PROGRAMME AND ABSTRACTS

6th CSDA International Conference on

Computational and Financial Econometrics (CFE 2012)

http://www.cfe-csda.org/cfe12

and

5th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computing & Statistics (ERCIM 2012)

http://www.cfe-csda.org/ercim12

Conference Center "Ciudad de Oviedo", Spain
1-3 December 2012

ERCIM
European Research Consortium
for Informatics and Mathematics

UNIVERSIDAD DE OVIEDO

http://www.uniovi.es

Queen Mary
University of London

http://www.qmul.ac.uk

Depósito Legal: AS-03972-2012
ISBN: 978-84-937822-2-1

International Organizing Committee:
Ana Colubi, Christophe Croux, Erricos J. Kontoghiorghes and Herman K. Van Dijk.

CFE 2012 Co-chairs:
Andrew Harvey, Yasuhiro Omori, Willi Semmler and Mark Steel.

CFE 2012 Programme Committee:

A. Amendola, G. Amisano, F. Audrino, M. Billio, F. Canova, S. Chaudhuri, C.W.S. Chen, C. Francq, A. M. Fuertes, L. Khalaf, G. Koop, S.J. Koopman, C.-M. Kuan, R. Leon-Gonzalez, H. Lopes, M. McAleer, M. Paolella, H. Pesaran, J.-Y. Pitarakis, D.S.G. Pollock, P. Poncela, T. Proietti, M. Reale, J. Rombouts, S. Sanfelici, K. Sheppard, M.K.P. So, L. Soegner, G. Storti, E. Tzavalis, J.-P. Urbain, D. Veredas, M. Wagner, Q. Yao, P. Zadrozny, J.M. Zakoian and Z. Zhang.

CSDA Annals of CFE Managing Editors:
David A. Belsley, Erricos J. Kontoghiorghes and Herman K. Van Dijk.

ERCIM 2012 Co-chairs:
Gil Gonzalez-Rodriguez, Geoff McLachlan, Hannu Oja, Marco Riani and Stephen Walker.

ERCIM 2012 Programme Committee:

E. Ahmed, A.M. Alonso, A.C. Atkinson, A. Christmann, A. Delaigle, T. Denoeux, M. Falk, P. Filzmoser, P. Foschi, C. Gatu, L.A. Garcia-Escudero, R. Gerlach, M.I. Gomes, A. Gordaliza, A. Grane, X. He, M.D. Jimenez-Gamero, M. Hubert, T. Kneib, V. Leiva, A. Lijoi, J. Lopez-Fidalgo, A. Luati, G. Mackenzie, A. Mayo, S. Meintanis, I. Molina, D. Morales, D. Paindaveine, M.C. Pardo, S. Paterlini, I. Pruenster, E. Ronchetti, L. Ugarte, S. Van Aelst, I. Van Keilegom, P. Vieu, R. Zamar and A. Zeileis.

Local Organizer:

Department of Statistics, OR and DM, University of Oviedo, Spain

Dear Friends and Colleagues,

We warmly welcome you to Oviedo, for the Sixth International Conference on *Computational and Financial Econometrics* (CFE 2012) and the Fifth International Conference of the ERCIM Working Group on *Computing & Statistics* (ERCIM 2012). As many of you know, this annual conference has been established as a leading joint international meeting for the interface of computing, empirical finance, econometrics and statistics, and it is endorsed by the journal of Computational Statistics & Data Analysis (CSDA).

The conference aims at bringing together researchers and practitioners to discuss recent developments in computational methods for economics, finance, and statistics. The CFE-ERCIM 2012 programme consists of almost 200 sessions, 5 plenary talks and over 800 presentations. Peer reviewed papers will be considered for publication in special issues of the journal Computational Statistics & Data Analysis and the CSDA Annals of Computational and Financial Econometrics.

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 university 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 Oviedo, Queen Mary, University of London, CSDA journal and ERCIM.

Looking forward, the CFE-ERCIM 2013 will be held at the Senate House, London, UK, 14-16 December 2013. You are invited and encouraged to actively participate in these events.

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

The CFE-ERCIM 2012 co-chairs and the International Organizing Committee.

ERCIM Working Group on
COMPUTATIONAL AND METHODOLOGICAL STATISTICS
http://www.dcs.bbk.ac.uk/ercim/

AIMS AND SCOPE

The working group (WG) focuses on all computational aspects of statistics. Of particular interest is research in important statistical application areas where both computational and methodological methods have a major impact. The aim is twofold: first, to consolidate the research in computational statistics that is scattered throughout Europe; second, to provide researchers with a network from which they can obtain an unrivalled source of information about the most recent developments in computational statistics and applications.

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

The WG comprises a number of tracks (subgroups, teams) in various research areas of computational statistics. The teams act autonomously within the framework of the WG in order to promote their own research agenda. The activities of the teams — including research proposals — are endorsed by the WG. The teams organize sessions and workshops during the annual WG meetings.

There is a strong link between the ERCIM WG, the ERS-IASC and the Journal of Computational Statistics & Data Analysis.

Specialized groups

Currently the ERCIM WG has approximately 350 members and the following specialized groups:

MCAS: Matrix Computations and Algorithms in Statistics.

EF: Econometrics and Finance.

SSEF: Statistical Signal Extraction and Filtering.

RACS: Robust Analysis of Complex Data Sets.

OHEM: Optimization Heuristics in Estimation and Modelling.

ISA: Imprecision in Statistical Data Analysis.

StatSoft: Statistical Software.

SFD: Statistics for Functional Data.

SEM: Latent Variable and Structural Equation Models.

MM: Mixture Models.

SEA: Statistics of Extremes and Applications.

SAET: Statistical Analysis of Event Times.

You are encouraged to become a member of the WG. For further information please contact the Chairs of the specialized groups (see the WG's web site), or by email at ercim@cfe-csda.org.

SCHEDULE

CFE 2012

Saturday, 1st December 2012

08:50 - 09:00	Opening (Auditorium)
09:00 - 09:50	Plenary Session A
09:50 - 10:20	Coffee Break
10:20 - 12:25	Parallel Sessions B
12:25 - 14:00	Lunch Break
14:00 - 15:40	Parallel Sessions C
15:40 - 16:15	Coffee Break
16:15 - 17:30	Parallel Sessions E
17:40 - 18:30	Plenary Session G
19:30 - 22:00	Reception

Sunday, 2nd December 2012

08:45 - 10:25	Parallel Sessions I
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions J
13:00 - 14:30	Lunch Break
14:30 - 16:10	Parallel Sessions K
16:10 - 16:40	Coffee Break
16:40 - 18:45	Parallel Sessions L
20:15 - 24:00	Conference Dinner

Monday, 3rd December 2012

08:30 - 10:10	Parallel Sessions M
10:10 - 10:40	Coffee Break
10:40 - 12:20	Parallel Sessions N
12:30 - 13:30	Parallel Sessions P
13:30 - 15:00	Lunch Break
15:00 - 16:40	Parallel Sessions Q
16:40 - 17:15	Coffee Break
17:15 - 18:05	Plenary Session R
18:15 - 18:25	Closing (Auditorium)

ERCIM 2012

Saturday, 1st December 2012

08:50 - 09:00	Opening (Auditorium)
09:00 - 09:50	Plenary Session A
09:50 - 10:20	Coffee Break
10:20 - 12:25	Parallel Sessions B
12:25 - 14:00	Lunch Break
14:00 - 16:05	Parallel Sessions D
16:05 - 16:35	Coffee Break
16:35 - 17:25	Plenary Session F
17:40 - 19:20	Parallel Sessions H
20:00 - 22:30	Reception

Sunday, 2nd December 2012

08:45 - 10:25	Parallel Sessions I
10:25 - 10:55	Coffee Break
10:55 - 13:00	Parallel Sessions J
13:00 - 14:30	Lunch Break
14:30 - 16:10	Parallel Sessions K
16:10 - 16:40	Coffee Break
16:40 - 18:45	Parallel Sessions L
20:00 - 24:00	Conference Dinner

Monday, 3rd December 2012

08:30 - 10:10	Parallel Sessions M
10:10 - 10:40	Coffee Break
10:40 - 12:20	Parallel Sessions N
12:30 - 13:20	Plenary Session O
13:20 - 15:00	Lunch Break
15:00 - 16:40	Parallel Sessions Q
16:40 - 17:10	Coffee Break
17:10 - 18:10	Parallel Session S
18:15 - 18:25	Closing (Auditorium)

TUTORIALS, MEETINGS AND SOCIAL EVENTS

TUTORIALS

Tutorials will be given on Friday the 30th of November 2012 at Cajastur premises. The first is given by Andrew Harvey (Dynamic models for volatility and heavy tails) at 09:00-14:00. The second tutorial is given by Hans-Georg Müller (Functional data analysis) at 15:15-20:00.

SPECIAL MEETINGS by invitation to group members

- CSDA Editorial Board reception, *Trascorrales, Old Fish Market*, Friday 30th of November 2012, 20:45-22:30.
- CFE meeting, *Auditorium of the Conference Center*, Saturday 1st of December 2012, 18:35-18:55.
- ERCIM meeting, *Multipurpose room*, Sunday 2nd of December 2012, 18:50-19:05.

SOCIAL EVENTS

- *The coffee breaks* will take place at the Hall of the conference center.
- *A guided city tour* is open to all the registrants and accompanying persons who have booked it on Friday 30th of November at 15:30.
- *Welcome Reception, Saturday 1st of December, from 19:30-20:00 to 22:00-22:30.* The reception is open to all registrants who booked it and accompanying persons who have purchased a reception ticket. It will take place at Latores restaurant. There will be buses outside the Conference Center for the 10-15 minutes transfer . The first bus will be at 19:30 and the last one at 20:00. The return is planned from 22:00 to 22:30. Conference registrants and accompanying persons should bring their reception tickets in order to attend the reception.
- *Conference Dinner, Sunday 2nd of December, 20:00.* The conference dinner is optional and registration is required. It will take place at Latores restaurant. There will be buses outside the Conference Center for the 10-15 minutes transfer. The first bus will be at 20:00 and the last one at 20:30. There will be buses departing after the dinner, and later on until 02:00. Conference registrants and accompanying persons should bring their conference dinner tickets in order to attend the conference dinner.

Venue:

Conference Center "Ciudad de Oviedo", Calatrava building. C/ Arturo Álvarez Buylla s/n 33005, Oviedo, Spain. E-mail: info@pec-oviedo.com

Registration, exhibitors and networking activities

The registration, exhibitors and networking area will be located in the Hall of the conference center.

Lecture rooms

The list of rooms for the paper presentations and their capacity is listed below.

Room	Capacity	Floor	Location	Room	Capacity	Floor	Location
Auditorium	2144	Ground	Back	Sala 1	77	First	Left wing
Sala 2	77	First	Left wing	Sala 3	82	First	Left wing
Sala 4	99	First	Left wing	Sala 5	87	First	Left wing
Sala 6	99	First	Left wing	Sala 7	77	First	Right wing
Sala 8	77	First	Right wing	Sala 9	82	First	Right wing
Sala 10	99	First	Right wing	Sala 11	87	First	Right wing
Sala 12	99	First	Right wing	Sala 13	300	Ground	Right wing
Multipurpose/Multiusos	300	Ground	Center				

Presentation instructions

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

Posters

The poster sessions will take place at the Hall of the Conference Center. The posters should be displayed during half day (in the morning, from the beginning until lunch time and in the afternoon from the lunch time until the end of the day). However the presenting authors have to be present only during the time assigned for their parallel session. The authors will be responsible for placing the posters in the poster panel displays and removing them. The maximum size of the poster is A0.

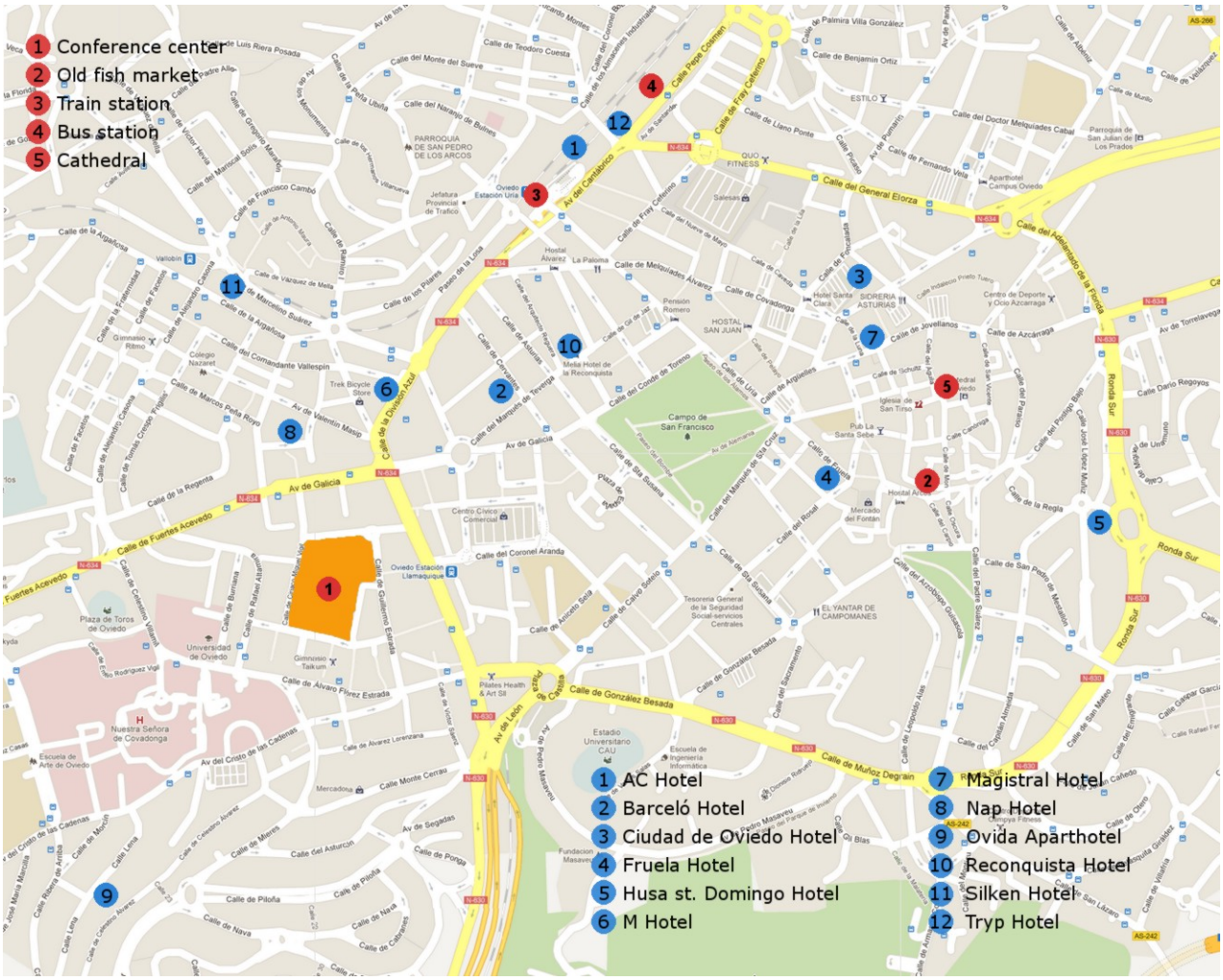
Internet

There will be wireless internet connection. You will need to have your own laptop in order to connect to the internet.

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 information desk.

Map of the venue and nearby area



PUBLICATION OUTLETS

Journal of Computational Statistics & Data Analysis (CSDA)

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

Selected peer-reviewed papers will be published in the CSDA Annals of Computational and Financial Econometrics. Submissions for the CSDA Annals of CFE should contain both a computational and an econometric or financial-econometric component.

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

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

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

SPONSORS

http://www.qmul.ac.uk

http://www.uniovi.es

http://www.elsevier.com

http://www.cajastur.es

Financiado por:

http://www.oviedo.es

EXHIBITORS

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

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

Taylor and Francis Group (URL http://www.taylorandfrancisgroup.com)

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

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

ENDORSED SOCIETIES & GROUPS

Journal of Computational Statistics & Data Analysis, Elsevier

ERCIM Working Group on *Computational and Methodological Statistics*

The Society for Computational Economics

International Association for Statistical Computing

Contents

Saturday 01.12.2012 09:00-09:50 Room: Auditorium Chair: Andrew Harvey Keynote talk CFE-ERCIM

Bootstrapping prediction intervals

Speaker: **Esther Ruiz, Universidad Carlos III de Madrid, Spain**

When predicting the future evolution of a given variable, there is an increasing interest in obtaining prediction intervals which incorporate the uncertainty associated with the predictions. Due to their flexibility, bootstrap procedures are often implemented with this purpose. First, they do not rely on severe distributional assumptions. Second, they are able to incorporate the parameter uncertainty. Finally, they are attractive from a computational point of view. Many bootstrap methods proposed in the literature rely on the backward representation which limits their application to models with this representation and reduces their advantages given that it complicates computationally the procedures and could limit the asymptotic results to Gaussian errors. However, this representation is not theoretically needed. Therefore, it is possible to simplify the bootstrap procedures implemented in practice without losing their good properties. The bootstrap procedures to construct prediction intervals that do not rely on the backward representation can be implemented in a very wide range of models without this representation. In particular, they can be implemented in univariate ARMA and GARCH models. Also, extensions to unobserved component models and multivariate VARMA models will be considered. Several applications with simulated and real data will be used to illustrate the procedures.

Saturday 01.12.2012 16:35-17:25 Room: Auditorium Chair: Gil Gonzalez-Rodriguez Keynote talk ERCIM 1

Nonlinear methods for the analysis of functional data

Speaker: **Hans-Georg Mueller, University of California Davis, United States**

For the analysis of samples of observed random functions, linear methods such as functional principal component analysis have proven very useful. As a consequence, in recent years much research effort has gone into exploring the properties of linear functional approaches, such as functional linear regression. We review some of these developments, especially in the context of demographic applications, and then demonstrate that for many applied problems, suitable nonlinear approaches provide a promising alternative. These approaches are particularly adequate for functional data lying on a nonlinear low-dimensional manifold. An interesting situation where this happens is the presence of random time variation. We will also explore applications to functional regression and classification.

Saturday 01.12.2012 17:40-18:30 Room: Auditorium Chair: Yasuhiro Omori Keynote talk CFE 1

Modeling structural changes in volatility

Speaker: **Luc Bauwens, Universite catholique de Louvain, Belgium**

The aim is to review univariate and multivariate ARCH models for structural changes in volatility and correlations and their Bayesian estimation by MCMC methods. Several algorithms will be explained and compared. The issue of detecting the number of breaks will be a central theme of the lecture.

Monday 03.12.2012 12:30-13:20 Room: Auditorium Chair: Xuming He Keynote talk ERCIM 2

Assigning statistical significance in high-dimensional problems

Speaker: **Peter Buhlmann, ETH Zurich, Switzerland**

High-dimensional data, where the number of variables is much larger than the sample size, occur in many applications nowadays. During the last decade, remarkable progress has been achieved in terms of point estimation and computation. However, one of the core statistical tasks, namely to quantify uncertainty or to assign statistical significance, is still in its infancy for many problems and models. After a brief review of some main concepts for high-dimensional statistical inference, we present procedures and corresponding theory for quantifying significance in single and multiple testing problems. Illustration on various examples highlights the methods' user-friendly nature for high-dimensional data analysis.

Monday 03.12.2012 17:15-18:05 Room: Auditorium Chair: Raquel Prado Keynote talk CFE 2

Portfolio–implied risk assessment

Speaker: **Stefan Mittnik, University of Munich, Germany**

The aim is to discuss methods for estimating quantities that are relevant for assessing market risk associated with specified sets of financial instruments or risk drivers. Such quantities are, for example, volatilities, correlations and tail–conditioned correlations. The estimates are derived by analyzing the behavior of portfolios constructed from these assets. We demonstrate how such portfolio–implied methods can be used to derive value–at–risk–implied correlation measures—as specified, for example, in the proposed Solvency II regulatory framework for European insurance companies—and related tail–dependence measures. Estimation strategies, such as pairwise, exactly–identified and overidentified approaches, and their properties are discussed. Extensions to the dynamic case, assuming multivariate GARCH structures, are presented, where portfolio–implied methods allow us, for example, to derive nondiagonal VEC–GARCH–type models, using only trivial, univariate modeling techniques.

Saturday 01.12.2012 10:20 - 12:25 Parallel Session B – CFE

CSI03 Room Auditorium BAYESIAN ECONOMETRICS Chair: Herman van Dijk

C489: **Multivariate realized stochastic volatility**
Presenter: **Yasuhiro Omori**, University of Tokyo, Japan
The multivariate realized stochastic volatility models with dynamic correlations are proposed. The realized stochastic volatility (RSV) model jointly models the daily financial asset returns and its realized volatility computed by using high frequency data. It is known to adjust the possible bias of the realized volatility due to the presence of non-trading hours and market microstructure noise in transaction prices. An MCMC algorithm is proposed to estimate model parameters where we have many latent variables.

