period of "great stability" and a recent period of near zero short-term interest rates in the UK.

CC1657: **How many bonds to efficiently diversify credit risk?**
Presenter: **Sara Razmpa**, LOIM, Switzerland
Co-authors: Florian Ielpo
A now old literature has focused on the diversification effects obtained from increasing the number of securities in a portfolio. When the situation in the case of stocks is clear, the credit bonds one has received little attention so far. Using a dataset of 4 years of returns on individual credit bonds issued in Euro and in US Dollar, across the investment grade and the high yield world, we present results regarding the optimal number of bonds necessary when an investor needs to mitigate idiosyncratic risk efficiently. Many investors are using ETFs to gain an efficient access to the credit world as the asset class is usually found to be complex. Our results show that with 90 bonds, about 90% of an investment grade index's performance is explained by the market factor. In the case of High Yield bonds, an investor would need twice as many bonds to obtain similar results. When then question the stability of these results across the different market environments, highlighting how the risk concentration in each specific market should have an impact over the number of bonds to be used to generate a decent index replication.

CG711 Room no assigned CONTRIBUTIONS IN CREDIT RISK Chair: Florian Ielpo

CC1425: **Distribution function for cumulative intensity of SSRJD and its applications to CVA of CDS**
Presenter: **Toshinao Yoshiba**, Bank of Japan, Japan
Co-authors: Tetsuya Adachi, Takumi Sueshige
The shifted square root jump diffusion (SSRJD) is a tractable process for the stochastic default intensity to valuate CDS (credit default swap) incorporating its surge under stress. While a copula approach is a promising way to evaluate the CVA (credit valuation adjustment) of a CDS with a wrong-way risk, the distribution function for cumulative intensity of SSRJD is needed to apply the copula approach. We show the derivation of the distribution function. Since the SSRJD is an affine process, the characteristic function of cumulative intensity follows Riccati-type ordinary differential equations. The analytic solution can be obtained; however, the solution defined on the space of complex-value is multivalued. We show the way to solve this problem with reducing the multilayered Riemann surface of the solution to a single layer. Applying the fractional fast Fourier transform and numerical integration to the characteristic function of SSRJD cumulative intensity, the distribution function is derived. Using the result, we also show comparative analyses between the copula approach and a co-jump model for counterparty and reference default intensities.

CC1522: **Default and recovery: The copula-based sample selection model approach**
Presenter: **Hui-Ching Chuang**, Yuan Ze University, Taiwan
A copula-based self-selection model is proposed to examine the determinants in the recovery rate at the time of default. The copula-based self-section model can adopt a flexible beta-regression model as the marginal model for the recovery rates and consider the dependence structure between default and recovery by exploiting various copula functions. We apply the model to a sample obtained from Moody's Default and Recovery Database.

CC1638: **Sovereign credit rating determinants: The impact of the European debt crisis**
Presenter: **Peter Reusens**, KU Leuven, Belgium
Co-authors: Christophe Croux
The aim is to compare the importance of different sovereign credit rating determinants over time, using a sample of 90 countries for the years 2002-2015. Applying the composite marginal likelihood approach, we estimate a multi-year ordered probit model for each of the three major credit rating agencies. After the start of the European debt crisis in 2009, the importance of the financial balance, the economic development and the external debt increased substantially and the effect of Eurozone membership switched from positive to negative. In addition, GDP growth gained a lot of importance for highly indebted sovereigns and government debt became much more important for countries with a low GDP growth rate. These findings provide empirical evidence that the credit rating agencies changed their sovereign credit rating assessment after the start of the European debt crisis.

CC1580: **Extracting risk neutral distributions using option prices and CDS spreads**
Presenter: **Mohammad Jahan-Parvar**, Federal Reserve Baord of Governors, United States
Co-authors: Sirio Aramonte, Samuel Rosen, John Schindler
A methodology is proposed to estimate the risk-neutral distribution of a firm's expected stock returns by blending option prices and CDS spreads, with the former providing information about the central part of the distribution, and the latter determining the left tail. We apply the methodology to a sample of risky U.S. firms. We assess the economic value of estimating risk-neutral distributions with both options and CDS by forming a long-short portfolio based on the lagged difference between the option-CDS skewness and the option-only skewness. The strategy generates significant abnormal returns, gross of transaction costs, after controlling for a large set of relevant factors. The results are especially strong in the 2008-2011 period, which includes the 2008 financial crisis, the Greek debt crisis, and the 2011 debt ceiling event.

CG486 Room 112 CONTRIBUTIONS IN NOWCASTING AND FORECASTING MACROECONOMIC TRENDS Chair: Jasper de Winter

CC1441: **Testing nowcast monotonicity**
Presenter: **Daniel Gutknecht**, University of Mannheim, Germany
Co-authors: Jack Fosten
Nowcasting has become an important tool to many public and private institutions in obtaining timely predictions of low-frequency variables such as Gross Domestic Product (GDP) using current information. Nowcasters often report that a nowcasting method is successful if its predictions improve monotonically as we move towards the publication date of the target variable. We develop a novel testing approach to formally evaluate the monotonicity of a nowcasting method. In order to highlight the usefulness of this new test, we provide various analytical examples where nowcast monotonicity fails or where it helps to assess the trade-off between timeliness and accuracy of the predictor variables. Formally, we extend a methodology for testing many moment inequalities, which allows the number of inequalities to be large relative to the sample size, to the case of nowcast monotonicity testing. We show that rolling parameter estimation in the pseudo out-of-sample approach can be accommodated in this setting and illustrate the performance of our test with a detailed set of Monte Carlo simulations. We conclude with an empirical application of nowcasting U.S. real GDP growth and five GDP sub-components.

CC1604: **GDP trend-cycle decompositions using state-level data**
Presenter: **Manuel Gonzalez-Astudillo**, Board of Governors of the Federal Reserve System, United States
The aim is to develop a method for trend-cycle decomposition of GDP exploiting the cross-sectional variation of state-level GDP and unemployment rate data. In the model, each state's (log of) GDP is the sum of state-specific trend and cycle components, whereas the state unemployment rate is the sum of a trend and a cycle that is a function of the cycle of GDP in a way resembling an Okun's law. The state-specific GDP trend is a linear combination of a trend common across states and an idiosyncratic trend. The same is true for the state-specific unemployment rate trend. In a similar way, each state's GDP cycle is a linear combination of a cycle common to all the states and an idiosyncratic innovation. We estimate the model with Bayesian methods using quarterly data of GDP and the unemployment rate from 2005:Q1 to 2015:Q3 for the 50 states and the District

of Columbia. Results show that the U.S. output gap reached about -6% during the Great Recession, as opposed to conventional estimates of about -7.5% (CBO-implied output gap), suggesting that the output trend was affected more than what conventional estimates obtain. Results also show that the output gap as of 2016:Q1 is about -1.5%, compared to the CBO estimate of -2%.

CC1707: **Measuring economic and financial conditions using a unified framework**
Presenter: **Hamza Demircan**, KOC University, Turkey
Co-authors: Cem Cakmakli
We propose a general framework where we estimate the economic conditions, or in other words, business cycle together with financial conditions using a big dataset of macroeconomic and financial variables. We do this by employing a dynamic factor model together with (Markov) switching mean and variance to capture different phases of economic financial conditions. The novel feature of the model is that we estimate a unique dynamics of cycles with varying phase shifts. Specifically, we allow for imperfect synchronization of cyclical regimes in economic and financial conditions indicators, due to phase shifts of a single common cycle. This enables us to estimate the indicators of economic and financial conditions together with their lead-lag relations in terms of time periods financial conditions indicators leading the business cycle indicator in a unified framework jointly.

CC1594: **Temporal disaggregation of short time series with structural breaks: Estimating quarterly data for emerging economies**
Presenter: **Jerome Trinh**, THEMA - CREST, France
Official data disclosed by emerging national administrations are sparse, especially time series such as national accounts which are mostly published once a year with from three to seven quarters of latency. Higher-frequency indicators or components can help disaggregating annual data into up-to-date quarterly data by using a well known method of temporal distribution. However, the small sample size and the instability of emerging countries official data prevent a good fit of the target time series and its indicator. Incorporating a procedure of structural break detection, such as a test of cointegration with structural breaks, into the method of temporal disaggregation method can improve the fit, hence the prediction of higher-frequency estimations. We detail disaggregation formulas that considers different types of structural break in the parameters of the model linking the target time series and its higher-frequency indicator. Small sample null distribution of the statistics for the test of cointegration with structural breaks are also computed. Then the predictive performance of the method is assessed by using empirical advanced countries data and simulated time series. Finally, an example of an application to macroeconomic business cycles analysis is initiated by studying the response of the Chinese macroeconomic fluctuations to aggregate supply and aggregate demand shocks.

CG568 Room 109 CONTRIBUTIONS IN OIL PRICE ANALYSIS Chair: Aleksander Welfe

CC1515: **Oil volatility pass-through and and real exchange rate misalignment in commodity exporting countries**
Presenter: **Nicola Rubino**, University of Barcelona, Spain
We investigate the short and long run relationship between leading commodity prices and real effective exchange rates in a group of thirty countries over the period 1980-2015, paying attention to non-linearities in three aspects. First, we tested for unit roots using tests that account for the presence of structural breaks, which do not consider the possibility of a structural break under the alternative hypothesis, and other previously proposed tests, which unambiguously imply trend stationarity. We thus proceeded to look for a possible co-integrating relationship through the C/T test, complemented by other previous test. Second, cointegrating relationships were estimated through DOLS, in order to calculate a Real Effective Exchange Rate misalignment measure. Third, we fitted the measure into a smooth transition regression model to evaluate the impact of oil price variations on the relationship between REER and leading commodity prices. Our results show that a behavioral co-integrating relationship between REER and leading commodity prices is indeed present in a fraction of the countries considered. After testing for non-linearity, we show that the existence of short run adjustment depends on the extent of the volatility of the oil prices in a fraction of countries, while in the residual ones mean reversion remains consistent across regimes or is totally absent.

CC0179: **The importance of external and internal shocks for real exchange rate and industrial production in Russia: SVAR approach**
Presenter: **Andrey Zubarev**, Russian Presidential Academy of National Economy and Public Administration, Russia
Co-authors: Andrey Polbin, Anton Skrobotov
The main sources of macroeconomic fluctuations are studies in Russia, which is an oil-exporting economy. An SVARX approach with long run restrictions was used to identify oil price shocks (identified by the oil price as an exogenous variable), nominal shocks and two types of productivity shocks. Particularly, one is a general productivity shock (in the whole economy), while the other is a Balassa-Samuelson-type productivity shock in the tradable sector of economy. With the help of impulse response functions we found that the productivity shock in the tradable sector leads to a permanent increase in real exchange rate, while a general productivity shock has only short-run negative effect on real exchange rate. This result is in line with New-Keynesian models with nominal rigidities. Contrary to previous studies, we found that the Balassa-Samuelson effect was not the main driver of industrial production and real exchange dynamics in Russia. Oil price dynamics was an important source of industrial production and real exchange rate fluctuations which supports the widespread opinion that the Russian economy is poorly diversified. We also found a huge contribution of nominal shock to the decomposition of real exchange rate dynamics in the second half of 2014 which could be attributed to insufficient tightening of monetary policy.

CC1734: **A financialization model of crude oil markets**
Presenter: **Takashi Kanamura**, Kyoto University, Japan
The aim is to investigate theoretically and empirically the financialization of crude oil markets. We propose a simple correlation model between stock and crude oil prices based on the supply and demand relationship taking into account the impact of stock markets on crude oil markets. Based on the correlation model, the optimal allocations of stock, crude oil and a risk free asset are obtained by solving the Merton's problem under the assumption of a log utility function. Using the price correlation model it is empirically shown that the correlations between S& P 500 and WTI crude oil prices are positive and increasing. In contrast, the correlations between S& P 500 and Brent crude oil prices are close to zero, which are lower than the correlations between S&P 500 and WTI crude oil prices. It may suggest that the financialization of crude oil markets be limited to WTI crude oil markets. The optimal allocations of S& P 500, WTI or Brent crude oil and a risk free asset are empirically obtained based on the correlation model. It is shown that after the 2008 financial turmoil the optimal WTI crude oil positions decrease in line with the optimal S& P 500 positions while the optimal Brent crude oil positions are not relevant to the optimal S& P 500 positions. The results may suggest that after the 2008 financial turmoil WTI crude oil be financialized and do not show the diversification effect to S& P 500 while Brent crude oil not be the case.

CC1509: **Real exchange rates, US dollar and crude oil price in the tripolar model**
Presenter: **Piotr Keblowski**, University of Lodz, Poland
Co-authors: Aleksander Welfe, Katarzyna Leszkiewicz-Kedzior
The foreign exchange rates of the non-euro European Union member states show strong dependence on the behaviour of the euro, because the states balance of payments is determined by a high share of intra-EU trade. The Polish currency (the zloty) is used to exemplify that as the exchange rates are additionally influenced by credit default risk premiums and the euro-dollar behaviour, the tripolar model is an appropriate analytical framework. On the other hand, it is found that the euro-dollar exchange rate, real crude oil prices and the parity of real risk-free interest rates define two stabile

long-run relationships. The propagation of shocks is demonstrated by means of an impulse-response analysis based on the conditional CVAR model with four cointegrating vectors.

CG364 Room 006 CONTRIBUTIONS IN COINTEGRATION ANALYSIS Chair: Alain Hecq

CC1595: **Identification of long-run effects in near-integrated systems**
Presenter: **Peter Boswijk**, University of Amsterdam, Netherlands
It has been previously shown that inference on cointegrating coefficients is not robust to local deviations from the unit root. When unit roots are incorrectly imposed, the resulting vector error correction model is misspecified. Quasi-maximum likelihood in this misspecified model leads to asymptotically mixed normal estimators, but centered around the pseudo-true value, which is related to the concept of long-run multiplier. However, in near-integrated systems, there is no unique stationary linear combination, so that the identification of appropriate long-run effects has to be addressed. We follow a previous approach and focus on the long-run relative impulse response. A new parametrisation of the unrestricted VAR model is developed, in terms of both this long-run parameter and the largest autoregressive roots. Next, we address possibilities to make asymptotically valid inference on the long-run relative impulse response. Analogously to an earlier result, we establish that the standard implementation of the bootstrap does not deliver valid inference. Therefore, we investigate alternative versions of the bootstrap, focussing in particular on the double bootstrap.

CC1721: **Structural breaks and instabilities at the end of sample**
Presenter: **Josep Lluis Carrion-i-Silvestre**, Universitat de Barcelona, Spain
The focus is on cointegration relationships that are affected by the presence of parameter instabilities that can appear in the whole time period. Usually, the estimation of structural breaks for cointegrated relationships requires excluding some observations at the extremes of the time period, avoiding the possibility of capturing parameter instabilities at the end-of-sample (EOS). We propose a procedure to estimate both structural breaks at the non-end-of-sample (NEOS) and parameter instabilities at the EOS. The new procedure is designed to improve the performance of the existing EOS instabilities tests.

CC1489: **Analysing cross-currency basis spreads**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic
Co-authors: Jaroslav Baran
The aim is to investigate drivers of cross-currency basis spreads that were historically close to zero but widened significantly since the start of the financial crisis. By constructing arbitrage-free boundaries we discover that for the example of EUR/USD and EUR/CZK FX currency pairs, the supply and demand imbalances may push basis spreads outside the boundaries. We also build a multiple regression and a cointegration model explaining drivers behind the EUR/USD basis swap spreads. The most important drivers of the cross-currency basis spreads appear to be short-term and long-term US and EU financial sectors credit risk indicators, Fed/ECB balance sheet ratio, EUR/USD exchange rate, or the market volatility.

CC1185: **Investigating nonlinear purchasing power parity for EUR/PLN within Bayesian STVECM framework**
Presenter: **Adrian Burda**, Uniwersytet Ekonomiczny w Krakowie, Poland
Existence of strong-form purchasing power (PPP), both for developed and emerging economies remains a widely investigated topics in international finance and macroeconomics. In case of Polish zloty, before the global financial crisis the existence strong-form PPP had been rejected in most studies, due to its appreciation against to main currencies, while after 2008 occurred more and more arguments supporting this hypothesis. The aim is to develop and apply to EUR/PLN a Bayesian procedure to investigate the purchasing power parity (PPP) in the exponential smooth transition vector error correction (ESTVECM) framework. The cointegrating relationship along with the nonlinearities caused by the departures from the long-run equilibrium is jointly estimated, by Gibbs sampler. For comparison purposes also simple linear vector error correction (VECM) with linear mechanism of error correction. The comparison of models is performed by posterior probability, and predictive density score, as well as a previous test. Usage of mentioned methods allows us to assess if taking into account complex structure of deviations from PPP could improve forecast accuracy of PPP model.

CG430 Room 002 CONTRIBUTIONS IN TIME-VARYING PARAMETER MODELS Chair: Simona Sanfelici

CC1745: **Nonparametric estimation of time-varying parameters in nonlinear models**
Presenter: **Young Jun Lee**, University College London, United Kingdom
Co-authors: Dennis Kristensen
Asymptotic properties are developed for nonparametric estimators of time-varying parameters for a general class of dynamic models. We take as given that the parameters of interest are identified as the maximizer of some population moment. The time-varying parameters are then estimated using local versions of standard M-estimators based on this population moment. Under high-level conditions including local stationarity, we show that the estimators are consistent and asymptotically normally distributed. We provide primitive conditions for our high-level conditions for Markov models. Our methodology and theory generalize existing nonparametric methods for time-varying parameters. To demonstrate the usefulness of our general set-up, we give explicit conditions to hold the established asymptotic properties for several specific nonlinear time series models.

CC1593: **Time-varying mixed frequency forecasting: A real-time experiment**
Presenter: **Stefan Neuwirth**, ETH Zurich - KOF Swiss Economic Institute, Switzerland
In order to test the usefulness of time-varying parameters when forecasting with mixed-frequency data we compare the forecast performance of bridge equations and unrestriced MIDAS models with constant and time-varying parameters. An out-of-sample forecasting exercise with US real-time data shows that the use of time-varying parameters does not improve forecasts significantly over all vintages. However, since the Great Recession, forecast errors are smaller when forecasting with bridge equations, especially for longer forecast horizons, due to the ability of time-varying parameters to incorporate gradual structural changes faster.

CC1257: **Dynamic Bayesian estimation of time-varying cointegration parameters**
Presenter: **Basile Marquier**, University of Sheffield, United Kingdom
Many models involving financial or macroeconomic models are changing over time and cointegration offers the possibility of identifying relationships in a set of economic time series. The aim is to estimate a time-varying cointegrated model. We discuss the possibility of time-varying cointegration by proposing a Forward Filtering Backward Sampling scheme for the parameters of the Vector Error correction model. The time-varying Vector Error Correction Model (VECM) is first derived from the dynamic VAR model. The long-run relationships matrix becomes a dynamic parameter and its rank, the cointegration rank, which is estimated from the cointegration matrix, is then also time-varying. Therefore we have the possibility of seeing the independent cointegration relations evolving over different time periods. This new dynamic approach is shown with several synthetic data sets, in which we change the cointegration relations and the rank over different time periods. Finally an application on different sectors of the Dow Jones is studied.

CC1379: **Semi-analytical method for pricing barrier options with time-dependent parameters**
Presenter: **Simona Sanfelici**, University of Parma, Italy
Co-authors: Chiara Guardasoni
Barrier options are path-dependent derivatives that have become increasingly popular and frequently traded in the recent years. Closed-form pricing formulas for these contracts are available only under very special cases, such as in the Black-Scholes framework. Pricing of barrier option is traditionally based on Monte Carlo methods that are affected by high computational costs and inaccuracy due to their slow convergence or on PDE methods. We propose a stable, accurate and efficient numerical method for pricing barrier options in a model with time-dependent parameters. The new approach is based on the Boundary Element Method that was introduced in the Engineering field in 1970. Especially when the differential problem is defined in an unbounded domain and the data are assigned on a limited boundary (as for barrier options), the method is particularly advantageous for its high accuracy, for the implicit satisfaction of the far-field conditions at infinity and for the low discretization costs. This method has already been tested for the Black-Scholes and other stochastic volatility and jump-diffusion models, giving very good results in terms of accuracy and computational time. We think that the method could have a considerable range of applications that in these last years we started investigating.

CG306 Room 104 CONTRIBUTIONS IN CO-MOVEMENTS IN ECONOMIC AND FINANCIAL TIME SERIES Chair: Gianluca Cubadda

CC1315: **The joint distribution of domestic indexes: An approach using conditional copulas**
Presenter: **Eva Ferreira**, University of the Basque Country, Spain
Co-authors: Susan Orbe, Jone Ascorbebeitia
It is well known that the comovements between portfolios are time-varying. Our interest is to detect whether the comovements variation between domestic indexes can be explained by some global risk factors. Concretely, we will study the dependence of EuroStoxx. Moreover, we are interested in measures of dependence beyond linear Pearson's correlation. This coefficient suits for normal variables, but financial variables have more complicated distributions. To overcome this fact, we propose the use of copulas to analyze the relation between domestic European indexes, conditional to the Eurostoxx. The use of copulas allows us to model the dependence better than with elliptic distributions. We estimate conditional copulas using nonparametric methods to obtain the joint probability distribution function between indexes. This method allows us to relate daily data with monthly data in a very simple manner, and therefore to relate stock indexes with macroeconomic variables such as inflation or GDP. We will measure the dependence and the conditional dependence using the Kendall's tau. We provide a statistic to test the significance of tau and its empirical distribution using jacknife.

CC1655: **Standard and alternative risk premiums in a world of lower nominal growth**
Presenter: **Florian Ielpo**, Centre Economie de la Sorbonne, France
Co-authors: Ling-Ni Boon
Nominal Gross Domestic Product growth has been declining over the past 30 years across the globe. The most recent forecasts for both Developed and Emerging countries show that this structural trend is here to stay: both growth and inflation forecasts are pointing in that direction. The aim is to investigate how this decline in overall nominal growth dynamics has impacted the returns coming from standard and alternative risk premiums over the 1976-2016 period. Using a bootstrap methodology, we use overlapping samples econometrics to assess the explanatory power of nominal growth over the returns from which multi-asset portfolios are made. Our results show that a large portion of standard assets such as bonds and equities show a significant dependency over the Worlds GDP growth and especially over the US one. In the case of alternative risk premiums, such as trend following, equity risk parity or carry strategies, this dependency is found to be weaker. Our findings are consistent with the current trend in the asset management industry to add more alternative risk premiums to balanced portfolios.

CC1682: **Cross-correlation analysis of international foreign exchange markets: An EEMD-based approach**
Presenter: **Milan Csaba Badics**, Corvinus University of Budapest, Hungary
The aim is to investigate the synchronization and contagion of international foreign exchange market co-movements especially before, during, and after the recent subprime crisis, by researching the interconnections in time-frequency domain. We test changes in correlations for different time scales with ensemble empirical mode decomposition (EEMD), a modified method of EMD. With clustering the different frequencies IMFs, we grouped thoese into high-, medium-, and low-frequency components, representing the short-, medium-, and long term volatilities of the foreign exchange time series. Our results indicate that correlation between different foreign exchange markets tends to be stable in low volatility periods, but this changes dramatically during turbulent periods both at higher and lower frequencies. Therefore, all these findings should be considered for risk managers from an international portfolio diversification perspective.

CC1461: **Credit default swaps: Does the traded volume influence academiass interest?**
Presenter: **Karin Martin-Bujack**, Universidad Pontificia Comillas, Spain
Co-authors: M Teresa Corzo
The 21st century has witnessed an extraordinary boom in the CDS market. This financial instrument has gained popularity to deal with credit risks both in the market place and in the academy. By examining the links between CDS trading volumes and research pieces about or using CDS data during the period 2001-2015 we study the lead-lag relationship of this two variables using a Vector Autoregressive model. We find that the variation in the volumes traded in the CDS market plays a leading role in the variation of the amount of publications that use this instrument, but although CDS notional amounts have decreased since 2007 Q4, we find that researchers still give a key role to the information that CDS markets incorporate.