C660: **Bayesian state-space models with structured priors for multiple time series**
Presenter: **Raquel Prado**, UC Santa Cruz, United States
Two classes of time-domain models that are useful for describing the relationships across multiple time series are presented. In particular, we consider hierarchical factor models with autoregressive factors and structured prior distributions that allow us to infer the common structured and often scientifically interpretable factors that underlie multiple time series in many practical scenarios. We also consider state-space models with sparsity priors on the parameters that describe the coupling relationship between any two of the multiple time series. We discuss several aspects of the proposed models, as well as posterior inference through Markov chain Monte Carlo and sequential Monte Carlo methods. We illustrate the two modeling approaches in real applications and explore connections with frequency-domain methods.

C1043: **Flexible mixture-of-experts models for longitudinal survival data**
Presenter: **Mattias Villani**, Linkoping University, Sweden
Co-authors: Matias Quiroz
Flexible mixture models for analyzing discrete-time survival data with time-varying covariates are proposed. The models are a mixture-of-experts models where the subjects' transitions between mixture components over time is explained by covariates that vary over the subject's lifetime. Our focus is on models and algorithms that can be used for longitudinal data sets with a large number of subjects. We apply the approach to a data set to model the annual bankruptcy risks for more than 228,000 Swedish corporations as a function of firms' financial statements and macroeconomic variables.

CS01 Room 4 MODELLING AND FORECASTING FINANCIAL TIME SERIES Chair: Alessandra Amendola

C433: **Penalized smoothed probability distance clustering of financial time series**
Presenter: **Gianluca Frasso**, University of Naples Federico II, Italy
Co-authors: Roberta Siciliano, Antonio D'Ambrosio
A new clustering approach for time series is introduced. We move from the idea to combine a probability distance clustering method with a penalized smoothing procedure of time series. We take advantage from the flexible class of probability distance clustering. It allows for probabilistic allocation of series to groups and it is based on the principle that probability and distance are inversely related. On the other hand the class of penalized smoothing procedure allows to efficiently separate the signal and the noise component of a generic time series. We will also evaluate the performances of this clustering procedure using different distance measures. Comparisons of our proposal with already known procedures are provided. We discuss the applicability of Penalized Smoothed Probability Distance Clustering (PSPDC) to financial time series emphasizing its fuzzy nature.

C451: **A simple volatility risk premium model**
Presenter: **Francesco Violante**, Maastricht University, Netherlands
Co-authors: Jeroen Rombouts, Lars Stentoft
A class of semiparametric models for daily asset return volatility based on multiple sources of information is proposed. The return volatility dynamics is constructed from low and high frequency data and uses information from option data as well. Using the definition of variance swap contracts we extract the volatility risk premium, or investor risk aversion and thus we can disentangle the volatility dynamics under the statistical and the risk neutral measure. The model is intuitive and simple to implement, and can be estimated using a quasi-maximum likelihood approach. Implementing the model on a number of stock market indices shows that there is clear evidence of non negligible volatility risk premia and indicates significant temporal dependencies in the estimated premia. The model works well in practice and helps predict future stock market returns providing important gains in terms of option pricing accuracy. Finally we relate the volatility risk premium to a set of macro-finance state variables.

C459: **Indirect estimation of alpha-stable GARCH models**
Presenter: **Alessandro Parrini**, University of Florence, Italy
Co-authors: Giorgio Calzolari, Roxana Halbleib
It is a well-known fact that financial returns exhibit conditional heteroscedasticity and fat tails. While the GARCH-type models are very popular in depicting the conditional heteroscedasticity, the α-stable distribution is a natural candidate for the conditional distribution of financial returns. The α-stable distribution is a generalization of the normal distribution and is described by four parameters, two of which deal with tail-thickness and asymmetry. However, practical implementation of α-stable distribution in finance applications has been limited by its estimation difficulties. We propose an indirect approach of estimating GARCH models with α-stable innovations by using as auxiliary models GARCH-type models with Student's t distributed innovations. We provide comprehensive empirical evidence on the performance of the method within a series of Monte Carlo simulation studies and an empirical application to financial index returns.

C821: **Long memory and asymmetry for matrix-exponential dynamic correlation processes**
Presenter: **Manabu Asai**, Soka University, Japan
Co-authors: Mike So
A fractionally-integrated matrix-exponential dynamic conditional correlation (FIEDCC) model is suggested in order to capture the asymmetric effects and long- and short-range dependencies of a correlation process. We also propose to employ an inverse Wishart distribution for the disturbance of covariance structure, which gives an alternative interpretation for multivariate t conditional distribution. By using the inverse Wishart distribution, we present a three-step procedure to obtain initial values for estimating high dimensional conditional covariance model with the multivariate t distribution. We investigated the finite sample properties of the ML estimator. Empirical results for 9 assets from chemical, banks, and oil and gas producers in US indicate that the new FIEDCC model outperforms the other dynamic correlation models for the AIC and BIC and for forecasting Value-at-Risk thresholds. Furthermore, the new FIEDCC model captures the stronger connection of the assets for the period right after the global financial crisis.

C455: **Model uncertainty and forecast combination in high dimensional multivariate volatility prediction**
Presenter: **Alessandra Amendola**, University of Salerno, Italy
Co-authors: Giuseppe Storti
The identification of the optimal forecasting model for multivariate volatility prediction is not always feasible due to the curse of dimensionality typically affecting multivariate volatility models. In practice only a subset of the potentially available models can be effectively estimated after imposing severe constraints on the dynamic structure of the volatility process. This situation leaves scope for the application of forecast combination strategies as a tool for improving the predictive accuracy. The aim is to propose some alternative combination strategies and compare their performances in forecasting high dimensional multivariate conditional covariance matrices for a portfolio of US stock returns. In particular, we will consider the combination of volatility predictions generated by multivariate GARCH models, based on daily returns, and dynamic models for realized covariance matrices, built from intra-daily returns.

CS35 Room 6 ESTIMATING AGENT BASED MODELS Chair: Peter Winker

C250: **Embedding financial market data on networks**
Presenter: **Tiziana Di Matteo**, Kings College London, United Kingdom
Co-authors: Won-Min Song, Tomaso Aste
The aim is to introduce a graph-theoretic approach to extract clusters and hierarchies in complex data- sets in an unsupervised and deterministic manner, without the use of any prior information. This is achieved by building topologically embedded networks containing the subset of most significant links and analyzing the network structure. For a planar embedding this method provides both the intra-cluster hierarchy, which describes the way clusters are composed, and the inter-cluster hierarchy which describes how clusters gather together. The performance, robustness and reliability of this method is discussed by investigating several synthetic data-sets, finding that it can outperform significantly other established approaches. Applications to financial data-sets show that industrial sectors and specific activities can be extracted and meaningfully identified from the analysis of the collective fluctuations of prices in an equity market.

C775: **Behavioural breaks in the heterogeneous agent model**
Presenter: **Jiri Kukacka**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik
The main aim is to incorporate selected findings from behavioural finance into a Heterogeneous Agent Model using a previous framework. In particular, we analyse the dynamics of the model around the so-called 'Break Point Date', when behavioural elements are injected into the system and compare it to our empirical benchmark sample. Behavioural patterns are thus embedded into an asset pricing framework, which allows to examine their direct impact. Price behaviour of 30 Dow Jones Industrial Average constituents covering five particularly turbulent U.S. stock market periods reveals interesting patterns. To replicate it, we apply numerical analysis using the Heterogeneous Agent Model extended with the selected findings from behavioural finance: herding, overconfidence, and market sentiment. We show that these behavioural breaks can be well modelled via the Heterogeneous Agent Model framework and they extend the original model considerably. Various modifications lead to significantly different results, and model with behavioural breaks is also able to partially replicate price behaviour found in the data during turbulent stock market periods.

C891: **Behavioral heterogeneity in futures markets**
Presenter: **Chueh-Yung Tsao**, Chang Gung University, Taiwan
Co-authors: Li Ling Tai
A heterogenous agent model for futures markets is developed. We assume that traders have heterogenous expectations about futures prices. To implement the model, we consider three types of traders, namely, the fundamentalists, the chartists, and the arbitragers. The chartists predict future prices based on historical prices. The fundamentalists trade on logn-run mean-reversion of futures prices to the fundamental values. In our model, the fundamental values are established based on the cost-of-carry model. Alternatively, the arbitragers believe that the trading volumes, the basis spreads, the open interests, and the time to maturities can provide useful information when predicting futures prices. We apply the heterogenous model to TAIEX futures market. We also estimate the three types of agents, respectively, as a representative agent in the market to be compared with. It is found from the empirical analysis that the heterogenous model outperforms the three representative agent models in terms of the explanatory ability (R^2). We further study the weight attracted to the three agents in the heterogenous model. It is found that the weights are adjusted over time and the chartists and the arbitragers are more active during the 2008 Financial Tsunami.

C941: **Agent-based evaluation for signaling effects in markets with asymmetric information**
Presenter: **Nisuo Du**, Kansai University, Japan
Co-authors: Tadahiko Murata
Effects of signaling in a market in which there is asymmetric information between sellers and buyers are evaluated. In such a market, only sellers know more about their products than buyers. Therefore buyers do not want to pay for the product more than the price for the average quality they expect in the market. Those buyers action make sellers hesitate to submit products with better quality in the market. Then Shortage of better products reduces the average quality in the market. In turn, buyers expect to pay for the product at lower price. Such iterations make the market with lower quality products. In order to cope with the problem, sellers are encouraged to show appropriate signals about the products. In our simulation, we show how signals have effects on the quality of products in a market by an agent-based approach. A preceding study shows a signal whose cost depends on the quality of products is effective. However, through our simulations and analysis, we find that the balance between the cost of quality- related signals and the product of the amount of products and their producing cost should be carefully considered in the market. Our simulation shows that the market with high quality products can be attained when the cost of quality-related signals is specified appropriately.

C942: **Agent-based pension simulation for widely diverse employment patterns**
Presenter: **Tadahiko Murata**, Kansai University, Japan
Co-authors: Hiroshi Arikawa
An agent-based approach for pension simulation in Japan is described. As one of the industrially advanced countries, Japan faces an aging population and a decreasing birth rate. In such a depopulating society, how to support retirees by national pension program is really important. The government tries to show the effectiveness of the current pension system using a model case of a family with a husband, who works for the same company for 40 years, and a wife, who is a full-time housewife for 40 years. Although they used to employ this model case for their explanation, employment patterns widely diversify nowadays. In order to show a pension plan for each employment pattern, an agent-based approach is one of promising way. However, implementing agent-based simulation with a huge number of agents, we should solve how to keep the internal information properties for each agent. In this paper, we employed a memory-mapped file I/O system call to keep the information of the huge number of agents. Using this method, we come to see pension plans for various employment patterns.

CS46 Room 1 DIMENSION REDUCTION IN TIME SERIES Chair: Pilar Poncela

C279: **Evidence of excess of co-movement in commodity prices at high frequencies**
Presenter: **Lya Sierra**, Pontificia Universidad Javerana, Colombia
Co-authors: Pilar Poncela, Eva Senra
The importance of common factors in the evolution of commodity prices is analyzed. We test if prices of different commodities tend to move together in excess of what can be explained by fundamentals. For this purpose, we consider monthly data of 51 commodities from 1980 to 2012 in order to confirm whether prices of raw resources show excess of co-movement at high frequencies. A dynamic factor model is estimated using principal components. We also test if macroeconomic and financial variables are linked to the common factor as well as whether such variable and factor affects commodity prices. As fundamentals we explore the role of variables such as interest rates, US dollar Exchange rate, world industrial production and Down Jones Stock Market Index.

C308: **Detecting big structural breaks in large factor models**
Presenter: **Liang Chen**, University Carlos III de Madrid, Spain
Co-authors: Juan Jose Dolado, Jesus Gonzalo
Time invariance of factor loadings is a standard assumption in the analysis of large factor models. Yet, this assumption may be restrictive unless parameter shifts are mild (i.e., local to zero). We develop a new testing procedure to detect big breaks in these loadings at either known or unknown dates. It relies upon testing for parameter breaks in a regression of the first of the r factors estimated by PCA on the remaining r-1 factors, where r is chosen according to a previously introduced information criterion. The test fares well in terms of power relative to other recently proposed tests on this issue, and can be easily implemented to avoid forecasting failures in standard factor-augmented (FAR, FAVAR) models where the number of factors is a priori imposed on the basis of theoretical considerations.

C261: **Dynamic sparse factor models**
Presenter: **Sylvia Kaufmann**, Study Center Gerzensee, Switzerland
Co-authors: Christian Schumacher
The analysis of large panel data sets (with N variables) involves methods of dimension reduction and optimal information extraction. Dimension reduction is usually achieved by extracting the common variation in the data into few factors (k, where $k \ll N$). Factors are proposed to be estimated within a state space framework. To achieve a parsimonious representation, the $N \times k$ factor loading matrix is estimated under a sparse prior, which assumes that either many zeros may be present in each column of the matrix, or many rows may contain zeros. The significant factor loadings in columns define the variables driven by specific factors and offer an explicit interpretation of the factors. Zeros in rows indicate irrelevant variables which do not add much information to the inference. The contribution also includes a new way of identification which is independent of variable ordering.

C475: **Factor models to assess consensus and disagreement in forecasting**
Presenter: **Eva Senra**, Universidad de Alcala, Spain
Co-authors: Pilar Poncela
The combination of individual forecasts is often a useful tool to improve forecast accuracy. Principal component regression is considered to produce a single forecast from several individual forecasts provided by the Survey of Professional Forecasters for the main US macroeconomic aggregates. We find that one principal component provides good adjustments and improvements in forecasting accuracy with respect to the average of the forecasts, but also that a second one might also improve the results obtained. There is also empirical evidence on high correlation between the first principal component and the average of the forecasts and the second principal component and the variance of them, which allows for an interpretation of the principal components as measures related to consensus and disagreement among the forecasters. We develop conditions based on the structure of the correlation matrix among the panelists, in order to obtain a related interpretation of the principal components considered as prediction factors. We also check through a small simulation to assess robustness.

C543: **Finite sample performance of small versus large scale dynamic factor models**
Presenter: **Maximo Camacho**, University of Murcia, Spain
Co-authors: Gabriel Perez-Quiros, Rocio Alvarez
The finite-sample performance of small versus large scale dynamic factor models is examined. Monte Carlo analysis reveals that small scale factor models outperform large scale models in factor estimation and forecasting for high level of cross-correlation across the idiosyncratic errors of series from the same category, for oversampled categories, and specially for high persistence in either the common factor series or the idiosyncratic errors. In these situations, there is a superior accuracy of pre-selecting the series to be used in a small scale dynamic factor model even for arbitrary selections of the original set of indicators.

CS53 Room 3 FINANCIAL MARKET AND MACRO Chair: Willi Semmler

C372: **Labour dynamics in macro models. Shock transmission through unemployment flows**
Presenter: **Ekkehard Ernst**, ILO, Switzerland
The aim is to develop and estimate a macroeconomic model with labour flows to understand the shock transmission of financial shocks to employment and wages. The model is based on a standard New-Keynesian framework but augmented with interactions with job creation and destruction dynamics. Policy shocks include monetary and fiscal policies but reforms that affect the pass-through of these shocks on labour markets are also considered. The model is fully estimated using GMM on a panel of 14 OECD countries to allow for realistic parameter values. The model shows realistic dynamics following productivity and financial market shocks. The latter, in particular, involve large deviations of output and employment from the steady state. In order to identify the nature of the transmission of shocks, various scenarios are analysed, including changes to employment protection and collective bargaining. The fact that all parameters are estimated allows for a precise breakdown of different transmission channels. In particular, it can be shown that changes in the transmission of a financial market shock to employment are mainly triggered by the working of the interest rate channel, more than through wage or investment growth.

C742: **A comparison of numerical methods for the solution of continuous time DSGE models**
Presenter: **Juan Carlos Parra Alvarez**, Aarhus University, Denmark
The accuracy and computing time of a set of techniques that approximate the solution of continuous-time DSGE models is assessed. Using the stochastic neoclassical growth model, the aim is to compare time-discretization methods based on a Markov chains approximation, first and second order perturbations and projection methods with collocation and least squares residual functions. All the approximation techniques are applied on the concentrated HJB equation and the first order conditions that describe the general equilibrium of the economy. Two cases are studied depending on whether a closed form solution can be derived. When no analytical solution is available it is studied the effect of different degrees of non-linearities on the approximated solution through changes in the risk aversion and the variance of the shocks. Projection methods are found to be more accurate than perturbation though they are more computational intensive. The accuracy of the time-discretization approach is sensible

to the design of the state-space, exhibits fitting problems at the boundaries of the grid and it suffers from the curse of dimensionality. The results encourage the use of perturbations for reasonable values of the structural parameters of the model and suggests the use of projection methods when a high degree of accuracy is required.

C714: **Fiscal policy rules, the term structure and macroeconomic stability in monetary unions: A behavioral macro-finance approach**
Presenter: **Christian Proano**, The New School for Social Research, United States

Using a stylized no-arbitrage macro-finance two-country model of a monetary union, the performance of various types of fiscal policy rules with respect to national and union-wide macroeconomic stabilization is investigated. In particular the effect of fiscal and external imbalances for the term structure of interest rates in the two countries is analyzed under the assumption that the economic agents use behavioral heuristics for the assessment of macroeconomic and intertemporal risk.

C121: **Specification analysis of international treasury yield curve factors**
Presenter: **Luca Tiozzo Pezzoli**, Banque de France - Paris Dauphine University, France
Co-authors: Fulvio Pegoraro, Andrew F. Siegel

The multi-country term structure literature is characterized by a relevant lack of agreement, not only about the number of latent factors that are required to explain the joint dynamics of two or more yield curves, but also about their nature, namely how many of them are common (i.e., they affect yields in all countries) and how many are local factors (i.e. they affect yields in one country only). The purpose is to provide, adopting a state-space approach and working with a new international Treasury zero-coupon bond database, a reliable detection of the optimal number of common and local factors required to jointly explain multi-country yield curves, and to obtain efficient Kalman-based estimation of the factor scores in a panel data with small cross-sectional and large time-series dimensions, and featuring strong serial and cross-sectional dependence. The international yield curve database is built at high (daily) frequency, applying across countries (U.S., Germany, U.K. and Japan) common criteria to filter coupon bond Treasury raw data, to guarantee a uniform level of liquidity and to interpolate the discount functions.

C590: **Financial market efficiency during crisis periods: A new perspective based on multifractality**
Presenter: **Cristina Sattarhoff**, University of Hamburg, Germany

The losses in financial market efficiency during the current global economic crisis are analyzed. We propose a new statistical test for the type of financial market efficiency as well as a proxy for the degree of financial market efficiency, based on the Multifractal Random Walk model. We can differentiate between Random Walk and multifractal martingale price processes as well as between different degrees of martingale efficiency. Our study on 22 international financial markets spans two time periods separated by the collapse of the US housing bubble. The results indicate efficiency losses within 16 countries during the crisis period. This is in line with the economic intuition that larger amounts of information enter the markets during crisis, with corresponding delays in the information processing.

CS64 Room 5 INFERENCE IN TIME SERIES VOLATILITY MODELS Chair: Jean-Michel Zakoian

C429: **Likelihood-based inference in the ACR cointegrated model**
Presenter: **Emil Nejstgaard**, University of Copenhagen, Denmark
Co-authors: Paolo Paruolo, Anders Rahbek

We consider likelihood-based inference in a general class of regime-switching cointegrated vector equilibrium correction (ECM) models. This framework allows for epochs of non-stationary behavior, asymmetric error correction and switching error covariance. Allowing for non-stationarity and asymmetric error correction gives enhanced flexibility in modeling persistent series, whereas the regime dependent error covariance provides a way of modeling simple time-changing volatility patterns in series with heteroskedasticity. We discuss consequences for inference when cointegration vectors are estimated and when a constant is included in the cointegration relations.

C810: **Monitoring procedure for parameter change in causal time series**
Presenter: **William Kengne**, University of Cergy Pontoise, France
Co-authors: Jean-Marc Bardet

A new sequential procedure is proposed to detect changes in the parameters of a process $X = (X_t)_{t\in\mathbb{Z}}$ belonging to a large class of causal models (such as AR(∞), ARCH(∞), TARCH(∞), ARMA-GARCH processes). The procedure is based on a difference between the historical parameter estimator and the updated parameter estimator, where both these estimators are based on the quasi-likelihood of the model. Unlike classical recursive fluctuation test, the updated estimator is computed without the historical observations. The asymptotic behavior of the test is studied and the consistency in power as well as an upper bound of the detection delay are obtained. The procedure is applied to the daily closing values of the Nikkei 225, S&P 500 and FTSE 100 stock index. We show in this real-data applications how the procedure can be used to solve off-line multiple breaks detection.

C434: **Quasi likelihood estimation for continuously invertible models**
Presenter: **Olivier Wintenberger**, Universite Paris Dauphine, France
Co-authors: Cai Sixiang

The inference of volatility models that are not everywhere invertible is considered. The concept of continuous invertibility is introduced to deal with such models as the EGARCH(1,1) model. A new methodology is proposed to constrain the model to be continuously invertible such that inference and forecasting are reliable. Monte Carlo studies are presented.

C591: **Estimating stochastic volatility models using prediction-based estimating functions**
Presenter: **Anne Floor Brix**, Aarhus University, Denmark
Co-authors: Asger Lunde

Prediction-based estimating functions (PBEFs) are reviewed and PBEFs for the Heston stochastic volatility model are derived. The link between PBEFs, regression and GMM is also explained. The finite sample performance of the PBEF based estimator is investigated in a Monte Carlo study, and compared to that of a GMM estimator based on first and second order conditional moments of integrated volatility. A feasible GMM estimator is computed by using realized variance. In contrast, PBEFs are computed using squared high-frequency returns directly. This informational advantage of PBEFs is confirmed by the Monte Carlo results. The two competing estimation methods are then extended to handle the case where the observed log-price process is contaminated by i.i.d. market microstructure (MMS) noise. First, the impact of MMS noise on the parameter estimates from the two estimation methods without noise correction is studied. Second, a noise robust GMM estimator is constructed by replacing integrated volatility with a realized kernel. The PBEFs are also recalculated in the noise setting, and the two estimation methods ability to correctly account for the noise is investigated. The Monte Carlo study shows that the estimator based on PBEFs also outperforms the GMM estimator in the setting of MMS noise.

C661: **Tests for volatility shifts in GARCH against long-range dependence**
Presenter: **Taewook Lee**, Hankuk University of Foreign Studies, Korea, South
Co-authors: Changryong Baek, Moosup Kim
Many empirical findings show that volatility in financial time series exhibits high persistence. Some researchers argue that such a persistency is due to volatility shifts in the market. Others believe that this is a natural fluctuation that can be explained by a stationary long range dependence model. These two approaches confuse many practitioners and forecasts for future volatility are dramatically different depending on which model to use. We consider a statistical testing procedure to distinguish volatility shifts in GARCH model against long-range dependence. Our testing procedure is based on the residual-based CUSUM test which is designed to correct the size distortion observed for GARCH models. We examine the validity of our methods by providing asymptotic distributions of test statistic under both volatility shifts and LRD models. Also, a Monte Carlo simulation study shows that our proposed methods achieve good size while provide a reasonable power against LRD. It is also observed that our test is quite robust to the misspecified GARCH models.

CS63 Room 2 REAL-TIME ANALYSIS WITH REVISIONS AND VOLATILITY Chair: Peter Zadrozny

C871: **The effects of banks' funding costs on interest rates and loan volumes**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Bjørn Helge Vatne
A dynamic factor model and detailed panel data from Norwegian banks are used to study the effects of banks' funding costs on interest rates and loan volumes. The data set consists of quarterly data for 2002Q1-2011Q3 and includes information on loan volumes and interest rates for firms and households. The cost of market funding is represented in the model by the three-month Norwegian Inter Bank Offered Rate (NIBOR) and the spread on unsecured senior bonds issued by Norwegian banks. Our estimates show clear evidence of incomplete pass-through: a ten basis point increase in NIBOR leads to an approximately 9 basis point increase in retail loan rates. Our findings are consistent with the view that banks face downward sloping demand curves for loans. We find that credit demand from households is more elastic with regard to the loan rate than demand from businesses. On the other hand, the credit supply to firms is more influenced by market risk than the supply of loans to households.

C767: **Modeling multivariate data revisions**
Presenter: **Samad Sarferaz**, ETH Zurich, Switzerland
Co-authors: Jan Jacobs, Jan-Egbert Sturm, Simon van Norden
Although many official statistics are jointly produced by statistical agencies, data revisions in individual time series are typically studied in isolation ignoring information in other related series. A previous modeling framework is extended to multivariate data revisions. We consider systems of variables, which may be related by one or more identities. Measurement errors in each variable may be composed of news and noise errors. We show how to model such systems with standard linear state space models. We motivate and illustrate the multivariate modeling framework with Swiss current account data using Bayesian econometric methods for estimation and inference, and thereby produce estimates of the underlying true values together with their respective error bands.

C730: **Now-casting and the real-time data flow**
Presenter: **Marta Banbura**, European Central Bank, Germany
Co-authors: Domenico Giannone, Michele Modugno, Lucrezia Reichlin
The term now-casting is a contraction for now and forecasting and has been used for a long time in meteorology and recently also in economics. We survey recent developments in economic now-casting with special focus on those models that formalize key features of how market participants and policy makers read macroeconomic data releases in real time, which involves: monitoring many data, forming expectations about them and revising the assessment on the state of the economy whenever realizations diverge sizeably from those expectations.

C817: **A Cholesky-factor regularization method for estimating a VARMA model of volatility**
Presenter: **Klaus Wohlrabe**, Ifo Institute, Germany
Co-authors: Stefan Mittnik, Peter Zadrozny
A Cholesky-factor regularization method for estimating a VARMA volatility model for single- or multi-frequency time-series data is described. The method is applied to single- and multi-frequency, financial and macroeconomic data. The key step in the method is adding a positive diagonal matrix to the outer-product matrix of residuals in every period, in order to make the matrix positive definite and, hence, Cholesky factorizable. The step is analogous to adding a positive diagonal matrix to an ill-conditioned X'X matrix in a ridge regression, in order to make X'X well conditioned, where X denotes the matrix of observations on explanatory variables. A main advantage of the present method appears to be that it allows prior data analysis, which leads to avoiding unnecessary, costly, nonlinear estimation of models and should make feasible the estimation of multivariate volatility models for more than a few series. By contrast, nondiagonal, multivariate, BEKK and VEC, volatility models can generally be successfully estimated only for a few series.

C446: **Real-time state-space method for computing smoothed estimates of future revisions of U.S. monthly chained CPIU**
Presenter: **Peter Zadrozny**, Bureau of Labor Statistics, United States
Well known CPI of urban consumers is never revised. Recently initiated chained CPI is initially released every month (ICPI), for that month without delay within BLS and for the previous month with 1 month delay to the public. Final estimates of chained CPI (FCPI) are released every February for January to December of the calendar year 2 years before. Every month, simultaneously with the release of ICPI, we would like to have a best estimate, given current information, of FCPI for that month, which will not be released until 2 calendar years later. ICPI and FCPI data may be indexed in historical time by months of occurrence or in current or real time by months of observation or release. The essence of the solution method is to use data indexed in historical time to estimate models and, then, for an estimated model, to use data indexed in real time to estimate FCPI. We illustrate the method with regression and VARMA models. Using a regression model, estimated FCPI is given directly by an estimated regression line; and, using a VARMA model, estimated FCPI is computed using a Kalman filter.

Saturday 01.12.2012 10:20 - 12:25 Parallel Session B – ERCIM

ESI05 Room Multipurpose CIRCULAR AND SPHERICAL DATA Chair: Miguel Fernandez

E473: **Directional statistics strikes back**
Presenter: **Kanti Mardia**, Leeds University, United Kingdom
Directional Statistics has a curious history. It may be said that its heyday was 1970-1980 when the subject came into the limelight, partly with my book and the discussion paper published in the Journal of the Royal Statistical Society. Then in 1990-1999 there was almost a lull in the subject. But then, around 2000 or so, a resurgence of interest led to new strides being taken. This now continues and, partly through the recognition by Image Analysts and Life Scientists of its importance, many new challenges have come to the fore. For example, in life sciences, protein plays a key role. The proteins are biological macromolecules that are of primary importance to all living organisms and there are various open problems including the Nobel-Prize-type problem related to protein folding. All these questions mainly depend on the shape of the protein in 3-D, which can be summarized in terms of either the configuration of points (landmarks) or more compactly by conformational angles. These types of new applications have led to novel methods in directional data analysis such as circular skew distributions, angular order restricted inference, distributions on torus, Laplace approximations, dedicated simulations, and shape connections. We will describe these methodologies and discuss their impact.

E275: **Advances in the modelling of circular data**
Presenter: **Arthur Pewsey**, University of Extremadura, Spain
Recent years have seen a renewed interest in the modelling of circular data, and numerous new distributions have been proposed in the literature. We will provide a review of the most important developments and illustrate how some of the new models can be fitted using R.

E554: **Circular order restricted inference methods for the analysis of data from cell cycle genes**
Presenter: **Cristina Rueda**, University of Valladolid, Spain
Co-authors: Miguel Fernandez, Sandra Barragan, Shyamal Peddada
A cell division cycle is a well-coordinated process in eukaryotes with cell-cycle genes exhibiting a periodic expression over time. There is considerable interest among cell-biologists to determine genes that are periodic in multiple organisms and whether such genes are also evolutionarily conserved in their time to peak expression. A novel statistical methodology is developed to deal with this problem using circular models and the concept of circular ordering. The idea is to use the ordering information to improve the efficiency of the statistical procedures. In particular, in this first stage, estimators and hypothesis testing procedures have been derived for Von-Mises circular models with ordered parameters. Using this methodology we discover several genes that are evolutionarily conserved in their temporal order in S. pombe, S. cerevisiae and H. sapiens. The new procedures are implemented in a software package in R called ISOCIR. The Circular Order Restricted Inference (CORI) procedures that have been developed, and future developments in the same line, could help biologists to test other interesting hypotheses about circular expressed genes.

ES08 Room 7 STATISTICAL ANALYSIS OF EVENT TIMES I Chair: Jacobo de Una

E143: **The copula-graphic estimator in censored nonparametric location-scale regression models**
Presenter: **Aleksandar Sujica**, Universite catholique de Louvain, Belgium
Co-authors: Ingrid Van Keilegom
A common assumption when working with randomly right censored data, is the independence between the variable of interest Y (the "survival time") and the censoring variable C. This assumption, which is not testable, is however unrealistic in certain situations. We assume that for a given X, the dependence between the variables Y and C is described via a known copula. Additionally, we assume that Y is the response variable of a heteroscedastic regression model $Y = m(X) + \sigma(X)\varepsilon$, where the explanatory variable X is independent of the error term ε, and the functions m and σ are 'smooth'. We propose an estimator of the conditional distribution of Y given X under this model, and show the asymptotic normality of this estimator. When estimating from right censored data there is always an area in the right tail where the survival function can not be estimated consistently. In our approach we are forced to estimate that area for pre-estimation of the conditional survival function. We also study the small sample performance of the estimator, and discuss the advantages/drawbacks of this estimator with respect to competing estimators.

E129: **Modelling unbalanced clustered multivariate survival data via Archimedean copula functions**
Presenter: **Roel Braekers**, Hasselt University, Belgium
Co-authors: Leen Prenen, Luc Duchateau
For clustered multivariate survival data, two different models are commonly used that take the association between the lifetimes into account: frailty models and copula models. Frailty models assume that conditional on a common unknown frailty term, the hazard function of each individual is independent. Copula models assume that the joint survival function of individuals within a cluster is given by a copula function, evaluated in the marginal survival functions. A major disadvantage of the present copula models over the frailty models is that the size of the different clusters must be small and equal to set up manageable estimation procedures for the different parameters. We describe a copula model for unbalanced clustered survival data. Hereby we focus on the class of Archimedean copula functions with completely monotone generators and exploit the Laplace transform-expression of these generators. We develop one- and two-stage procedures to estimate the association parameter. As results, we show the consistency and asymptotic normality of this estimator. We perform a simulation study to investigate the finite sample properties and illustrate this copula model on a real life data set on the time until first insemination for cows which are clustered within different herds.

E135: **Testing for covariate effects in a semiparametric generalized logit-based proportional hazards model**
Presenter: **Martha Lorena Avendano**, Complutense University of Madrid, Spain
Co-authors: M. Carmen Pardo
In clinical trials and observational studies, it is often interesting to evaluate the effects of covariates. We develop tests for covariate effect in a semiparametric proportional hazards model previously introduced. Valid statistical inferences about the corresponding covariate effects can be drawn based on the asymptotic distribution of the maximum partial likelihood estimator for the vector of regression coefficients. Our tests are based on Kullback Leibler divergence as well as Renyi's divergence. Extensive studies demonstrate that the proposed tests are appropriate for practical use.

E136: **Modified Aalen-Johansen estimator of the cumulative incidence function for left-truncated competing risks data**
Presenter: **Arthur Allignol**, University of Freiburg, Germany
This work stems from an observational study for which the goal was to assess the safety of exposure to statins during pregnancy, including the risk of spontaneous abortion. As women enter the cohort several weeks after conception, the data are left-truncated. Apart from spontaneous abortion, a pregnancy may end in an induced abortion or a life birth. The presence of both left-truncation and competing risks require the use of survival analysis techniques. The cumulative incidence function (CIF) is the method of choice for estimating the probability of spontaneous abortion in this setting. However, as abortions happen early in pregnancy, estimation might become problematic due to very small risk sets at the beginning

of the follow-up, leading to highly variable and thus potentially meaningless estimates. And due to the cumulative nature of the CIF, the problems introduced at these early time points propagate. For instance, an induced abortion observed early in the control group while three women were in the risk set led to a CIF of induced abortion of 0.37 (95% confidence interval [0.08; 0.93]) at the plateau. An estimate much higher than that in the exposed group. Estimating the CIF conditional on being event-free at some later time point is one suggested ad hoc method for obtaining reasonable estimates. Alternatively, a modified version of the Kaplan-Meier estimator in the usual survival setting tackles these issues. The consistency and weak convergence of the estimator on the entire range of the distribution function has been shown. We extend that estimator to the competing risks setting, study finite-sample properties of the modified estimator in a simulation study and illustrate its usefulness in the statin study.

E127: **Recent advances in the statistical analysis of doubly truncated data**
Presenter: **Jacobo de Una-Alvarez**, University of Vigo, Spain
Co-authors: Carla Moreira, Rosa Crujeiras
Doubly truncated data appear in a number of applications, including Epidemiology, Survival Analysis, Astronomy, and Economy. An obvious situation with double truncation arises when the sampled individuals are those with events between two specific dates. Under double truncation, ordinary estimators of the time-to-event distribution or the regression function may be systematically biased, and proper corrections are needed. Several methods for the analysis of doubly truncated data have been proposed in recent years. We will review some of these, including semiparametric estimation of a cumulative distribution function, kernel density estimation, and nonparametric regression. Real data illustrations will be provided. Some open problems in this area of research will be presented too.

ES19 Room 13 ROBUST METHODS BASED ON TRIMMING Chair: Andrea Cerioli

E125: **Robust constrained clustering in presence of component-wise outliers**
Presenter: **Alessio Farcomeni**, Sapienza - University of Rome, Italy
With the aim of clustering data in k groups, we adopt a partially spoiled independent contamination model, where two kinds of outliers are present: on one hand, structural outliers (i.e., spoiled observations); on the other hand, observations for which contamination in one dimension is independent of contamination in other dimensions. This general setting is particularly interesting in higher dimensions, where most (if not all) observations will contain at least one outlying entry. To tackle this contamination model, we propose to combine the widely used impartial trimming with a technique we call snipping: for non-trimmed observations we allow removal of some dimensions, while the remaining are used for clustering. We define and solve a TCLUST-type problem where we control the relative sizes of the eigenvalues of the cluster specific scatter matrices, guaranteeing that the problem is well defined and estimators are consistent.

E535: **Flexible multivariate tolerance zones based on trimmed $k-$means**
Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Agustin Mayo-Iscar
The use of the non-discarded observations after applying a trimming procedure is a well-known way to obtain tolerance zones. This is the case, for instance, of multivariate asymptotic distribution free tolerance ellipsoids based on the non-discarded observations after computing the Minimum Covariance Determinant estimator. However, in situations where the underlying population is far from being elliptical, tolerance zones based on ellipsoids can be completely meaningless. In these situations, more flexible procedures are needed to catch the population structure in an accurate way. We study the use of the set of non-discarded observations after applying the trimmed $k-$means methodology to built asymptotic distribution free tolerance zones made of the union of an appropriate number of small balls that capture the shape of the level sets of the underlying probability density function. The proposed tolerance zones arise from trying to summarize the whole distribution through the use of a set of quantizers instead of merely considering summary functionals like the mean and the scatter matrix estimators. The use of large values of k provides very flexible and adaptive tolerance zones that turn out to be very useful in complex problems in Statistical Quality Control.

E547: **Semiparametric tolerance regions based on trimming**
Presenter: **Eustasio del Barrio**, Universidad de Valladolid, Spain
A semiparametric tolerance region of a parametric model based on trimming is introduced. Such semiparametric tolerance regions allow us to use a parametric model as an approximation to the true distribution. Our approach relies on the choice of a metric. We consider several choices, including transportation cost, Kolmogorov and integral metrics. We present also a procedure for testing the null hypothesis that the true distribution belongs to the semiparametric tolerance region. Our proposal generalizes earlier works dealing with multinomial models, but the new procedure allows us to consider the case of continuous data.

E573: **Likelihood ratio type tests in robust regression**
Presenter: **Stefan Van Aelst**, Ghent University, Belgium
Co-authors: Matias Salibian-Barrera, Victor Yohai
ANOVA tests are the standard tests to compare nested linear models fitted by least squares. These tests are equivalent to likelihood ratio tests and thus they are very powerful. However, least squares estimators are very vulnerable to outliers, and thus also the related ANOVA type tests are also extremely sensitive to outliers. Therefore, we consider regression τ -estimators to estimate the parameters in the linear models. Regression τ-estimators combine high robustness with high efficiency which makes them a suitable choice for robust inference beyond parameter estimation. We introduce robust likelihood ratio type test statistics based on the τ estimates of the error scale in the linear models. To determine the null distribution of the test statistics we either use an asymptotic approximation or the fast and robust bootstrap. We study the robustness and power of the resulting robust likelihood ratio type tests for nested linear models.

ES27 Room 8 FUZZINESS IN STATISTICS Chair: M. Brigida Ferraro

E207: **Gradual quasi-continuous histograms**
Presenter: **Olivier Strauss**, University Montpellier II - LIRMM, France
Co-authors: Ines Couso
Quasi-continuous histograms (QCH) framework allows designing techniques that enable us to manipulate the probability density underlying a finite set of observations with simple computation. Several statistical tools have been created based on QCH framework including mode estimation and rank statistics. QCH framework is based on extending the classical histogram density estimation technique in two steps. The first step consists in replacing the crisp partition of the reference interval by a fuzzy partition. The second step consists in transferring the counts associated to each cell of the partition to any subset W of the reference interval. This count transfer provides an estimation of #(W), the number of observations that belong to W. Among the different methods for achieving this count transfer, the so-called possibilistic transfer aims at providing specific bounds of #(W). However, in some rare cases, especially when the counts associated to the cells are low, the possibilistic transfer does not work properly, i.e. the upper estimation can be lower than #(W). We propose a modification of the QCH framework that solves this problem. This modification involves associating a gradual number, instead of a real number, for representing the counts associated with each cell of the partition. We prove the validity of the approach. Some practical implementations are proposed to keep the complexity of the method as low as possible.

E422: **Density regression trees**
Presenter: **Antonio D'Ambrosio**, University of Naples Federico II, Italy
Co-authors: Roberta Siciliano, Massimo Aria
Symbolic data analysis was introduced in the last decade in order to provide a generalization of standard data analysis tools. This class of data allows each individual (or class) to take a finite set of values, or categories, intervals or distribution on each variable. In literature this kind of data are known as multi-valued variables. We focus on the first category mentioned above (quantitative multi-valued variables). Several parametric and non-parametric data analysis procedures have been already proposed in symbolic data literature (linear regression, clustering methods, factor analysis, etc.). As a matter of fact, supervised classification and non-parametric regression methods are not yet available. We propose a new non-parametric and non-linear regression approach dealing with quantitative multi-valued response variable and heterogeneous classical and symbolic predictors, by emphasizing the fuzzy nature of this method.