CG633 Room 107 CONTRIBUTIONS IN EMPIRICAL MACROECONOMICS Chair: Daniel Kaufmann

CC1401: **Extracting fiscal policy expectations from daily stock returns**
Presenter: **Etsuro Shioji**, Hitotsubashi, Japan
The "fiscal foresight" problem refers to the fact that most fiscal policy changes are announced well before their implementation. As a result, innovations in the actual amount of government spending may not capture the true timing or magnitude of the surprises. To overcome this difficulty, we construct new daily indicators of surprises about public investment spending for Japan. This novel approach combines a detailed analysis of newspaper articles on future policies with information on a cross section of excess returns on stocks of construction companies. In a previous work, we constructed an indicator which was essentially a weighted average of those returns on the days that the news arrived. In contrast, the two new indicators take advantage of heterogeneity within the industry. Degrees of dependence on governmental contracts differ markedly between construction companies. The first indicator simply takes the difference in the average excess returns between the firms that are more dependent on public procurement and the less dependent ones. The second indicator is more elaborate, and is based on the "Target Rotation" approach in the factor analysis. It is shown that both serve as good leading indicators of future public investment spending.

CC1540: **Determinants of economic development: SEM approach**
Presenter: **Vladimir Potashnikov**, RANEPA, Russia
Co-authors: Andrey Zubarev, Oleg Lugovoy
The study is devoted to finding main determinants of economic development. The key idea of the study is to select potential variables not by the views of a particular researcher, but by some automatic procedure. Criteria for variable selection should be unified. We use SEM (structural equation modeling) technique for this purpose. Contrary to many other studies we model development as a latent variable. We use preliminary data transformation procedure and constructed "correlation baskets" (groups of variables, which explain the variation of a particular variable) to find clusters of variables that potentially indicate and explain economic development. Our database consists of more than 1400 variables from WGI, WDI, Frazer Institute database, and PWT. We chose GDP per capita (PPP) and different factors that describe healthcare and education system (e.g. childe and neonatal mortality, several types of school enrolment and etc.) as indicators of economic development. Amongst many potential clusters of explanatory variables that was characteristics of institution quality showed strong persistence in being important factors for economic development. Using some robustness checks we found the two most important institutional variables are rule of law and political stability. We also found that involvement into international trade enhances development level which is in line with Rodriks papers.

CC1636: **Market and political power interactions in Greece: An empirical investigation**
Presenter: **Vanghelis Vassilatos**, Athens University of Economics and Business, Greece
Co-authors: Tryphon Kollintzas, Dimitris Papageorgiou, Mike Tsionas
Using a dynamic panel of 21 OECD countries, we find that, unlike the other OECD countries in the sample, wage setting institutions, competition conditions, public finances and external imbalances can account for the behavior of the public sector wage premium (WPR) and the self employed taxation gap (TSL) in Greece, and to a lesser extent in Spain and Portugal, in a manner that is consistent with an insiders-outsiders society. That is, a politicoeconomic system characterized by groups of selfish elites that enjoy market power, but at the same time cooperate in influencing government in protecting and promoting their collective self interests. Then, we find that for Greece as well as Spain and Portugal, WPR and TSL have an adverse effect on both TFP and output growth. Finally, the effect of WPR and TSL on the business cycle (shock propagation mechanism) is investigated via a panel VAR analysis. Again, impulse response function analysis suggests that the shock propagation mechanism of WPR and TSL for Greece and to a lesser extent for Spain and Portugal, are quite different from the rest of the OECD countries. These results are important in order to understand the Greek crisis.

CC1658: **Escaping the great recession**
Presenter: **Leonardo Melosi**, Federal Reserve Bank of Chicago, United States
Co-authors: Francesco Bianchi
The aim is to show that policy uncertainty about how the rising public debt will be stabilized accounts for the lack of deflation in the US economy at the zero lower bound. We first estimate a Markov-switching VAR to highlight that a zero-lower-bound regime captures most of the comovements during the Great Recession: a deep recession, no deflation, and large fiscal imbalances. We then show that a micro-founded model that features policy uncertainty accounts for these stylized facts. Finally, we highlight that policy uncertainty arises at the zero lower bound because of a trade-off between mitigating the recession and preserving long-run macroeconomic stability.

CG566 Room 111 CONTRIBUTIONS IN JUMPS AND VOLATILITY Chair: Hans Manner

CC1565: **Flat activity: Revisiting Blumenthal-Getoor index estimation**
Presenter: **Aleksey Kolokolov**, Goethe University Frankfurt, Germany
Co-authors: Roberto Reno
Realized multipower variations, originally introduced to eliminate jumps, can be extremely useful for inference in pure-jump models. In particular, we show how to build a superior estimator of the jump activity index of a semimartingale observed at high-frequency by comparing different multipower variations. Our methodology remains valid for both Brownian and pure-jump semimartingales, hence allowing to distinguish the two classes of models, but is biased (as alternative estimators) in the presence of flat trading, which is viewed as a microstructural contamination to the semimartingale hypothesis. The implementation on S& P500 and VIX indexes clearly indicates that, after removing for the bias due to flat trading, there is no necessity of pure-jump models in both cases.

CC1567: **Modelling and forecasting realised volatility in German-Austrian continuous intraday electricity prices**
Presenter: **Ainhoa Zarraga**, University of the Basque Country, Spain
Co-authors: Aitor Ciarreta, Peru Muniain
The aim is to use high-frequency continuous intraday electricity price data from the EPEX market to estimate and forecast realised volatility. Three different jump tests are used to break down the variation into jump and continuous components using quadratic variation theory. Several heterogeneous autoregressive models are then estimated for the logarithmic and standard deviation transformations. GARCH structures are included in the error terms of the models when evidence of conditional heteroscedasticity is found. Model selection is based on various out-of-sample criteria. Results show that decomposition of realised volatility is important for forecasting and that the decision whether to include GARCH-type innovations might depend on the transformation selected. Finally, results are sensitive to the jump test used in the case of the standard deviation transformation.

CC1649: **Efficient parameter estimation for multivariate jump-diffusions**
Presenter: **Gustavo Schwenkler**, Boston University, United States
Co-authors: Francois Guay
The aim is to develop an unbiased Monte Carlo estimator of the transition density of a multivariate jump-diffusion process. The drift, volatility, jump intensity, and jump magnitude are allowed to be state-dependent and non-affine. It is not necessary that the volatility matrix can be diagonalized using a change of variable or change of time. Our density estimator facilitates the parametric estimation of multivariate jump-diffusion models based on discretely observed data. Under conditions that can be verified with our density estimator, the parameter estimators we propose have the same asymptotic behavior as maximum likelihood estimators as the number of data points grows, but the observation frequency of the data is kept fixed. In a numerical case study of practical relevance, our density and parameter estimators are found to be highly accurate and computationally efficient.

CC0228: **Improved forecasting of realized variance measures**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Jeremias Bekierman
The problem of forecasting realized volatility measures is considered. These measures are known to be highly persistent, but also to be noisy estimates of the underlying integrated variance. This fact has been recently exploited to extend the heterogeneous autoregressive (HAR) model by letting the model parameters vary over time depending on estimated measurement errors and show that their model leads to improved forecasts. We propose an alternative specification that allows the autoregressive parameter of the HAR model for log-volatilities to be driven by a latent

gaussian autoregressive process that potentially also depends on the estimated measurement error. The model can be estimated straightforwardly using the Kalman filter. Our empirical analysis considers the realized volatilities of 40 stocks from the the S&P 500 estimated using three different observation frequencies. Our in-sample results show that the time-varying parameters resulting from our specification are more flexible than those of the previous model and consequently provides a better model fit. Furthermore, our model generates superior forecasts and consistently outperforms the competing models in terms of different loss functions and for various subsamples of the forecasting period.

CG294 Room 110 CONTRIBUTIONS IN MONETARY POLICY Chair: William Dupor

CC1663: **Monetary policy transmission during the global economic crisis: A case of small open economies**
Presenter: **Przemyslaw Wlodarczyk**, University of Lodz, Poland

Monetary policy is one of the key elements of undertaken macroeconomic policy, however, despite having relatively strong proof of the existence of short-run money-output nexus, we still lack adequate evidence concerning the cyclical characteristics of this relationship. Existing economic literature points towards apparent asymmetry of reaction of economies to changes in the character of monetary policy during the business cycle. It is however not clear whether these changes are a result of structurally different reaction of economic entities or a change in the volatilities of economic variables. We aim to fill this gap in existing research delivering new evidence on the money-output nexus in chosen small open economies. In our analyses we use Structural VAR (SVAR) model of monetary transmission mechanism for chosen small open economies (including e.g. non-eurozone Visegrad group countries). Model is estimated using data from the 2000-2015 period. We compare the results based on the short-run restrictions set using heteroscedasticity assumptions with the results of both traditional SVAR models with the restrictions based on either Cholesky decomposition or exclusion restrictions, as well as Threshold Structural VAR (TSVAR) models using Cholesky decomposition.

CC1362: **International house price cycles, monetary policy and risk premiums**
Presenter: **Gregory Bauer**, Bank of Canada, Canada

Three alternative causes of house price corrections are evaluated: anticipated tightenings of monetary policy, deviations of house prices from fundamentals, and rapid credit growth. A new cross-country measure of monetary policy expectations based on an international term structure model with time-varying risk premiums is constructed. House price overvaluation is estimated via an asset pricing model. The variables are incorporated into a panel logit regression model that estimates the likelihood of a large house price correction in 18 OECD countries. The results show that corrections are triggered by increases in the market's forecast of higher policy rates. The estimated degree of house price overvaluation also contains significant information about subsequent price reversals. In contrast to the financial crisis literature, credit growth is less important. All of these variables help to forecast recessions.

CC1499: **Uncertainty shocks and monetary policies**
Presenter: **Alessia Paccagnini**, University College Dublin, Ireland
Co-authors: Valentina Colombo

The interaction between uncertainty shocks and conventional and unconventional U.S. monetary policies is investigated. Uncertainty is captured by appealing to some indicators recently developed. Relying on a non-linear VAR, we isolate the effects of uncertainty shocks in recessions versus expansions. Uncertainty shocks trigger negative macroeconomic fluctuations across the business cycle. To offset such fluctuations, the Federal Reserve reacts relying on conventional monetary policy tools. However, when shocks to uncertainty occur during deep recessions, the Federal Reserve reacts switching from the conventional to the unconventional monetary tools.

CC0944: **A term structure of interest rates model with zero lower bound and non-standard monetary policy measures**
Presenter: **Viktors Ajevskis**, Bank of Latvia and Riga Technical University, Latvia

It is proposed a ZLB/shadow rate term structure of interest rates model with both unobservable factors and those of non-standard monetary policy measures. The non-standard factors include the ECB's holdings of APPs and LTROs as well as their weighted average maturities. The model is approximated by the Taylor series expansion and estimated by the extended Kalman filter, using the sample from July 2009 to September 2015. The results show that the 5-year OIS rate at the end of September 2015 was about 60 basis points lower than it would have been in the case of the absence of the non-standard monetary policy measures.

CG342 Room 007 CONTRIBUTIONS IN FORECASTING AND TIME VARIATION Chair: Helmut Luetkepohl

CC0938: **Forecasting the real price of oil: Time-variation and forecast uncertainty**
Presenter: **Christoph Funk**, Justus-Liebig-University Giessen, Germany

The price of oil is a driving factor of the world economy and gives a first hint on the potential economic development over time. Therefore, the purpose is to examine whether it is possible to generate an accurate forecast of the oil price and the associated uncertainty in this context. We contribute to the literature in two ways. First, the real-time out-of-sample performance of twelve individual forecasting models is investigated. This is a larger set of models than typically used in similar forecasting studies. In addition, the use of recursively constructed RMSE ratios discover potential weaknesses of the used models. Thus, several different combination approaches are tested with the result that a combination of individual models is beneficial for the forecasting performance. The second contribution is related to the forecast uncertainty of the investigated models. The forecast density provides an estimate of the probability distribution of future values and offers a complete description of the associated uncertainty. Thereby, the density is based on previous out-of-sample forecast errors and is obtained by bootstrap sampling. The model calibration will then be evaluated by checking whether the probability integral transform is uniformly distributed.

CC1629: **GARCH model for income time series data and forecasting income inequality**
Presenter: **Haruhisa Nishino**, Hiroshima University, Japan

Income inequality is a great issue also in the Japanese economy as well as the world economy. We take an approach by use of a parametric model for analyzing income inequality. Lognormal distribution, which is used, is better fitted to Japanese income data and useful for extracting inequality from income data because its scale parameter only represents inequality. We propose GARCH models including income inequality for income time series data. 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 quasi-likelihood of the GARCH model from quantile income data. The proposed model has an inequality structure and a time series structure. It is useful for analyzing persistent income inequality and forecasting income inequality. The models are estimated from Japanese quantile income data (Family Income and Expenditure Survey by Ministry of Internal Affairs and Communications, Statistics Bureau, Japan) and the various GARCH models are compared and examined from the point of view of forecasting.

CC1684: **Forecasting during the recent financial crisis: Automatic versus adaptive exponential smoothing methods**
Presenter: **Mohaimen Mansur**, The University of Manchester, United Kingdom

Central banks today need to produce reliable forecasts of a large volume of macroeconomic and financial time series in order to make well-informed forward-looking policy decisions. Accurate real time forecasting of these series, however, proved to be particularly challenging following the recent global financial crisis which induced moderate to large structural breaks to many of these series. We advocate use of two types of exponentially smoothing methods which are known to be robust to different types of structural changes. An automatic method chooses the best

forecasting model based on in-sample performance and the adaptive method selects the down-weighting rate of past data based via cross-validation. A systematic forecasting exercise using the FED-MD dataset, an up-to-date data consisting of various macroeconomic and financial time series of the US demonstrates that forecasts of advised exponential methods are comparable to well performing no-change forecasts during the pre-crisis stable period, but the exponential methods outperform the benchmark for many of the series during the crisis and following recovery periods. We further investigate properties of those series for which the gains from exponential methods are particularly large.

CO1434: **Forecasting global temperature with time-series methods**
Presenter: **Alessandro Giovannelli**, University of Venice, Italy
Co-authors: Marco Lippi, Umberto Triacca, Antonello Pasini, Alessandro Attanasio
The impact of climate change on territories, ecosystems and humans has been dramatic in the last fifty years and is likely to become heavier in the next decades, this making modeling and forecasting climate indicators, Global Temperature in particular, of the utmost importance. We propose the use of time-series methods, which are only weakly based on physical knowledge but provide an efficient use of the data available. We study and compare the forecasting performance of two models. The first is a standard Vector AutoRegression (VAR), in which Global Temperature is predicted using past values of (1) Global Temperature itself, (2) the greenhouse gases radiative forcing, GHG-RF, (3) the Southern Oscillation Index, SOI. The second is a Large-Dimensional Dynamic Factor Model (DFM). We use the information contained in 140 time series of local temperatures, corresponding to a grid with spacial resolution of 2.5×2.5 degrees. Our main findings are: (a) The cointegrated VAR, including GHG-RF, and SOI, performs better than the Factor Model at all horizons from 1 to 10 years.(b) However, augmenting the data in the VAR with factors (FAVAR) we obtain a competitive model. Moreover, averaging the forecasts of the FAVAR and the cointegrated VAR we obtain the best results.

CC672 Room 105 CONTRIBUTIONS TO COMPUTATIONAL AND EMPIRICAL ECONOMETRICS Chair: Lorenzo Mercuri

CC0279: **Pricing strategy optimization considering customer sensitivity with Monte Carlo simulations**
Presenter: **Elena Krasheninnikova**, BBVA Data and Analytics, Spain
Co-authors: Roberto Maestre
While setting a price for a given customer and product can be solved with standard regression models like Generalized Lineal Models (GLM), the pricing strategy optimization deals with two main components: on the one hand, the pricing model and, on the other, the customer's degree of acceptance of a given price (sensitivity to the price). Its main goal is to deal with two central objectives: (a) increase retention (probability of acceptance of a given price) and (b) increase revenue. The purpose of this research is to provide a simple framework to calculate the Pareto frontier of several pricing strategies through Random optimization (RO) driven by a probabilistic model, Maximum decay points (MDP) and business constraints. This methodology yields expected retentions and revenues and allows for the testing of new prices. In the experimental section we compare the results obtained in several scenarios taking into account the set of different options presented in the proposed framework (e.g., distributions used by RO, best expected retention). We also focus on computational aspects such as parallel computation, which provides the advantage to independently compute different pricing scenarios through RO.

CC0622: **Google data in bridge equation models for GDP**
Presenter: **Thomas Goetz**, Deutsche Bundesbank, Germany
Co-authors: Thomas Knetsch
There has been increased interest in the use of Big Data in general, and of Google Search data in particular, when it comes to forecasting macroeconomic time series such as private consumption, unemployment or inflation. However, applications on forecasting aggregate GDP, one of the variables predominantly focused on in central Banks, are rather rare. We incorporate Google Search data into a set of Bridge Equations, one of the workhorse models for short-term predictions of German GDP in the Deutsche Bundesbank. We show precisely how to integrate these big data information, emphasizing the appeal of the underlying model for this application. Naturally, the choice of which Google Search terms to add to which equation is crucial not only for the forecasting performance itself, but also for the economic consistency of the implied relationships. We use different ways of selecting the Google Search terms: subjectively, Google Correlate, based on a recursive out-of-sample forecast exercise, by extracting common unobserved factors, and a LASSO-approach. In a pseudo-real time forecast analysis we compare these alternatives in terms of their forecast accuracy for aggregate GDP and various disaggregate time series. We find that there are indeed sizeable gains possible from using Google Search data, which are likely to increase only further in the future, i.e. with ever more data.

CC0594: **Feedback trading and index option prices**
Presenter: **Thorsten Lehnert**, Luxembourg School of Finance, Luxembourg
The channel through which mutual fund flows and stock return volatility are related is studied. Mutual fund investors are known to react on recent market conditions (past performance) and invest (withdraw) their money in (out of) mutual funds. Fund managers are forced to trade accordingly and, therefore, act as feedback traders. We examine whether S& P 500 index option prices are affected by feedback trading. Our testing framework is a heterogeneous agents option pricing model, where some agents are not fully rational. Hence, our agents form different beliefs about the future level of market volatility, and trade options accordingly. We introduce feedback traders, who incorporate noisy signals into their volatility beliefs, next to agents who trade on long-term mean reversion in market volatility and agents who trade on exogenous shocks from the underlying market. The proposed option valuation framework is similar to a stochastic volatility model and is implemented using an augmented filtered historical simulation approach. We find that feedback trading appears to be an important determinant of index option prices. In addition, we find that feedback trading partly explains the term structure of the index option smile, because it primarily affects short-term options.

CC1695: **Estimation of the CAPM with measurement error**
Presenter: **Anastasia Morozova**, University of Konstanz, Germany
Co-authors: Winfried Pohlmeier
The aim is to analyze the impact of measurement errors in the market portfolio on the estimates of the CAPM. In our theoretical set-up, we consider different types of measurement error and derive the consequences for the estimation of the CAPM parameters. Unlike traditional errors-in variable models, the different types of measurement error we introduce are theory based by and result directly from misspecifications of the market supply. We show that besides the well-known attenuation bias for the CAPM betas the CAPM alphas are upward biased leading to spurious excess returns. By exploding cross-equation restriction, we can show that the parameters CAPM with measurement error are identifiable without further instruments. In our Monte Carlo study, we investigate the bias for the CAPM parameters depending on type and size of the measurement error. We derive the conditions under which the true model parameters are feasible. As one of the identification strategies, we consider Instrumental Variable regression in case of many and/or weak instruments constructed from past asset returns. Since the dimensionality of this instrument set may be comparable to the sample size, the regularization and shrinkage techniques are used to obtain and improve the estimates of the model.

EG602 Room 205 CONTRIBUTIONS IN NONPARAMETRIC ESTIMATION Chair: Byeong Park

EC0322: **Strong approximations for a class of integrated empirical processes withapplications to statistical tests**
Presenter: **Salim Bouzebda**, Universite de Technologie de Compiegne, France

The main purpose is to investigate the strong approximation of a class of integrated empirical processes. More precisely, we obtain the exact rate of the approximations by a sequence of weighted Brownian bridges and a weighted Kiefer process. Applications include the two-sample testing procedures together with the change-point problems. We also consider the strong approximation of integrated empirical processes when the parameters are estimated.

EC1549: **Smooth backfitting in errors-in-variables additive models**
Presenter: **Kyunghee Han**, Seoul National University, Korea, South
Co-authors: Byeong Park

Nonparametric additive regression models are studied when noisy covariates are observed within measurement errors. Based on deconvolution techniques, we construct an iterative algorithm for smooth backfitting of additive function in the presence of errors-in-variables. We show that the smooth backfitting achieves univariate accuracy of the standard deconvolution for estimating each component function. Deconvolving noise on backfitting is confined into negligible magnitude that rate of convergence of the proposed estimator is accelerated when the smoothness of measurement errors falls into a certain range. In this case, each component function estimator enjoys asymptotic normality with the oracle variance that can be obtained under the knowledge of the other components. We also present finite sample performance of the deconvolution smooth backfitting in comparison with a naive application of the standard smooth backfitting ignoring measurement errors. Monte Carlo simulation is demonstrated that our method gives smaller mean integrated squared errors than the naive one in average.

EC1644: **Comparison of mean integrated squared errors of parametric and nonparametric estimates**
Presenter: **Dimitrios Bagkavos**, University of Crete, Greece
Co-authors: Prakash Patil

The aim is to investigate the performance of nonparametric and parametric estimates in terms of the Mean Integrated Squared Error when estimating mixtures of normal densities. Particular focus is given on the performance of the estimates when the number of the Gaussian components in the mixture increases.

EC1442: **Non-parametric Bayesian estimation of the diffusion coefficient**
Presenter: **Shota Gugushvili**, Leiden University, Netherlands

A non-parametric Bayesian approach is considered to estimate the deterministic diffusion coefficient of a stochastic differential equation based on discrete time observations on its solution. On the theoretical side, we justify our approach by demonstrating that the posterior distribution asymptotically, as the sample size grows to infinity, concentrates around the 'true' diffusion coefficient and derive the corresponding posterior contraction rate. On the implementational side, we show that our approach is straightforward to implement, requires little fine-tuning from the user, and leads to good practical results.

EG008 Room 202 CONTRIBUTIONS IN MODEL SELECTION IN HIGH DIMENSIONS Chair: Christophe Biernacki

EC1202: **Incorporating pathway information for prediction in omic applications**
Presenter: **Renaud Tissier**, Leiden University Medical Centre, Netherlands
Co-authors: Jeanine Houwing-Duistermaat, Mar Rodriguez Girondo

Nowadays, omics datasets such as transcriptomics, proteomics, and metabolomics are available for building prediction models. To deal with correlation between the omics variables and the large number of variables, regularized regression techniques are typically used. A drawback of these methods is that the results are hard to interpret. To obtain more interpretable prediction models while keeping a good predictive ability, we propose to incorporate pathways information on the omics variables in the prediction models. We use a three-step approach: 1) network construction, 2) clustering to empirically derive modules or pathways, and 3) building a prediction model. For the first step we use two methods, one based on weighted correlation and one based on Gaussian graphical modeling. Identification of modules (groups) of features is performed by hierarchical clustering. To incorporate the grouping information in a prediction model, we adopt two different strategies: group-based variable selection and group-specific penalization. We compare the performance of our new approaches (combinations of network and strategies) with standard regularized regression via simulations. Finally, our approaches are applied to two sets of omic sources (metabolomics and transcriptomics) in the prediction of body mass index (BMI) using longitudinal data from the Dietary, Lifestyle, and Genetic determinants of Obesity and Metabolic syndrome (DILGOM) study, a population-based cohort from Finland.