E453: **Fuzzy HPD-Region**
Presenter: **Shohreh Mirzaei Yeganeh**, TU wien, Austria
Co-authors: Reinhard Viertl
From generalized (fuzzy) a-posteriori densities of parameters in fuzzy Bayesian inference it is necessary to construct generalized (fuzzy) confidence regions for these parameters. In standard Bayesian inference this is done by so-called highest a-posteriori density regions (HPD-Regions). In case of fuzzy information Bayes' theorem can be generalized in order to take care of fuzziness. This generalization yields a so-called fuzzy a-posteriori density $\pi^*(.)$ on the parameters space Θ. These fuzzy a-posteriori densities generate so-called δ-level functions $\underline{\pi}_\delta(.)$ and $\overline{\pi}_\delta(.)$ which are classical real-valued functions on the parameter space, defined by the δ-cuts of $C_\delta[\pi^*(\theta)] = [\underline{\pi}_\delta(\theta); \overline{\pi}_\delta(\theta)]$ for all $\delta \in (0;1]$. In order to construct generalized confidence regions for θ at specified confidence level $1-\alpha$ the family $\mathcal{D}_\delta$ of all classical densities $g(.)$ on Θ, subject to $\underline{\pi}_\delta(\theta) \leq g(\theta) \leq \overline{\pi}_\delta(\theta)\ \forall \theta \in \Theta$ is considered. Denoting by ${}^\delta HPD_{1-\alpha}(g)$ the HPD-Region for θ based on a density $g(.) \in \mathcal{D}_\delta$ the concept of HPD-Regions for θ can be generalized. Now for fuzzy a-posteriori densities the generalization of HPD-Regions is possible by generating HPD-Regions for all classical densities $g(.)$ between the δ-level functions of the fuzzy a-posteriori density. Based on these HPD-Regions, ${}^\delta HPD_{1-\alpha}(g)$ for all $\delta \in (0;1]$ a generating family of subsets $(A_\delta; \delta \in (0;1])$ of Θ is defined and by the construction lemma for fuzzy sets the membership function of the generalized fuzzy HPD-Region, $HPD^*_{1-\alpha}$ is obtained.

E482: **The computation of the sample d_θ-median for random intervals**
Presenter: **Beatriz Sinova**, University of Oviedo, Spain
Co-authors: Maria Angeles Gil, Gil Gonzalez-Rodriguez, Stefan Van Aelst
Random intervals are used to model random experiments which generate intervals, that is to say, experiments for which the interest is focused either on intrinsically imprecise-valued random elements, like variables measuring ranges or fluctuations, or on real-valued random variables which are imprecisely measured (because of technical limits). The most commonly used central tendency measure to summarize the information given by a random interval is the Aumann expectation. This measure is very much influenced (like the mean in the real-valued settings) by data changes or the existence of outliers. In searching measures with a more robust behavior in the interval-valued case, some different approaches to extend the notion of median have been suggested. Specifically, the d_θ-median is based on the concept of the spatial median, defined through the L^2 metric coined as d_θ. In contrast to other previous medians defined through L^1 distances, the computation is not as easy-to-determine: indeed, there are no explicit formulas to calculate it. Therefore, differents algorithms to compute the sample d_θ-median of a random interval will be now compared.

E961: **Fuzzy sets as a tool for protecting sensitive information**
Presenter: **Pelayo Quiros**, University of Oviedo, Spain
Co-authors: Pedro Alonso, Irene Diaz, Susana Montes
Protecting sensitive information is an important issue when data are publicly released. Different techniques have been investigated and proposed in order to measure the risk of linking sensitive information to individuals and/or organization. Among them, k-anonymity, l-diversity and t-closeness gained popularity as they are able to measure the linkage between sensitive and identity information in tabular data. The main limitation of these approaches relies on the fact that close values appears different, although they express a similar concept. This drawback could be overcome by introducing a fuzzy model. A priori, the fuzzy released data provides with a first step of privacy because some uncertainty is introduced. At the same time the protected data become more interpretable. The goal is to provide a formal framework to re-define the concepts of the k-anonymity, l-diversity and t-closeness in terms of fuzzy sets and to check if these new approaches provide more data protection.

ES64 Room 11 STATISTICS OF EXTREMES AND RISK ASSESSMENT Chair: Ivette Gomes

E219: **Estimating upper bounds for ruin probabilities in the Sparre Andersen model**
Presenter: **Ana Freitas**, University of Porto, Portugal
Co-authors: Margarida Brito
The problem of estimating the adjustment coefficient R in the Sparre Andersen model is considered. Such a model allows us to estimate upper bounds for the probability of ruin. We have proposed a geometric-type estimator for R and proved a result about its asymptotic normality. We also have shown that it is possible to construct confidence intervals for R based on that estimator, using the tail bootstrap procedure. More recently, we established an Edgeworth expansion for the distribution function of an estimator (conveniently normalized), directly related to the geometric-type estimator. We present Cornish-Fisher expansions for the quantiles of the distribution function and, in order to investigate the finite sample behavior of the corresponding confidence intervals, we perform a simulation study in some particular cases.

E396: **Models and inference for spatial extremes**
Presenter: **Jennifer Wadsworth**, Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: Jonathan Tawn
Appropriate modelling of the dependence in the extremes of spatial processes aids accurate assessment of the risk posed by environmental factors. We discuss some models recently proposed for spatial extremes, considering the cases of both asymptotic independence and asymptotic dependence. We also consider some ideas for improving inference for spatial extreme value models.

E442: **Extreme value estimation based on interval censored and non-censored data**
Presenter: **Juerg Huesler**, University of Bern, Switzerland
Co-authors: Jasmin Wandel
The aim is to discuss several approaches for the estimation of the parameters of the tail of the distribution in the case that not all data are from exact measurements, that means that some are from inaccurate measurements. In this case we have interval-censored data which should be also used for the estimation of the parameters. We investigate different estimation methods and compare them analytically and by simulations. This work is motivated by a real data example.

E542: **Interval forecasts for the range: An extreme value approach**
Presenter: **Paulo Araujo Santos**, ESGTS and CEAUL, Portugal
Co-authors: Hammoudeh Shawkat
The widely used Value-at-Risk measure is a quantile of the return distribution. We suggest a quantile of the price range distribution as an alternative risk measure that can be useful for some strategies in the financial markets. The price range is used by several authors as an proxy for volatility and is widely used by traders in financial markets for example for position sizing. We propose a method based on the extreme value theory for estimating a quantile of the price range, and compare, the performance of this method with other approaches using backtesting of historical daily price range series.

E562: **Extreme value analysis with non-stationary observations**
Presenter: **Chen Zhou**, Erasmus University Rotterdam, Netherlands
Co-authors: Laurens de Haan
Extreme value analysis based on independent observations from non-identical distributions with comparable tails is considered. We first establish the asymptotic behavior on the weighted tail empirical process based on the non-stationary observations. As an application, we show that if the tail distributions of the observations are heavy-tailed, then the asymptotic normality of the Hill estimators remain valid even if the observations are non-stationary.

ES38 Room 10 SKEW DISTRIBUTIONS: PROBABILITY AND INFERENCE Chair: M. Dolores Jimenez-Gamero

E388: **Improved approximation of the sum of random vectors by the skew-normal distribution**
Presenter: **Marcus Christiansen**, University of Ulm, Germany
Co-authors: Nicola Loperfido
The properties of the multivariate skew-normal distribution as an approximation to the distribution of the sum of n i.i.d. random vectors will be studied. More precisely, we establish conditions ensuring that the uniform distance between the two distribution functions converges to zero at a rate of $n^{-2/3}$. The advantage over the corresponding normal approximation is particularly relevant when the summands are skewed and n is small, as illustrated for the special case of exponentially distributed random variables. Applications to some well-known multivariate distributions are also discussed.

E847: **Exact and bootstrap control charts based on the skew-normal distribution**
Presenter: **Fernanda Figueiredo**, Porto University, FEP and CEAUL, Portugal
Co-authors: Ivette Gomes
Modeling real data sets, even when we have some potential models for the underlying data distribution, is always a very hard task due to some uncontrollable perturbation factors. Classes of distributions that allow departures from the most commonly used models, for instance, by the introduction of extra parameters that control the skew and the tail weight, will provide better models to fit the data and the application of inferential methods will be carried out in a more robust way. We consider a class of skew-normal distributions, and after enhancing some properties of this class, we develop exact and bootstrap control charts.

E814: **Geometric properties of skewed continuous $l_{n,p}$-symmetric distributions**
Presenter: **Wolf-Dieter Richter**, University of Rostock, Germany
The general methods from skewed distributions theory and from the theory of geometric and stochastic representations of $l_{n,p}$-symmetric distributions were combined previously to introduce skewed continuous $l_{n,p}$-symmetric distributions. Several representations of the skewed normal distribution and of skewed elliptically contoured distributions were also considered previously from a unified geometric point of view. The present talk combines the main aspects of both theories.

E844: **An asymmetric family of unimodal distributions with mode-invariance**
Presenter: **Toshihiro Abe**, Tokyo University of Science, Japan
Co-authors: Hironori Fujisawa
An asymmetric family of unimodal distributions from a family of the symmetric distributions is considered by using a specific class of transformation controlling the skewness. Its mode and normalizing constant remain unchanged for the transformation. Monotonicity of skewness can be shown easily under a weak condition for a transformation due to its mode invariance property.

E402: **Testing for skew-symmetric models**
Presenter: **M. Dolores Jimenez-Gamero**, Universidad de Sevilla, Spain
Skew-symmetric (ss) models are semiparametric models for continuous random vectors. Starting with a symmetric probability density function (pdf), a ss family is generated by multiplying the symmetric pdf by a skewing function. A characterization of this skewing function allows us to propose a test for the goodness of fit to a ss model. The test statistic is based on comparing two estimators of the population pdf: a consistent kernel-based estimator and another estimator which is built by assuming that the ss in the null hypothesis is true.

ES53 Room 12 COUNTING PROCESSES Chair: Paula R. Bouzas

E206: **Modeling multivariate extreme events using self-exciting point processes**
Presenter: **Oliver Grothe**, University of Cologne, Germany
Co-authors: Volodymyr Korniichuk, Hans Manner
A new model is proposed to capture the typical features of multivariate extreme events observed in financial time series, namely clustering behavior in magnitudes and arrival times of multivariate extreme events, and time-varying dependence. The model is developed in the framework of the peaks-over-threshold approach in extreme value theory and relies on a Poisson process with self-exciting intensity. We discuss the properties of the model, treat its estimation, deal with testing goodness-of-fit, and develop a simulation algorithm. The model is applied to return data of two stock markets and four major European banks.

E244: **Preventive and corrective maintenance for a device subject to internal failures and external shocks**
Presenter: **Juan Eloy Ruiz-Castro**, University of Granada, Spain
A discrete complex reliability system subject to internal failures and external shocks is modeled in an algorithmic form. We assume two types of internal failures: repairable and non-repairable. When a repairable failure occurs the unit goes to corrective repair and a non-repairable failure produces a total failure. Also, the unit can undergo external shocks that can produce a cumulative damage in the unit; this one can not be observed. When a threshold is reached the unit must be removed and it is replaced by a new and identical one. On the other hand, inspections occur randomly. When it happens, the internal performance of the system is observed and the unit could go to preventive repair. In this case, the external damage

keeps in memory. The system is modeled in an algorithmic form. A transient and stationary analysis is performed and some measures of interest are worked out. It has been implemented computationally with Matlab and one numerical example is shown.

E255: **A Monte Carlo test to check the independence of Poisson processes and a study of its power**
Presenter: **Ana C. Cebrian**, University of Zaragoza, Spain
Co-authors: Jesus Abaurrea, Jesus Asin

A test to check the independence between three or more non homogeneous point processes, conditionally on their corresponding intensities, is proposed. It is intended for Poisson processes (PPs) but it can be applied to any point process which can be simulated. The test is based on the notion of close point relation, and the underlying idea is to compare the behavior of the sets of close points in the observed vector process, and in simulated processes with the same characteristics and mutually independent components. The distribution of the statistic is not known, but its empirical distribution is obtained using a Monte Carlo approach designed to keep the computing time cost low. To study the power of the test, a method to generate dependent point processes is needed. Two approaches are considered, the first one generates three dependent PPs using a tandem of $M \setminus M \setminus 1$ queues, with exponential services. The second approach generates three dependent point processes (not Poisson) marking a PP with marks generated by a Markov chain of three states. The power of the test according to the degree of the induced dependence and the homogeneous or non homogeneous character of the Poisson intensities is studied in both cases.

E417: **Simulating and forecasting Cox processes with MATLAB**
Presenter: **Nuria Ruiz-Fuentes**, University of Jaen, Spain
Co-authors: Paula R. Bouzas

Within counting processes, Cox processes are very flexible due to the stochastic nature of their intensity, so they are suitable for representing many real phenomena. Although they are very useful and the estimation of their statistics and forecasting is basic, there is scarce statistical software to deal with these tasks. We propose an implementation in MATLAB to simulate counting data of such processes under several situations, mainly depending on the type of intensity process. The code also deals with compound Cox processes with different kinds of randomly distributed marks. For simulated and real data our proposal processes data without restrictive assumptions on the statistics associated to the process. The user can estimate the mean process, the intensity process and even some statistics for a compound Cox process in a subset of the marks space. The estimation of the mean and of the intensity processes is carried out by an ad hoc functional data analysis procedure, preserving the theoretical property of monotonicity of the mean process.

E420: **Estimation of the count-conditional intensity of a self-exciting process**
Presenter: **Paula R. Bouzas**, University of Granada, Spain
Co-authors: Nuria Ruiz-Fuentes

A self-exciting process is a counting process which evolves with after-effects, i.e. its intensity process depends on the past of the counting one. The count-conditional intensity is the expected value of the intensity given the counts. This process characterizes the self-exciting process so its estimation is an important task. A point estimator for the count-conditional intensity process is proposed when the self-exciting process is observed as recurrent event data. For each number of counts, the count-conditional intensity is a function of time that is estimated in a finite set of time points by the proposed estimator. Then, their functional forms are approximated and finally used to estimate statistics of the self-exciting process.

ES54 Room 9 FUNCTIONAL DATA WITH SPATIAL DEPENDENCE Chair: Piercesare Secchi

E251: **On mixing conditions for asymptotics of functional time series**
Presenter: **Shahin Tavakoli**, EPFL, Switzerland
Co-authors: Victor Panaretos

Motivated by DNA minicircle data from Molecular Dynamics, we investigate mixing conditions that enable us to draw statistical inferences for stationary functional data. We are interested in general stationary processes -as opposed to linear processes. We review existing functional mixing conditions, examples of processes that satisfy them, and asymptotic results they allow for. We then consider moment-based functional mixing conditions, and show how these can be used to recover or extend existing asymptotic results. We also consider examples of functional processes satisfying our mixing conditions, and probe the stability of our conclusions under discrete observation of the functional time series.

E456: **Constrained nonparametric covariance function estimation for spatially dependent functional data**
Presenter: **Heather Battey**, University of Bristol, United Kingdom

Spatial and spatio-temporal patterns arise in a variety of disciplines such as brain imaging, seismology and economics. Much applied research adopts the modelling strategy of assuming a parametric form for the covariance function. However, deducing parametric specifications for the covariance function that are able to capture specific features is very difficult, especially when $\mathcal{V} \subset \mathbb{R}^d$. When a random field has a high-dimensional continuous domain like $\mathbb{R}^d$, random fields possessing a simple stochastic representation often have an intractable covariance function, hence it may be impossible to parameterise the covariance function. Let $\{Y(v) : v \in \mathcal{V}\}$ be a real-valued weakly stationary random field over a spatial or spatio temporal domain $\mathcal{V} \subset \mathbb{R}^d$ ($d \geq 1$) with jointly continuous covariance function $\{C(u,v) : u,v \in \mathcal{V}\}$. We construct a nonparametric estimator for C based on discretised realisations of the i.i.d. random functions $(Y_i)_{i\in\mathbb{N}}$ that automatically imposes the positive definiteness requirement. Although the estimator is defined as the solution to an infinite dimensional minimisation problem, our specification is such that it admits a finite dimensional representation and may be computed efficiently. We will discuss the properties of this estimator and present simulation studies comparing it to a range of alternatives.

E493: **Static clustering for country population pyramids that evolves on time**
Presenter: **Pedro Delicado**, Universitat Politecnica de Catalunya, Spain

Consider n individuals for which a functional random variable χ is observed at T consecutive instants of times. Let $\{\chi_i^t : 1 \leq i \leq n, 1 \leq t \leq T\}$ be the resulting functional data set. The following problem is considered: to identify k prototypes (i.e., k clusters: $C_j, 1 \leq j \leq k$) of functional data that are stable in time, and assign each observed functional data χ_i^t to one of these clusters. This way a given individual i may spend the first periods in a cluster, then it can evolve to another one and so on. The motivating example for this research is the set of population pyramids from year 1950 to year 2010 for each country in the world ($n = 197$, $T = 61$). The objective is to do a clustering analysis such that the same cluster could include, for example, the Swedish population pyramid for 1950 and that corresponding to Ireland 2010. This way each cluster (population pyramids with similar shape) can be identified with a different stage of demographic development, which is achieved by each country at different times. There are two equally important problems: how to do the clustering and how to visualize the results.

E561: **Detecting temporal-patterns in spatially-dependent high-dimensional data**
Presenter: **Simone Vantini**, Politecnico di Milano, Italy

Geo-referenced high-dimensional data are analyzed describing the use over time of the mobile-phone network in the urban area of Milan, Italy. The aim is the identification of hidden temporal-patterns possibly related to specific locations and to different activities taking place within the city. We used different dimension reduction techniques able to disclose different patterns. As a reference analysis we performed a treelet analysis. Treelets

are a data-driven basis obtained by a hierarchical implementation of pair-wise principal component analysis. We improved over the latter analysis in two directions. Firstly, by integrating it with a novel bagging Voronoi procedure based on random Voronoi tessellations of the area that is able to deal with spatial dependence in a fully non-parametric perspective. Secondly, we propose a novel hierarchical independent component analysis procedure that removes unphysical constraints imposed to the temporal-patterns by the treelet analysis (e.g., orthogonality) but still preserving their hierarchical structure. The results clearly show different temporal patterns interpretable, for example, as daily work activities in the financial district, leisure activities in residential areas in the evenings and in the weekend, commuting along the highways and at train stations during rush hours, shopping in commercial streets and malls on Saturday afternoon.

E601: **Non-parametric spatial functional regresion estimation from a RKHS approach**
Presenter: **Elvira Romano**, Seconda Universita degli Studi di Napoli, Italy
Co-authors: Emilio Porcu, Maria Dolores Ruiz-Medina
In the context of non-linear regression, we extend previous results to the spatial functional context in terms of Hilbert-valued spatial random variables. Specifically, a spatial functional non-linear regression estimator is derived by its projection into the Reproducing Kernel Hilbert Space generated by the auto-covariance operator of the Hilbert-valued spatial random regression function. Estimation of the corresponding random projections, approximating the functional temporal values of such a random regression function over a spatial grid, is achieved by minimization of the $H-$norm of the mean quadratic integrated error. A simulation study is developed to illustrate the performance of the non-linear spatial regression estimator proposed in a Hilbert-valued context.

EP01 Room Hall POSTER SESSION I Chair: Cristian Gatu

E681: **A simulation study to compare several population means under heteroscedasticity**
Presenter: **Berna Yazici**, Anadolu university, Turkey
Co-authors: Ahmet Sezer, Evren Ozkip
If the assumption of equal variances is violated, the comparison of means of different groups can not be made using the classical F test. There are some alternative tests. The most common ones are the Welch test, the Scott-Smith test, the Brown-Forsythe test, the Weerahandi's Generalized F test and Xu and Wang's Generalized F test. Those tests are reviewed. Then a simulation study is conducted in order to compare their power for different sample sizes under homoscedasticy and heteroscedasticy. Also normal and nonnormal distributed variables are compared for different scenarios. Finally, all the tests are applied on a real life data set with heteroscedasticity and the results are interpreted.

E707: **Application of neural network models to predict air quality**
Presenter: **Carmen Capilla**, Universidad Politecnica de Valencia, Spain
An application of artificial neural network models is presented to predict hourly nitrogen dioxide concentrations from local pollutants concentrations, traffic, meteorological data, and periodic components (sine and cosine terms for the daily and weekly cycles). The predictions produced by three different methods are compared: feed-forward multilayer perceptron networks, radial basis function networks and generalized regression networks. The multilayer perceptron model is applied using the Levenberg-Marquard and the scaled conjugate gradient learning algorithms. The number of neurons in the hidden layer is the optimum found by experimentation. The transfer functions selected are the sigmoid and the hyperbolic tangent for the hidden layer, and linear for the output layer. The output of the models is the pollutant prediction 24 hours in advance. The models are trained on data from 2003. Data from 2004 are used as the validation set and observations from 2005 are the test data set. The model evaluation criteria are the mean absolute error, the root mean square error, the mean absolute percentage error and the correlation coefficient between the pollutant observations and the predictions. Bootstrap estimates of standard errors of these criteria are calculated. Overall, the radial basis function network gives the best results.