EC1307: **Model selection in discrete clustering: The EM-MML algorithm**
Presenter: **Claudia Silvestre**, Escola superior de Comunicacao Social, Portugal
Co-authors: Margarida G M S Cardoso, Mario Figueiredo

Finite mixture models are widely used for cluster analysis in several areas of application. They are commonly estimated through likelihood maximization (using diverse variants of the expectation-maximization algorithm) and the number of components (or clusters) is determined resorting to information criteria: the EM algorithm is run several times and then one of the pre-estimated candidate models is selected (e.g. using the BIC criterion). We propose a new clustering approach to deal with the clustering of categorical data (quite common in social sciences) and simultaneously identify the number of clusters - the EM-MML algorithm. This approach assumes that the data comes from a finite mixture of multinomials and uses a variant of EM to estimate the model parameters and a minimum message length (MML) criterion to estimate the number of clusters. EM-MML thus seamlessly integrates estimation and model selection in a single algorithm. The EM-MML is compared with traditional EM approaches, using alternative information criteria. Comparisons rely on synthetic datasets and also on a real dataset (data from the European Social Survey). The results obtained illustrate the parsimony of the EM-MML solutions as well as their clusters cohesion-separation and stability. A clear advantage of EM-MML is also the computation time.

EC1632: **A focused information criterion for high-dimensional data**
Presenter: **Thomas Gueuning**, KU Leuven, Belgium
Co-authors: Gerda Claeskens

Most variable selection procedures have in common that they select one single best model that is used to estimate all the quantities of interest related to the data. This is the case for popular information criteria such as the AIC and the BIC and for penalization procedures such as the LASSO and the SCAD. The Focused Information Criterion (FIC) departs from these methods by performing focused driven variable selection. The FIC selects the model that best estimates a particular quantity of interest (the focus) in terms of mean squared error (MSE). Consequently, different models can be selected for different quantities of interest and the FIC can provide estimators with smaller MSE. An example of such a quantity of interest is the prediction for a new particular observation of the covariates. The current FIC literature is restricted to the low-dimensional case. We show that the FIC idea can be extended to high-dimensional data. We distinguish two cases: (i) the considered submodel is of low-dimension and

(ii) the considered submodel is of high-dimension. In the former case, we obtain an alternative low-dimensional FIC formula that can directly be applied. In the latter case we use a desparsified estimator that allows us to derive the MSE of the focus estimator. We illustrate the performance of the high-dimensional FIC with a numerical study and a real dataset example.

EC1672: **Greedy variable selection on the Lasso solution grid**
Presenter: **Piotr Pokarowski**, University of Warsaw, Poland

The Lasso estimator is a very popular tool for fitting sparse models to high-dimensional data. However, theoretical studies and simulations established that the model selected by the Lasso is usually too large. The concave regularizations (SCAD, MCP or capped-l1) are closer to the maximum l0-penalized likelihood estimator that is to the Generalized Information Criterion (GIC) than the Lasso and correct its intrinsic estimation bias. That methods use the Lasso as a starting set of models and try to improve it using local optimization. We propose a greedy method of improving the Lasso solution grid for Generalized Linear Models. For a given penalty the algorithm orders the Lasso non-zero coordinates according to the Wald statistics and then selects the model from a small family by GIC. We derive an upper bound on the selection error of the method and show in numerical experiments on synthetic and real-world data sets that the algorithm is more accurate than concave regularizations.

EG052 Room 203 CONTRIBUTIONS ON FUNCTIONAL DATA ANALYSIS I Chair: Donatello Telesca

EC1298: **Asymptotic properties of a componentwise ARH(1) plug-in predictor: Comparative case study**
Presenter: **Javier Alvarez-Liebana**, University of Granada, Spain
Co-authors: Maria Dolores Ruiz-Medina

New convergence results on prediction of linear processes in function spaces, in the autoregressive Hilbertian process framework of order one (ARH(1) framework), are presented. Mean-square convergence, in the space of Hilbert-Schmidt operators, of the proposed componentwise estimator of the autocorrelation operator to the theoretical operator is derived, under the assumption that the eigenvectors of the autocovariance operator C are known. Mean absolute convergence of the plug-in ARH(1) predictor, in the Hilbert space considered, is obtained as well. A simulation study is undertaken to illustrate the large sample behavior of the formulated functional parameter estimator and its associated ARH(1) plug-in predictor, in the Gaussian case. In addition, a comparative study is developed to test the performance of the proposed approach in comparison with alternative ARH(1) prediction techniques, based on componentwise estimators of the autocorrelation operator, in the case of known and unknown eigenvectors of C, as well as in terms of functional nonparametric kernel estimators, and spline and wavelet based penalized estimators. The numerical results obtained illustrate the fact that a similar performance of the approaches compared is obtained for small and large sample sizes, and for the different truncation rules tested.

EC1503: **Functional estimation of log-Gaussian Cox process for prediction of risk maps in disease mapping**
Presenter: **Maria Pilar Frias Bustamante**, University of Granada, Spain
Co-authors: Antoni Torres, Maria Dolores Ruiz-Medina

The aim is to present new results on classical and Bayesian estimation and prediction of log-Gaussian Cox processes, whose stochastic intensity is defined by an autoregressive Hilbertian process of order one (ARH(1) process). The functional parameter estimation of the stochastic Hilbert-valued intensity is performed in terms of classical and Bayesian componentwise estimators. Their asymptotic efficiency and equivalence is proved under certain conditions on the class of ARH(1) processes considered. The prediction of the functional values of the random intensity is achieved by the implementation of Kalman filtering in terms of the componentwise parameter estimators formulated. Conditional prediction of the log-Gaussian Cox process is then derived. A conditional simulation study is performed from different types of risk cancer data. EOF-based nonparametric, and nonlinear parametric estimation of the trend of the random intensity, is undertaken to illustrate the properties of the formulated functional parameter estimators, and associated plug-in predictors, comparing results and evaluating the relative performance, in terms of efficiency. Their application to the estimation and prediction of risk maps in disease mapping is contemplated as well, comparing both estimators in terms of the associated functional analysis of variance, and the correlation analysis of the associated functional residuals

EC1554: **Recursive strategies in functional data analysis**
Presenter: **Stella Hadjiantoni**, University of Kent, UK and University of Oviedo, Spain, United Kingdom
Co-authors: Ana Colubi, Erricos John Kontoghiorghes

Choosing the number of basis elements is part of the analysis when dealing with functional data. The dimension of the basis expansion is considered an unknown parameter and investigation is required to determine its value. A recursive numerical method is investigated for choosing the number of basis elements within the context of model selection.

EC1550: **Conditional density estimation of categorical data given functional regressors**
Presenter: **Lena Reichmann**, University of Mannheim, Germany
Co-authors: Melanie Birke, Carsten Jentsch

Nonparametric density and conditional density estimation of mixed continuous and categorical data play a role when logit or probit models are generalised to a nonparametric setting. We follow a previous approach for categorical and real vector of continuous observations and generalise the conditional density estimation for categorical response given functional data. We start with observations $(\mathcal{X}_i, Y_i)$, $i = 1, \ldots, n$ independent identically distributed with the same distribution as the tuple $(\mathcal{X}, Y)$ where Y is a categorical variable with G unordered categories and $\mathcal{X}$ a functional variable located in a (semi-) metric space (E, d) such that $P(Y = y|\mathcal{X} = \chi) =: f_{Y|\mathcal{X}}(y|\chi)$ exists. Our estimator differs from the standard frequency estimator by the possibility of smoothing also over the categorical variable using a discrete kernel and an additional bandwidth. We prove strong consistency, derive convergence rates and show asymptotic normality. Generally, for kernel estimation the choice of bandwidths plays a crucial role to obtain a good finite sample performance. In our setup we adopt a parametric bootstrap approach to select the bandwidths automatically. We compare this approach to a cross validation bandwidth choice.

EG004 Room 217 CONTRIBUTIONS IN EMPIRICAL STATISTICS AND SIMULATIONS Chair: Helle Sorensen

EC1346: **Asymptotic inferences for the odds ratio under the double binomial model**
Presenter: **Maria Alvarez Hernandez**, University of Vigo, Spain
Co-authors: Antonio Martin Andres, Inmaculada Herranz Tejedor

Various asymptotic methods have been proposed for obtaining a two-tailed confidence interval for the odds ratio based on two independent samples. Different methods are evaluated with the following conclusion: a) None of the classic methods has been selected, although the Agresti adjusted logit method is an acceptable and conservative option; b) the best methods are based on the likelihood ratio test and on the arcsine transformation, after adding 0.5 to all the data (their confidence intervals are obtained iteratively); and c) a good and simple alternative to the two previous methods consists in applying the Sterne method to the chi-squared statistic, so that one can obtain an explicit confidence interval. The best methods are also selected to make the conclusions for the CI and the independence test compatible; the conclusions are similar to those above, but a very good and simple alternative is the adjusted Sterne method for chi-squared (the Sterne method above, applied to the increased data in the Agresti values).

EC1669: **Control charts for monitoring the parameters of a logistic distribution**
Presenter: **Ivette Gomes**, FFCUL, Universidade de Lisboa and CEAUL, Portugal
Co-authors: Fernanda Otilia Figueiredo, Mukherjee Amitava
Modeling and monitoring real data from several industrial processes can be a very difficult task. However, if we are able to get a good data fit, the implementation of a parametric control chart for process monitoring is in general better than using a nonparametric procedure in terms of efficiency. We develop control charts for monitoring the parameters of a logistic distribution and its performance is analyzed. A practical application concerning the monitoring of the diameters and lengths of stopper-corks taken from the daily production is also provided to illustrate the adequacy of the logistic distribution to model such type of data and the performance of the previous developed control charts.

EC1585: **Using latent Dirichlet allocation on a large commercial scale**
Presenter: **Erik Mathiesen**, Rekrvt Ltd, United Kingdom
Latent Dirichlet Allocation (LDA) is not just an academic artefact but a very powerful tool in practical applications. We present and share some of our insights and experiences from working with large scale LDA models in a commercial setting. Our applications involve large and diverse corpora that require frequent retraining and adjustments of hyper parameters in a time sensitive manner. We will address a range of challenges encountered, such as: How do you efficiently train and calibrate your model? How do you determine your hyper parameters? How many of them should you have? How often should you update them? We aim to provide a practical guide to commercial usage of LDA with predominant use of heuristics.

EC1605: **Random growth rate and carrying capacity in Verhulst population dynamics**
Presenter: **Maria Brilhante**, Fundacao FCUL, Portugal
Co-authors: Dinis Pestana, Maria Rocha
Either in biological or in economic settings it makes sense to consider that the carrying capacity M and the growth rate ρ in Verhulst population model $\mathrm{d}N(t)/\mathrm{d}t = \rho N(t)[M-N(t)]$ are random. The difference equation $x_{n+1} = r x_n[1-x_n]$, $x_n \in [0,1]$, $n = 1,2,\ldots$ obtained from the discretization of $\mathrm{d}\alpha(t)/\mathrm{d}t = \rho M \alpha(t)[1-\alpha(t)]$, $\alpha(t) = N(t)/M \in [0,1]$, with $\alpha(n) = (1+\rho M)x_n/(\rho M)$, $r = 1+\rho M$ has been thoroughly investigated, namely in what concerns the fixed point search of the stability rate ($x_{n+1} \approx x_n$), with overwhelming success in explaining gross fluctuations of generations size. We investigate the dynamics of the "randomized" logistic parabola $Rx(1-x)$, $x \in [0,1]$, $R = 1+\rho M$, namely the possibility instabilities or of extinction when the growth rate is high. Interesting models, such as generalized Pareto, Morris natural exponential family, extended Panjer basic count models are used to fit R from available time series, and this is the basis for a simulation study on the evolution dynamics and namely on the settings bringing in chaos.

EG250 Room 212 CONTRIBUTIONS IN COPULA MODELING AND ITS APPLICATIONS Chair: Paul Janssen

EC1458: **Measuring dependence between dimensions of welfare using multivariate copula-based coefficients**
Presenter: **Mercedes Prieto-Alaiz**, University of Valladolid, Spain
Co-authors: Ana Perez Espartero
Welfare is multidimensional as it involves not only income, but also education, health or labor. The composite indicators of welfare are usually based on aggregating somehow the information across dimensions and individuals. However, this approach ignores the relationship between the dimensions being aggregated. To face this goal, we analyse the multivariate dependence between the dimensions included in the Human Development Index (HDI), namely income, health and schooling, through three copula-based measures of multivariate association: Spearman's footrule, Gini's gamma and Spearman's rho. We discuss their properties and prove new results on Spearman's footrule. The copula approach focuses on the positions of the individuals across dimensions, rather than the values that the variables attain for such individuals. Thus, it allows for more general types of dependence than the linear correlation. We base our study on data from 1980 till 2014 for the countries included in the 2015 Human Development Report. We find out that though the overall HDI has increased over this period, the dependence between the dimensions included in the HDI (income, health and schooling) remains high and nearly unchanged so that the richest countries tend to be also the best ranked in both health and education.

EC1613: **Modelling maximum and minimum temperature with a bivariate model and a copula to capture dependence across locations**
Presenter: **Xiaochun Meng**, Oxford Unuversity, United Kingdom
Co-authors: James Taylor
The impact of climate change on the probability distribution of daily temperature is investigated. Understanding changes in the frequency, severity and seasonality of daily temperature extremes is important for public policy decisions concerning heat and cold waves. We model the daily maximum and minimum temperature using a bivariate vector ARMA-GARCH model, with conditional dependency modelled using a bivariate copula with a dynamic structure for the dependency parameter. Previous work has focussed on univariate models. Although there are notable differences in the evolution of the daily maximum and minimum temperature series, their simultaneous modelling provides a richer understanding of the time series dynamics, and a useful by-product is the implicit modelling of the diurnal temperature range, which has been used as an index of climate change, and has been identified as a factor affecting human health, such as cardiovascular and respiratory disease. To gain insight into the spatial characteristics of the daily maximum and minimum, we use a pair copula construction to produce the joint distribution of the two variables observed at a set of locations. We model temperature time series from multiple European locations recorded over a 60-year-period. Among the methods that we use to assess goodness-of-fit, we consider the use of the scoring rules to evaluate multivariate density estimation.

EC1630: **Mixture D-vine model based approach to clustering three-way data**
Presenter: **Marta Nai Ruscone**, LIUC, Italy
In real life applications data with a complex structure can be often arranged in a three way data structure, these include data coming from longitudinal studies of multiple responses, spatiotemporal data or data collecting multivariate repeated measures. A three-way data set is characterized by three modes and namely rows, columns and layers. In these type of data there are two types of dependence: between variable and between temporal (or spatial) dependence. Finite mixtures are often used to perform model based clustering of multivariate datasets, but none of the existing methods are developed to reveal simultaneously these two different type of dependences in three-way data. In order to reveal and fully understand the complex and hidden dependence patterns in a wide class of continuous three-way data structure, we propose a mixture model of multivariate densities having D-vine representations. This model decouples the margins and their dependence structure, making it possible to describe the margins by using different distribution families, including non-Gaussian ones. Again, many possible dependence structures can be studied using different copulas.Parameter estimates from simulated and real datasets finally show the suitability of the proposed procedure.

EC1590: **A bivariate semiparametric chart based on the number of observations lying between the control limits**
Presenter: **Elisavet Sofikitou**, University of Piraeus, Greece
Co-authors: Markos Koutras
A new bivariate, Shewhart-type semiparametric control chart is intorduced which takes into account not only the location of a single pair of order statistics, but also the number of specific observations of the test sample that lie between appropriately chosen control limits of the historical sample.

The major advantage of this chart is the fact that the *FAR* and the ARL_{in} values are not affected by the choice of the marginal distributions; although these quantities are typically affected by the dependence structure of the monitored characteristics, they are practically almost the same when different copulas are used. Expressions for the operating characteristic function of the new control chart and the FAR are obtained. In addition, an exact formula is provided for the alarm rate when the bivariate observations follow a specific bivariate distribution. Finally, a detailed numerical study is carried out to assess the performance of the new control chart. The work has been partly supported by the University of Piraeus Research Center.

EG038 Room 204 CONTRIBUTIONS IN STATISTICAL SOFTWARE Chair: Cristian Gatu

EC1511: **IndTestPP: An R package for testing independence between point processes in time**
Presenter: **Ana C Cebrian**, University of Zaragoza, Spain
Co-authors: Jesus Abaurrea, Jesus Asin
IndTestPP (Tests of Independence Between Point Processes in Time) is an R package aiming to provide different parametric and nonparametric tests to check the independence between two or more homogeneous or nonhomogeneous point processes. The following tools are implemented: two tests for Poisson processes, a conditional test and a parametric bootstrap test based on the close point set. The last one is also applicable to any parametric point process which can be simulated. A nonparametric Lotwick-Silverman test, also based on the close point set, is developed and it is applicable to any homogeneous point process, without any distribution model assumption. Versions of the spatial K and J-function adapted to time processes are developed, together with nonparametric Lotwick-Silverman tests based on those measures. A graphical procedure, the Dutilleul plot for two point processes, is also implemented. The package provides tools for generating trajectories of a vector of time point processes with different types of dependence: common Poisson shock processes, networks of queues, multivariate Neyman-Scott process with dependent cluster centers, and marked Poisson process with dependent marks generated by a Markov chain. Different examples of use will be shown.

EC1738: **A controlled error approximate algorithm for regression model selection. A comparison with related packages in R**
Presenter: **Cristian Gatu**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Georgiana-Elena Pascaru, Erricos John Kontoghiorghes
Algorithms for regression model selection are compared in terms of execution times and quality of obtained solution. Specifically, the recently introduced heuristic algorithm (HBBA) and implemented in the R package "lmSubsets" is compared to the genetic and approximate algorithms implemented in R packages. The HBAA yields solutions having relative errors with respect to the optimum that lie within a given tolerance. Thus the quality of HBAA solutions can be properly assessed. Approximate algorithms obtain solutions that are in general not optimal and do not provide any information as how far they are from the optimum. The aim is to compare the HBBA with competitor approximate algorithms. Specifically, the comparison (a) investigates the maximum problem size that can be tackled by these algorithms within a given reasonable computing time; (b) determines the average relative errors of the approximate solutions, say τ, when compared to the optimum solutions which are obtained by HBBA with zero tolerance; and (c) assess the execution times of the approximate algorithms and the HBBA with error tolerance controlled by τ.

EC1582: **Zscore an R package for MCA based score development**
Presenter: **Jone Lazaro**, University of the Basque Country (UPV/EHU), Spain
Co-authors: Maider Mateo-Abad, Arantza Urkaregi, Irantzu Barrio, Inmaculada Arostegui
Surveys are one of the most used assessment tools. The rating scales used for this purpose are mainly ordinal scales such as Likert type scales, where a total score is made by summing up the numerical values corresponding to a response of the multiple items in the questionnaire. However, this commonly used approach assumes that the responses of each item have an underlying linear quantitative measurement. We propose a methodology to develope an score based on Multiple Correspondence Analysis (MCA) which does not assume that all the items have the same importance or relation to the aim under examination. This methodology has been implemented into an R package called Zscore. By providing the questionnaire data set together with the active variables the user is interested in, the main function zscore calculates the score and fits out a table where the first dimensions coordinates of the MCA and the weights of each modality of the active variables are shown. The package handles missing values when it is specified so the user can also decide whether to work with missing values, and how to refer to them. This package provides researchers an easy tool to obtain the overall score of evaluation from the responses to a questionnaire.

EC1676: **EPLMIX: Extended Plackett-Luce models for modeling and clustering ranking data in R**
Presenter: **Cristina Mollica**, Sapienza Universita di Roma, Italy
Co-authors: Luca Tardella
Ranking data represent a peculiar form of multivariate ordinal data taking values in the set of permutations. Despite the numerous methodological contributions to increase the flexibility of ranking data modeling, the application of more sophisticated models is limited by the related computational issues.The EPLMIX package offers a comprehensive framework aiming at adjusting the R environment to the recent methodological advancements. The usefulness of the novel EPLMIX package can be motivated from several perspectives: (i) it contributes to fill the gap concerning the Bayesian estimation of ranking models in R, by focusing on the Plackett-Luce model as generative distribution and its extension within the finite mixture approach; (ii) it combines the flexibility of R routines and the speed of compiled C code, with possible parallel execution; (iii) it covers the fundamental phases of ranking data analysis, allowing for a more careful and critical application of ranking models in real experiments. The functionality of the novel package is illustrated with some real data examples.

EC664 Room 206 CONTRIBUTIONS IN ROBUST METHODS Chair: Ana Maria Bianco

EC1562: **A unified framework of robust PCA: Use of robust unit approach**
Presenter: **Jungeum Kim**, Seoul National University, Korea, South
Co-authors: Hee-Seok Oh
A new framework of robust PCA is proposed for improving the robustness and for reflecting various outlier types as well as skewed data. This framework is composed of two concepts: 1) robust unit that is induced by a combination of any PCA procedure and a restriction function as an outlier filter and 2) two-stage strategy that divides and conquers outliers. As a practical application of the proposed framework, we develop a robust PCA procedure termed robust pair PCA (RP-PCA) by coupling a t-distribution-based probabilistic PCA (T-PCA) with our framework. Moreover, for missing data, we suggest a new procedure for handling missing values that fully exploits EM algorithm of T-PCA under the robust unit. Empirical performance of the proposed method is evaluated through numerical studies including simulation study and real data analysis, which demonstrates promising results of the proposed robust method.

EC1459: **Robust inference in a heteroskedastic multilevel model with structural change**
Presenter: **Ronrick Da-ano**, University of the Philippines-Diliman, Philippines
Co-authors: Erniel Barrios, Joseph Ryan Lansangan
A multilevel model is estimated with cross-sectional interactions in higher and lower levels with structural change by a hybrid procedure of the forward search algorithm preceding bootstrap method. The simulation study exhibits the ability of the hybrid procedure to produce narrower

confidence interval even when there is model misspecification error and structural change. Moreover, it has a comparable predictive ability with the classical restricted maximum likelihood (REML) estimation. However, the hybrid method yield estimates of the parameters with lower bias relative to REML. The hybrid of forward search and bootstrap method can further robustify estimates of fixed and random coefficients under various levels of interclass correlation or in the presence of structural change in a multilevel model.

EC1608: **A new robust quality control chart**
Presenter: **Baris Surucu**, Orta Dogu Teknik Universitesi, Turkey
Co-authors: Didem Egemen

Quality control charts are widely used in manufacturing industry to monitor characteristics of a process. The most popular of these control charts is due to Shewhart. Its derivatives and other control charts proposed in the literature are mainly based on simple random sampling and the assumption of normality. However, the normality assumption fails very much in practice for many data sets. A recent sampling method named as ranked set sampling is also known to be superior over simple random sampling in terms of its efficiency. We propose control charts for nonnormal symmetric distributions based on ranked set sampling. We construct quality control charts for short and long-tailed symmetric distributions under ranked set sampling scheme especially when parameters are unknown. We firstly make use of a robust estimation technique to estimate unknown parameters of the distributions. Secondly, we propose a new ranked set sampling scheme for the estimation of scale parameters. Then, we construct the control charts accordingly. Finally, we conduct a simulation study to show the performances of the newly proposed charts and also give real life examples to explain the subject.

EC1524: **Size on sets of indistinguishable models**
Presenter: **Anne Balter**, Tilburg University, Netherlands
Co-authors: Antoon Pelsser

Models can be wrong and recognising their limitations is important in financial and economic decision making under uncertainty. Finding the explicit specification of the uncertainty set has been difficult so far. We develop a method that provides a plausible set of models to use in robust decision making. The choice of the specific size of the uncertainty region is what we will focus on. We use the Neyman-Pearson Lemma to characterise a set of models that cannot be distinguished statistically from a baseline model. The set of indistinguishable models can explicitly be obtained for a given probability for the Type I and II error.