E722: **Multilevel functional clustering analysis of probabilistic sleep microstate curves**
Presenter: **Roman Rosipal**, Slovak Academy of Sciences, Slovakia
The purpose is introduce and validate an EEG data-based model of the sleep process with an arbitrary number of different sleep states and a high time resolution allowing modeling of sleep microstructure. The proposed probabilistic sleep model describes sleep via posterior probabilities of a finite number of microstates. Using the model, we extracted objective sleep parameters describing quantitative and qualitative characteristics of the probabilistic sleep microstate curves and proved their usefulness when assessing selected aspects of sleep quality. We are investigating functional data clustering methods applied to sleep microstate posterior curves. The hierarchical structure of the data given by the repeated visits of subjects in the sleep lab motivates our focus on recently proposed multilevel functional clustering analysis approaches. We are applying the multilevel functional principal component analysis to the sleep posterior curves and we are comparing hard and soft clustering methods on extracted scores. The obtained clustering structure is validated by means of information it delivers about the daytime measured objective sleep quality indices and is compared with information obtained with the previously extracted sleep microstate parameters.

E750: **Efficiency of methods for combining information from independent interlaboratory comparisons with systematic effects**
Presenter: **Viktor Witkovsky**, Slovak Academy of Sciences, Slovakia
The statistical inference about the common mean of several normal distributions with unknown and possibly unequal variances has attracted research for a long time. Models and methods for point and interval estimation of the common mean appear to be fundamental also for combining and interpreting independent measurement results of an unknown quantity from several laboratories or measurement methods. In metrology, such experiments of international scope are carried out by the National Metrology Institutes in order to establish the degree of equivalence of national measurement standards known as the key interlaboratory comparisons. However, the measurement results are typically (more or less) influenced by the unobservable systematic effects, specific for each participating laboratory. The information on the systematic effects should be provided by the participating laboratories based on their expert knowledge. Subsequently, it should be used for derivation of the point and interval estimators of the common mean in order to deliver and characterize the metrological key comparison reference value (KCRV) for the considered quantity. We will study and compare exactness and efficiency of several selected interval estimators for common mean in models suitable for interlaboratory comparisons with the estimator based on combining information from independent sources through confidence distributions (CDs), a methodology recently suggested.

E728: **Comparing the means for unbalanced several skewed data with unequal variances**
Presenter: **Evren Ozkip**, Anadolu University, Turkey
Co-authors: Berna Yazici, Ahmet Sezer
Multiple comparison methods (MCMs) are widely used to investigate differences between population means or, more generally, between subsets of population means using sample data. This study presents procedures for hypothesis testing for the mean of several log- normal populations. The methods are based on the concepts of generalized $p-$value, Welch test and likelihood ratio test. The merits of the proposed methods are evaluated based on the power and compared in the case of unequal variances and unequal sample sizes. Our simulation results indicate that when the number of populations is four or more, the generalized variable method performs much better than the other methods regardless of the sample sizes.

E816: **Log-ratio analysis of sets of compositional data**
Presenter: **Zerrin Asan**, Anadolu University, Turkey
Co-authors: Anna Dustira, Mikko Vihtakari, Michael Greenacre
In logratio analysis the pairwise ratios between all elements are analysed on a logarithmic scale. The results are given in the form of a factorial map showing the elements being analysed jointly with the samples - the links between two elements represent the corresponding logratio. This method is routinely applied to a single matrix of homogenous data for which taking ratios of the elements makes sense. In the present application there are three sets of elements, homogenous within each set but not between one another. In this case we have adapted logratio analysis by making the logratio transformation within each table separately, and then computing the total variance within each table. The tables were then normalized with respect to one another by dividing each table by the square root of its respective variance, leading to each table having equal variance. Then the usual dimension reduction algorithm was applied, using the singular value decomposition. In the resultant figure, therefore, links only between elements of the same table should be considered in the interpretation. The application is concerned with the identification and quantification of geochemical differences of Permian dark- and light chert facies from across Svalbard island.

E834: **Optimal quadratic estimators for linear discrete-time stochastic systems with multi-sensor correlated missing measurements**
Presenter: **Irene Garcia-Garrido**, Universidad de Granada, Spain
Co-authors: Josefa Linares-Perez, Raquel Caballero-Aguila
The least-squares quadratic estimation problem in linear discrete-time stochastic systems with missing measurements coming from multiple sensors is addressed. Missing measurements (that is, observations containing only noise) occur randomly and, for each sensor, a different Bernoulli sequence is used to model this uncertainty about the state being present or missing in the observations. Such variables are assumed to be correlated at instants that differ m units of time; this special form of correlation can model situations where the state cannot be missing in $m+1$ consecutive observations. To address the quadratic estimation problem, augmented state and observation vectors are defined by assembling the original vectors with their second-order powers defined by the Kronecker product. Then, a new augmented system is obtained and the quadratic estimation problem for the original system is reduced to the linear estimation problem for the augmented system. By using an innovation approach, recursive algorithms for the linear filter and fixed-point smoother of the augmented state based on the augmented observations are derived, and the required quadratic estimators are established. The performance of the proposed estimators is illustrated by a simulation example where the linear and quadratic estimation error variances are compared.

EP02 Room Hall POSTER SESSION II Chair: Cristian Gatu

E989: **Multivariate quantile regression based on additive models: An application to prediction of SO_2 pollution incidents**
Presenter: **Marta Sestelo**, University of Vigo, Spain
Co-authors: Isabel Matinez-Silva, Javier Roca-Pardinas
In applications, regression models are frequently used to evaluate the relationship between a response variable and its covariates. In this framework, it is usually estimated the conditional mean of the response, however this prediction measures only the "center" of the conditional distribution. In some circumstances, it could be interesting to know the complete distribution of the response, and this is provided by its quantiles. The quantile regression models allow us the prediction of a given quantile of the conditional distribution using the covariates' value. Several approaches to estimating these type of models have been suggested in the statistical literature, e.g., methods based on splines or Lambda Mean Standard deviation (LMS) method. We propose the implementation of an algorithm based on local polynomial kernel smoothers to estimate Additive Models. These models were applied on the prediction of atmospheric SO_2 pollution incidents where we have computed several different regression curves corresponding to various percentage points of the distribution, getting a more complete picture of these emissions.

E1008: **Almost sure central limit theorem for nonstationary normal random fields**
Presenter: **Luisa Pereira**, University of Beira Interior, Portugal
Recently, a new field of limit theorems has been discovered, which is called almost sure central limit theorem (ASCLT). Many results have been obtained for independent and dependent random variables. The first results on ASCLT dealt with partial sums of independent and identically distributed (i.i.d.) random variables. Later on, the ASCLT was extended to the partial maximum of an i.i.d. random sequence. We obtain an almost sure version of a maximum limit theorem for nonstationary normal random fields, under some covariance conditions. Normal random fields play an important role since the specification of their finite-dimensional distributions is simple, they are reasonable models for many natural phenomena, estimation and inference are simple, and the model is specified by expectations and covariances.

E995: **Methods for estimating the upcrossings index**
Presenter: **Ana Paula Martins**, University of Beira Interior, Portugal
Co-authors: Joao Sebastiao
The upcrossings index, η, $0 \leq \eta \leq 1$, as a measure of the clustering of upcrossings of a high level by the variables of a stationary sequence is an important parameter when studying extreme events. This parameter provides different and complementary information to that provided by the key parameter for extending extreme value theory for independent and identically distributed random variables to stationary sequences, the extremal index θ, $0 \leq \theta \leq 1$. Relations between the upcrossings index and other dependence coefficients, such as the extremal index and upcrossings-tail dependence coefficients can be found in the literature. We propose new estimators for the upcrossings index and prove their consistency and asymptotic normality. The performance of the estimators is examined through simulations. Finally, the estimators are applied to real data sets.

E1033: **Isotropic turbulence and periodicity: An alternative comprehensive study**
Presenter: **Evangelos Akylas**, CUT, Cyprus
Co-authors: Elias Gravanis
Turbulence theory predicts specific laws for the evolution of the turbulence statistics during the period of decay. Under isotropic conditions, statistical arguments allow for the extraction of the well–known Karman–Howarth (KH) equation, which connects the double and triple velocity correlations, averaged over space. Direct Numerical Simulations (DNS) of isotropic turbulence solve massively the velocity field in a bounded periodic box. Although the most common and convenient tool for studying accurately turbulence characteristics, through solving directly the Navier-Stokes equations, the presence of the periodic boundary breaks the isotropy and affects in general the turbulence statistics in a, more or less, unknown way. Periodical boundary conditions are introduced for the first time to KH equation in order to mimic the DNS logic, but keeping, by construction, the isotropy conditions always true. Furthermore, the Oberlack and Peter's model is analyzed and used in order to close the system of the equations and produce the evolution of the statistics of isotropic decaying turbulence. The results are compared with highly resolved DNS of isotropic turbulence which have been performed in CUT using a parallelized code. The comparison reveals excellent qualitative agreement enlightening the basic mechanism through which the boundary affects the turbulence.

Saturday 01.12.2012 14:00 - 15:40 Parallel Session C – CFE

CS06 Room 5 STATISTICAL SIGNAL EXTRACTION AND FILTERING I **Chair: Stephen Pollock**

C303: **Single versus multiple-source error models for signal extraction**
Presenter: **Miguel Jerez**, Universidad Complutense de Madrid, Spain
Co-authors: Jose Casals, Sonia Sotoca
Results obtained when applying fixed-interval smoother to two linear and time-invariant state-space forms are discussed. The first model has different error terms in the state and observation equation, while the second has a unique perturbation term which coincides with the one-step-ahead forecast error. When observational equivalence is imposed, both forms provide similar smoothed states which, surprisingly, have drastically different variances. This happens because the smoothed states for the single-source error representation have the same dynamics as the system, while those coming from multiple-error model depend also on an additional forward autoregressive structure. The practical implications of this result are discussed through an example, in which the unadjusted US GDP is decomposed into trend-cycle, seasonal and irregular components using both types of models. We find that single error representations have clear advantages when one wants to adjust for seasonality, while multiple error models are better suited to build coincident trend-cycle indicators. Building on this analysis, it is natural to adopt a 'best of both worlds' approach, which applies each representation to the task in which it has comparative advantage.

C494: **Multivariate stochastic volatility estimation using particle filters**
Presenter: **Kostas Triantafyllopoulos**, University of Sheffield, United Kingdom
Multivariate methods for analyzing volatility over a number of stocks, or indices, have been a central topic of interest in econometrics. In particular, recently simulation-based methods, such as MCMC have been proposed. We develop an estimation methodology, based on particle filters, which, unlike MCMC, has a sequential mode of application. The proposed algorithm is suitable for high dimensional data and it deploys an inverse Wishart autoregressive stochastic process for the volatility covariance matrix, while the returns are modelled with a skew-t distribution. Using the convolution of the Gaussian / Wishart distributions, a novel approach is adopted for the selection of the importance function, for the particle filter. We make new proposals for the efficient estimation of the hyper-parameters of the model (such as the autoregressive matrix or the skewness parameters). Some properties of the model, in relation to models in the existing literature, are given. The methodology is put into practice using 2 data sets, the common constituents of the Dow Jones index (a 30 dimensional data set) and the common constituents of the FTSE-100 (a 92 dimensional data set).

C519: **Additive model selection**
Presenter: **Italia De Feis**, National Research Council of Italy, Italy
Co-authors: Umberto Amato, Anestis Antoniadis
A new sparsity-smoothness penalty for high-dimensional additive models is proposed. Several existing algorithms already developed for this problem are reviewed. A computationally efficient algorithm for fitting such models is introduced. Using reasonable assumptions, exploiting recent results on grouped lasso-like procedures and taking advantage of several oracle results, asymptotic optimality of the proposed estimator is proved for high-dimensional yet sparse additive models. Finally, a variable selection procedure is proposed for testing the presence of additive components and it is compared with some high dimensional testing procedures available in the literature.

C830: **A new frame based de-noising procedure for fast oscillating signals**
Presenter: **Daniela De Canditiis**, IAC (CNR), Italy
In recent years there has been a growing interest in frame based de-noising procedures. The advantage of frames with respect to classical orthonormal bases (e.g. wavelet, Fourier, polynomial) is that they can furnish an efficient representation of a more broad class of signals. For example, signals which have fast oscillating behavior as sonar, radar, EEG, stock market, audio and speech are much more well represented by a frame (with similar oscillating characteristic) than by a classical wavelet basis, although the frame representation for such kind of signals can be not properly sparse. In literature the frame based de-noising procedures can be divided into two classes: Bayesian approaches and variational approaches: both types promote sparseness through specific prior hypothesis or penalization term. A new frame based de-noising procedure is presented where no sparseness hypothesis is required on frame coefficients. In particular, the estimator is derived as the empirical version of the Wiener filter generalized to the frame operator. An analytic expression of it is furnished so no searching strategy is required for the implementation. Results on standard and real test signals are presented.

CS70 Room 4 MODELLING CREDIT RISK IN FINANCIAL MARKETS **Chair: Alessandra Canepa**

C317: **Credit and crises: An investigation of housing markets and credit growth in the OECD**
Presenter: **Dilruba Karim**, Brunel University, United Kingdom
Co-authors: Ray Barrel
The role of credit growth in the time series process for house prices in a panel of countries is tested. We test for structural breaks in the relationship at times of significant liberalisation. We also look for unspecified breaks using recent developments in econometric techniques. In addition, we investigate whether credit growth is a useful tool for predicting financial crises in OECD countries as compared to house prices and other potential indicators. We exploit a unique data set to test the role of credit as a percent of GDP, and discuss our results in the light of microeconmetric evidence.

C424: **Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios**
Presenter: **Angelos Vouldis**, UADPhilEcon Univ of Athens and Bank of Greece, Greece
Co-authors: Dimitrios Louzis, Vasilios Metaxas
Dynamic panel data methods are used to examine the determinants of non-performing loans (NPLs) in the Greek banking sector, separately for each loan category (consumer loans, business loans and mortgages). The study is motivated by the hypothesis that both macroeconomic and bank-specific variables have an effect on loan quality and that these effects vary between different loan categories. The results show that, for all loan categories, NPLs in the Greek banking system can be explained mainly by macroeconomic variables (GDP, unemployment, interest rates, public debt) and management quality. Differences in the quantitative impact of macroeconomic factors among loan categories are evident, with non-performing mortgages being the least responsive to changes in the macroeconomic conditions.

C1013: **Profitability uncertainty, capital investment and the cross-section of stock returns**
Presenter: **Marco Della Seta**, University of Lausanne, Switzerland
Co-authors: Sebastian Gryglewicz
Recent theoretical and empirical studies have established the existence of a negative relationship between corporate real investment and future stock returns. We investigate the effects of uncertainty about firm profitability on this relationship. In a real options model, we show that uncertainty

weakens the negative investment-return relationship, and that the relationship may turn positive for sufficiently high levels of uncertainty. Our empirical analysis supports the main prediction of the model.

C476: **Real estate prices and credit risk: An investigation on the US non-performing-loans**
Presenter: **Alessandra Canepa**, Brunel University, United Kingdom
Co-authors: Mauro Costantini, Mohammad Tajik
The sharp increase in US mortgage default rate has recently drawn the attention on the crucial role played by credit risk management on financial stability. In particular, the booming of house price fuelled by growth in subprime mortgage lending and low interest rates has lead to unprecedented credit losses and has affected the global economy. In general, the ex-post risk for financial institutions takes the form of non-performing loans (NPLs). Investigating the determinants of the causes of loan delinquencies is of crucial importance for regulatory authorities concerned with financial stability of the banking system. The link between non-performing loans, real estate prices and banking system is examined. Using stochastic dominance analysis we show that there is a strong relationship between house prices and non-performing loans.

CS16 Room Multipurpose MASSIVELY PARALLEL COMPUTATIONS ON GPUS Chair: Michael Creel

C117: **Trading volume in general equilibrium with complete markets**
Presenter: **Eric Aldrich**, University of California, Santa Cruz, United States
The aim is to investigate asset trade in a general-equilibrium complete-markets endowment economy with heterogeneous agents. It is shown that standard no-trade results cease to hold when agents have heterogeneous beliefs and that substantial trade volume is generated, even in the presence of a spanning set of assets. Further, trade volume and price movements have a positive relationship in the model, as is well documented in the empirical literature. A computational algorithm is developed which solves finite-period heterogeneous-beliefs economies and it is demonstrated how the problem is well suited for large-scale parallel computing methods, such as GPU computing.

C188: **Adaptive sequential posterior simulators for massively parallel computing environments**
Presenter: **Garland Durham**, University of Colorado, United States
Co-authors: John Geweke
Massively parallel desktop computing capabilities now well within the reach of individual academics modify the environment for posterior simulation in fundamental and potentially quite advantageous ways. But to fully exploit these benefits algorithms that conform to parallel computing environments are needed. Sequential Monte Carlo comes very close to this ideal whereas other approaches like Markov chain Monte Carlo do not. This paper presents a sequential posterior simulator well suited to this computing environment. The simulator makes fewer analytical and programming demands on investigators, and is faster, more reliable and more complete than conventional posterior simulators. The paper extends existing sequential Monte Carlo methods, filling gaps in the theory to provide a thorough and practical foundation for sequential posterior simulation that is well suited to massively parallel computing environments. It provides a specific algorithm that requires only code for simulation from the prior and evaluation of prior and data densities and works well in a variety of applications representative of serious empirical work in economics and finance. The algorithm is robust to pathological posterior distributions, and provides accurate marginal likelihood simulation approximations. The paper concludes with an application that illustrates the potential of these simulators for applied Bayesian inference.

C741: **Heterogeneous computing in economics: A simplified approach**
Presenter: **Stefano Grassi**, Aalborg University and CREATES, Denmark
Co-authors: Matt P. Dziubinski
The potential of heterogeneous computing in solving dynamic equilibrium models in economics is shown. We illustrate the power and simplicity of C++ Accelerated Massive Parallelism (C++ AMP) recently introduced by Microsoft. Solving a simple real business cycle model using value function iteration we achieve a speed gain compared to the previously documented approaches together with a simplified programming style that naturally enables parallelization.

C190: **Econometrics on GPUs**
Presenter: **Michael Creel**, Universitat Autonoma de Barcelona, Spain
Co-authors: Mohammad Zubair
A graphical processing unit (GPU) is a hardware device normally used to manipulate computer memory for the display of images. GPU computing, also known as general purpose GPU computing, is the practice of using a GPU device for scientific or general purpose computations that are not necessarily related to the display of images. The ability of a GPU to render rapidly changing complex images depends on the ability of the GPU to perform floating point operations related to rotations, scaling, shading, etc. Such operations exhibit data parallelism where different computational units operate simultaneously on independent data. GPUs have become extremely powerful for this sort of work. A number of scientific applications have a data parallel structure, and realizing this, vendors and other parties have developed programming frameworks to make it relatively easy to conduct general purpose computation using GPUs. We show how some problems in econometrics have the data parallel structure that allows for successful use of GPU computing. We find speedups from 9.5 up to 55.4 times, compared to computations done on a single CPU core. These speedups can be obtained with very little expense, energy consumption, and time dedicated to system maintenance, compared to equivalent performance solutions using CPUs. The paper is accompanied by example code that illustrates how econometric computations can be done using one or a number of GPUs. The main intent of the paper is to contribute to the limited number of examples of GPU computing in economics and econometrics, and to provide example code that can be adapted with relative ease to other problems.

CS33 Room 1 COMPUTATIONAL METHODS FOR FLEXIBLE BAYESIAN MODELS Chair: Roberto Leon-Gonzalez

C179: **Estimation of the continuous piecewise linear model**
Presenter: **Rodney Strachan**, The Australian National University, Australia
Co-authors: Fabrizio Carmignani, Rabee Tourky
A number of important economic models imply continuous piecewise linear (CPL) functions. A general method for estimating multivariate CPL models with uncertainty over the number of 'pieces' is presented. Challenges in estimating CPL models include deciding the number and locations of knots for CPL functions. We avoid such issues by proposing hyper-planes that pass through the $n-$polytope described by the data. A clustering technique is used to draw candidate planes which define the contiguous subsets of each piece. These candidates are proposed for a Metropolis-Hastings sampler of the CPL model. Earlier attempts at estimating such models concern either low dimensional cases or limited forms for the function. The suggested approach estimates a completely general CPL function. The technique is applied to estimation of the U.S. monetary policy reaction function.

C544: **Bayesian analysis of anchoring vignette data**
Presenter: **Markus Jochmann**, Newcastle University, United Kingdom
One important problem associated with survey questions (such as, "How healthy are you? Excellent, good, fair, or poor.") is the fact that respondents

interpret ordinal response categories in different ways. Thus, the validity of analyses based on the resulting data can be biased. Anchoring vignettes is a survey design technique that can be used to take account of this problem. One commonly used model for anchoring vignette data is the "compound hierarchical ordered probit" (chopit) model. This paper discusses how to perform Bayesian inference for this model using a state of the art massively parallel algorithm on the gpu.