EG018 Room 214 CONTRIBUTIONS IN GENERALIZED LINEAR MODELS AND BEYOND Chair: Peter Song

EC1610: **On the estimation of the vector generalized linear model**
Presenter: **Panagiotis Paoullis**, University of Oviedo, Spain
Co-authors: Ana Colubi, Erricos John Kontoghiorghes

Vector Generalized Linear Models (VGLMs) is a class of regression models that are limited only by the assumption that the regression coefficients enter through a set of linear predictors. The VGLM class is very large and encompasses a wide range of multivariate response types and models, e.g. it includes univariate and multivariate distributions, categorical data analysis, time series, survival analysis, generalized estimating equations, correlated binary data and nonlinear least squares problems. Models such as Generalised Linear Model, zero-inflated Poisson regression, zero-altered Poisson regression, positive-Poisson regression, and negative binomial regression are all special cases of VGLMs. The algorithm that is employed to find the Maximum Likelihood Estimate (MLE) of VGLM is based on iteratively reweighted least squares (IRLS) and Fisher scoring. Three methods for computing the IRLS of VGLM are presented. The first method transforms the VGLM in each iteration to an ordinary linear model and uses the QR decomposition to find the estimate. The other two employ the generalized QR decomposition to estimate the MLE of VGLM, formulated as iterative generalized linear least-squares problems. Strategies to exploit the special characteristics and properties of the problem are discussed.

EC1266: **Generalized linear mixed models for binary responses: An application to resignation from the Brazilian Army**
Presenter: **Cleber Iack**, Faculdade de Ciencias - Universidade de Lisboa / Ministerio da Defesa - Brasil / Bolsista do CNPQ N Brasil, Portugal
Co-authors: Helena Mourino

Modeling the relationship between explanatory variables and a response variable is a crucial task in Statistics. When the response variable is binary, the generalized linear model is used. However, this model assumes that observations are independent of each other. For clustered and/or longitudinal data, the fixed effects model cannot be applied any more. In this context, there is the need to incorporate the clustered (observations are nested within larger units) and/or longitudinal (repeated observations are nested within subjects) nature of the data. The resulting model is a generalized linear mixed model, which can incorporate both fixed and random effects for the regressors. For longitudinal data, the random effects models allow the regression coefficients to vary between subjects, which describe the effect of each individual on its own repeated observations. An application of the model to resignation from the Brazilian Army is given. It aims at identifying the variables intrinsically related to early exit. From 2011 to 2014, 16540 militaries were analysed. The sampled data was clustered by military's background (IME, AMAN, EsSEx and EsAEx). Also, each military was evaluated by 5-points Likert scale on a semester basis. Models with different random effects specification were compared using the likelihood ratio tests.

EC1354: **Generalized additive models with flexible link function**
Presenter: **Elmar Spiegel**, University of Goettingen, Germany
Co-authors: Fabian Sobotka, Thomas Kneib

Ordinary generalized linear models (GLM) depend on several assumptions: (i) the specified linear predictor and (ii) the pre-specified likelihood/link function. In order to avoid the restriction of linear predictors, generalized additive models (GAM) with semiparametric predictors have been considered. The covariates may be included as flexible nonlinear or spatial functions to avoid biased estimates. The influence of misspecified link functions has been shown before and single index models are a common solution which results in a GLM with a flexibly estimated link function. However these single index models are usually restricted to a linear predictor and aim to compensate for the non-linear structure with the estimated link. We show that this is insufficient and present a solution by combining a flexible link estimation with GAM. The link functions are estimated with (strictly) monotone p-splines. Due to the monotonicity constraint, the estimations are identifiable and the results interpretable.

EC1406: **Balancing disclosure risk with utility for generalized linear model output in a remote analysis system**
Presenter: **Atikur Khan**, Commonwealth Scientific and Industrial Research Organisation, Australia
Co-authors: Christine OKeefe

A remote analysis system (RAS) allows analysts to obtain statistical analysis results without providing direct access to confidential data stored in a secure server. However statistical analysis results can reveal information about the underlying confidential data. Statistical disclosure control (SDC) methods are used to confidentialize RAS outputs in order to protect the confidential data. Confidentiality protection through perturbation is one of the most commonly adopted SDC methods. This method adds random noise to the estimated coefficients or to the associated estimating equation. Thus the application of any perturbation based SDC method could result in an inefficient estimator, with the danger that the inefficient estimator may result in worthless inferences. To date, little attention has been given to systematically controlling both the disclosure risk and utility

in SDC methods for RAS. We develop a framework for the perturbation of estimating equations that enables a RAS to release confidentialized generalized linear model (GLM) output in such a way that the disclosure risk does not exceed a pre-selected threshold level whilst maintaining a good utility of data. Finally, we present some empirical results demonstrating the application of our framework for confidentialization of GLM coefficients.

EC658 Room 201 CONTRIBUTIONS IN BAYESIAN METHODS Chair: Michael Daniels

EC1575: **A Bayesian interpretation of data-driven estimation by model selection**
Presenter: **Xavier Loizeau**, Ruprecht-Karls-Universitat Heidelberg, Germany
Co-authors: Jan Johannes
Considering an indirect Gaussian sequence space model and a hierarchical prior, oracle/minimax-optimal concentration and convergence of the associated posterior and Bayes estimator are shown. Notably, the hierarchical prior does not depend neither on the true parameter value nor on the given class. The posterior is taken iteratively as a new prior and the associated posterior is calculated for the same observation again and again. Thereby, a family, indexed by the iteration parameter, of fully data driven priors is constructed. Each element of this family leads to an oracle/minimax-optimal concentration and convergence of the associated posterior and Bayes estimator as the noise level tends to zero. For a fixed noise level letting the iteration parameter tend to infinity, the associated posterior distribution shrinks to a point measure. The limit distribution is degenerated on the value of a projection estimator with fully-daten driven choice of the dimension parameter using a model selection approach where a penalized contrast criterium is minimised. Thereby, the classical model selection approach gives in some sense an infinitely increasing weight to the information contained in the observations in comparison to the prior distribution. It is further shown that the limit distribution and the associated Bayes estimator converges with oracle and minimax-optimal rates as the noise level tends to zero.

EC0924: **The use of homogeneous scoring rules in Bayesian model selection**
Presenter: **Monica Musio**, University of Cagliari, Italy
Co-authors: Philip Dawid
The desire for an "objective Bayesian" approach to model selection has produced a wide variety of suggested methods, none entirely satisfactory from a principled perspective. We explore the implications, for Bayesian model selection, of replacing the log score by some other proper scoring rules. These can be used as benchmarks for comparing the quality of statistical models. In particular we will se that using homogeneous proper scoring rules, we bypass the problems link to the use of improper priors. When applied prequentially, these will typically enable consistent selection of the true model. Examples in the continuous and discrete cases will be given.

EC1426: **Comparison of parametric and semi-parametric binary response models**
Presenter: **Xiangjin Shen**, Bank of Canada, Canada
A Bayesian semi-parametric binary choice model is proposed using the quasilikelihood function as the likelihood part of the posterior distribution. We compare the performances of the Bayesian semi-parametric model with the sample theory semi-parametric model. Also we compare the semi-parametric models with probit and logit parametric models. The comparisons are based on simulated data and Monte Carlo experiments. As the criteria of comparison we use the marginal effect, mean squared error (MSE) and receiver operating characteristic (ROC) curve. What we find are (i) when the data is balanced the performances of the semi-parametric models are indistinguishable from the performances of the parametric models (i.e. probit and logit models) (ii) However, when the data is extremely unbalanced (the yes response rate being less than 3%), the maximum likelihood estimation of the semi-parametric as well as the parametric models does not converge, whereas the Bayesian estimation converges. After the simulated data and Monte Carlo experiments, we use the US Panel Study of Income Dynamics (PSID) and Canadian Socio-economic Information Management System (CANSIM) data, and test the robustness of the Bayesian semi-parametric binary choice model and other binary choice models.

EC1716: **Approximate Bayesian computation methods for phase-type distributions**
Presenter: **Concepcion Ausin**, Universidad Carlos III de Madrid, Spain
Co-authors: Pedro Galeano, Simon Wilson
Phase type (PH) distributions are a very flexible class of models to describe failure times. This family includes the exponential, Erlang, mixtures of exponentials and Coxian distributions as particular cases. Given a set of observed failure times, classical procedures based on the method of moments and maximum likelihood estimation have been proposed to fit a PH distribution. However, it is not easy how to derive confidence intervals for any quantities of interest depending on the PH estimated parameters. Alternatively, this can be done in a natural way from the Bayesian perspective using Markov Chain Monte Carlo (MCMC) methods. Nevertheless, classical and Bayesian methods for PH distributions are usually based on sophisticated computational algorithms which are very time consuming, mainly due to the difficulties in the evaluation of the likelihood. In order to solve these problems, our proposal is to make use of the recently developed Approximate Bayesian Computation (ABC) methods, which offer the advantage of avoiding the evaluation of the likelihood and are mainly based on simulation. We believe that the ABC methodology can provide a potential alternative to MCMC for PH distributions since these models are very easy to simulate but their likelihood is usually difficult and/or computationally expensive to evaluate.

EG098 Room 213 CONTRIBUTIONS IN QUANTILE REGRESSION Chair: Juan Carlos Escanciano

EC1653: **A new lack-of-fit test for quantile regression models using logistic regression**
Presenter: **Mercedes Conde-Amboage**, University of Santiago de Compostela, Spain
Co-authors: Valentin Patilea, Cesar Sanchez-Sellero
A new lack-of-fit test for parametric quantile regression models is proposed. The test is based on interpreting the residuals from the quantile regression model fit as response values of a logistic regression, the predictors of the logistic regression being functions of the covariates of the quantile model. Then a correct quantile model implies the nullity of all the coefficients but the constant in the logistic model. Given this property, we use a likelihood ratio test in the logistic regression to check the quantile regression model. In the case of multivariate quantile regressions, to avoid working with a large dimension logistic regression, we use predictors obtained as functions of univariate projections of the covariates from the quantile model. Finally, we look for a 'least favourable' projection for the null hypothesis of the likelihood ratio test. Our test can detect general departures from the parametric quantile model. To approximate the critical values of the test, a wild bootstrap mechanism is used. A simulation study shows the good properties of the new test versus other nonparametric tests available in the literature.

EC1200: **The method of simulated quantiles for regression parameters estimation.**
Presenter: **Paola Stolfi**, Roma Tre University, Italy
Co-authors: Lea Petrella
A method of simulated quantile (MSQ) is developed which generalise the matching sample quantile methods and the method of simulated moments, in order to estimate parameters in all the situations where the standard methods like ML, GMM and SMM may be difficult or impossible to apply. We will consider regression models $y|x \sim F(\theta)$ where F may not have an analytic form and/or moments in which x can be both a vector of

exogenous or lagged variables. In order to implement the MSQ method we introduce a statistic $T(y,x)$ which is informative for the parameters of the models considered, and we match its sample quantiles and theoretical ones, the last simulated from $F(\theta)$, to find the values of the unknown parameters. To show the performances of the method we will consider regression and autoregressive models when different error term distributions are assumed. In addition to test the robustness of the method we also consider the case when contaminated data are observed. We extend the results to conditional variance models in order to take into account for more general non linear and heroscedasticity situations.

EC1409: **A joint (value-at-risk, expected shortfall) regression framework with financial applications**
Presenter: **Timo Dimitriadis**, University of Konstanz, Germany
Co-authors: Sebastian Bayer
The aim is to propose a joint regression framework which extends quantile regression to the pair Value-at-Risk (VaR) and Expected Shortfall (ES). It is well known that ES alone is not elicitable and can thus not be modeled through a regression. However, a recent result in the literature shows that the pair comprising VaR and ES is jointly elicitable. This result allows for a joint regression of the pair (VaR, ES) on covariates using M-estimation. We show consistency and asymptotic normality of the parameter estimates using nonstandard conditions as the loss function has discontinuity points. A Monte Carlo simulation study verifies the asymptotic distribution of the parameter estimates for homoscedastic and heteroscedastic regression designs. We propose a new backtest for ES based on the joint regression framework. Our backtest exhibits almost correct size properties and dominates existing ES backtests in terms of its size-corrected power. Our general regression framework allows for further interesting applications such as predicting the pair (VaR, ES) by daily realized measures as predictor variables or an extension of the CAViaR approach to ES.

EC0437: **Least angle quantile lasso regression**
Presenter: **Jing Huang**, European School of Management Technology, Germany
The least angle method is used to select variables in quantile regression. Quantile regression has large applications in understanding tail distribution, and variable selection is helpful to study the regression under limited resources. The least angle method chooses variables by selecting the minimum angle between dependent variables and regressors rather than minimizing the distance, in which case least angle method is believed to have fast speed. Furthermore, we found that the current algorithm for selecting variables in quantile regression is fast in speed, but not sufficient to find the optimal variables set, and an exhaustive method is prohibitive. Therefore, we used an approximation to transform the minimization problem into a simple OLS problem turn the selecting problem into an outlier detecting problem. The we suggest several ways, like least trimmed square and PCA (factor analysis), to improve the algorithm to selection variables at each searching step more efficiently. At last, we conduct a simulation study, which shows PCA method has the best performance among all the algorithms we tested.

Sunday 11.12.2016 16:20 - 18:00 Parallel Session P – CFE-CMStatistics

CG198 Room 110 CONTRIBUTIONS IN FINANCIAL ECONOMETRICS Chair: Walter Distaso

CC0169: **A generalised model of typical and atypical news transmission**
Presenter: **Sarantis Tsiaplias**, Melbourne University, Australia
Co-authors: Chew Lian Chua
The purpose is to examnine the impact of information cross-flows on stock prices, and how information from a given stock affects the prices of other stocks. A model is developed that categorizes a stock's news into typical and atypical news, and allows both types of news to be filtered through to other stocks. We explore the relationship between news and the expected volatility surfaces and show how the concavity of the surfaces depends on time, news magnitude and news type. The surfaces change significantly in the presence of atypical news such that a stock can be both a substitute and a complement.

CC1581: **A new kurtosis matrix, with statistical applications**
Presenter: **Nicola Loperfido**, University of Urbino, Italy
The number of fourth-order moments which can be obtained from a random vector rapidly increases with the vector's dimension. Scalar measures of multivariate kurtosis may not satisfactorily capture the fourth-order structure, and matrix measures of multivariate kurtosis are called for. We propose a kurtosis matrix derived from the dominant eigenpair of the fourth standardized moment. We show that it is the best symmetric, positive semidefinite Kronecker square root approximation to the fourth standardized moment. Additional properties are derived for realizations from GARCH and reversible random processes. Statistical applications include independent component analysis and projection pursuit. The star product of matrices highlights the connection between the proposed kurtosis matrix and other kurtosis matrices which appeared in the statistical literature. A simulation study assesses the practical relevance of theoretical results.

CC1597: **Detecting Granger causality with a nonparametric information-based statistic**
Presenter: **Hao Fang**, University of Amsterdam, Netherlands
Co-authors: Cees Diks
Testing causal effects has attracted much attention in the domains of many disciplines since Grangers pioneering work. The recent literature shows an increasing interest in testing for Granger non-causality in a general sense by nonparametric evaluation of conditional dependence. We introduce a novel nonparametric test based on the first order Taylor expansion of an information theoretic measure: transfer entropy. This new test statistic is shown to have an information-based interpretation for Granger non-causality. The proposed test avoids the impotence of the frequently-used test previously proposed as a result of the lack of the positive definiteness under some alternative circumstances. Asymptotic normality of the test statistic is achieved when all densities are estimated with appropriate sample-size dependent bandwidth, and practical guidelines for choosing bandwidth are formulated for specific cases. Simulation result confirms the usefulness of this test. Finally an application to financial data indicates the existence of bidirectional Granger causality.

CC0184: **Likelihood ratio tests for explosive financial bubble with application of ruble/dollar exchange rate**
Presenter: **Elena Sinelnikova-Muryleva**, RANEPA, Russia
Co-authors: Anton Skrobotov
New likelihood ratio tests are proposed for explosive financial bubbles. The null hypothesis is a unit root behavior throughout while under the alternative there is explosive behavior in some subsample of the series. For unknown dates of origin and collapse, the sup, average and exponential test statistics are proposed. These statistics reveal higher power than other existing tests based on Dickey-Fuller regression for almost all cases considered; and these statistics are robust to the four-regime specification (unit root, then explosive process, then stationary process, then unit root) and multiple bubble specifications. An empirical application to Russian exchange rate shows existence of the bubble in the 2nd half of 2014 and detects the earlier date of bubble originating than is usually supposed by the specialists.

CC0895: **Partial copula methods for models with multiple discrete endogenous explanatory variables and sample selection**
Presenter: **Myoung-Jin Keay**, South Dakota State University, United States
A flexible parametric approach is presented for models with multiple discrete endogenous explanatory variables (EEV) with finite support. The joint distributions of each EEV and structural error are modeled by using copulas and their marginal distributions, but the ones among the EEV's are left unspecified. Our partial copula approach can be applied in any models with discrete EEV's. It can be also used for correcting selection bias and finding average treatment effects.

CG304 Room 102 CONTRIBUTIONS ON NETWORK ANALYSIS Chair: Carsten Jentsch

CC0257: **News and network structures in equity market volatility**
Presenter: **Yin Liao**, Queensland University of Technology, Australia
Co-authors: Adam Clements
An understanding of the linkages between assets is important for understanding the stability of markets. Network analysis provides a natural framework within which to examine such linkages. The impact of firm specific news arrivals on the interconnections is examined at an individual firm and overall portfolio level. While a great deal of research has focused on the impact of news on the volatility of a single asset, much less attention has been paid to the role of news in explaining the links between assets. It is found that the both the volume of news and its associated sentiment are important drivers the connectedness between individual stocks and the overall market structure. Firms that experience negative news arrivals during periods of market stress become more centrally important in the market structure.

CC1589: **A new approach to building the interindustry input-output table using block estimation techniques**
Presenter: **Ryohei Hisano**, University of Tokyo, Japan
A new approach is presented to estimate the interdependence of industries in an economy by applying data science solutions. By exploiting interfirm buyer-seller network data, we show that the problem of estimating the interdependence of industries is similar to the problem of uncovering the latent block structure in network science literature. To estimate the underlying structure with greater accuracy, we propose an extension of the sparse block model that incorporates node textual information and an unbounded number of industries and interactions among them. The latter task is accomplished by extending the well-known Chinese restaurant process to two dimensions. Inference is based on collapsed Gibbs sampling, and the model is evaluated on both synthetic and real-world datasets. We show that the proposed model improves in predictive accuracy and successfully provides a satisfactory solution to the motivated problem. We also discuss issues that affect the future performance of this approach.

CC1586: **Analysis of multivariate time series using Wavelet dependence graphs**
Presenter: **Maria Grith**, HU Berlin, Germany
Co-authors: Matthias Eckardt
Localized partial correlation and Granger causality graphs are defined for possibly nonstationary multivariate time series using wavelet-based methods. These graphs describe the pairwise conditional time-varying linear dependence structure and account for the contemporaneous and lagged influences. The edges are characterized in terms of a local partial coherence measure, defined by the time-series pairwise correlations decomposed in the time-frequency domain. This measure is interpreted as a partial error localized spectral correlation function, and can be computed without prior fit of any linear filter. We estimate weighted and sparse graphs for observed data and illustrate our method in an empirical study.

CC1532: **Statistical inference for financial connectedness**
Presenter: **Ruben Hipp**, University of Mannheim, Germany
Co-authors: Carsten Jentsch
By using forecast error variance decompositions to identify networks, some authors established a new standard in estimating networks by introducing a concept of financial connectedness. Although many papers devoted their application to this concept of financial connectedness, the corresponding literature is lacking rigorous asymptotic theory that allows for statistical inference. To fill this gap, we make use of the framework of locally stationary processes. For this purpose we propose a general class of local-linear estimators for the time-varying VAR coefficient matrices and the innovations variance matrix that covers e.g. the previous approach. We derive explicit expressions for the limiting bias and variance and prove a CLT for these estimators to deduce corresponding limiting results for various connectedness measures. As the limiting distributions turn out to be complicated, we propose a model-based bootstrap procedure that builds on our estimates. In an application on the U.S. financial market we adopt this bootstrap method to construct confidence intervals. Following these results, we point out practical issues in the setting of financial connectedness and advice on how to handle bandwidth selection and estimation imprecision in periods of financial stress.

CC1502: **Dynamic correlation network analysis of Japanese stock returns**
Presenter: **Takashi Isogai**, Tokyo Metropolitan University, Japan
The dynamic correlation network of Japanese stock returns is analyzed to study the correlation structure as well as correlation dynamics of the stock market empirically. Stock groups are generated by correlation network clustering in order to work around the high dimensionality problem due to the large number of stocks. Such data-oriented group definition is more reliable than the existing sector classification. Homogeneous groups of stocks in a balanced size are created by segregating the whole stock returns by recursive modularity optimization; a single portfolio that comprises group portfolio returns is also created. Thus, within- and between-group dynamic correlation networks are built, respectively. We, first, confirmed that a higher level of correlation is observed during the crisis periods, namely after the Lehman shock and the Great East Japan Earthquake. We also identify significant differences in the pattern of correlation dynamics between groups. Then, dynamic changes in the network topology measures including density, centrality, and heterogeneity are examined in each correlation network; significant changes in network topologies are detected.

CG637 Room 103 CONTRIBUTIONS IN HIGH FREQUENCY ECONOMETRICS Chair: Kim Christensen

CC0272: **The drift burst hypothesis**
Presenter: **Kim Christensen**, Aarhus University, Denmark
The usual tenet that volatility dominates over the drift over short time intervals is not necessarily true when the drift term is locally explosive. The Drift Burst Hypothesis postulates the existence of such locally explosive drifts in the price dynamics. We first show that drift explosions can arise naturally in markets with heterogeneous information in which partially informed agents learn from prices. We then lay down a nonparametric identification strategy for drift bursts, which is embedded in the paradigm of traditional continuous-time finance. The empirical application of the test shows that the occurrence rate of drift bursts is quite large, and that they can most often be associated to flash crashes, since the return during a drift burst tends to revert immediately thereafter. This allows to contrast existing theories describing the impact of illiquidity frictions on price dynamic formation.

CC1596: **Understanding the distribution of volatility**
Presenter: **Martin Thyrsgaard**, Aarhus University, Denmark
Co-authors: Kim Christensen, Bezirgen Veliyev
A noise robust estimator is constructed for the cumulative distribution function of the invariant distribution of the latent spot volatility process of an asset price. As a first step towards constructing this estimator, we derive a noise robust estimator of the volatility occupation time over a fixed time span. This step relies on the asset price being observed over a grid with mesh going to zero. Noise robust estimators typically converge at a slower rate, however, a Monte Carlo study suggests that this cost is greatly outweighed by the benefits from being able to use more observations. In the second step, we let the time span tend to infinity while still letting the distance between observations tend to zero, thereby obtaining a consistent estimator of the cumulative distribution function of the invariant distribution. We use this estimator to construct a Kolmogorov-Smirnov type test for the invariant distribution of the latent volatility process. Finally, we apply the newly developed methods to a set of ultra high-frequency equity tick data.

CC1659: **Dependent microstructure noise and integrated volatility estimation from high-frequency data**
Presenter: **Merrick Zhen Li**, University of Amsterdam, Netherlands
Co-authors: Roger Laeven, Michel Vellekoop
We develop econometric tools to study integrated volatility with potentially time-dependent microstructure noise in high-frequency data. In this context, we first develop consistent estimators of the variance and covariances of noise using a variant of realized volatility. Next, we adapt the pre-averaging method and derive a consistent estimator of the integrated volatility, which converges stably with optimal rate $n^{(-1/4)}$ to a mixed Gaussian distribution. In a finite sample analysis, we find that the second moments of noise and the integrated volatility induce biases to each other, and we propose novel two-step estimators to correct the "interlocked" bias. Our extensive simulation studies demonstrate the excellent performance of our estimators. Empirically, we find strong evidence of positively autocorrelated noise in the examined stocks and show the considerable accuracy gains achieved by our new estimators.

CC1675: **Forecasting the daily spot volatility paths of equity indices via functional autoregressive models: An empirical study**
Presenter: **Isao Ishida**, Konan University, Japan
The performance of functional autoregressive models in forecasting daily spot volatility paths is empirically investigated using high frequency intraday data of S& P 500 and Nikkei 225 equity indices. The functional data analysis involves smoothing in obtaining functional representations from discretely observed data. In this step, we apply some of the methods, such as the Fourier transform method, developed specifically for estimating daily spot volatility paths to the observations of high frequency equity index returns, and treat the obtained daily functions as observed realizations of the daily spot volatility paths (unlike in the case of the functional ARCH/GARCH models). We then use the functional autoregressive models for modeling and forecasting the daily spot volatility paths, and find some improvement in forecast accuracy over several alternative procedures for spot volatility forecasting.

CC1474: **Forecasting volatility in stock market: The gains from using intraday data**
Presenter: **Xingyi Li**, University of Agder, Norway
Co-authors: Valeriy Zakamulin
The aim is to comprehensively assess the gains in volatility forecasting accuracy provided by realized measures of daily volatility versus that provided by using daily returns. We extend previous studies on forecasting stock market volatility in several directions. First, we use an extensive set of intraday data on prices of single stocks and stock market indices. Second, we assess the gains in forecast accuracy provided by using intraday data over multiple horizons, ranging from 1 day to 6 months. Third, we compare forecasting abilities of several competing models. Our results indicate that there are marginal differences between forecast accuracies provided by alternative models. The major finding of our empirical study is that, regardless of the length of the forecasting horizon, intraday data allow one to reduce the mean squared forecasting error up to 35%. Thus, the gains from using intraday data are highly economically significant.

CG382 Room 109 CONTRIBUTIONS ON HYPOTHESIS TESTING FOR ECONOMETRIC MODELS Chair: Anders Kock

CC0358: **Testing for prospect and Markowitz stochastic dominance efficiency**
Presenter: **Nikolas Topaloglou**, Athens University of Economics and Business Research Center, Greece
Co-authors: Stelios Arvanitis
Non-parametric tests are developed for prospect stochastic dominance Efficiency (PSDE) and Markowitz stochastic dominance efficiency (MSDE) with rejection regions determined by block bootstrap re-sampling techniques. Under the appropriate conditions we show that they are asymptotically conservative and consistent. We engage into Monte Carlo experiments to assess the finite sample size and power of the tests allowing for the presence of numerical errors. We use the tests to empirically establish whether the value-weighted market portfolio is the best choice of individuals with preferences satisfying locally convex utility schemes. Our results indicate that we cannot reject the hypothesis of prospect stochastic dominance efficiency for the market portfolio.This is supportive of the claim that the particular portfolio can be rationalized as the optimal choice for any S-shaped utility function. Instead, we reject the hypothesis for Markowitz stochastic dominance,which could imply that there exist reverse S-shaped utility functions that do not rationalize the market portfolio.

CC0417: **Backtesting marginal expected shortfall and related systemic risk measures**
Presenter: **Jeremy Leymarie**, LEO - CNRS - University of Orleans, France
Co-authors: Denisa Georgiana Banulescu, Christophe Hurlin, Olivier Scaillet
Two backtesting tests are proposed to assess the validity of the systemic risk measure forecasts. This new tool meets the need of financial regulators of evaluating the quality of systemic risk measures generally used to identify the financial institutions contributing the most to the total risk of the financial system (SIFIs). The tests are based on the concept of cumulative violations and it is built up in analogy with the recent backtesting procedure proposed for ES (Expected Shortfall). First, we introduce two backtests that apply for the case of the MES (Marginal Expected Shortfall) forecasts. The backtesting methodology is then generalised to MES-based systemic risk measures (SES, SRISK) and to the Delta CoVaR. Second, we study the asymptotic properties of the tests in presence of estimation risk and we investigate their finite sample performances via Monte Carlo simulations. Finally, we use our backtests to asses the validity of the MES, SRISK and Delta CoVaR forecasts on a panel of EU financial institutions.

CC1571: **Testing for explosive bubbles in the presence of autocorrelated innovations**
Presenter: **Erik Christian Montes Schutte**, Aarhus University, Denmark
The aim is to analyze an empirically important issue with recently developed recursive right-tailed unit root tests for bubbles in asset prices. First, we show that serially correlated innovations, which is a feature that is present in most financial series used to test for bubbles, can lead to severe size distortions when using either fixed or automatic (based on information criteria) lag-length selection in the auxiliary regressions underlying the tests. Second, we propose a sieve bootstrap version of these tests and show that this results in more or less perfectly sized test statistics at virtually no cost in power. Finally, we show an empirical application of the bootstrap versions of the tests on the housing markets of OECD countries.

CC0428: **New goodness-of-fit diagnostics for conditional discrete response models**
Presenter: **Igor Kheifets**, ITAM, Mexico
Co-authors: Carlos Velasco
New specification tests are proposed for conditional models with discrete responses, which are key to apply efficient maximum likelihood methods, to obtain consistent estimates of partial effects and to get appropriate predictions of the probability of future events. In particular, we test the static and dynamic ordered choice model specifications and can cover infinite support distributions for e.g. count data. The traditional approach for specification testing of discrete response models is based on probability integral transforms of a jittered discrete data which leads to continuous uniform iid series under the true conditional distribution. Then, standard specification testing techniques for continuous variables could be applied to the transformed series, but the extra randomness from jitters affects the power properties of these methods. We investigate an alternative transformation based only on original discrete data that avoids any randomization. We analyze the asymptotic properties of goodness-of-fit tests based on this new transformation and explore the properties in finite samples of a bootstrap algorithm to approximate the critical values of test statistics which are model and parameter dependent. We show analytically and in simulations that our approach dominates the methods based on randomization in terms of power. We apply the new tests to models of the monetary policy conducted by the Federal Reserve.

CC1365: **Model-free tests for the null hypothesis of stochastic trendless**
Presenter: **Julio Angel Afonso-Rodriguez**, University of la Laguna, Spain
There exists large evidence indicating that many macroeconomic and financial time series can be well characterized as containing a stochastic trend component, but exhibiting some periods of exuberance or high instability. Among some existing alternatives, this departure from the pure I(1) behaviour, or even from the local-level model, could be explained by the inclusion of an additional component allowing to capture these periods of excessive variability. The resulting generating mechanism is called stochastically integrated, and nests both conventional integration and stationarity, and the so-called case of heteroskedastic integration (HI). We study the theoretical properties of some commonly used testing procedures for the null hypothesis of conventional integration and stationarity under stochastic integration, and found that their outcomes can lead to wrong spurious evidence of stationarity or integration, respectively. Given these results, we adapt two existing statistics in the context of stochastic cointegration to test for the null of a stochastically trendless (ST) process against the alternative of I(1) and to test for the null of standard stationarity against the alternative of ST. Finally, we also consider the possibility of using modified versions of these procedures to test for the null of I(1) against the alternative of HI.

CC666 Room 101 CONTRIBUTIONS IN BAYESIAN ECONOMETRICS Chair: Roberto Leon-Gonzalez

CC0499: **Investment dynamics in advanced economies: Evidence from a Bayesian panel VAR**
Presenter: **Bjoern van Roye**, European Central Bank, Germany
Co-authors: Ramon Gomez-Salvador
The importance of business investment dynamics is shown to vary across in Advanced Economies (AE). To account for heterogeneity across AEs, we use a Bayesian Panel VAR in linear hierarchical form. While business investment is not particularly low in the United States, the United Kingdom and Japan, growth has been persistently low in a number of European countries. We show that although there is a common component in political uncertainty, financial conditions, demand expectations and firm profitability across countries, the importance of the drivers remain substantially heterogeneous and depend on the individual country's structure.

CC1560: **Bayesian analysis of alternative long memory stochastic volatility models using realized volatility**
Presenter: **Manabu Asai**, Soka University, Japan
In recent years fractionally differenced processes have received a great deal of attention due to its exibility in nancial applications with long memory. We consider a class of models generated by Gegenbauer polynomials, incorporating the long memory in stochastic volatility (SV) components in order to develop the General Long Memory SV (GLMSV) model. We examine the statistical properties of the new model, suggest using the spectral likelihood estimation for long memory processes, and investigate the nite sample properties via Monte Carlo experiments. We apply the model to three exchange rate return series. Overall, the results of the out-of-sample forecasts show the adequacy of the new GLMSV model.

CO0865: **Consumer clustering model based on the time of new product adoption using ID-POS data**
Presenter: **Masataka Ban**, Nihon University College of Economics, Japan
In marketing literature, the consumer behavior of selecting one brand from various goods is called "brand choice". For consumer heterogeneity, in general, brand choice behavior is modeled by hierarchical Bayes discrete choice model like logit or probit which have consumer's individual-level parameter. A brand choice model is proposed for consumer clustering in terms of a new product adoption. In particular, the model is constructed by hierarchical Bayes probit model having a Dirichlet process (DP) prior with time ordering clustering constraint. Features of this model is that (1) the model enables the estimation of the number of clusters, and then it is not necessary to set that before analysis. (2) Time ordering clustering leads to estimation of breakpoints among consumer clusters. The consumers are categorized into an adequate time ordering cluster based on the similarity of their market response. The model estimates provide useful information corresponding to the marketing concepts containing time ordering clusters (e.g. Roger's diffusion of innovation theory, product life cycle management). The model is estimated by Markov Chain Monte Carlo sampling method. A Metropolis-Hastings-based algorithm modified to fulfill the constraint is used.

CO0937: **An analysis of TV viewing behavior using a direct utility model**
Presenter: **Shohei Hasegawa**, Hosei University, Japan
Co-authors: Satoshi Nakano, Sang-gil Lee
The purpose is to understand TV viewer behavior. In marketing literature, TV viewing data are usually used as viewing frequencies of TV advertisement and independent variables for measuring advertising effects. The authors use TV viewing data as dependent variables and analyze TV viewer's behavior using a direct utility model. The proposed model assumes that an individual allocates viewing time to different type of TV program genres by maximizing his or her utility subject to a time budget constraint. The model has two parameters, the first is baseline parameter which means a preference and the other is satiation parameter which means a rate of diminishing marginal utility. Heterogeneity into the model is introduced using a hierarchical Bayes model and estimate individual parameters. The empirical analysis using actual data shows that (1) viewing behaviors are different between weekdays and weekends, (2) news show genre has high preference on weekdays and variety show genre on weekends, (3) educational show genre is the most satiating both on weekdays and weekends. We apply the results to clustering viewers according to their estimates.

CC0238: **Identifying shocks in structural VAR models via heteroskedasticity: A Bayesian approach**
Presenter: **Dmitry Kulikov**, Eesti Pank, Estonia
Co-authors: Aleksei Netsunajev
A contribution to the literature on statistical identification of macroeconomic shocks is made by proposing a Bayesian VAR with time-varying volatility of the residuals that depends on a hidden Markov process, referred to as an MS-SVAR. With sufficient statistical information in the data and certain identifying conditions on the variance-covariance structure of the innovations, distinct volatility regimes of the reduced form residuals enable all structural SVAR matrices and impulse response functions to be estimated without the need for conventional prior identifying restrictions. We give mathematical identification conditions and propose a novel combination of the Gibbs sampler with a Bayesian clustering of impulse responses for the posterior inference on the MS-SVAR parameters. The new methodology is applied to the US data on output, inflation, real money and policy rates, where we demonstrate that the effects of two real and two nominal shocks are clearly identified by the new methodology.

CG334 Room 104 CONTRIBUTIONS IN FORECASTING Chair: Justinas Pelenis

CC1107: **Comparing and combining neural networks for stock market direction prediction**
Presenter: **Daniel Grabowski**, Universitaet Giessen, Germany
Artificial neural networks (ANN) have recently proven extremely successful in a variety of tasks, including forecasting. Their appearance in the econometrics and financial literature is, however, still relatively sparse. To fill this gap, different types of neural networks are applied to stock market direction prediction. The performance of these neural networks is compared to standard econometric as well as machine learning methods. Different neural network architectures and different activation functions are considered. This includes simple single-layer ANN as well as recurrent neural networks, which are better able to capture the time series structure of financial data. The performance of nonlinear models has been shown to benefit from forecast combination. Consequently, the combined forecasts of the different neural networks are also evaluated.

CC1166: **Forecast elicitation with weighted scoring rules**
Presenter: **Justinas Pelenis**, Institute for Advanced Studies, Vienna, Austria
The aim is to investigate the possible advantages of matching the loss function (scoring rule) used for the estimation with the loss function used for the evaluation of density forecasts. We focus on weighted scoring rules for density forecasts which reward forecasting performance on specific regions of support. When forecasting models are correctly specified the choice of the specific scoring rule lead to asymptotically identical results. However, if the models are mis-specified density forecasts elicited under different scoring rules might diverge and consequently lead to different decisions. We examine the benefit of this approach in the context of forecasting downside risk in the equity markets.

CC0521: **Equilibrium error and expected industry portfolio returns**
Presenter: **Victor Troster**, Universitat de les Illes Balears, Spain
Co-authors: Jose Penalva, Abderrahim Taamouti
The equilibrium error, the error term from the cointegration relationship between industry portfolio cumulative returns and excess stock market cumulative returns, is found to have strong predictive power for future industry portfolio returns. Since only the unexpected component of a state variable should command a risk premium, we take deviations from the common long-term relationship between industry portfolio cumulative returns and excess stock market cumulative returns, which proxy for changes in the investment opportunity set. In line with gradual information diffusion across connected industries, these changes in the investment set will lead to stock return predictability by informed investors. We also show that the out-of-sample explanatory power is economically meaningful for investors. Simple trading strategies implied by the proposed predictability provide portfolios with higher mean returns and Sharpe ratios than a buy-and-hold or a benchmark strategy does.

CC1551: **Measuring underlying inflation using dynamic model averaging**
Presenter: **Yuto Iwasaki**, Bank of Japan, Japan
The aim is to present a new framework for measuring underlying inflation with multiple core indicators for Japans consumer price index (CPI). Specifically, a combined core indicator is constructed by applying an econometric method based on dynamic model averaging as a weighted average of individual core indicators. The combined core indicator has time-varying combination weights reflecting changes in the predictive performance of each individual core indicator on a real time basis. Thus, the combined core indicator has the potential to adapt to changes in the nature and sources of price movements. Empirical evidence indicates that the combined core indicator firmly outperforms the individual core indicators over time. In addition, the combination weights for the exclusion-based indicators (e.g. the CPI excluding fresh food) tend to be high when aggregate shocks drive the overall inflation. In contrast, combination weights for the distribution-based indicators (e.g. trimmed mean) tend to be high when idiosyncratic shocks are dominant.

CC1525: **Density forecasting in nonlinear models with stochastic volatility**
Presenter: **Peter Exterkate**, University of Sydney, Australia
Kernel ridge regression is a technique to perform ridge regression with a potentially infinite number of nonlinear transformations of the independent variables as regressors. This makes it a powerful forecasting tool, which is applicable in many different contexts. However, it is usually applied only to independent and identically distributed observations. We introduce a variant of kernel ridge regression for time series with stochastic volatility. The conditional mean and volatility are both modelled as nonlinear functions of observed variables. We set up the estimation problem in a Bayesian manner and derive a Gibbs sampler to obtain draws from the predictive distribution. A simulation study and an application to forecasting the distribution of returns on the S& P500 index are presented, and we find that our method outperforms most popular GARCH variants in terms of one-day-ahead predictive ability. Notably, most of this improvement comes from a more adequate approximation to the tails of the distribution.

CG362 Room 106 CONTRIBUTIONS IN PANEL DATA ECONOMETRICS Chair: Jan Kiviet

CC1373: **Specification of dynamic panel data models: An empirical application to corporate capital structure**
Presenter: **Jan Kiviet**, University of Amsterdam, Netherlands
Co-authors: Julio Pindado, Ignacio Requijo
In corporate finance capital structure is modeled simply by partial adjustment. We demonstrate that this is rejected against more complete forms of dynamic adjustment. Also, usually only the endogeneity problems associated with the joint occurrence of unobserved firm effects and the lagged-dependent variable as explanatory factor are taken into account, without paying attention to possible endogeneity due to genuine simultaneity of the remaining regressors. We develop a model specification search strategy in which the actual form of the dynamic adjustments and possible relevance of instantaneous and lagged feedbacks from the dependent variable, with respect to the explanatory variables, are assessed from the data. We do this by applying the generalized method of moments technique to an ordered set of sequentially augmented restrictions on the parameters of a general dynamic model specification derived from the main capital structure theories. In the same process, we exploit and verify a sequentially augmented set of orthogonality conditions. At the various stages of this search strategy, diagnostic tests are used to decide on either the termination or the direction of the sequential further articulation of the imposed coefficient restrictions and adopted orthogonality conditions. For US firms, we obtain results that shed new light on the actual pattern and speed of dynamic adjustments and on the relevance of lagged and instantaneous feedbacks via other variables.

CC1678: **Estimating linear dynamic panel data models using nonlinear moment conditions**
Presenter: **Markus Fritsch**, University of Passau, Germany
Co-authors: Andrew Adrian Yu Pua, Joachim Schnurbus
GMM-estimation of linear dynamic panel data models based on nonlinear moment conditions that arise when there is no serial correlation in the error terms for balanced and unbalanced panels is implemented in R. Additionally, the theoretical properties of an IV-estimator based on these moment conditions are derived when both, cross-section and time series dimension are large. Efficiency gains are demonstrated in a Monte Carlo simulation and for a well-known household demand data set.

CC1374: **Flexible panel regression model for bivariate count/continuous data with insurance applications**
Presenter: **Yang Lu**, Aix-Marseille University, France
A semi-parametric bivariate panel regression model is proposed that is suitable for mixed count/continuous variables. We develop a polynomial expansion approach for the distribution of the correlated bivariate random effect. Besides the distributional flexibility it offers, the model allows for closed form forecast updating formulas on an individual basis. This property is especially important for insurance applications, in which the future individual insurance premium should be regularly updated according to ones own past claim history. An application on vehicle insurance claims is provided.

CC0922: **EM estimation of dynamic panels with heteroskedastic random coefficients**
Presenter: **Andrea Nocera**, Birkbeck, University of London, United Kingdom
How to combine the EM algorithm with the Restricted Maximum Likelihood method to estimate dynamic heterogeneous panels is shown. The EM-REML approach allows us to estimate iteratively both the average effects and the unit-specific coefficients. Compared to existing methods, it leads to an unbiased estimation of the variances of the random coefficients. Second, our approach allows the random coefficients residuals to have heteroskedasticity of unknown functional form and thus can be seen as a generalization of the one-way error component models where both the random effects and the regression disturbances are heteroskedastic. The estimation procedure can also be adapted to allow for cross-section dependence. An interesting feature of the EM algorithm is that it allows us to make inference on the random coefficients population. Monte Carlo simulations reveal that the proposed estimator has good properties even in small samples and therefore, should be regarded as a valid alternative to Bayesian estimation whenever the researcher wishes to make inference on the coefficients distribution while having little knowledge on what a sensible prior might be. Finally, a novel approach to investigate heterogeneity of the sensitivity of sovereign spreads to government debt is presented.

CC1764: **A panel ordered response model for banks credit ratings estimation**
Presenter: **Patrycja Chodnicka - Jaworska**, University of Warsaw, Poland

The basic goal is to analyse factors influencing the European banks credit ratings. It has been put a following research question: Are recognizable and smaller credit rating agencies use the same factors for estimation banks default risk? Two hypotheses are proposed: Countries credit ratings have a significant influence on banks credit ratings estimation process; The banks capital adequacy, profitability, liquidity and management quality have the significant influence on the banks credit ratings. For estimation the mentioned hypotheses there were used ordered logit panel data models. The analysis has been prepared by using the quarterly data form the Thomson Reuters database for 1998 to 2015 period of time. As dependent variables are used the European banks long term issuer credit ratings proposed by smaller and bigger credit rating agencies. The sample has been divided into subsamples according to the type of credit rating and the size of credit rating agency.

CG380 Room 112 CONTRIBUTIONS ON ESTIMATION OF ECONOMETRIC MODELS Chair: Antonio Montanes

CC1462: **A note on estimation of a mixture of multivariate Tobit models**
Presenter: **Jan Bruha**, Czech National Bank, Czech Republic

The aim is to propose a technique to estimate the latent-class mixture of multivariate Tobit models using the approximate Bayesian computation. Using the computational experiments, we compare the proposed algorithm to alternatives, such as the Gibbs sampler. We apply the proposed algorithm to analyze latent classes of expenditure behaviour of Czech households.

CC1634: **Detection and recovery from inconsistencies in the general linear model with singular dispersion matrix**
Presenter: **Marc Hofmann**, University of Oviedo, Spain
Co-authors: Ana Colubi, Erricos John Kontoghiorghes

A new method to recover from an insconsistent GLM is proposed. The GLM is reformulated as a GLLSP. The minimal set of observations that explain the inconsistencies in the model can be identified by solving a combinatorial sparse approximation problem. An exhaustive algorithm is proposed. Gram-Schmidt orthogonalization is used as the main computational tool. When the number of observations is large, non-exhaustive algorithms can be employed instead.

CC1661: **Estimation and prediction via group SLOPE (gSLOPE)**
Presenter: **Damian Brzyski**, Indiana University Bloomington, United States

The penalized method, group SLOPE (gSLOPE), is presented, which can be used to select entire groups of explanatory variables in classical multiple linear regression model. Such groups could be for example defined as different levels of explanatory factor. Our method could be treated as a generalization of widely known group LASSO. We focus on some theoretical results such as the gSLOPE property of controlling the group false discovery rate (gFDR) under orthogonal case. This property says that in the idealistic case, when all columns in design matrix are orthogonal, we can select tuning parameters such as the control over the expected proportion of falsely discovered groups among all discovered groups (which we define as gFDR) is guaranteed for any predefined level $q \in (0,1)$. The extension to near-orthogonal situation and the algorithm of parameters selection will be also discussed. Moreover, we present the result that our method adapts to unknown sparsity and is asymptotically minimax which means, in some sense, that gSLOPE yields the best possible prediction.

CC1430: **Simple estimators for GARCH models**
Presenter: **Todd Prono**, Federal Reseve Board, United States

The aim is to propose closed-form and variance-targeted two stage least squares (VTTSLS) estimators for the popular GARCH(1,1), threshold GARCH(1,1), and general GARCH(p,q) models, where identification depends either on skewness in the rescaled errors or asymmetry in the conditional variance function. Limit theory for these estimators is established in the empirically relevant case of an ill-defined fourth moment for the GARCH process. The resulting distributional limits, determined using point process theory developed for regularly varying and (weakly) dependent sequences, are highly non-normal though stable, with ill-defined variances. The rate of convergence of these estimators depends on the tail index of the GARCH process, and tend to be quite a bit slower than the usual root n case. Relative to kurtosis-focused, closed-form estimators, the VTTSLS estimators only require a well-defined third moment for the GARCH process and so are better aligned with empirical findings for historical asset returns. In a Monte Carlo study, the VTE benchmarks the VTTSLS, where the VTE has a comparable limit to the VTTSLS in the case of an ill-defined fourth moment.

CC1444: **Unbiased estimation of autoregressive models in bounded series**
Presenter: **Lola Gadea**, University of Zaragoza, Spain
Co-authors: Antonio Montanes, Josep Lluis Carrion-i-Silvestre

The standard analysis of time series consider that an stochastic process can vary freely in the limit . However, there are some important macroeconomic variables that are bounded by definition. In this case, the OLS estimation generates a biased estimate of the autoregressive parameters, which might affect the estimation of other relevant statistics, if the bounded nature of the time series is not accounted for. In order to overcome undesirable effects, we propose to implement different approaches in the literature that correct the estimation bias of autoregressive processes, taking into account the existence of bounds. First, we focus on the median-unbiased estimation procedure, which requires the computation of look-up tables to obtain a correspondence between the value of the OLS estimation of the autoregressive parameter. We have computed similar look-up tables for different values of the bounds to adapt his procedure for bounded stochastic processes. Secondly, we essay the performance of a modified estimator in bounded series and show, both theoretically and through a Monte Carlo experiment, its adequacy depending on the value of the limits. These results open a wide range of theoretical and empirical possibilities, given the abundance of economic bounded series and the multitude of procedures that require accurate estimation of the autoregressive parameters.