C245: **Bayesian stochastic search for best predictors: Nowcasting GDP growth**
Presenter: **Fuyu Yang**, University of East Anglia, United Kingdom
Co-authors: Nikolaus Hautsch, Dieter Hess

A Bayesian framework for nowcasting GDP growth in real time is proposed. A Gibbs sampling scheme can be applied to filter out daily GDP growth rates in a state space system based on high dimensional vintage macroeconomic announcements. Estimations of the growth rate in real time using all available macroeconomic information are considered. A posterior simulator is designed to make Bayesian inference. The latent economic conditions can be tracked in a Bayesian framework. Forecast density and nowcast density of the growth rate using the sample draws from the posterior sampler are available as soon as the estimation is achieved. Using the updated macroeconomic variable announcements at a variety of mixed frequencies, we can provide not only an estimate for the current GDP growth rate on a daily basis, but also we can evaluate the uncertainties.

C180: **Fat-tailed gamma autoregressive processes for stochastic volatility with jumps**
Presenter: **Roberto Leon**, GRIPS, Japan

In times of macroeconomic or financial turmoil volatility typically jumps suddenly to extreme values. We develop a distribution for volatility that allows for fat tails and that generalizes the Gamma Autoregressive Process. We show that this specification allows for a simple and efficient MCMC algorithm. By conditioning on some auxiliary variables, all volatilities can be sampled jointly from the joint conditional posterior. Because of its simplicity and efficiency, the MCMC algorithm can be used to tackle long time series. In applications to real data we show that the extension to fat tails is empirically relevant.

CS40 Room 3 ASSETS LINKAGES, MULTIVARIATE DENSITY PREDICTION AND PORTFOLIO OPTIMIZATION Chair: Florian Ielpo

C469: **Likelihood-based independent component analysis and mixture models**
Presenter: **Jochen Krause**, University of Zurich, Switzerland
Co-authors: Marc Paolella

A new and general likelihood-based approach to independent component analysis (LICA) is proposed which allows for dimension reduction by decomposing N source signals into M, $M \leq N$, independent components. Fast maximum likelihood estimators are derived, and a new statistical test for the optimal number of independent components, testing the joint in independence of M time series, is introduced. For the financial context of portfolio selection, LICA is extended to incorporate (mixture) GARCH, and the LICA-MixNormal-GARCH model is devised. The tractability of the relevant characteristic functions facilitates portfolio optimization by expected shortfall as the downside risk measure. Using augmented likelihood estimation (for avoiding estimation issues related to problems associated with mixture models), the model is investigated in an empirical out-of-sample forecast study based on the Dow Jones 30 Industrial Average.

C526: **MARC-MARS: Modeling asset returns via conditional multivariate asymmetric regime-switching**
Presenter: **Pawel Polak**, Swiss Finance Institute, Switzerland
Co-authors: Marc Paolella

A new multivariate asymmetric regime-switching model for asset returns is proposed. It extends the Regime Switching Dynamic Correlation model to (possibly special case of a) multivariate generalized hyperbolic distribution. In doing so the model accommodates the stylized fact that the correlations are not constant through time, and allows for volatility clustering, asymmetry and excess kurtosis in the returns. Due to the mixture property of the predictive distribution, the sums of marginals, as required for portfolio optimization, is tractable. A new two-stage EM algorithm, coupled with shrinkage via a quasi-Bayesian prior, is developed for estimation and is demonstrated to be far faster than existing methods, and thus crucial for use with a large number of assets. The model nests several models previously proposed in the literature, and, using daily returns on the 30 DJIA stocks, is demonstrated to outperform all of them in terms of in-sample fit and out-of-sample forecast.

C360: **Risk spillovers in international equity portfolios**
Presenter: **Matteo Bonato**, UBS AG, Switzerland
Co-authors: Massimiliano Caporin, Angelo Ranaldo

Risk spillover is defined as the dependence of a given asset variance on the past covariances and variances of other assets. Building on this idea, we propose the use of a highly flexible and tractable model to forecast the volatility of an international equity portfolio. According to the risk management strategy proposed, portfolio risk is seen as a specific combination of daily realized variances and covariances extracted from a high frequency dataset, which includes equities and currencies. In this framework, we focus on the risk spillovers across equities within the same sector (sector spillover), and from currencies to international equities (currency spillover). We compare these specific risk spillovers to a more general framework (full spillover) whereby we allow for lagged dependence across all variances and covariances. The forecasting analysis shows that considering only sector- and currency-risk spillovers, rather than full spillovers, improves performance, both in economic and statistical terms.

C421: **Determining the maximum number of uncorrelated strategies in a global portfolio**
Presenter: **Ling-Ni Boon**, University Paris One Pantheon Sorbonne, France
Co-authors: Florian Ielpo

The number of uncorrelated strategies to be included in a portfolio is a pertinent issue to hedge fund managers who desire to improve risk-adjusted returns. This question can be formulated as the search for the optimal number of factors in a factor model, a task for which various criteria have been proposed, many of them in the framework of Principal Component Analysis. Three criteria based upon information criteria (IC), refined IC and comparison of variances are implemented on five sets of financial returns data between 1997 till 2012. Refined IC is judged to be best in terms of accuracy and precision. The number of factors estimated, in parenthesis, by dataset, is Global Macro Hedge Fund (5), US Treasury Bond Rates (3), US Credit Spreads (1), Commodity Prices (2) and Currencies (2). The results conform to intuition and previous research, if any. Factors are identified, interpreted and their stability over time is investigated via testing of the significance of correlation between factors over a rolling window. While certain datasets, such as US Treasury Bonds, indicate that the number of uncorrelated factors is lower during recessions, the correlation matrices may not be robust due to outliers, especially during economic crises.

CS43 Room 6 MODELLING AND FORECASTING VOLATILITY AND RISK Chair: Ana Perez-Espartero

C139: **Estimation of the volatility in local long memory in stochastic volatility models**
Presenter: **Josu Arteche**, University of the Basque Country UPV/EHU, Spain

Long Memory in Stochastic Volatility (LMSV) models are flexible tools for the modelling of persistent dynamic volatility, which is a typical characteristic of financial time series. Empirical researchers however have usually preferred more rigid ARCH based models where the conditional variances are exact functions of past observations, which makes it very easy to obtain a series of estimated volatilities. In SV models the conditional variance is a latent variable driven by innovations different from those driving the levels and its estimation is not a simple task. This is especially arduous if the volatility component shows long memory, which makes more traditional techniques, such as the Kalman filter, difficult to use. A new technique of volatility extraction, based on a semiparametric version of the optimal Wiener-Kolmogorov filter in the frequency domain, is proposed. A feasible plug-in version is implemented under a partial specification of the volatility component. To that end an estimator of the (pseudo) spectral density function is required. A pre-whitened estimator in the frequency domain is considered and its consistency is shown under stationary and non stationary but mean reverting long memory, for constant and shrinking to zero Fourier frequencies. Finally, an application to the daily Dow Jones Industrial index is included.

C266: **Bayesian non-parametric portfolio allocation and hedging risk with multivariate asymmetric GARCH**
Presenter: **Audrone Virbickaite**, Universidad Carlos III de Madrid, Spain
Co-authors: Concepcion Ausin, Pedro Galeano

GARCH models have been extensively researched for over 30 years. Lately, multivariate GARCH models and various extensions to incorporate asymmetric variance and correlation behavior are becoming more popular in explaining time-varying volatilities. We use a Generalized Asymmetric Dynamic Conditional Correlation model, which is able to explain asset specific dynamics and asymmetry in correlations. Also, in order to assume a more flexible distribution for the errors, we model them as an infinite scale mixture of multivariate Gaussian distributions and make use of Bayesian non-parametric approach, in particular, Dirichlet Process mixture models. Moreover, we present some financial applications. We solve portfolio allocation and hedging problems using real data and show that Bayesian non-parametric approach is more flexible and informative in two ways: we are able to relax the distributional assumptions of the errors and also obtain posterior distributions of optimal portfolio and hedging weights of interest, which provide a more complete evaluation of financial risk.

C301: **EGARCH models with fat tails, skewness and leverage**
Presenter: **Genaro Sucarrat**, BI Norwegian Business School, Norway
Co-authors: Andrew Harvey

An EGARCH model in which the conditional distribution is heavy-tailed and skewed is proposed. The properties of the model, including unconditional moments, autocorrelations and the asymptotic distribution of the maximum likelihood estimator, are obtained. Evidence for skewness in conditional t-distribution is found for a range of returns series and the model is shown to give a better fit than the corresponding skewed-t GARCH model.

C922: **Revisiting the fractional cointegrating dynamics of implied-realized volatility relation with wavelet band spectrum regression**
Presenter: **Michaela Barunikova**, IES FSV Charles University Prague, Czech Republic
Co-authors: Jozef Barunik

The fractional cointegrating relationship between ex-ante implied volatility (IV) and ex-post realized volatility (RV) is revisited. It is argued that the concept of corridor implied volatility (CIV) should be used instead of the model-free option-implied volatility when assessing relationship. For RV, recently proposed methods, robust to noise and jumps are used. It is found that it does not affect the implied-realized volatility relation. A new tool for the estimation of fractional cointegrating relation between IV and RV based on wavelets is developed, a wavelet band least squares (WBLS). The main advantage of WBLS in comparison to other frequency domain methods is that it allows us to work with potential non-stationary volatility. The dynamics of the relationship in the time-frequency domain with the wavelet coherence is studied. It is confirmed that dependence comes solely from the lower frequencies of the spectra. The relationship is estimated (using S&P 500 and DAX option prices covering the recent financial crisis) only on this part of the spectra and the results are compared to the fully modified narrow-band least squares based on the Fourier frequencies. It is concluded that in the long-run, volatility inferred from the option prices using CIV provides an unbiased forecast of RV.

CS44 Room 2 THEORY AND APPLICATIONS OF REGIME SPECIFIC BEHAVIOUR Chair: Jean-Yves Pitarakis

C423: **Testing for structural breaks in predictive regressions**
Presenter: **Lukasz Prochownik**, University of Southampton, United Kingdom

The concept of predictive regressions has been studied for the past 20 years and its application is particularly present in applied economics, finance and econometrics. The basic set-up in the predictive regression framework associates the noisy explained variable with the lagged persistent regressor, which can be characterized as a first-order autoregression with a root local to unity. Most papers define it as a NI(1) process as the degree of persistence is not known and they do not rule out the possibility of the process being stationary or even mildly explosive. A vast literature on predictive regression aimed to derive robust inferences and analysed the predictive power of regressors. However, the possibility of structural instability has been predominantly ignored in the 20th century. Given the nature of predictive regressions, Andrews test may underperform when one wants to test for coefficient instability. Given more realistic assumptions, the simulations indicate that the SupWald test is likely to over-reject the null hypothesis of no structural break. The remedy is to use the IVX method which fixes the problems arising from both, local-to-unity characteristics of explanatory variable and the violation of standard exogeneity assumption. Even when the degree of persistence of the regressor is not known, the IVX can help to deliver valid decision regarding the presence of the break in the predictive regression model.

C334: **The reaction of stock market returns to anticipated unemployment**
Presenter: **Jesus Gonzalo**, University Carlos III de Madrid, Spain
Co-authors: Abderrahim Taamouti

The short-run impact of anticipated and unanticipated unemployment rates on stock prices is empirically investigated. We particularly examine the nonlinearity in stock market's reaction to unemployment rate and study the effect at each individual point (quantile) of stock return distribution. Using nonparametric Granger causality and quantile regression based tests, we find that, contrary to the general findings in the literature, only anticipated unemployment rate has a strong impact on stock prices. Quantile regression analysis shows that the causal effects of anticipated unemployment rate on stock return are usually heterogeneous across quantiles. For quantile range (0.35, 0.80), an increase in the anticipated unemployment rate leads to an increase in the stock market price. For the other quantiles the impact is statistically insignificant. Thus, an increase in the anticipated unemployment rate is in general good news for stock prices. Finally, we offer a reasonable explanation of why unemployment rate should affect stock prices and how it affects them. Using Fisher and Phillips curve equations, we show that high unemployment rate is followed by monetary policy action of Federal Reserve (Fed). When unemployment rate is high, the Fed decreases the interest rate, which in turn increases the stock market prices.

C338: **Co-summability: From linear to non-linear co-integration**

Presenter: **Vanessa Berenguer-Rico**, University of Oxford, United Kingdom
Co-authors: Jesus Gonzalo

Co-integration theory is an ideal framework to study linear relationships among persistent economic time series. Nevertheless, the intrinsic linearity in the concepts of integration and co-integration makes them unsuitable to study non-linear long run relationships among persistent processes. This drawback hinders the empirical analysis of modern macroeconomics which often deals with asymmetric responses to policy interventions, multiplicity of equilibria, transition between regimes or even log-linearized equilibria. We formalize the idea of co-summability built upon a concept order of summability previously introduced which, in turn, is conceived to deal with non-linear transformations of persistent processes. Theoretically, a co-summable relationship is balanced and describes a long run equilibrium that can be non-linear. To test for these type of equilibria, inference tools for balancedness and co-summability are designed and their asymptotic properties are analyzed. Their finite sample performance is studied through Monte Carlo experiments. The practical strength of co-summability theory is shown through two empirical applications. Specifically, asymmetric preferences of central bankers and the environmental Kuznets curve hypothesis are studied through the lens of co-summability.

C330: **Inferring the predictability induced by a persistent regressor in a predictive threshold model**

Presenter: **Jean-Yves Pitarakis**, University of Southampton, United Kingdom

A distributional theory for detecting predictability induced by a persistent variable is developed. Our framework is that of a predictive regression model with threshold effects and our goal is to develop operational and easily implementable inferences when one does not want to impose a priori restrictions on the parameters other than the slopes corresponding to the persistent predictor. Differently put our tests for the null hypothesis of no predictability against threshold style predictability across two regimes remain valid without the need to know whether the remaining parameters of the model are characterised by threshold effects or not (e.g. shifting versus non-shifting intercepts). One interesting feature of our setting is that our test statistic remains unaffected by whether some nuisance parameters are identified or not.

Saturday 01.12.2012 14:00 - 16:05 Parallel Session D – ERCIM

ESI01 Room Auditorium ROBUST STATISTICS **Chair: Mia Hubert**

E664: **M-estimation of shape matrices under incomplete and serially dependent data**
Presenter: **Gabriel Frahm**, Helmut Schmidt University, Germany
Co-authors: Hannu Oja, Klaus Nordhausen
The shape matrix of a multivariate distribution is an essential parameter which is frequently required in many statistical applications like, e.g., principal components analysis, canonical correlation analysis, linear discriminant analysis, and linear regression. The asymptotic distribution of M-estimators for the shape matrix under incomplete and serially dependent data is presented. It is shown that under incomplete data, a particular scaling condition must be satisfied to guarantee Fisher consistency. Finally, a fixed-point algorithm, which is fast and reliable even in high dimensions, is briefly discussed.

E484: **Local multiple-output quantile regression**
Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium
Co-authors: Marc Hallin, Zudi Lu, Miroslav Siman
A new quantile regression concept, based on a directional version of Koenker and Bassett's traditional single-output one, has recently been introduced for multiple-output location/regression problems. The polyhedral contours provided by the empirical counterpart of that concept, however, cannot adapt to unknown nonlinear and/or heteroskedastic dependencies. Local constant and local linear (actually, bilinear) versions of those contours are introduced, which both allow us to asymptotically recover the conditional halfspace depth contours that completely characterize the response's conditional distributions. Bahadur representation and asymptotic normality results are established. Illustrations are provided both on simulated and real data.

E449: **Robust estimation under a nested error model**
Presenter: **Isabel Molina**, Universidad Carlos III de Madrid, Spain
Co-authors: Roland Fried, Daniel Pena, Betsabe Perez, Anita Thieler
Robust methods for fitting linear mixed models with application to small area estimation are investigated. In particular, we study the nested error model that includes random effects associated with a grouping variable. Several methods based on M estimation have appeared in the literature. These methods are based on the joint estimation of the regression coefficients and the variance components. We consider also stepwise methods in which the variance components are robustly estimated in a first step and, in a second step, estimates of regression coefficients are obtained. The different approaches are compared and an application to small area estimation is discussed. The selection of starting values and tuning constants under this framework are critical issues that will be also discussed.

ES07 Room 9 TIME SERIES MODELING AND COMPUTATION **Chair: Andres M. Alonso**

E588: **Moving average and other models for integer-valued time series**
Presenter: **Maria Eduarda Silva**, University of Porto; CIDMA, Portugal
Time series of counts are available in a wide variety of fields and the need to analyse such data adequately led to a multiplicity of approaches and a diversification of models that explicitly account for the discreteness of the data. One of those approaches consists in replacing the multiplication in the conventional ARMA models by an appropriate random operator, denominated thinning operator, originating the so called INARMA models. Within the INARMA models, the INAR class has been widely studied. We focus on the INMA models. Moving-average models play an important role in time series analysis, especially in economics and finance. They can be viewed as the superposition of the random impulses or shocks according to some mechanism. In the context of time series of counts, several classes of INMA models may be defined, depending on the specification of within and between thinning operations dependence. We review several types of INMA models and introduce a class of integer-valued moving average models with feedback: the Self-Exciting Threshold INteger Moving Average models.

E1014: **Threshold models in the analysis of the time series of counts**
Presenter: **Manuel Scotto**, Universidade de Aveiro, Portugal
Co-authors: Raquel Nicolette, Isabel Pereira
Continuous-valued threshold autoregressive models have been extensively investigated. In contrast, in the field of integer-valued time series modeling little research have been carried out so far to develop models for dealing with time series of counts exhibiting piecewise-type patterns. This work aims to give a contribution towards this direction. The self-exciting threshold integer-valued autoregressive model with two regimes driven by sequences of independent Poisson-distributed random variables is introduced and studied in some detail. Basic probabilistic and statistical properties of this model are discussed. Moreover, parameter estimation and model selection are also addressed. The results are illustrated through a simulation study.

E758: **Vertical and horizontal asymmetries for moving averages driven by Laplace motion**
Presenter: **Anastassia Baxevani**, University of Cyprus, Cyprus
Co-authors: Krzysztof Podgorski, Jorg Wegener
It is discussed how a class of second order models built on the skewed Laplace distributions can account for asymmetries observed in spatio-temporal records. It is shown that these models are capable of mimicking not only distributional skewness but also more complex geometrical asymmetries of the sample path such as tilting, front-back slope asymmetry and time irreversibility. Several measures intending to capture such asymmetries have been discussed in the past. We propose a family of measures that is motivated by crossing level distributions of the slope obtained via the Rice formula. As an application, the Matern covariance model with asymmetric kernels is discussed. Moreover, a complete parametric fit of such a model to records featuring asymmetries is provided, based on the presented asymmetry measures.

E560: **Nonparametric pseudo-LM stationarity testing. Asymptotics and some bootstrap and panel improvements**
Presenter: **Manuel Landajo**, University of Oviedo, Spain
Co-authors: Maria Jose Presno, Paula Fernandez
A nonparametric test for the null of a stationary-around-a-trend data generation process against the unit root alternative is proposed. The test provides a generalization to classical Lagrange-multiplier stationarity testing. The model includes an arbitrary unknown trend function, which is estimated nonparametrically by using OLS regressions on a sieve of trigonometric polynomials. The test statistic is computed upon the basis of the residuals of nonparametric regression. After a suitable rescaling, the limiting null distribution of the test is standard normal, unlike those of most parametric stationarity tests, whose distributions typically are nonstandard. Under the alternative the test rejects with probability approaching one as sample size goes to infinity. A number of bootstrap and panel-based extensions of the basic test are also outlined, which improve the finite

sample performance of the nonparametric test and enable its application to panel data affected by both time-series and cross-sectional dependence. Theoretical analysis is complemented with Monte Carlo simulation studies and some empirical applications to economic and financial time series.

E997: **Data smoothing subject to diminishing marginal returns**
Presenter: **Ioannis Demetriou**, University of Athens, Greece
Let data of a univariate process be given. If the data are related by a sigmoid curve, but the sigmoid property has been lost due to the errors of the measuring process, then the least sum of squares change to the data that provides nonnegative third divided differences may be required. It is a structured quadratic programming calculation, which is solved very efficiently by a special least squares algorithm that takes into account the form of the constraints. The algorithm is outlined and two examples on real economic data are considered. The first is an application to the U.S.A. renewable energy consumption data during the period 1980-2010, which exhibit a sigmoid pattern. The second is an application to technological substitutions among the PDP computers to the VAX computers between the years 1984 and 1991. The results are briefly analyzed and the modeling capability of the method is demonstrated.