CG276 Room 107 CONTRIBUTIONS IN FACTOR MODEL ANALYSIS Chair: Pierre Guerin

CC1614: **The ETF-index volatility spread**
Presenter: **Jaideep Oberoi**, University of Kent, United Kingdom
Co-authors: Hitesh Doshi

Major exchange traded funds (ETFs) that track an underlying index are known to track it closely due to the manner in which they are organized. Deviations of ETF share prices from the underlying basket of stocks lead to arbitrage opportunities that institutions can profit from. Options written on the ETFs and the benchmark indices offer an opportunity to study the difference between cash settled and delivery-based contracts, where other features of the contracts can be adjusted for in standard ways. We document systematic and predictable patterns in the difference between the prices (implied volatilities) of ETFs and the underlying benchmark indices they are designed to track. We explore this predictability and seek explanations for its time variation among factors such as institutional liquidity, the liquidity of underlying securities and the liquidity of the contracts themselves. We argue that the delivery-based nature of ETF options offers a unique opportunity to identify changes in the liquidity of index constituents in times of financial market stress.

CC1643: **Efficient estimation of large approximate factor models using constrained principal components regression**
Presenter: **Rachida Ouysse**, University of New South Wales, Australia

The aim is to develop a new approach to the estimation of the number of factors, factors and factor loadings in large dimensional factor models. Principal components analysis (PCA) provides consistent estimation of the factor structure. For efficient estimation it is essential to estimate a large error covariance matrix when N is large. The proposed method does not assume conditional sparsity, and proposes two approaches to estimating the common factors and factor loadings; both are based on solving a constrained principal component regression problem. The method solves a PCA problem under the constraint of bounded $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(e_{it}e_{jt})|$, which identifies the factor space. Regularisation is achieved by shrinking the off diagonal elements of the covariance matrix to zero. We present results for the main constrained problem and its dual problem which can be viewed as a lasso applied to the off diagonal elements of the covariance matrix. The results from a series of Monte carlo simulations are appreciable for estimating the factor space, estimating the number of factors, but less significant in terms of forecasts performance. The performance of a one-step estimator, which endogenizes the shrinkage estimator and the cross-section bound, are illustrated in a series of monte carlo simulations. We illustrate the method in the context of factor-augmented forecasts of U.S. inflation and growth rate of output.

CC1377: **Markov-switching three-pass regression filter**
Presenter: **Pierre Guerin**, Bank of Canada, Canada
Co-authors: Massimiliano Marcellino, Danilo Leiva-Leon

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

CC1635: **Exploring effects of conventional and unconventional monetary policies: An infinite VAR in data-rich environment**
Presenter: **Emanuele De Meo**, Prometeia SpA, Italy
Co-authors: Giacomo Tizzanini, Lorenzo Prosperi, Lea Zicchino

The aim is to extend the empirical literature on domestic and international effects of conventional (CMPs) and unconventional monetary policies (UMPs, i.e. asset purchase programs and helicopter money) in 11 advanced and emerging countries, proposing a new econometric framework in a data-rich environment. Our approach is based on four building blocks. First, we explore the inter-connectedness of international real and financial variables with network analysis by estimating a global adjacency matrix. Second, we model each country as a small-open economy by means of local factor-augmented vector error-correction models with weakly exogenous foreign variables derived from the global adjacency matrix. Third, we estimate the global solution of the model with an infinite-dimensional VAR (IVAR). Fourth, we identify monetary policy shocks relying on a combination of sign restrictions and generalized impulse-response functions. Our findings are the following. First, UMPs are generally more effective in sustaining domestic business cycles and inflation rates than CMPs. Second, helicopter money appears preferable as a boost to domestic output growth, whereas evidence on its ability to sustain domestic inflation is mixed across countries. Third, both UMPs exhibit higher spillover to foreign countries compared to CMPs. Finally, among EMU countries, helicopter money would especially benefit output growth in Germany and Italy.

CC1715: **The (adaptive) Lasso in the Zoo - Firm characteristic selection in the cross-section of expected returns**
Presenter: **Marcial Messmer**, University of St.Gallen, Switzerland
Co-authors: Francesco Audrino

The adaptive Lasso is shown to be superior to both the Lasso and OLS in most panel specifications with low signal-to-noise ratio based on Monte Carlo Simulations. The results are robust to heteroskedastic, cross-sectionally correlated and non-gaussian errors. Based on the results of the simulation, we find that cross-sectional returns are highly dimensional. However, most published firm characteristics are rejected as predictors for returns when considered in a multivariate selection analysis. The empirical application, which comprises more than 70 published firm characteristics, constructed based on CRSP/Computstat data from 1962-2014, shows that price related FC, namely, the one and twelve month(s) momentum are among the most robustly selected coefficients for explaining differences in average/expected cross-sectional returns. The results are consistent, along large, mid and small cap stocks and for most sub periods considered.

CC671 Room 111 CONTRIBUTIONS IN APPLIED ECONOMETRICS AND FINANCE Chair: Christopher Baum

CC0208: **The effects of minimum wages on immigrants' employment: Evidence from the Swedish economy**
Presenter: **Christopher Baum**, Boston College, United States
Co-authors: Hans Loof, Pardis Nabavi, Andreas Stephan

Neoclassical economic theory predicts that higher minimum wages lead to lower employment. From a theoretical standpoint, the impact of minimum wages is likely to be larger among the foreign-born than among natives: If immigrants are less productive than natives within the low-skilled group, then standard economic theories predict that immigrants should experience more adverse employment effects than natives when minimum wages increase. The effects of minimum wages on immigrants' employment in Sweden will be evaluated. Sweden does not have statutory minimum wage laws. Instead, minimum wages are determined separately in each industry in the contracts between the unions and the employer organizations. These contracts are extended to all workers in each sector and are therefore also binding for non-union workers. In addition to varying across sectors, these minimum provisions vary across regions and seniority. We apply a structural estimation approach to the estimation of the relationship between minimum wages and employment using Swedish data that focuses on the potentially crucial heterogeneity across sectors as well as across regions. Our estimation approach is based on a generalized structural equation model (GSEM). Stata's GSEM extends that framework to incorporate multiple equation systems and latent variables.

CC0196: **Leading indicators of fiscal distress: Evidence from the extreme bound analysis**
Presenter: **Martin Bruns**, DIW Berlin and Free University Berlin, Germany
Co-authors: Tigran Poghosyan

Early warning systems (EWS) are widely used for assessing countries vulnerability to fiscal distress. Most EWS employ a specific set of only fiscal leading indicators predetermined by the researchers, which casts doubt on their robustness. We revisit this issue by using the Extreme Bound Analysis, which allows identifying robust leading indicators of fiscal distress from a large set. Consistent with the theoretical predictions of latest generation crisis models, we find that both fiscal (e.g. fiscal balance, foreign exchange debt) and non-fiscal leading indicators (e.g. output, FX reserves, current account balance, and openness) are robust. In addition, we find that a fiscal vulnerability indicator based on fiscal and non-fiscal leading indicators offers a 29% gain in predictive power compared to a traditional one based on fiscal leading indicators only. It also has good

predictive power out of sample, with 78 percent of crises predicted correctly and only 34 percent false alarms issued for the period 2008 to 2015. This suggests that both fiscal and non-fiscal leading indicators should be taken into account when assessing country's vulnerability to fiscal distress.

CC1528: **Connectedness and spillovers in recession and boom**
Presenter: **Mirela Sorina Miescu**, Queen Mary University of London, United Kingdom
The aim is to investigate the state-dependent nature of spillover effects and the connectedness among countries. Employing the Diebold-Yilmaz connectedness measure we estimate a non-linear VAR model for a group of 7 advanced economies using post WW II monthly data. We find that the group connectedness and the spillover effects tend to be higher in times of recession compared to normal times. The directional results show that European countries are more susceptible to shocks originated in USA and Japan, while USA remains quite insulated from developments that occur in Europe. Finally we show that on top of growth surprises, financial and inflationary shocks are also relevant in explaining the spillover effects among countries. This analysis is of more than academic interest as policymakers are concerned with the domestic vulnerabilities to external shocks. A better understanding of the international business cycle would increase their ability to deal with sudden rises in the economic comovements in the years to come.

CC0298: **FAVAR models for mixed-frequency data**
Presenter: **Franz Ramsauer**, Technical University of Munich, Germany
Co-authors: Michael Lingauer
The previous Factor-Augmented Vector Autoregression (FAVAR) Model is extended to mixed-frequency and incomplete panel data. Within the scope of a fully parametric two-step approach, the alternating application of two expectation-maximization algorithms jointly estimates the model parameters and missing observations. Furthermore, it addresses the selection of the factor dimension and autoregressive order. In contrast to non-parametric two-step estimation methods comprising principal component analyses and linear regressions, we use maximum likelihood estimation. Thereby, we derive equations for the Kalman filter and smoother, which explicitly take into account that the factors consist of latent and observed components. To eliminate any identification problem of the model parameters we constrain the loadings matrix. Our empirical study applies the presented framework to U.S. data for measuring the effects of the monetary policy on the economy and the financial markets. In this context, the consequences for the quarterly growth rates of the Gross Domestic Product (GDP) are of particular importance.

CC1556: **Day-of-the-Week effect in MILA stock markets: A relative distribution approach**
Presenter: **Julio Cesar Alonso**, Universidad Icesi, Colombia
Co-authors: Juan David Martin, Beatriz Gallo
The Day-of-the-Week effect has been extensively studied due to its importance for both investors and researchers. Most of this evidence is bolstered by estimations of time-series models, especially in Latin American Markets. However, some studies question this evidence because of its reliance on error distributional assumptions. We use the Relative Distribution Methods to investigate the existence of the Day-of-the-Week effect in Latin American Integrated Market (MILA) countries. This non-parametric method allows searching for this calendar anomaly in the location, scale, and shape of the returns distributions. We find that there is evidence in favor of a Day-of-the-Week effect in MILA.

CC669 Room 105 CONTRIBUTIONS IN TIME SERIES Chair: Joshua Chan

CC1705: **The impact of international linkages for public debt dynamics**
Presenter: **Carlos Cuerpo**, Universidad Autonoma de Madrid, Spain
Co-authors: Pilar Poncela
Public debt dynamics is generally assessed in the literature based on a simple accounting application involving the debt accumulation equation, whereby future changes in the debt-to-GDP ratio are determined by the foreseen interest rate-GDP growth gap, primary balances and one-off deficit-debt adjustments, such as financial sector bail-outs. This analysis is simple and transparent but presents important caveats; (i) its deterministic nature does not properly capture the existing uncertainty about future economic conditions; and (ii) it ignores potential interdependencies between the different determinants and amongst countries. In order to overcome these caveats, we present a Global VAR model that tracks changes in existing bilateral linkages by optimally weighting the relative importance of trade, financial and contagion channels. The three channels are found to be relevant in maximizing the forecasting accuracy of the GVAR model. Accounting for spillover potential also positively affects the uncertainty surrounding future debt paths as obtained via bootstrapping techniques. Finally, the relative strength or optimal weighting of the different channels is directly linked to the business cycle, as the crisis brought an increase in the relevance of contagion through the sovereign bond markets.

CC0872: **Restrictions search for panel VARs**
Presenter: **Annika Schnuecker**, DIW Berlin, Germany
As panel vector autoregressive (PVAR) models can include several countries and variables in one system, they are well suited for global spillover analyses. However, PVARs require restrictions to ensure the feasibility of the estimation. The stochastic search variable selection for PVAR models (SSVSP) is introduced as an alternative estimation procedure for PVARs. This extends the stochastic search specification selection (S^4) to a restriction search on single elements. The SSVSP allows for incorporating dynamic and static interdependencies as well as cross-country heterogeneities. It uses a hierarchical prior to search for data-supported restrictions. The prior differentiates between domestic and foreign variables, thereby allowing a less restrictive panel structure. Absent a matrix structure for restrictions, a Monte Carlo simulation shows that SSVSP outperforms S^4. Furthermore, this is validated by performing a forecast exercise for G7 countries.

CC1579: **Space-time autoregressive models**
Presenter: **Charles Saunders**, University of Western Ontario, Canada
Spatial econometric models are estimated via MLE and GMM methods, since fixed-effects and OLS approaches are biased. Introducing time dynamics into a spatial model leads to an additional layer of ML bias and more complicated moment conditions for GMM. The indirect inference estimator is implemented as an alternative, which employs a relatively simple estimator for complex models. The distance between estimates from the data and simulated spatial model data is minimized. We show that indirect inference methods can provide suitable bias correction when both spatial and time dynamics are present. The two-stage indirect inference is applied to spatial econometric models to construct finite-sample exact confidence sets. The resulting estimates and confidence set are able to side-step complicated likelihood functions and moment conditions.

CC0431: **Seasonal adjustment without revisions**
Presenter: **Barend Abeln**, Investment consultant, Netherlands
Seasonality in macroeconomic time series can 'obscure' movements of other components in a series that are operationally more important for economic and econometric analyses. A seasonal adjustment program called CAMPLET, an acronym of its tuning parameters, is presented, which consists of a simple adaptive procedure to separate the seasonal and the non-seasonal component from an observed time series. Once this process is carried out there will be no need to revise these components at a later stage when new observations become available. We compare and contrast CAMPLET with X-13ARIMA-SEATS. The main features of CAMPLET are described, and a brief review of X-13ARIMA-SEATS was provided. We evaluate the outcomes of both methods in a controlled simulation framework using a variety of processes. We apply CAMPLET and X-

13ARIMA-SEATS to three time series: U.S.non-farm payroll employment, operational income of Ahold and real GDP in the Netherlands. The main findings are that both methods generally produce similar seasonal adjusted figures.

EI679 Room Graduation hall THE INTERFACE BETWEEN MACHINE LEARNING AND STATISTICS Chair: Mattias Villani

EI1412: **Data-efficient machine learning for autonomous robots**
Presenter: **Marc Deisenroth**, Imperial College London, United Kingdom
One of the fundamental challenges in fully autonomous robots is learning from data directly. This requires data-driven statistical methods for modeling, predicting, and decision making, while taking uncertainty into account, e.g. due to measurement noise, sparse data, or stochasticity in the environment. A practical challenge that comes with autonomous robots is that learning needs to be efficient performing millions of experiments with robots is time consuming and wears out the hardware. We identify practical algorithms for data-efficient learning in robotics and discusses further challenges and opportunities in the field.

EI1735: **Inference in probabilistic graphical models using sequential Monte Carlo**
Presenter: **Thomas Schon**, Uppsala University, Sweden
The aim is to introduce some of our new methods for inference in general probabilistic graphical models (PGMs). The key is a sequential decomposition of the PGM which provides a sequence of auxiliary distributions defined on a monotonically increasing sequence of probability spaces. By targeting these auxiliary distributions using Sequential Monte Carlo (SMC) methods we are able to approximate the full joint distribution defined by the PGM. We also (briefly) introduce the underlying Sequential Monte Carlo methods (e.g. particle filters/smoothers), which are computational methods primarily used to deal with the state inference problem in nonlinear state space models. We are (since a few years back) seeing these methods finding new applications in more and more general model classes, for example PGMs. The systematic combination of SMC and MCMC, referred to as particle MCMC provides another powerful family of algorithms that we will touch upon. The first algorithms of this type were published in 2010 and since then we have (for very good reasons) witnessed a rapidly growing interest in these algorithms.

EI1743: **Inducing meaning representations from text with little or no supervision**
Presenter: **Ivan Titov**, University of Amsterdam, Netherlands
The lack of accurate methods for predicting meaning representations of texts is the key bottleneck for many natural language processing applications such as question answering or text summarization. Although state-of-the-art semantic analyzers work fairly well on closed domains (e.g. interpreting natural language queries to databases), accurately predicting even shallow forms of semantic representations (e.g. underlying predicate-argument structure) for less restricted texts remains a challenge. The reason for the unsatisfactory performance is reliance on supervised learning, with the amounts of annotation required for accurate open-domain parsing exceeding what is practically feasible. We consider approaches which induce semantic representations primarily from unannotated texts. Unlike semantically-annotated data, unannotated texts are plentiful and available for many languages and many domains which makes our approach particularly promising. We contrast our non-parametric Bayesian model and a new approach called reconstruction-error minimization (REM) for semantics. We show that REM achieves state-of-the-art results on semantic representation induction tasks, across languages without any language-specific tuning. Moreover, the REM framework lets us specialize the semantic representations to be useful for (basic forms of) semantic inference and integrate various forms of prior linguistic knowledge.

EO455 Room 215 SURVIVAL ANALYSIS AND LONGITUDINAL DATA ANALYSIS Chair: Peng Zhang

EO0693: **Variable selection and structure identification for varying coefficient Cox models**
Presenter: **Toshio Honda**, Hitotsubashi Univeristy, Japan
Co-authors: Ryota Yabe
Varying coefficient Cox models with high-dimensional covariates are considered. We apply the group Lasso method to these models and propose a variable selection procedure. Our procedure copes with variable selection and structure identification from a high dimensional varying coefficient model to a semivarying coefficient model simultaneously. We derive an oracle inequality and closely examine restrictive eigenvalue conditions, too. We give the details for Cox models with time-varying coefficients. The theoretical results on variable selection can be easily extended to some other important models and we briefly mention those models since those models can be treated in the same way. The considered models are the most popular models among structured nonparametric regression models.

EO1211: **Efficient estimation of the Cox model with auxiliary subgroup survival information**
Presenter: **Chiung-Yu Huang**, Johns Hopkins University, United States
Co-authors: Jing Qin, Huei-Ting Tsai
With the rapidly increasing availability of data in the public domain, combining information from different sources to infer about associations or differences of interest has become an emerging challenge to researchers. We present a novel approach to improve efficiency in estimating the survival time distribution by synthesizing information from the individual-level data with t-year survival probabilities from external sources such as disease registries. While disease registries provide accurate and reliable overall survival statistics for the disease population, critical pieces of information that influence both choice of treatment and clinical outcomes usually are not available in the registry database. To combine with the published information, we propose to summarize the external survival information via a system of nonlinear population moments and estimate the survival time model using empirical likelihood methods. The proposed approach is more flexible than the conventional meta-analysis in the sense that it can automatically combine survival information for different subgroups and the information may be derived from different studies. Moreover, an extended estimator that allows for a different baseline risk in the aggregate data is also studied. Simulation studies show that the proposed estimators yield a substantial gain in efficiency over the conventional partial likelihood approach.

EO1216: **Utilizing high-throughput imaging analysis to predict survival in adrenocortical carcinoma patients**
Presenter: **Peng Zhang**, University of Michigan, United States
Understanding tempo of disease and predicting timing of disease progression is difficult in adrenocortical carcinoma (ACC) patients. We conducted high-throughput imaging analysis on computed tomography (CT) scans, which provides precise measurements of organs and tissues. The new high-dimensional phenotype data are called morphomics. We have shown that morphomics as longitudinal markers predict the overall survival for ACC, independent of the ENSAT staging system. CT scans can be assessed repetitively at any time during the course of disease, and hence have the ability to provide more personalized risk prediction on disease progression and mortality.

EO1766: **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 occur over time, often important covariates we wish to consider also change over time. Such covariates are referred as time-dependent covariates. Quantile regression offers a flexible survival data modeling by allowing the covariates to vary with quantiles. We will present a novel quantile regression model accommodating time-dependent covariates, for analysing survival data subject to right censoring. The simple

estimation technique assumes the existence of instrumental variables. In addition, we will present a doubly-robust estimator. The utility of the proposed methodology will be demonstrated using the Stanford heart transplant dataset.

EO459 Room 002 STATISTICS OF MULTIVARIATE AND SPATIAL EXTREMES Chair: Andrea Krajina

EO0835: **Asymptotically distribution-free goodness-of-fit testing for tail copulas**
Presenter: **Umut Can**, University of Amsterdam, Netherlands
Co-authors: John Einmahl

Consider an i.i.d. sample generated from a multivariate distribution F that lies in the max-domain of attraction of an extreme value distribution G. Then, G describes the asymptotic joint behavior of the (standardized) component-wise maxima, and it is characterized by the marginal extreme value indices and the tail copula R. We propose a procedure for constructing asymptotically distribution-free goodness-of-fit tests for the tail copula R. The procedure is based on a transformation of a suitable empirical process derived from a semi-parametric estimator of R. The transformed empirical process converges weakly to a standard Wiener process, paving the way for a multitude of asymptotically distribution-free goodness-of-fit tests. In a simulation study we show that the limit theorems provide good approximations for finite samples and that tests based on our approach have high power.

EO0832: **Full likelihood inference for max-stable data**
Presenter: **Raphael Huser**, King Abdullah University of Science and Technology, Saudi Arabia
Co-authors: Clement Dombry, Marc Genton, Mathieu Ribatet

How to perform full likelihood inference for max-stable multivariate distributions or processes based on a stochastic Expectation-Maximization (EM) algorithm is shown. In contrast to current approaches, such as pairwise likelihoods or the Stephenson-Tawn likelihood, our method combines statistical and computational efficiency in high-dimensions, and it is not subject to bias entailed by lack of convergence of the underlying partition. The good performance of this methodology is demonstrated by simulation based on the logistic model, and it is shown to provide dramatic computational time improvements with respect to a direct computation of the likelihood. Strategies to further reduce the computational burden are also discussed.

EO1139: **Estimation of extreme bivariate quantiles and extreme quantiles for functional data**
Presenter: **Andrea Krajina**, University of Goettingen, Germany
Co-authors: John Einmahl, Laurens de Haan

Bivariate quantile regions of the form $(x,y) : f(x,y) < c$ are considered, where f, the joint density, is decreasing in both coordinates. Such a region has the property that it consists of the less likely points and hence that its complement is as small as possible. These extreme quantile regions contain hardly any or no data and therefore statistical inference is difficult. In particular when we want to protect ourselves against a calamity that has not yet occurred, we need to deal with probabilities $p < 1/n$, with n the sample size. Using extreme value theory, we construct an estimator of such a quantile region and prove a refined form of consistency. Adapting a suitable definition of the data depth enables us to define extreme quantile regions for functional data as well. Estimation follows using extreme value theory. We illustrate on examples the adapted definition of data depth and the performance of the estimator of the quantile regions.

EO1438: **Conditional independence among max-stable laws**
Presenter: **Kirstin Strokorb**, University of Mannheim, Germany
Co-authors: Ioannis Papastathopoulos

Independence and conditional independence are key concepts in the theory of probability and statistical inference. In general, neither does independence imply conditional independence nor does conditional independence imply independence of the subvectors of a random vector. However, we show that, if the distribution of a max-stable random vector has a positive continuous density, then conditional independence of any two subvectors conditioned on the remaining components implies already their independence. We conclude that a broad class of commonly used tractable max-stable models cannot exhibit an interesting Markov structure. This result also complements recent developments, where it is shown that, up to time reversal, only max-auto-regressive processes of order one can appear as discrete time stationary max-stable Markov chains.

EO714 Room 007 BAYESIAN SEMI-AND NONPARAMETRIC MODELLING III Chair: Raffaele Argiento

EO0970: **Non-standard approaches to nonparametric Bayes**
Presenter: **Jeff Miller**, Harvard University, United States

The "fully Bayesian nonparametric" approach is to construct a complete probabilistic model and perform Bayesian inference using algorithms that are exactly correct (possibly up to MCMC sampling error). However, this can be computationally burdensome and time-consuming to implement. By compromising on certain aspects of this fully BNP approach, it is sometimes possible to obtain "semi-BNP" methods which are faster and easier to use, and behave like fully BNP methods. Work in this direction will be discussed, centered around three themes: (a) hybrid frequentist-Bayesian methods, (b) analytical approximations, and (c) partial models. Examples from work-in-progress will be used to illustrate.

EO1022: **Nonparametric network models for link prediction**
Presenter: **Sinead Williamson**, University of Texas at Austin, United Kingdom

Many datasets can be represented as a sequence of interactions between entities for example communications between individuals in a social network, protein-protein interactions or DNA-protein interactions in a biological context, or vehicles journeys between cities. In these contexts, there is often interest in making predictions about future interactions, such as who will message whom. A popular approach to network modeling in a Bayesian context is to assume that the observed interactions can be explained in terms of some latent structure. For example, traffic patterns might be explained by the size and importance of cities, and social network interactions might be explained by the social groups and interests of individuals. Unfortunately, while elucidating this structure can be useful, it often does not directly translate into an effective predictive tool. Further, many existing approaches are not appropriate for sparse networks, a class that includes many interesting real-world situations. We will describe models for sparse networks that combine structure elucidation with predictive performance. We use a Bayesian nonparametric approach, which allows us to predict interactions with entities outside our training set, and allows the both the latent dimensionality of the model and the number of nodes in the network to grow in expectation as we see more data.

EO1134: **Edge-exchangeable graphs, sparsity, and power laws**
Presenter: **Tamara Broderick**, MIT, United States
Co-authors: Diana Cai, Trevor Campbell

Many popular network models rely on the assumption of (vertex) exchangeability, in which the distribution of the graph is invariant to relabelings of the vertices. However, the Aldous-Hoover theorem guarantees that these graphs are dense or empty with probability one, whereas many real-world graphs are sparse. We present an alternative notion of exchangeability for random graphs, which wecall edge exchangeability, in which the distribution of a graph sequence is invariant to the order of the edges. We characterize the class of edge exchangeable models with a paintbox

construction, and we demonstrate that edge-exchangeable models, unlike models that are traditionally vertex exchangeable, can exhibit sparsity and power laws. To do so, we outline a general framework for graph generative models; by contrast to a previous pioneering work, models within our framework are stationary across steps of the graph sequence. In particular, our model grows the graph by instantiating more latent atoms of a single random measure as the dataset size increases, rather than adding new atoms to the measure.

EO0483: **Time-dependent feature allocation models via Poisson Random Fields**
Presenter: **Paul Jenkins**, University of Warwick, United Kingdom
Co-authors: Valerio Perrone, Dario Spano, Yee Whye Teh
In a feature allocation model, each data point depends on a collection of unobserved latent features. For example, we might classify a corpus of texts by describing each document via a set of topics; the topics then determine a distribution over words for that document. In a Bayesian nonparametric setting, the Indian Buffet Process (IBP) is a popular prior model in which the number of topics is unknown a priori. However, the IBP is static in that it does not account for the change in popularity of topics over time. Here we present the *Poisson random field Indian Buffet Process* (PRF-IBP), a probabilistic model for collections of time-stamped documents. By adapting the Poisson random field model from population genetics, we derive a stochastic process with appealing properties including that (i) each feature popularity evolves independently as a diffusion and (ii) marginal observations at a fixed timepoint are given by the IBP. We describe a Markov Chain Monte Carlo algorithm for exact posterior simulation and illustrate our construction by analysing the topics of NIPS conference papers over 12 years. This is joint work with Valerio Perrone, Dario Spano, and Yee Whye Teh.

EO245 Room 006 ANALYSIS OF SATELLITE IMAGERY Chair: Armin Schwartzman

EC1703: **Creating daily, fine spatial resolution Sentinel-2 time-series**
Presenter: **Pete Atkinson**, Lancaster University, United Kingdom
Co-authors: Wang Qunming, Wenzhong Shi
Image fusion is a common problem in the field of remote sensing and there are many different approaches for this. Space-time (ST) image fusion is potentially of novel benefit for fusing Sentinel-2 and Sentinel-3 images. These two sensors have very different spatial and temporal resolutions (Sentinel-2 MSI sensor 10 m, 20 m and 60 m, 10 days, albeit 5 days with 2 sensors; Sentinel-3 OLCI sensor 300 m, less than two days, conditional upon clear skies). The problem is that for local monitoring one either has the desired spatial (e.g. 20 m) or temporal (e.g. nearly daily) resolution, but not both. While the OLCI instrument is suitable for monitoring over large areas, it is not suitable for monitoring at the local scale. Area-to-point regression kriging (ATPRK) was, therefore, used to provide a very fine spatial resolution, frequent time-series of Sentinel-based images for the first time. The effectiveness of ATPRK-based ST fusion is demonstrated with examples. The approach can be applied globally, offering the prospect, given the appropriate infrastructure, of a global fine resolution Sentinel product for local applications.

EO1730: **Monitoring vegetation from space with Gaussian processes and latent force models**
Presenter: **Gustau Camps-Valls**, Universitat de Valencia, Spain
The aim is to summarize our experience in the last decade on developing algorithms in the interplay between Physics and Statistical Inference to analyze Earth Observation (EO) satellite data. Some of them are currently adopted by ESA and EUMETSAT. We pay attention to machine learning models that help to monitor land, oceans, and atmosphere through the analysis and estimation of climate and biophysical variables. In particular, Bayesian non-parametric methods (such as Gaussian Processes and latent force models) provide an adequate framework to design models with high prediction accuracy and able to cope with uncertainties, to encode physical knowledge about the problem, perform domain adaptation through time and space, and to attain self-explanatory models learned from empirical data. The theoretical developments will be guided by the challenging problems of estimating biophysical parameters and learning causal relations at both local and global planetary scales.

EO1732: **Geospatial data in support of agricultural and agri-environmental statistics, advantages and warnings**
Presenter: **Elisabetta Carfagna**, University of Bologna, Italy
Due to the technological development, in the last decades, different kinds of geospatial data have become easily accessible at decreasing prices and have started to be used for producing statistics. We focus on spatial resolution of data, on change of support and on some kinds of transformations like aggregation and disaggregation of spatial data when remote sensing data, Global Positioning Systems and Geographic Information Systems (GIS) are used for producing agricultural and agri-environmental statistics. Particular attention is devoted to the impact of above-mentioned characteristics and transformations of geospatial data on sampling frame construction and sample design, stratification, use of remote sensing data for agricultural and agri-environmental statistics, small area estimation and yield forecasting.

EO1723: **Glacier terminus estimation from Landsat image intensity profiles**
Presenter: **Armin Schwartzman**, University of California, San Diego, United States
Co-authors: Joseph Usset, Arnab Maity, Ana-Maria Staicu
Mountain glacier retreat is an important problem related to temperature increase caused by global climate change. The retreat of mountain glaciers has been studied from the ground, but there exists a need for automated methods to catalog glacial change where ground measurements are not available. We propose a statistical processing and analysis pipeline to track glacial termini over time from Landsat images. First, intensity profiles are obtained from the Landsat images along the glacier flow line and are smoothed using regression splines. Based on the assumption that the glacial terminus location lies near a point of high negative change in each smoothed profile, an algorithm is proposed that seeks a minimum of aggregated first derivative values across the profiles, to obtain an approximate path of termini locations over time. Then spline smoothing is applied to this approximate path for estimation of long-term terminus movement. The predictions from the method are evaluated on simulated data and compared to ground measurements obtained for the Franz Josef, Gorner, Rhone, and Nigardsbreen glaciers.

EC662 Room 204 CONTRIBUTIONS IN METHODOLOGICAL STATISTICS Chair: Ricardo Cao

EC0176: **Factorial, raw and central moments**
Presenter: **Idika Okorie**, University of Manchester, United Kingdom
General relations are derived expressing factorial moments in terms of raw moments, raw moments in terms of factorial moments, factorial moments in terms of central moments and central moments in terms of factorial moments.

EC1497: **Intervalar location of Wald-type confidence intervals for a linear combination of binomial proportions**
Presenter: **Adelaide Freitas**, University of Aveiro, Portugal
Co-authors: Sara Escudeiro, Vera Afreixo
Focusing on asymptotic Wald-type confidence intervals (CIs) for a linear combination of three or more independent binomial proportions, we discuss the interpretation of the interval location based on the mesial and distal non-coverage probabilities and introduce the notion of satisfactory mesially (distally) located intervals for situations in which extremal observations exist. A comparative simulation study was carried out to investigate the performance of the classical version and four adjusted versions of Wald CIs with respect to the coverage probability and the mesial and

distal non-coverage probabilities. Two of these adjusted versions, based on shrinkage estimators for proportions, were recently established in the specialized literature and have the advantage of being able to handle extremal observations. Although these two versions are very conservative and their mesial and distal non-coverage probabilities are not balanced, which is not the case for the other two adjusted versions considered, the location of their CIs for almost all settings considered in our simulation study was mesially satisfactory, having a very low probability of being distally located, which is consistent with the ability of these two variants to handle extremal observations.

EC1357: **New approaches for bootstrap bandwidth selection in density estimation under dependence**
Presenter: **Ines Barbeito**, Universidade da Coruna, Spain
Co-authors: Ricardo Cao
Smoothed versions of the Stationary Bootstrap and Moving Blocks Bootstrap are established for bandwidth selection in density estimation for dependent data. Exact expressions for the bootstrap version of the mean integrated squared error under dependence are obtained in both contexts. Those expressions are very useful, since the implementation of the bootstrap selector does not require Monte Carlo approximations. The empirical choice of the block length in these two situations is also analyzed. Finally, the good practical performance of both new bootstrap bandwidth selectors is shown in an extensive simulation study. The methods are illustrated by applying them to two real data sets.

EC1422: **A two-sample test for the error distribution in nonparametric regression**
Presenter: **Gustavo Rivas**, National University of Asuncion, Paraguay
Co-authors: Maria Dolores Jimenez-Gamero
A test for the equality of error distributions in two nonparametric regression models is proposed. The test statistic is based on comparing the empirical characteristic functions of the residuals calculated from independent samples of the models. The null distribution of the test statistic is unknown. To approximate it, a weighted bootstrap estimator is studied, providing a consistent estimator. The finite sample performance of this approximation as well as the power of the resulting test are evaluated by means of a simulation study.

EC1270: **A new multivariate two-sample test based on distributions of interpoint distances**
Presenter: **Pablo Montero-Manso**, Universidade da Coruna, Spain
Co-authors: Jose Vilar
A new non-parametric test for the multivariate two-sample problem is proposed. The test statistic is based on comparing empirical distributions of within- and between-sample interpoint distances. The method inherits the benefits of distance-based approaches, including applicability to high dimension, low sample size (HDLSS) scenarios and to complex or categorical data. Flexibility of the distributional approach is particularly useful when considering the possibility of using specific distances in fields such as ecology or time series analysis. The proposed method is compared to alternative distance-based, kernel-based, binary classifier-based and parametric two-sample tests in a wide range of simulated and real datasets. The results give our method the distinctive characteristic of achieving superior sensitivity in several popular scenarios and competitive performance in the rest.

EC665 Room 202 CONTRIBUTIONS IN TIME SERIES ANALYSIS Chair: Holger Drees

EC0754: **Limit theorems for residuals from VARMAX models with potentially serially correlated errors**
Presenter: **Benjamin Holcblat**, University of Luxembourg, Luxembourg
Co-authors: Steffen Gronneberg
In time series analysis, numerous inference procedures need to use residuals instead of the unobserved error terms. Allowing for serial correlation between error terms, we prove limit theorems for partial sums of smooth functions of VARMAX residuals. Process limits based on residuals do not correspond to process limits based on the error terms, generally. However, the former correspond to the process limits of average-corrected error terms. Applications of our results include estimation of log-GARCH models, generalization of Jarque-Bera normality tests, and CUSUM-type of tests.

EC1400: **The sufficient conditions for the consistency of generalized resampling schemes**
Presenter: **Lukasz Lenart**, Cracow University of Economics, Poland
Sufficient conditions for the consistency of Generalized Resampling Schemes (GRS) for nonstationary time series are formulated. Under appropriate conditions GRS reduces to usual subsampling procedures, and it is simply related to the well-known Moving Block Bootstrap. In simple words, the GRS is based on the definition of the estimator (for the parameter of interest) which is based on so-called generalized subsample. These sufficient conditions are related to existing sufficient conditions for the consistency of subsampling procedures for non-stationary time series. In proving the consistency, the mixing condition, adequate size of generalized subsample (in comparison with the length of sample) and the existence of a limit distribution play a central role. The properties of GRS are examined in a simulation study. In particular, the usual MC simulations are carried out in order to examine the sizes and powers of the test concerning Fourier coefficient in the class of (nonstationary) Almost Periodically Correlated time series.

EC1671: **Time series anomaly detection with patterns and structural breaks identification: A constrained clustering approach**
Presenter: **Carlo Drago**, University of Rome Niccolo Cusano, Italy
Time series anomaly detection is an approach useful to identifying deviations from a specific pattern over time. A relevant problem is to identify correctly the structure of the time series and the structural change which have to be considered as important by itself. We combine the analysis of the time series components and the structural change identification in order to characterize adequately the time series considered. At this point we consider an approach based on constrained clustering in order to detects the specific deviations of the time series from the identified patterns. In particular a validation of the clusters found over time can be useful to the aim to find the relevant anomalies which can occur on the data. Finally we will discuss on the sensitivity analysis which can be necessary to identify the anomalies which are not sensitive to the method used. In order to present the approach we will use both simulated time series and real data.

EC0265: **Nonparametric estimation of the distribution of the autoregressive coefficient from panel random-coefficient AR(1) data**
Presenter: **Vytaute Pilipauskaite**, Vilnius University - Universite de Nantes, Lithuania
Co-authors: Remigijus Leipus, Anne Philippe, Donatas Surgailis
Nonparametric estimation is discussed for the distribution function $G(x)$ of the autoregressive coefficient $a \in (-1,1)$ from a panel of N random-coefficient AR(1) data, each of length n, by the empirical distribution function of lag 1 sample autocorrelations of individual AR(1) processes. Consistency and asymptotic normality of the empirical distribution function and a class of kernel density estimators is established under some regularity conditions on $G(x)$ as N and n increase to infinity. The Kolmogorov-Smirnov goodness-of-fit test for simple and composite hypotheses of Beta distributed a is discussed. A simulation study for goodness-of-fit testing compares the finite-sample performance of our nonparametric estimator to the performance of its parametric analogue discussed in Beran et al. (2010).

EC1617: **Time-series filtering for replicated observations via a kernel approximate Bayesian computation**
Presenter: **Takanori Hasegawa**, The Institute of Medical Science, The University of Tokyo, Japan
Co-authors: Kaname Kojima, Yosuke Kawai, Masao Nagasaki

In time-series analysis, state-space models have been widely used in estimating the conditional distributions of hidden variables and parameter values, and in elucidating the structure that can generate the observation data. Kalman filter is utilized to calculate the theoretically exact distributions on linear modes, and several extensions, e.g. unscented Kalman filter and particle filter, have been applied to approximately calculate the distributions on nonlinear models. Recently, approximate Bayesian computation (ABC) has been applied to such time-series filtering to handle observations with intractable noises; however, it remains problems in (i) reducing estimation bias, (ii) handling replicated observations, and (iii) selecting appropriate models in practice. To address such limitations, we propose a novel method combined with the kernel Bayes' rule approach for the filtering problem in state space models. Simulation studies showed that the proposed method has power to estimate the parameter values of stochastic differential equations on state space models even when using time series with intractable noises.

EG462 Room 217 CONTRIBUTIONS IN MATRIX METHODS AND MULTIVARIATE ANALYSIS Chair: Dietrich von Rosen

EC1080: **Approximation of a covariance matrix by Kronecker product of compound symmetric matrix and arbitrary matrix**
Presenter: **Katarzyna Filipiak**, Poznan University of Technology, Poland
Co-authors: Daniel Klein, Martin Singull

Statistical modeling of doubly multivariate data has often been based on a separable covariance matrix, that is, covariances that can be written as a Kronecker product of two covariance matrices, say A and B. If one of those matrices is structured as a compound symmetry correlation matrix, $CS = (1-a)I_p - aJ_p$ with $1/(p-1) < a < 1$, then the approximation of a non-separable doubly multivariate data covariance matrix by a Kronecker product of CS and B can be applied to calculate the power of the test for verifying the structure of the covariance matrix. Therefore, approximation of a positive definite matrix by a Kronecker product $CS \otimes B$ in the sense of Frobenius norm, as well as its properties, will be given.

EC1602: **Orthogonal nonnegative matrix factorization based on the Tweedie family**
Presenter: **Hiroyasu Abe**, Doshisha University, Japan
Co-authors: Hiroshi Yadohisa

Orthogonal nonnegative matrix factorization (ONMF) is a multivariate analysis technique for approximating a given nonnegative data matrix using the product of two nonnegative factor matrices, one of which is column orthogonal. There are mainly two types of estimation algorithm for ONMF: multiplicative updating algorithms and k-means-based algorithms. In the former type of algorithm, the factor matrix with the orthogonal constraint cannot be perfectly obtained as an orthogonal matrix but must be approximately obtained, and the objective function is not monotonically non-increasing. In the latter type of algorithm, by contrast, the matrix is iteratively updated with orthogonality maintained, and the objective function is monotonically non-increasing.We propose two new k-means-based algorithms for ONMF based on Poisson and compound Poisson-gamma, both of which belong to the Tweedie family distribution. ONMF based on the two distribution has not been proposed yet. We derive an update equation for the non-orthogonal factor matrix by a new auxiliary function method using an inequality of a bivariate concave function. The advantage of an ONMF based on the compound Poisson distribution is its robustness, as shown by our simulation study.

EC1620: **Testing dimensionality of multivariate variance components**
Presenter: **Satoshi Kuriki**, The Institute of Statistical Mathematics, Japan
Co-authors: Tomoyuki Shirai, Khanh Duy Trinh

Let A be a standard Gaussian random matrix in the space Sym(n) of n by n real-symmetric or Hermitian matrices (i.e., GOE or GUE, respectively). Let PD(n) be the cone of positive semidefinite matrices in Sym(n). We derive the distribution of the squared distance between the random matrix A and the cone PD(n). This distribution appears in balanced multivariate variance components model. In this model, within and between sum of squares matrices (H and G, say) are independent Wishart matrices. The difference of matrix parameters of H and G is a positive semidefinite matrix referred to as variance components. The rank of the variance components matrix is typically much smaller than the size of the matrices H and G, and its inference is crucial in modeling. When the number of groups goes to infinity, the degrees of freedom of H and G go to infinity, and then the LRT for the rank of the variance components matrix converges to the squared distance between A and PD(n). In GOE and GUE cases, the distributions are shown to be mixtures of chi-square distributions with weights expressed in terms of the Pfaffian or the determinant. Moreover, when the matrix size n goes to infinity, the limiting distribution is proved to be Gaussian. This Gaussianity was conjectured in previous literature. Based on the obtained distributions, we propose a multiple testing procedure to estimate the rank of the variance components matrix. Mouse growth data are analyzed as an example.

EC1384: **Multivariate imputation by data depth**
Presenter: **Pavlo Mozharovskyi**, CREST-ENSAI, France
Co-authors: Julie Josse, Francois Husson

Single imputation is an appropriate technique to handle missing data if one simply needs to complete a single data set, when no inference is required, when the applied statistical method is computationally too demanding for multiple data sets, or when a few values are missing only but one seeks an alternative to the list-wise deletion. The presented methodology for single imputation of missing values borrows the idea from data depth - a measure of centrality defined for an arbitrary point of the space with respect to a probability distribution or a data cloud. This consists in iterative maximization of the depth of each observation with missing values, and can be employed with any properly defined statistical depth function. Being able to grasp the underlying data topology, the procedure is distribution free, allows to impute close to the data, preserves prediction possibilities different to local methods (nearest neighbor imputation, random forest), and has attractive robustness and asymptotic properties under elliptical symmetry. It is shown that its particular case - when using Mahalanobis depth - has direct connection to well known treatments for multivariate normal model, such as iterated regression or regularized PCA. The methodology has been extended to the multiple imputation proper for data stemming from an elliptically symmetric distribution. Simulation and real data studies contrast the suggested framework with existing popular alternatives.

EC1454: **Advances in depth-based classification**
Presenter: **Ondrej Vencalek**, Palacky University in Olomouc, Czech Republic
Co-authors: Oleksii Pokotylo

Concept of data depth can be used for solving a broad class of standard multivariate statistical problems including problem of classification. We introduce a new way to evaluate classifiers' performance. Subsequently we derive two new depth-based classifiers depth-weighted classifier and rank-weighted classifier. We investigate their properties, concentrating mainly on their similarities/differences to the Bayes classifier which minimizes total probability of misclassification.

EC659 Room 214 CONTRIBUTIONS IN SEMI- AND NON-PARAMETRIC STATISTICS Chair: Noel Veraverbeke

EC1421: **Nonparametric predictive inference for diagnostic test thresholds**
Presenter: **Manal Alabdulhadi**, Durham Uinverstiy, United Kingdom
Co-authors: Frank Coolen, Tahani Coolen-Maturi
The accuracy of diagnostic test relates to the ability of the test to distinguish between diseased and healthy individuals. Providing good methods for defining the accuracy of diagnostic tests assist physicians to detect the probability of disease for their patients. In 2-Group and 3-Group ROC analysis, setting thresholds for classification is often the most important decision. The standard uses the maximisation of the Youden index, a global measurement of diagnostic accuracy. We consider an alternative to the maximisation of the Youden index, by explicitly considering the use of the classification procedure for a specific number of future patients. We consider nonparametric predictive inference (NPI), which is a powerful statistical framework that yields direct probabilities for one or m future observations, based on n observations for related random quantities. We introduce 2-Group and 3-Group predictive method to select optimal diagnostic thresholds in order to have the best classification of one or more future persons on the basis of their test results. We find that the optimal thresholds sometimes lead to other value of thresholds with different number of m future patients. We generalize the Youden index by applying our method to the Youden index and maximising the sum of the probabilities of correct classification for the different groups. Comparison between our method and generalization of Youden index is discussed.

EC1413: **The power of tests for signal detection in high-dimensional data**
Presenter: **Marc Ditzhaus**, Heinrich-Heine University Duesseldorf, Germany
Co-authors: Arnold Janssen
In modern studies the detection of signals becomes more and more important. Especially, if they are weak and rare it is quite difficult to detect them. In this context mixture models are often used to model the signal detection problem. It has been previously discussed the behaviour of the log-likelihood ratio test for the sparse normal mixture model. It has been suggested to use Tukey's higher criticism (HC) for this model. In contrast to LLR, HC does not depend on the unknown signal strength and the unknown proportion of signals within the data. Nevertheless, the area of complete detection are the same for both tests. Recently, the results were extended to more general models. We give a short overview of the literature and present results concerning a new general model, which includes all of the above mentioned models.

EC1584: **Jackknife-type ridge estimator in semiparametric regression models**
Presenter: **Esra Akdeniz**, Marmara University, Turkey
A ridge estimator is introduced for the vector of parameters β in a semiparametric regression model (SPRM). To reduce the bias, the standard jackknife and weighted jackknife techniques are proposed in SPRM. Jackknife-type ridge estimator is defined. Performances of the proposed estimator is examined with respect to the mean squared error (MSE) criterion. Through simulation the jackknife estimators are compared in terms of mean squared error criterion when the sample size is small to moderate. The proposed estimator is also applied on a real data set.

EC0300: **Semiparametrics: A typical misspecification in life-satisfaction studies**
Presenter: **Stefan Sperlich**, Univserity of Geneva, Switzerland
Co-authors: Setareh Ranjbar
Functional form specification is an important issue in econometrics. While non-parametric techniques can partially solve this problem by relaxing the assumptions on any explicit functional form, it is at the cost of easy interpretation and also facing the curse of dimensionality. The semi-parametric techniques are introduced as a way of circumventing these shortcomings. However, the non-parametric part of the model is particularly sensitive to model assumptions in the parametric part. In the recent literature on life-satisfaction these methods have been used without taking sufficient care. It will be shown how this can easily, though not necessarily, lead to erroneous conclusions.

EC0723: **Copula-based estimation of conditional densities and hazard rate functions**
Presenter: **Noel Veraverbeke**, Hasselt University, Belgium
Co-authors: Paul Janssen, Jan Swanepoel
New copula-based smooth Bernstein estimators for the conditional density function and the related conditional hazard rate function are presented. The conditional density estimator is defined as a smoother of the copula-based Bernstein estimator of the conditional distribution function. In a similar way, the conditional hazard rate estimator is a smoother of the conditional cumulative hazard rate function estimator. We discuss the asymptotic properties of the bias, variance and distribution in terms of the smoothing parameters.