ES14 Room 12 BIOSTATNET: YOUNG RESEARCHERS FORUM I Chair: Maria Durban

E642: **Development and implementation of a methodology to select optimal cut-points to categorize continuous covariates in prediction models**
Presenter: **Irantzu Barrio**, Universidad del Pais Vasco UPV-EHU, Spain
Co-authors: Inmaculada Arostegui, Maria Xose Rodriguez-Alvarez, Jose Maria Quintana
In the development of clinical prediction models it is common to use categorical variables as predictors, especially when those models aim to be applied in daily clinical practice in order to support clinicians at decision time. The aim is to propose two methods, named AddFor and Genetic, to obtain an optimal number and location of cut-points to categorize continuous predictors so as to be used in clinical prediction models. Considering we have a dichotomous response variable Y and a continuous covariate X, our proposal consists on categorizing X in such a way that we obtain the best predictive logistic model (highest area under the receiver operating characteristic curve - AUC) for Y. We have applied the proposed methodology to the IRYSS-COPD study and categorized the continuous covariates arterial blood gas PCO2 and respiratory rate, into 4 and 3 categories respectively, and found out that there were not statistically significant differences between the AUC obtained with the categorical variable and the AUC given by the original continuous predictor. Finally, we have developed an R package which implements these methods and provides the user with the optimal cut-points and the categorized variable to be used in the prediction model.

E724: **OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests**
Presenter: **Monica Lopez-Raton**, University of Santiago de Compostela, Spain
Co-authors: Maria Xose Rodriguez-Alvarez, Carmen Cadarso-Suarez, Francisco Gude-Sampedro
Continuous diagnostic tests are often used for discriminating between healthy and diseased populations. For their application in clinical practice, it is useful to select a cutpoint c which defines positive and negative test results. In general, individuals with a diagnostic test value $T \geq c$ are classified as diseased (positive test), whereas patients with $T < c$ are classified as healthy (negative test). Several strategies for selecting optimal cutpoints in diagnostic tests have been proposed depending on the main goal. Two main statistical approaches to the problem of selecting an optimal value are usually considered. One of them uses the Receiver Operating Characteristic (ROC) curve, and the other one maximizes the Pearson Chi-squared test. An R package, OptimalCutpoints, for selecting optimal cutpoints in diagnostic tests, is presented. It incorporates criteria that take the costs of the different diagnostic decisions into account, as well as the prevalence of the target disease and several methods based on measures of diagnostic test accuracy. Moreover, it enables optimal values to be calculated according to levels of given (categorical) covariates. The graphical output includes the ROC curve of the test.

E745: **A statistical model for hospital admissions caused by seasonal diseases**
Presenter: **David Morina**, Universitat Autonoma de Barcelona, Spain
Co-authors: Pedro Puig, Jose Rios, Anna Vilella, Antoni Trilla
The aim is to present a model based on two-order integer-valued autoregressive time series to analyze the number of hospital emergency service arrivals caused by diseases that present seasonal behavior. We also introduce a method to describe this seasonality, on the basis of Poisson innovations with monthly means. We show parameter estimation by maximum likelihood and model validation and develop several methods for forecasting, on the basis of long-time means and short-time and long-time prediction regions. An application to model the number of hospital admissions per week caused by influenza will be analyzed.

E764: **Model building in non-proportional hazard regression**
Presenter: **Mar Rodriguez-Girondo**, University of Vigo, Spain
Co-authors: Thomas Kneib, Carmen Cadarso-Suarez, Emad Abu-Assi
To improve both prognosis and clinical management of acute myocardial infarction patients, an accurate assessment of the different prognostic factors is required. Recent developments of statistical methods allow for a flexible modeling of the variables affecting survival, including the inspection of possible time-dependent associations. Despite their immediate appeal in terms of flexibility, these models introduce additional difficulties when a subset of covariates and the corresponding modeling alternatives have to be chosen. However, a piecewise exponential representation of the original survival data enables us to use a Poisson likelihood based estimation scheme. We propose to conduct such a data transformation and adapt model building procedures proposed in generalized additive models regression settings to the survival context. Three different model building techniques are adapted and compared via an intensive simulation study. An application to prognosis after discharge for patients who suffered an acute myocardial infarction is presented.

E794: **Testing significant points in quantile regression**
Presenter: **Isabel Martinez-Silva**, University of Santiago de Compostela, Spain
Co-authors: Javier Roca-Pardinas, Rosaura Leis-Trabazo, Carmen Cadarso-Suarez
Quantile regression (QR) is a regression model to asses the relationships between a set of covariates vector and quantile curves of a response variable. In some instances, the effects of continuous covariates on the outcome are highly nonlinear. Consequently, appropriate modeling has to take such flexible smooth effects into account. The main goal is to identify significant points (e.g. maximum, minimum or inflection points) by studying the derivatives of the QR curves. For estimating the QR curve and its first derivative, we propose the use of local linear kernel smoothers. Furthermore, smoothing windows are automatically selected, using the cross-validation technique, and confidence intervals are constructed using bootstrap methods. One drawback of the procedures outlined above is that the choice of smoothing windows and the use of bootstrap resampling techniques may entail high computational cost. In this connection, reliance on local linear kernel estimators has enabled us to use binning techniques, which considerably reduce computation time and render our procedures operational in practical situations. Finally, the proposed methodology was used to identify significant points within the sex- and age-specific reference curves of anthropometric measures in a population of children from Galicia (Northwest Spain).

ES18 Room 10 MODELS AND GOODNESS-OF-FIT FOR TIME SERIES AND DEPENDENT DATA Chair: Alejandra Cabana

E226: **A nonparametric test to compare regression curves based on characteristic functions**
Presenter: **Juan-Carlos Pardo-Fernandez**, Universidade de Vigo, Spain
Co-authors: Anouar El Ghouch, M. Dolores Jimenez-Gamero
A new procedure to test for the equality of k regression curves in a fully nonparametric context will be presented. The test is based on the comparison of empirical estimators of the characteristic functions of the regression residuals in each population. The asymptotic behaviour of the test statistic is studied in detail. It is shown that under the null hypothesis the distribution of the test statistic converges to a combination of χ^2_1 random variables. Under certain population conditions, the asymptotic null distribution of the test statistic is proportional to a χ^2_{k-1}. Numerical studies reveal a good practical performance of the test based on the approximation of the asymptotic null distribution of the test statistic.

E215: **Testing between polynomial and exponential tails**
Presenter: **Joan del Castillo**, Universitat Autonoma de Barcelona, Spain
Two methods to distinguish between polynomial and exponential tails are introduced. The methods are mainly based on the properties of the residual coefficient of variation as a random process, for the exponential and non-exponential distributions. A graphical method, called CV-plot, shows departures from exponentiality in the tails. It is, in fact, the empirical coefficient of variation of the conditional excedance over a threshold. New statistics are introduced for testing the exponentiality of tails using multiple thresholds. Some simulation studies present the critical points and compare them with the corresponding asymptotic critical points. Moreover, the powers of new statistics have been compared with the powers of some others statistics for different sample size. The methods are applied to financial time series. In particular, the daily log-returns of exchange rates of US dollar and other currencies are considered, as well as the residuals of a GARCH model applied to the same series.

E386: **Modeling stationary data by a class of generalized Ornstein-Uhlenbeck processes**
Presenter: **Enrique M. Cabana**, Universidad de la Republica, Uruguay
Co-authors: Argimiro Arratia, Alejandra Cabana
A standard Wiener process w on $\mathbb{R}$, that is, a centred, Gaussian process with the properties $w(0)=0$, $\mathbb{E}(w(t)-w(s))^2=|t-s|$, is mapped onto the Ornstein Uhlenbeck process $x(t)=\mathrm{OU}_\kappa w(t)$ by the operator OU_κ, $\Re(\kappa)>0$, defined by the formula $\mathrm{OU}_\kappa y(t)=\int_{-\infty}^t e^{-\kappa(t-s)}dy(s)$, provided the integral makes sense. In particular, when y is replaced by w, the integral becomes a Wiener integral. Such an operator will be called *Ornstein-Uhlenbeck Operator with parameter* κ, in short: OU_κ. New stationary processes can be constructed by iterating the application of Ornstein-Uhlenbeck operators. The processes obtained by applying successively several OU operators to w and multiplying by a constant factor can be used as models for stationary Gaussian continuous parameter processes, and their evaluations at equally spaced epochs may be applied as models for stationary time series. The processes $x=\sigma\prod_{j=1}^p \mathrm{OU}_{\kappa_j} w$ where $\kappa_1,\kappa_2,\ldots,\kappa_p$ have positive real parts, $j=1,2,\ldots,p$, are real when the parameters are either real or appear in pairs of conjugated imaginary values. The parameters $\kappa=(\kappa_1,\ldots,\kappa_p)$, σ can be estimated by forcing the theoretical covariances to approach the empirical ones, and also by maximum likelihood. The performance of those estimation methods is discussed. A comparison between the abilities of OU and ARMA models to fit stationary data is attempted.

E585: **An asymptotically pivotal transform of the residuals sample autocorrelations with application to model checking**
Presenter: **Miguel A. Delgado**, Universidad Carlos III de Madrid, Spain
Co-authors: Carlos Velasco
An asymptotically distribution-free transform of the sample autocorrelations of residuals in general parametric time series models, possibly non-linear in variables, is proposed. The residuals autocorrelation function is the basic model checking tool in time series analysis, but it is not useful when its distribution is incorrectly approximated because the effects of parameter estimation and higher order serial dependence have not been taken into account. The limiting distribution of the residuals sample autocorrelations may be difficult to derive, particularly when the underlying innovations are uncorrelated but not independent. In contrast, our proposal is easily implemented in fairly general contexts and the resulting transformed sample autocorrelations are asymptotically distributed as independent standard normal distributions when innovations are uncorrelated, providing an useful and intuitive device for time series model checking in the presence of estimated parameters. We also discuss in detail alternatives to the classical Box-Pierce test, showing that our transform entails no efficiency loss under Gaussianity in the direction of MA and AR departures from the white noise hypothesis, as well as alternatives to Bartlett's T_p-process test. The finite sample performance of the procedures is examined in the context of a Monte Carlo experiment for the new goodness-of-fit tests are discussed.

E1044: **New goodness-of-fit diagnostics for dynamic discrete data**
Presenter: **Carlos Velasco**, Universidad Carlos III de Madrid, Spain
Co-authors: Igor Kheifetes
New specification tests for dynamic models of discrete data are proposed. We test the specification of the conditional distribution of multinomial and count data, which is key to apply efficient maximum likelihood methods, to obtain consistent estimates of partial effects and appropriate predictions of the probability of future events. The traditional approach is based on a continuation random transformation of discrete data which leads to continuous uniform iid series under the true conditional distribution. Then standard techniques can be applied to the transformed series, but the extra random noise involved may affect power properties of these methods. We investigate an alternative estimate of a cumulative distribution function (cdf) based on discrete data which can be compared directly to a continuous standard uniform cdf. We analyze the asymptotic properties of goodness-of-fit tests based on this new approach and explore the properties in finite samples of a bootstrap algorithm to approximate the critical values of test statistics. We find that in many relevant cases our new approach performs much better than random-continuation counterparts

ES24 Room 13 STATISTICAL PROBLEMS WITH COMPLEX DATA Chair: Jane-Ling Wang

E165: **Non parametric forecasting of a function-valued non stationary processes. Application to the electricity demand**
Presenter: **Jean-Michel Poggi**, University of Paris-Sud, Orsay, France
Co-authors: Anestis Antoniadis, Xavier Brossat, Jairo Cugliari
The problem of forecasting a functional valued stochastic process is studied. We first explore a previous model combining kernel estimation and wavelets, in the context of a practical application: the French electrical power demand for which the hypothesis of stationarity may fail. The departure from stationarity is twofold: an evolving mean level and the existence of groups that may be seen as classes of stationarity. We explore the adaptation of the original procedure to this special case to get a good reference starting point. Next, we propose some variants and corrections that enhance the prediction performance leading to a predictor that often outperforms alternative operational models. The corrections aim to take into account the presence of these nonstationary features, through a lot of method parameters, including a functional clustering. To restore stationarity, it suffices to handle the existence of groups by constraining the model to use the past data that belongs to the same group of the last available data.

E282: **Statistical problems in diffusion weighted MR (dMRI)**
Presenter: **Joerg Polzehl**, WIAS Berlin, Germany
Co-authors: Karsten Tabelow
An introduction into problems in diffusion weighted magnetic resonance tomography is provided. This includes a description of the experiment and a discussion of data pre-processing and properties of the observed data. Appropriate smoothing techniques for noise reduction in dMRI will be discussed. The second part of the talk addresses modeling of dMRI data, starting with the clinically established diffusion tensor model (DTI). Models for orientation distribution functions that overcome the limitations of DTI will be introduced. To conclude, some remarks on fiber tracking and connectivity analysis will be made. Several examples will be presented and it will be shown how the analysis of dMRI data can be performed within R.

E377: **Statistical analysis of damage evolution via images**
Presenter: **Christine Mueller**, TU University Dortmund, Germany
The surface damage evolution under stress is often analysed by images of long-distance microscopes. Usually hundreds of images are obtained during the fatigue process. To analyse this huge amount of images automatically, a new image tool implemented in R was developed. This new image tool allows the automatic detection of micro cracks and corresponding statistical analysis of crack quantities. It uses Dijkstra's shortest path algorithm to detect micro cracks in situations where the cracks are surrounded by plastic deformations and where a discrimination between cracks and plastic deformations is difficult. The new method was applied to over 2000 images. In particular several specific damage parameters like mean crack length or mean crack orientation can be calculated during the fatigue process. Some of these specific damage parameters are compared statistically with simple damage parameters using images of two specimens under different stress levels at different time points of the fatigue process. It is shown that the specific damage parameters discriminate between the two different damage evolutions in an earlier stage than the simple parameters. They are also less influenced by different brightness and scales of the images and show other desirable properties of a damage parameter.

E539: **Oscillations observed in multiple dimensions**
Presenter: **Sofia Olhede**, University College London, United Kingdom
Multivariate observations over time are an important form of data prevalent in neuroscience, oceanography and other applications. Often such data exhibits cycles or structure that is cyclically repetitive over time, but the nature of the repetition likewise evolves over time. Modelling such phenomena in multiple dimensions is challenging because the evolution across dimensions must be constrained in order for the full phenomenon to be simply described. If no such constraints are made the problem becomes overparameterized and ill posed. We discuss how to model such oscillations jointly across several cycles, and challenges presented in their analysis and estimation. The discussion will be illustrated by examples from Oceanography and Neuroscience.

ES32 Room 8 STATISTICS FOR FUZZY DATA Chair: Gil Gonzalez-Rodriguez

E300: **A decomposition theorem for fuzzy set-valued random variables**
Presenter: **Enea Bongiorno**, Universita degli Studi di Milano, Italy
Co-authors: Giacomo Aletti
Let X be a fuzzy set–valued random variable (FRV), and Θ_X the family of all (deterministic) fuzzy sets η for which the Hukuhara difference $X \ominus_H \eta$ exists $\mathbb{P}$–almost surely. We prove that X can be decomposed as $X(\omega) = C \oplus Y(\omega) \oplus \mathbb{I}_{\mathbf{Ste}(X(\omega))}$ for $\mathbb{P}$–almost every $\omega \in \Omega$, where C is the unique deterministic fuzzy set that minimizes $\mathbb{E}[\rho(X,\eta)^2]$ as η is varying in Θ_X, $\rho(X,\eta)$ is related to well–known distances, Y is a centered FRV (i.e. its generalized Steiner point is the origin), and $\mathbf{Ste}(X(\omega))$ is the generalized Steiner point of $X(\omega)$. This decomposition allows us to characterize all FRV translation (i.e. $X(\omega) = \nu \oplus \mathbb{I}_{\xi(\omega)}$ for some deterministic fuzzy convex set ν and some random element ξ in $\mathbb{R}^d$). In particular, X is a FRV translation if and only if the Aumann expectation $\mathbb{E}X$ is equal to C up to a translation.

E319: **Modeling fuzzy post-retirement financial strategies**
Presenter: **Arnold Shapiro**, Penn State University, United States
The purpose is to investigate and model the fuzziness inherent in post-retirement financial strategies. To this end, we focus is on an agent who has been a retirement savings plan participant, who has made consistent significant contributions to the plan, and who has followed a life-cycle approach to accumulation. It is assumed that only pension assets will be used to fund retirement benefits, and that the estate, if any, will be comprised of non-pension assets and residual pension assets. It also is assumed that there is some leeway with respect to the agent's actual retirement date. The topics addressed include a short overview of post-retirement financial issues, the conceptualization of fuzzy post-retirement financial strategies and related concepts and parameters, and the mechanics of their implementation.

E439: **Characterizing the distribution of fuzzy random variables**
Presenter: **Pedro Teran**, Oviedo, Spain
The notion of the cumulative distribution function of a random variable has been previously attempted to extend to the case of fuzzy random variables. However those attempts are typically naive in that the extended c.d.f. fails the essential property of characterizing the distribution of the random variable: given a c.d.f., the probability that the value of the random variable lies in an arbitrary Borel set is uniquely fixed. A more mathematically sophisticated approach was initiated in the last years noticing that the distribution of a fuzzy random variable can be recovered from the capacity functional of the random closed set defined as the hypograph of the fuzzy random variable. A disadvantage of this indirect approach is that the functional is then defined on a family of sets of a space of higher dimension, most of which do not correspond to fuzzy sets in the initial space. A direct approach in which the distribution of a fuzzy random variable X is characterized by the functional $A \mapsto P(X \subset A)$ defined on an appropriate family of fuzzy sets is presented. That is a natural analog of the c.d.f. from the non-fuzzy case.

E483: **Bagged fuzzy clustering**
Presenter: **Marta Disegna**, Free University of Bolzano, Italy
Co-authors: Pierpaolo D'Urso, Riccardo Massari
The aim is to propose a segmentation technique, the so-called Bagged Fuzzy Clustering (BFC), which represents a fuzzy version of the Bagged Clustering (BC) method. As BC, FBC it combines sequentially fuzzy partitioning and hierarchical clustering methods, to overcome some limitations of these two procedures. As a first step, B bootstrap samples with replacement are generated by drawing from the original sample. The Fuzzy C–Medoids Clustering (FCMdC) procedure is run on each bootstrap sample, obtaining $B \times K$ medoids and membership degrees of each unit to the different clusters suitably normalized. The second step consists in running a hierarchical clustering algorithm on the $B \times K$ medoids for determining a set of partitions of medoids. The best partition of medoids is obtained investigating properly the dendrogram. The membership degrees of the original units to each cluster are obtained as a sum of the normalized membership degrees of the units to all medoids in the cluster. In order to show the effectiveness of the procedure, an application on tourism data is finally provided.

E693: **Fuzzy coding in constrained ordinations**
Presenter: **Michael Greenacre**, Universitat Pompeu Fabra, Spain

Canonical correspondence analysis and redundancy analysis are two methods of constrained ordination regularly used in the analysis of ecological data when several response variables (for example, species abundances) are related linearly to several explanatory variables (for example, environmental variables, spatial positions of samples). Here the advantages of the fuzzy coding of explanatory variables are demonstrated. First, nonlinear relationships can be diagnosed; second, more variance in the responses can be explained; and third, in the presence of categorical explanatory variables (for example, years, regions) the interpretation of the resulting triplot ordination is unified because all explanatory variables are measured at a categorical level. The idea is applied to a data set on fish species surveyed at several locations in the Barents Sea.

ES45 Room 7 STATISTICAL ANALYSIS OF EVENT TIMES II Chair: M. Carmen Pardo

E217: **Estimation of conditional transition probabilities in non-Markov multi-state models**
Presenter: **Luis Machado**, University of Minho, Portugal
Co-authors: Jacobo de Una-Alvarez, Somnath Datta

The problem of estimation of the transition probabilities of an irreversible three-state illness-death model is revisited. We are interested in a regression setup where we estimate these probabilities given a continuous covariate that could either be a baseline covariate or a current covariate that is observed for an individual before the individual makes a particular transition of interest. There exist estimators of conditional transition probabilities but they are restricted to the Markov case, and to a parametric specification of the prediction function. We provide two competing non-parametric regression estimators of the transition probability matrix, both valid under mid regularity conditions even when the system is non-Markov. In both estimators, local smoothing is done by introducing regression weights that are either based on a local constant (i.e. Nadaraya-Watson) or a local linear regression. Right censoring is handled by appropriate reweighting. Right censoring is handled by appropriate reweighting. These approaches are evaluated through a simulation study, comparing the two different estimators. The proposed methods are illustrated using real data.