EG639 Room 205 CONTRIBUTIONS IN CLUSTERING Chair: Marco Riani

EC1467: **Logistic regression augmented community detection with application in identifying autism-related gene pathways**
Presenter: **Qing Pan**, George Washington University, United States
Co-authors: Yunpeng Zhao, Chengan Du
When searching for gene pathways leading to specific disease outcomes, we propose to take advantage of additional information on gene characteristics to differentiate genes of interests from irrelevant background ones when connections involving both types of genes are observed and their relationships to the disease are unknown. Novel generalized stochastic blockmodel are proposed that singles out irrelevant background genes with the help of auxiliary information, and clusters relevant genes into cohesive groups using the adjacency matrix. Expectation-maximization algorithm is modified to maximize a joint pseudo-likelihood assuming latent indicators for relevance to the disease and latent group memberships as well as Poisson or multinomial distributed link numbers within and between groups. Asymptotic consistency of label assignments are proven. Superior performance and robustness in finite samples are observed in simulation studies. The proposed method identifies previously missed gene sets underlying autism and related neurological diseases using diverse data sources including de novo mutations, gene expression and protein-protein interactions.

EC1364: **Paired indices for clustering evaluation: A typology**
Presenter: **Margarida G M S Cardoso**, Instituto Universitario de Lisboa-Business Research Unit-Lisboa, Portugal
Paired indices of agreement are commonly used to measure the accordance between two partitions of the same data set. They are generally determined based on a cross-classification table of counts of pairs of observations both partitions agree to join and/or separate in the clusters. However, there are still open issues regarding the specific thresholds one should consider for each index to conclude about the degree of agreement between the partitions. We analyze the distribution of 14 indices under the null hypothesis (H0) of agreement occurring by chance to acquire new insights on the indices behavior. We resort to the IADJUST method to generate cross-classification tables under H0. The experimental scenario considers 3 clusters, balanced or unbalanced, poorly, moderately or well separated. The analysis suggests a new typology of paired indices of agreement. This result resorts to the indices adjusted values (values deducted from agreement by chance) and also in the indices distributional characteristics intra-scenarios (average, quantiles, range, standard deviation, coefficient of variation, skewness and kurtosis).

EC1375: **Simultaneous dimension reduction and multi-objective clustering using probabilistic factorial discriminant analysis**
Presenter: **Vincent Vandewalle**, Inria, France

In model based clustering of quantitative data it is often supposed that only one clustering variable explains the heterogeneity of all the others variables. However, when variables come from different sources, it is often unrealistic to suppose that the heterogeneity of the data can only be explained by one variable. If such an assumption is made, this could lead to a high number of clusters which could be difficult to interpret. A model based multi-objective clustering is proposed, is assumes the existence of several latent clustering variables, each one explaining the heterogeneity of the data on some clustering projection. In order to estimate the parameters of the model an EM algorithm is proposed, it mainly relies on a reinterpretation of the standard factorial discriminant analysis in a probabilistic way. The obtained results are projections of the data on some principal clustering components allowing some synthetic interpretation of the principal clusters raised by the data. The behavior of the model is illustrated on simulated and real data.

EC1609: **Guided projections for analysising the structure of high dimensional data**
Presenter: **Thomas Ortner**, Vienna University of Technology, Austria
Co-authors: Christian Breiteneder, Sarka Brodinova, Peter Filzmoser, Maia Zaharieva

Guided projections are a powerful data transformation method, creating new possibilities to reveal the group structure for high dimensional data in the presence of noise variables. We project all observations onto the space spanned by a small subset of q observations. We can measure the similarity to these observations using a combination of orthogonal and score distances based on the spanned subspace. By exchanging observations one by one based on the similarity to the selected observations we create a sequence of projections which we call guided projections. Each projection represents a new variable and the full series the transformation space where certain structures are revealed which are not visible in the full data space. This sequence directly provides a possibility for diagnostic plots, revealing clusters as well as outliers. Based on simulated data we identify the strengths and limitations of guided projections compared to other data transformations by calculating various validity measures for cluster procedures as well as performing Wards clustering procedure on the transformed data. We further show the relevance of the transformation by applying it to a real-world data set (fruit data).

EC1639: **Clustering cancer evolutionary trees**
Presenter: **Yusuke Matsui**, Nagoya university graduate school of medicine, Japan
Co-authors: Satoru Miyano, Teppei Shimamura

Multi-regional sequencing provides new opportunities to investigate genetic heterogeneity within or between common tumors from an evolutionary perspective. Several state-of-the-art methods have been proposed for reconstructing cancer sub-clonal evolutionary trees based on multi-regional sequencing data to develop models of cancer evolution. However, the methods developed thus far are not sufficient to characterize and interpret the diversity of cancer sub-clonal evolutionary trees. We propose a clustering method (phyC) for cancer sub-clonal evolutionary trees, in which sub-groups of the trees are identified based on topology and edge length attributes. For interpretation, we also propose a method for evaluating the diversity of trees in the clusters, which provides insight into the acceleration of sub-clonal expansion. Simulation showed that the proposed method can detect true clusters with sufficient accuracy. Application of the method to actual multi-regional sequencing data of clear cell renal carcinoma and non-small cell lung cancer allowed for the detection of clusters related to cancer type or phenotype. The software phyC is available from https://github.com/ymatts/phyC.

EG674 Room 203 CONTRIBUTIONS ON FUNCTIONAL DATA ANALYSIS II Chair: Antonio Cuevas

EC1552: **Registration and classification for the X-ray image data**
Presenter: **Pengcheng Zeng**, Newcastle University, United Kingdom

Functional data analysis is widespread in every branch of modern science, such as medicine, biology, chesmitry and engineering. The registration part for multi-dimensional functional data is a significant step, including removing the translating, scaling, rotating and warping problems. It plays a key role in doing the later functional data analysis. We first show the data acquisation of multi-dimensional functional data from X-ray video clips. The pre-processing procedure, like normalization will be presented. After that, a new registration method is proposed, which is basically a mixture of Generalized Procrustes Analysis and self-modelling registration method. This new method can effectively resolve the problems, especially for rotation and warping. It is better than some other existing methods in terms of simulation study and real data analysis. We will end up with the classification for the X-ray image data.

EC1711: **Functional principal component analysis for derivatives of multivariate curves**
Presenter: **Heiko Wagner**, University of Bonn, Germany
Co-authors: Maria Grith, Alois Kneip, Wolfgang Haerdle

Two methods are presented based on functional principal component analysis (FPCA) for the estimation of smooth derivatives of a sample of random functions, which are observed in a more than one-dimensional domain. We apply eigenvalue decomposition to a) the dual covariance matrix of the derivatives, and b) the dual covariance matrix of the observed curves. To handle noisy data from discrete observations, we rely on local polynomial regressions. If curves are contained in a finite-dimensional function space, the second method performs better asymptotically. We apply our methodology in a simulation and empirical study, in which we estimate state price density (SPD) surfaces from call option prices. We identify three main components, which can be interpreted as volatility, skewness and tail factors. We also find evidence for term structure variation.

EC1435: **Functional regression analysis with compositional response**
Presenter: **Renata Talska**, Palacky University Olomouc, Czech Republic
Co-authors: Alessandra Menafoglio, Karel Hron, Eva Fiserova, Jitka Machalova

Regression analysis is a key statistical tool to model a linear relationship between a response variable and a set of covariates. In functional data analysis (FDA), methods to perform linear regression with functional response and scalar predictors have been widely discussed. More delicate appears the situation in which the response variable is represented as a probability density, since the L2 space (of square integrable functions), usually employed in FDA, does not account for the inherent constraints of densities. The aim is to introduce functional regression model with distributional response using the Bayes space approach, i.e. a geometric viewpoint that allows capturing the inherent features of distributional data. Indeed, densities primarily carry relative information, and the unit integral constraint represents just one of its possible equivalent representations. Accordingly, densities can be considered as elements of a Bayes Hilbert space, whose geometry is designed to precisely capture the specific properties of densities (e.g. scale invariance, relative scale). In order to apply functional regression tools for L2 data, particularly those based on B-spline representations, the centred logratio transformation - mapping the Bayes Hilbert space into L2 - is considered. The methodological developments are illustrated with a real-world example.

EC1376: **A consistent goodness-of-fit test in separable Hilbert spaces with applications to high dimensional data**
Presenter: **Daniel Gaigall**, Leibniz University Hannover, Germany
Co-authors: Marc Ditzhaus

A nonparametric goodness-of-fit test for random variables with values in a separable Hilbert space is considered. The test statistic based on the

Cramer-von-Mises statistic applied to projected data and is given by an integral over the projections. Applications include functional data in $\mathcal{L}^2$-spaces or observations in $\mathbb{R}^d$, where the dimension d may be fixed or it may depend on the sample size n, i.e. $d = d_n$ and $d_n \to \infty$ as $n \to \infty$. The convergence in distribution of the test statistic under the null hypothesis is shown and the consistency of the test is concluded.

EC1719: **On a regression model with constraints in Hilbert spaces**
Presenter: **Marta Garcia Barzana**, Universidad de Oviedo, Spain
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez
The least-squares estimation of linear regression models involves an optimization problem that may be subject to a certain group of constraints. The well-known constrained least-squares approach assumes that the number of inequality linear constrains is fixed. This framework is extended by removing such an assumption. Thus, the number of constrains can vary depending on the sample size. This problem has been addressed in the context of linear regression with interval data. However, the goal is to extend the problem to the abstract case of regression models in Hilbert spaces, which accommodates as well more complex data, such as functional data. An estimator is proposed and a case-based example is presented.

EC657 Room 213 CONTRIBUTIONS IN APPLIED STATISTICS Chair: Gil Gonzalez-Rodriguez

EC1066: **Proxy variables in a matching analysis of data from the Swedish social insurance agency**
Presenter: **Philip Fowler**, Department of Statistics, USBE, Umea University, Sweden
Co-authors: Xavier de Luna, Per Johansson, Petra Ornstein, Sofia Bill, Peje Bengtsson
An enhanced cooperation between the Public Employment Service (PES) and the Social Insurance Agency (SIA) in Sweden was implemented in 2012. The target group of the joint efforts were individuals identified in need of support in order to regain work ability. Such individuals partook in a joint assessment, i.e. meetings with both aforementioned agencies, with the aim to assess the individuals' work abilities. The idea being that such an assessment could lead to a better rehabilitation plan and thus quicker reintroduction to the labour market. To evaluate this, we perform a matching analysis on data from PES, SIA and Statistics Sweden in order to reduce covariate differences between the treatment and control groups, that otherwise could bias the results of the study. Furthermore, a prediction of the individuals' duration in sick leave without intervention was made by case workers at the Social Insurance Agency. This prediction is used as a proxy variable for potential unmeasured confounders in our analysis. Data on the outcome for the treated individuals was purposely not available to the researchers in the matching process and did thus not influence the choice of matching.

EC1493: **Forecasting Spanish electricity demand: Calendar modelling, using discrete intervals mobile seasonalities**
Presenter: **Juan Carlos Garcia-Diaz**, Universitat Politecnica de Valencia, Spain
Co-authors: Oscar Trull
Electricity demand prediction based on time series forecasting is an essential task for the distribution network operators that are in charge of the electricity distribution and planning. They use information of the observed demand to predict future estimations, both in short and long-term. Forecasting and prevision techniques must always deal with several irregularities on the time series. The main significant one is the calendar effect, since national or local holidays affect seriously to the series, and thus the forecasting process. Many previous works deal with this subject by modifying the series and applying dummy variable regressors. We present a new proposal in which the calendar effect is included within the model as a part of it, not as an external modifier. These irregularities are included as new seasonalities using discrete intervals in multiple seasonal Holt-Winters models. This methodology is applied to electricity demand forecasting in Spain, and in special to deal with Easter holidays. Results are shown and discussed.

EC1677: **Symbolic network analysis of bike sharing data**
Presenter: **Vladimir Batagelj**, IMFM, Slovenia
Co-authors: Anuska Ferligoj
Many cities around the world provide a bike sharing service. Some of them (San Francisco Bay Area, New York, Chicago, Boston, Philadelphia, Washington D.C., Melbourne, Paris, London, etc.) offered as open data the data sets about trip histories. They usually consist of the trip start and end (day, time, station) andmember's or bike's id. Sometimes they are augmented by additional data about stations and changes of their status. For some data sets also the weather data were collected. On the basis of these data, interesting insights into the dynamics of bike sharing systems can be obtained. We propose different symbolic networks that summarize selected aspects of a given bike sharing system. We present some analyses of selected bike sharing systems based on them.

EC1670: **Identifying biological modulators using local energy statistics**
Presenter: **Teppei Shimamura**, Nagoya University Graduate School of Medicine, Japan
Co-authors: Yusuke Matsui, Satoru Miyano
The gene expression programs that establish and maintain specific states in a cell are controlled by thousands of transcription factors. The activity of a transcription factor is tightly regulated by some proteins called modulators at the transcriptional, post-transcriptional and post-translational levels. Current approaches for identifying biological modulators from gene expression data are restricted to capture a particular type of three-way dependence between univariate random variables where the ability of a transcription factor to control its target gene is influenced by a modulator, and cannot be used in complex gene regulation when multiple transcription factors and modulators are functionally related. We propose a new method to identify biological modulators by capturing multivariate local dependence based on energy statistics which is a class of statistics based on distances. The method then assigns a measure of statistical significance to each candidate modulator by permutation test. We compare our approach with a leading competitor for modulator identification and illustrate the performance of our method through both simulation and real data analysis.

EC1084: **Using current and lagged explanatory variables to forecast the water discharge of the Mohawk river**
Presenter: **Khawla Mahmood**, University of Brighton, United Kingdom
Building forecasting model in the hydrological field could be one of the most difficult tasks. We study the water discharge, which is regarded one of the main reasons for flooding as soon as the amount of this discharge of water exceeds the normal limits. We construct a new model for forecasting this amount depending on some of the related variables. To precisely perform forecasting process, the decomposition technique has been applied for the studied series, which are water discharge, temperature, precipitation, ground water, tide, and wind speed. One of the most important and common digital filters in the environmental studies, which is the Kolmogorov-Zurbenko filter, has been used to decompose preceding variables. A three different scales series, which are long, seasonal, and short term components, have been extracted by applying this filtering method. Transfer Function-Noise model has been used to formalise the relationship between the response variable and each of the previously mentioned variables. Lagged and current explanatory variables have been incorporated in this TFN equation. Based on the results of the correlation analysis for the residuals of this multiple input transfer function-noise model, it has been revealed that modelling the studied variables using this function has successfully fitted them.

EC661 Room 201 CONTRIBUTIONS IN STATISTICAL MODELLING Chair: Inmaculada Barranco-Chamorro

EC1451: **The covariate balancing generalized propensity score for continuous treatment regimes in the presence of censoring**
Presenter: **Samantha Noreen**, Emory University, United States
Co-authors: Qi Long
The propensity score is widely used for causal inference in observational studies. The covariate balancing propensity score (CBPS) methodology for binary treatment assignments has been proposed to address potential misspecification of the propensity score by exploiting the covariate balancing property of the propensity score. Extending the CBPS, the covariate balancing generalized propensity score (CBGPS) considers general treatment regimes. While the CBGPS has several appealing features, our preliminary numerical studies showed that the CBGPS tends to be numerically unstable. We investigated refinements to the CBGPS approach (iCBGPS), which demonstrated superior performance over the CBGPS in our empirical studies. Extending this approach further, our goal was to develop the iCBGPS in the presence of censoring. Specifically, many observational studies include information on patients censored by death or dropout, and standard propensity score methods including the CBGPS use a complete-case analysis. The iCBGPS methodology in the presence of censoring takes advantage of this extra information previously unused in such analyses. In our subsequent empirical studies, the inclusion of censoring information improved performance over the iCBGPS, as well as standard generalized propensity score and CBGPS methods, in the absence of censoring.

EC1598: **Consistent estimation of random coefficient models with fixed effects**
Presenter: **Thi Thu Hien Pham**, University of Leuven, Belgium
The aggregate random coefficient model previously proposed is widely used for demand modeling where only aggregate market data is available. The model incorporates random coefficients to explain the heterogeneity of consumers' tastes for differentiated product characteristics and price. However, product characteristics which are invariant across markets lead to unobserved heterogeneity. We extend the model to include fixed effects. To solve the extended model, we make use of the limited information maximum likelihood method. Monte Carlo simulations show that the employed method produces consistent estimates of the random coefficients in the presence of fixed effects. An empirical example with a data set based on the European automobile industry is presented. The asymptotic properties of the limited information maximum likelihood estimator are also discussed.

EC1753: **On the choice of time scales in competing risks predictions**
Presenter: **Minjung Lee**, Kangwon National University, Korea, South
In the standard analysis of competing risks data, proportional hazards models are fit to the cause-specific hazard functions for all causes on the same time scale. These regression analyses are the foundation for predictions of cause-specific cumulative incidence functions based on combining the estimated cause-specific hazard functions. However, in predictions arising from disease registries, where only subjects with disease enter the database, disease related mortality may be more naturally modelled on the time since diagnosis time scale while death from other causes may be more naturally modelled on the age time scale. The single time scale methodology may be biased if an incorrect time scale is employed for one of the causes and alternative methodology is not available. We propose inferences for the cumulative incidence function in which regression models for the cause-specific hazard functions may be specified on different time scales. We establish that the covariate conditional predictions are consistent and asymptotically normal using empirical process techniquesand propose consistent variance estimators for constructing confidence intervals. Simulation studies show that the proposed two time scale methods perform well, outperforming the single time scale predictions when the time scale is misspecified.

EC1652: **Unconstrained estimation procedures for exploratory MIRT models**
Presenter: **Sara Fontanella**, The Open University, United Kingdom
Co-authors: Lara Fontanella, Pasquale Valentini, Nickolay Trendafilov
Within the social and behavioural sciences, item-level data are often categorical in nature and item factor analysis (IFA) represents an appropriate tool for their analysis. We consider only a specific class of factor analytic models, namely Multidimensional Item Response Theory (MIRT) models. These models can be defined in terms of both exploratory and confirmatory perspectives. In the former context, identification problems have to be considered. We focus on the rotational indeterminacy. In contrast to the common approach, where this issue is addressed by imposing constraints on the model parameters, we discuss two unconstrained version of Bayesian MIRT models. In the first proposal, the rotational indeterminacy is addressed in an ex-post procedure: the estimation of the model parameters is carried out by an unconstrained Gibbs sampler and the rotational invariance is eliminated in a post-processing step based on the Procrustes approach. The second model, instead, deals with this issue by introducing sparsity in the model. A very successful strategy is to specify prior probability distributions that favour shrinkage for the coefficients of the discrimination parameter matrix. We address the sparse MIRT problem by introducing the sparsity-inducing prior suggested in the Stochastic Search Variable Selection (SSVS) approach for regression models.

EC1393: **Balancing scores in causal diagrams and causal estimates for different data contexts**
Presenter: **Priyantha Wijayatunga**, Umea University, Sweden
Potential outcome model and graphical model are two major frameworks for causal inference tasks such as finding the effectiveness of a medicine for a given disease or that of a training program for unemployed people to find employments, etc. using past observed data. These two frameworks are related with each other and their logical equivalence is shown previously. We show how balancing scores found in the potential outcome model can be represented in the graphical models (causal diagrams). There have been some discussions on this so far, but we show that they are not quite correct. We discuss the diagrams and causal estimates for different contexts such as for matched data in cohort studies, case-control data, etc. These discussions that correct current literature show that one can define new causal estimates for matched case-control data. Note that in such contexts, due to data selection methods it is often not meaningful to use causal effect estimates used in, for example, cohort studies where different data selection methods are used. We also show how the estimates for matched data can be adjusted to obtain those for the population.

EC656 Room 206 CONTRIBUTIONS IN COMPUTATIONAL AND NUMERICAL METHODS IN STATISTICS Chair: Jean Jacod

EC1427: **Probability forecasts for serving competing renewal processes for resource allocation**
Presenter: **Samira Sadeghi**, University of Alberta, Canada
A probabilistic algorithm is constructed for resource allocation: several clients, with arrivals following processes of renewal type specific to each client, are competing for a resource; a reward for serving depends on the client. The motivation comes from television advertising: several networks (channels) broadcasting simultaneously insert commercial breaks, which are filled by a user-tailored content-resource. There is a limited number of resources: if more networks are to be served, some are rejected. The reward for serving the network depends on factors like the number of viewers. The current method is a first-come-first-serve algorithm serving requests in their incoming order. It is in general sub-optimal: if a lucrative network has a big probability of a commercial break in the near future, it may be more profitable to reserve the resource for it rather than serving the less rewarding ones. This raises a question of prediction: it turns out that in this situation, it is more appropriate to predict the probabilities rather than events themselves. The constructed algorithm is shown to yield better results for known probabilities; for estimated ones, experiments indicate that it outperforms the first-come-first-serve algorithm. We discuss several methods for predicting probability distributions relevant in this setting.

EC1728: **Primal path algorithm for compositional data analysis**
Presenter: **Jong-june Jeon**, University of Seoul, Korea, South

The compositional data has two distinguished characteristics compared to a general multivariate data: the values in the observations are nonnegative; the summand of the values is exactly 1. We propose an efficient path algorithm of lasso for the analysis of compositional data. The proposed algorithm has three advantages over the previously developed algorithms. First the algorithm gives the exact solution path based on checking the Karush-Kuhn-Tucker conditions of convex function with linear constraints. Second, the algorithm is easy to extend to the regularized regression problem with general loss functions such as Huberized loss. Third, the algorithm gives an exact ordinary least square estimator in the end of optimizations. We also develop the classification model for the compositional data with the proposed algorithm.

EC1592: **Edgeworth expansion for Euler approximation of continuous diffusion processes**
Presenter: **Bezirgen Veliyev**, Aarhus University, Denmark
Co-authors: Mark Podolskij, Nakahiro Yoshida

An Edgeworth expansion is presented for the Euler approximation scheme of a continuous diffusion process driven by a Brownian motion. Our methodology is based upon a recent work that establishes Edgeworth expansions associated with asymptotic mixed normality using elements of Malliavin calculus. Potential applications of our theoretical results include higher order expansions for weak and strong approximation errors associated to the Euler scheme, and for studentized version of the error process.

EC1437: **Speeding up parameter tuning for multi-class Classification: A Partial parametric path algorithm**
Presenter: **Belen Martin-Barragan**, The University of Edinburgh, United Kingdom

The objective functions of Support Vector Machine methods (SVMs) often include parameters to weigh the relative importance of margins and training accuracy. For multi-class classification problems, in the presence of different misclassification costs, identifying a desirable set of values for these parameters is key for a good performance. We propose a partial parametric path algorithm, based on the property that the path of optimal solutions of the SVMs with respect to the preceding parameters is piecewise linear. This partial parametric path algorithm requires the solution of just one quadratic programming problem, and a number of linear systems of equations. Thus, it can significantly reduce the computational requirements of the algorithm. To systematically explore the different weights to assign to the misclassification costs, we combine the partial parametric path algorithm with a variable neighborhood search method. Our numerical experiments show the efficiency and reliability of the proposed partial parametric path algorithm. Furthermore through our numerical experiments we also verify the combination of partial parametric path algorithm and a variable neighborhood search method helps us to find a good set of parameters systematically.

EC1536: **On numerical inversion of the empirical characteristic function and its applications**
Presenter: **Viktor Witkovsky**, Slovak Academy of Sciences, Slovakia

The methods for making the exact statistical inference frequently require evaluation of the probability density function (PDF), the cumulative distributions function (CDF), and/or the quantile function (QF) of a random variable from its (known) characteristic function (CF), which is defined as a Fourier transform of its probability distribution function. Working with CFs provides an alternative (frequently more simple) route, than working directly with PDFs and/or CDFs. However, the analytical derivation of the PDF and/or CDF by using the inverse Fourier transform is available only in special cases. Thus, in most practical situations, a numerical derivation of the PDF/CDF from the CF is an indispensable tool. The methods based on numerical inversion of the CFs can be used also in non-parametric settings. We shall present brief overview of selected approaches for numerical inversion of the CFs (including the empirical CFs) and selected examples to illustrate applicability of this approach for making the exact statistical inference, as e.g. for derivation the distribution of the Cramer-von Mises and the Anderson-Darling statistics and/or distribution of the test statistic for testing symmetry based on ECF.

Authors Index