E583: **Extended geometric processes for application to reliability**
Presenter: **Laurent Bordes**, University of Pau, France
Co-authors: Sophie Mercier

Renewal processes have been widely used in reliability, to describe successive failure times of systems submitted to perfect and instantaneous maintenance actions. In case of imperfect maintenance, different models have been developed to take this feature into account, among which geometric processes. In such a model, successive lifetimes are independent and identically distributed up to a multiplicative scale parameter $a > 0$, in a geometric fashion. We envision a more flexible progression, where the multiplicative scaling factor is not necessarily a geometric progression any more. The corresponding counting process is here named Extended Geometric Process (EGP). As a first step in the study of an EGP, we consider its semiparametric estimation based on the observation of the n first gap times. We start with the estimation of the Euclidean parameter a following the regression method. We next proceed to the estimation of the unknown distribution of the underlying renewal process. Several consistency results, including convergence rates, are obtained. We next turn to applications of EGPs to reliability, where successive arrival times stand for failure (and instantaneous maintenance) times. A first quantity of interest is the pseudo-renewal function associated to an EGP, which is proved to fulfill a pseudo-renewal equation. When the system is deteriorating (case $a < 1$), a preventive renewal policy is proposed: as soon as a lifetime is observed to be too short, under a predefined threshold, the system is considered as too deteriorated and replaced by a new one. This renewal policy is: as soon as a lifetime is observed to be too short (under a predefined threshold) the system is considered as too deteriorated and replaced by a new one. This renewal policy is assessed through a cost function, on a finite horizon time. Numerical experiments illustrate the study.

E600: **Dependence in multivariate competing risks data**
Presenter: **Thomas Scheike**, University of Copenhagen, Denmark

Multivariate competing risks data has recently received considerable attention. Some of the possible methods that are available are reviewed and illustrated for estimating the correlation in the occurrences of cancers in twins, based on the Nordic twin registries. There are two basic approaches for considering this problem. One approach is based on modeling the correlation on the hazard scale. The other one considers the absolute risk scale in terms of the probabilities of the model. These two approaches express two quite different aspects of the correlation.The focus will be on the absolute risk by considering the cumulative incidence function. A random effects model for describing such correlation in terms of the bivariate cumulative incidence will be presented. It will also been shown how other measures of dependence that can be estimated directly.

E780: **Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks**
Presenter: **Hein Putter**, Leiden University Medical Center, Netherlands
Co-authors: Alina Nicolaie, Hans van Houwelingen

A new approach to the problem of dynamic prediction of survival data in the presence of competing risks is presented as an extension of a previous landmark model for ordinary survival data. The key feature of our method is the introduction of dynamic pseudo-observations constructed from the prediction probabilities at different landmark prediction times. They specifically address the issue of estimating covariate effects directly on the cumulative incidence scale in competing risks. A flexible generalized linear model based on these dynamic pseudo-observations and a GEE approach to estimate the baseline and covariate effects will result in the desired dynamic predictions and robust standard errors. Our approach focuses directly on the prediction probabilities of interest, thus avoiding complex modeling of cause-specific hazards or subdistribution hazards. As a result, it is robust against departures from such models. From a computational point of view an advantage of our approach is that it can be fitted with existing statistical software and that a variety of link functions and regression models can be considered, once the dynamic pseudo-observations have been estimated. We illustrate our approach on a real data set of chronic myeloid leukemia patients after bone marrow transplantation.

E746: **Parametric regression models for bivariate zero-inflated left-censored survival data.**
Presenter: **Yves Grouwels**, I-BioStat - UHasselt, Belgium
Co-authors: Roel Braekers

In some clinical or environmental studies researchers observe a positive random response variable. Due to difficulties in the measuring mechanism, they see for some subjects only an upper bound of this response. This type of data is called left-censored. In some studies with left-censored data the response variable also attains a zero-value with a positive discrete probability. Parametric regression models are introduced to describe these bivariate zero-inflated left-censored survival data. The association between two non-zero responses is modeled by a one-parameter family of copulas. Since maximizing the full log-likelihood is computationally difficult, a two-stage estimation procedure can be considered. Firstly one estimates the parameters in the margins, ignoring the dependence of the two components. The second stage involves maximum likelihood of the dependence parameters with the univariate parameters held fixed from the first stage. A partitioned form of the asymptotic variance-covariance matrix of the two-stage estimator can be deduced, together with a subset-delete jackknife estimator for this matrix. The performance of the model is shown through a simulation study and an application on a practical data example.

ES37 Room 11 IMPRECISION AND GRAPHICAL MODELS Chair: Antonio Salmeron

E982: **Subset variable selection**
Presenter: **Klea Panayidou**, University of Frederick, Cyprus
The aim is to select the most informative subset of variables from a larger set of variables whose joint distribution factorizes according to a tree structure. For this purpose, a measure is proposed which facilitates the computations and is effectively an upper bound of the Shannon entropy. The selected subset is evaluated and the selected variables are used to predict the remaining ones. The proposed method is applied to a real dataset.

E375: **A cost-sensitive measure to quantify the performance of an imprecise classifier**
Presenter: **Andres R. Masegosa**, University of Granada, Spain
Co-authors: Joaquin Abellan
The aim is to introduce a measure to check the performance of an imprecise classifier, i.e. a classifier that returns set-valued predictions. An innovative aspect of our measure is that it penalizes wrong predictions using a cost matrix of the errors, given by an expert; and it is based on the Bayesian decision rule. To check our measure we have used two imprecise classifiers, both based on the Imprecise Dirichlet Model (IDM) and graphical models to represent the information from a data set: (i) the CNB classifier, based on the Naïve Bayes classifier; and (ii) the CDT classifier, based on decision trees. A practical example is also presented.

E376: **Learning the structure and the parameters of a credal network with the imprecise sample size Dirichlet model**
Presenter: **Serafin Moral**, University of Granada, Spain
Co-authors: Andres R. Masegosa
When learning Bayesian networks with Bayesian scores, it is known that the final result is significantly dependent of the equivalent sample size parameter. The aim is to address this problem by considering the imprecise sample size Dirichlet model which takes into account an interval of possible values for the parameter. We will focus on a local application of the model: an independent interval for each variable given its parents. The result will be a family of credal networks each one of them associated with different combinations of the different sample sizes for the problem variables. Special attention will be given to the algorithmic aspects of the learning, proposing to approximate the solutions that try to recover in an efficient way a representative set of credal networks. Inference (computation of upper and lower conditional probabilities and conditional dominance) will also be shown.

E124: **Efficient algorithms for inference in graphical models**
Presenter: **Nokuthaba Sibanda**, Victoria University of Wellington, New Zealand
A multi-stage manufacturing process or medical procedure with M stages has output variables Y_k, input variables X_k, and external variables Z_k, at stage k. Propagation of variation from an upstream output, Y_{k-1} to a downstream output Y_k induces correlation in key output variables observed at the final stage, M. In addition, upstream variables $\{X_j, Z_j\}, j \in 1,2,\ldots,k-1$ may have an additional direct effect on Y_k. We employ Bayesian networks to construct models for the outputs of the procedure that take into account propagation of variation from all upstream stages. Two components are required for the model: (i) the structure of the network in terms of conditional independence relationships; (ii) coefficient estimates for the model. Although part of the network structure is implicit from the process stages, we need to identify any additional sources of variation from input and external variables from stages preceding the parent stage. We use a two-step iterative algorithm for fitting the model. The first step proposes a candidate network structure. In the second step, we obtain parameter estimates conditional on the current network structure. Results comparing a number of estimation algorithms will be presented.

E326: **HUGIN architecture for propagating belief functions**
Presenter: **Antonio Salmeron**, University of Almeria, Spain
Co-authors: Finn V. Jensen
The use of a HUGIN like architecture for propagating Dempster-Shafer belief functions is studied. The main issue of the proposed scheme is that it allows to do the entire propagation with mass function. To achieve this, an inverse for mass functions is defined and an efficient way to compute it is proposed.

Saturday 01.12.2012 16:15 - 17:30 Parallel Session E – CFE

CS09 Room 7 RECENT ADVANCES IN THE APPLIED MACROECONOMICS Chair: Fabio Canova

C230: **Why are recessions associated with financial crises different**
Presenter: **Luca Benati**, Department of Economics, University of Bern, Switzerland
Bayesian time-varying parameters structural VARs with stochastic volatility is used to investigate the specific dimensions along which recessions associated with severe financial crises have historically been different from 'normal' recessions, and standard business-cycle fluctuations, in the post-WWII U.S., Japan, Euro area, Sweden, and Finland. We identify four structural shocks by combining a single long-run restriction to identify a permanent output shock, with three sign restrictions to identify demandand supply-side transitory shocks. Evidence suggests that severe financial crises have systematically been characterized by a negative impact on potential output dynamics, which in two cases–Japan, following the collapse of the asset prices bubble of the second half of the 1980s, and Finland, in the early 1990s–appears to have been nothing short of dramatic. On the other hand, with very few exceptions, neither the relative importance of the different types of shocks, the way they have been propagating through the economy, or the persistence of output growth, have exhibited any systematic difference compared to standard macroeconomic fluctuations. Our main conclusion is therefore that, once controlling for the size of the shocks, financial crises appear to have been broadly similar to standard macroeconomic fluctuations along most, but not all, dimensions.

C1057: **On the failure to predict financial crisis**
Presenter: **Gabriel Perez-Quiros**, Bank of Spain, Spain
Co-authors: Maria Dolores Gadea
Much has been written on why economists failed to predict the financial and real state crises. Reading recent literature, it seems that the crisis was so obvious that economists should have been blind when looking at data before the crisis to miss such an important call. We illustrate this failure looking at one of the most cited and relevant variable in this analysis. The now infamous credit to GDP chart. We compare the conclusions made in the literature a posteriori with the results that would have been drawn from an analysis a prioiri. The picture is different, showing that , even though credit affects the business cycle in both the expansion and the recession phases, this effect is almost negligible and impossible to exploit from a policy maker's point of view.

C679: **The impact of financial crises on the risk-return tradeoff and the leverage effect**
Presenter: **Bent Jesper Christensen**, Aarhus University, Denmark
Co-authors: Morten Nielsen, Jie Zhu
The aim is to investigate the impact of financial crises on two fundamental features of stock returns, namely, the risk-return tradeoff and the leverage effect. We apply the fractionally integrated exponential GARCH-in-mean (FIEGARCH-M) model for daily stock return data, which includes both features and allows the co-existence of long memory in volatility and short memory in returns. We extend this model to allow the financial parameters governing the volatility-in-mean effect and the leverage effect to change during financial crises. An application to the daily U.S. stock index return series from 1926 through 2010 shows that both financial effects increase significantly during crises. Strikingly, the risk-return tradeoff is significantly positive only during financial crises, and insignificant during non-crisis periods. The leverage effect is negative throughout, but increases significantly by about 50% in magnitude during financial crises. No such changes are observed during NBER recessions, so in this sense financial crises are special. Applications to a number of major developed and emerging international stock markets confirm the increase in the leverage effect, whereas the international evidence on the risk-return tradeoff is mixed.

CS18 Room 13 NEW DEVELOPMENTS IN GARCH MODELS AND FINANCIAL SERIES MODELLING Chair: Christian Francq

C154: **Forecasting GDP over the business cycle in a multi-frequency and data-rich environment**
Presenter: **Marie Bessec**, University Paris Dauphine, France
Co-authors: Othman Bouabdallah
Two specifications developed recently in the forecasting literature are merged: the MS-MIDAS model and the MIDAS-factor model. The MS-factor MIDAS model (MSFaMIDAS) that we introduce 1) incorporates the information provided by a large data-set, 2) takes into account mixed frequency variables, 3) captures regime-switching behaviors. Monte Carlo simulations show that this new specification tracks the dynamics of the process quite well and predicts the regime switches successfully, both in sample and out-of-sample. We apply this new model to US data from 1959 to 2010 and detect properly the US recessions by exploiting the link between GDP growth and higher frequency financial variables.

C238: **Dynamic correlation models based on volatilities**
Presenter: **Jean-David Fermanian**, ENSAE-CREST, France
Co-authors: Hassan Malongo
A lot of efforts have been devoted to model the dynamic correlations between financial asset returns. It has been recognized it is difficult to find their main drivers, either through exogenous macroeconomic variables or by latent (not observable) factors. Nonetheless, in asset pricing, some standard models invoke the individual stochastic volatilities as tools to explain the correlation moves, especially to price index options. Here, the authors try to fill the gap between this practice of options valuation, and financial econometrics. They explain to what extent the stochastic volatilities of asset returns can explain the moves of their correlations empirically. Following the previous intuition, several new versions of multivariate GARCH models are proposed and estimated, particularly in the DCC family. Extensions towards time-dependent copulas are detailed. The performances of these new model specifications to risk-manage financial portfolios dynamically (in a mean-variance framework) are evaluated.

C341: **GARCH models without positivity constraints: Exponential or log GARCH**
Presenter: **Jean-Michel Zakoian**, CREST and University Lille 3, France
Co-authors: Christian Francq, Olivier Wintenberger
The probabilistic properties and the estimation of the asymmetric log-GARCH(p,q) model are studied. In this model, the log-volatility is written as a linear function of past values of the log-squared observations, with coefficients depending on the sign of such observations, and past log-volatility values. The model shares interesting features with the EGARCH model, in particular the absence of positivity constraints on the volatility coefficients. Conditions are obtained for the existence of solutions and finiteness of their log-moments. It is shown that the quasi-maximum likelihood estimation (QMLE) of the parameters is strongly consistent and asymptotically normal. A Lagrange-Multiplyer (LM) test is derived for testing the null assumption of a log-GARCH against more general formulations including the EGARCH. Simulations illustrating the theoretical results and an application to real financial data are proposed.

CS66 Room 8 CREDIT RATING MODELS **Chair: Silvia Figini**

C239: **Determinants of multiple states business exit in Europe**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Luca Sensini
The difficulties experienced by firms and institutions during the Global Financial Crisis (GFC) demonstrated the importance of understanding the determinants of financial default risk and to further investigate contagion between industries, regions and countries. The aim is to identify the fundamental variables that drive the financial distress across the European countries paying attention to the different reasons that may cause the exit from the market. A competing risk approach has been implemented at both national and European levels, allowing us to study the single-country specificities as well as the between-country interdependencies. The competing risk model is able to estimate the probability that a firm has to exit the market for each reason according to a preselected subset of variables that would have different effects on the risks due to multiple causes. The significant variables and their sign are compared for each country model in order to assess the differences in the determinants of financial distress and in the predictive ability of the model set-ups. Moreover, the predictive ability of the models has been evaluated by means of some accuracy measures, at different time horizons.

C246: **Model uncertainty in credit rating models**
Presenter: **Silvia Figini**, University of Pavia, Italy
Co-authors: Paolo Giudici
Model uncertainty remains a challenge to researchers in different applications. When many competing models are available for estimation and without enough guidance from theory, model averaging represents an alternative to model selection. In credit risk, statistical models are usually chosen according to a model selection procedure that aims at selecting the most performing structure. The chosen model is, once selected, taken as the basis for further actions, such as parameter estimation, default prediction and predictive classification. We investigate whether the usage of more models, in a model averaging perspective, improves the performance of credit risk models taking into account model uncertainty. To achieve this aim we focus on the role of the prior assumptions in the Bayesian Model Averaging framework. In order to compare the results obtained by single models and model averaged ones, we propose a set of performance measures. In order to show how our proposal works, we have used a real data set provided by a credit rating agency composed of small and medium enterprises, that has already been considered in the statistical literature.

C359: **Default prediction of SMEs by a generalized extreme value additive model**
Presenter: **Raffaella Calabrese**, University of Milano-Bicocca, Italy
Co-authors: Silvia Osmetti
A new model is proposed for default prediction of Small and Medium Enterprises (SMEs). The main weakness of the scoring models proposed in the literature is not to consider the default as a rare event. To take into account this characteristic, the quantile function of the Generalized Extreme Value (GEV) distribution is suggested as link function in a Generalized Linear Model (GLMs). In the GLMs, the relationship between the independent variable and the predictor is constrained to be linear. Since this assumption is not usually satisfied by scoring models, a Generalized Additive Model (GAM) is suggested with the quantile function of the GEV distribution as link function. Hence, the Generalized Extreme Value Additive (GEVA) model is proposed. Finally, our proposal is applied to empirical data on Italian SMEs. It is obtained that the GEVA model shows a high accuracy for predicting defaults.

CS21 Room 3 RISKS MEASURES **Chair: Hasinavonizaka Rahantamialisoa**

C181: **Aggregation of market risks using pair-copulas**
Presenter: **Fatima Jouad**, Paris 1 University - AXA GRM, France
Co-authors: Dominique Guegan
The advent of the Internal Model Approval Process within Solvency II and the desirability of many insurance companies to gain approval has increased the importance of some topics such as risk aggregation in determining overall economic capital level. The most widely used approach for aggregating risks is the variance-covariance matrix approach. Although being a relatively well-known concept that is computationally convenient, linear correlations fail to model every particularity of the dependence pattern between risks. We apply different pair-copula models for aggregating market risks that represent usually an important part of an insurer risk profile. We then calculate the economic capital needed to withstand unexpected future losses and the associated diversification benefits. The economic capital will be determined by computing both 99.5thVaR and 99.5thES following the requirements of Solvency II and SST.

C480: **A maximum (non-extensive) entropy approach to equity options bid-ask spread**
Presenter: **Oren Tapiero**, Bar-Ilan Univeristy, Israel
The behavior of equity options bid-ask spread across strike prices is modelled by maximizing Kaniadakis entropy. The implied true option prices relative to their respective bid and ask prices are derived. We find that the true value of option prices is on average closer to bid prices. Also, we verify the validity of two different theories explaining equity options bid-ask spreads: that of asymmetric information in options market explaining the options bid-ask spread, and that of asymmetric information in the underlying market causing the latter. Using the parameters derived from maximizing Kaniadakis Entropy and daily data calculated from the TAQ database (the underlying average daily relative bid-ask spread and net order flow), we find a feedback relationship between κ (the parameter determining the overall behavior and size of equity options bid-ask spread across strike prices) and the latter two variables. Nevertheless, the transfer entropy metric between the latter two variables and κ suggests that the directional information flow is from the options to the underlying market. Thus, to an extent, making κ a possible measure of information asymmetry in the equity options market.

C294: **Concepts of risk measure**
Presenter: **Hasinavonizaka Rahantamialisoa**, CES, France
Co-authors: Dominique Guegan, Bertrand Hassani
Recent global financial crises has trigged up the need for developing a standard framework for risk measurement. For instance, one of the challenging tasks of modern risk measurement is to prevent liquidity risk. To be more precise, in practice, every institution needs levels of liquidity high enough to meet its payment obligations and low enough to take advantage of any investment opportunities. Due to this fact, there is a growing interest for the risk manager to focus on the tail behaviour and the use of a tail conditional expectation (TCE), since it shares properties that are considered desirable and applicable in a variety of situations. In fact for an institution faced with a random loss, the tail conditional expectation represents the conditional average amount of loss that can be incurred in a given period, given that the loss exceeds a specified value. This value is usually based on the quantile of the loss distribution, the so-called value-at-risk (VaR). In this study , we attempt to deal with the following issues : discuss the limitation of the VaR, establish a link between VaR, expected shortfall (ES) and TCE, understand the extension to spectral risk measure and the distortion measure, and then construct some intuition for developing a standard framework for risk measurement. In addition, we provide a comprehensive analysis of the feasibility of this approach using some well-known distributions in finance.

CS24 Room 4 TESTING FOR COMMON CYCLES IN MACROECONOMIC TIME SERIES Chair: Alain Hecq

C369: **DSGE models and cyclical co-movements in VARs**
Presenter: **Massimo Franchi**, University of Rome La Sapienza, Italy
Co-authors: Paolo Paruolo
Currently-employed conditions for the existence of Vector Auto Regressive (VAR) representations for Dynamic Stochastic General Equilibrium (DGSE) models are reviewed. We find that the conditions put forth in previous works imply the existence of various forms of cyclical co-movements in the VAR representations. These cyclical co-movements provide implications of the DGSE model that are testable on their VAR representation.

C273: **Macroeconomic forecasting through regularized reduced-rank regression**
Presenter: **Gianluca Cubadda**, University of Rome Tor Vergata, Italy
Co-authors: Emmanuela Bernardini
A strategy to detect and impose reduced-rank restrictions in large multivariate time series models is proposed. In this framework, it has recently been shown that Canonical Correlation Analysis (CCA) does not perform well. We propose a method that combines the richness of reduced-rank regression with the simplicity of univariate auto-regressive models. In particular, we suggest to use proper shrinkage estimators of the variance-covariance matrices that are involved in CCA, thus obtaining a method that is asymptotically equivalent to CCA, but it is numerically more stable in finite samples. Simulations and empirical applications document the merits of the proposed approach both in forecasting and in structural analysis.

C147: **Testing for common cycles with varied frequency data**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Jean-Pierre Urbain, Thomas Goetz
A way to detect the presence of common cyclical features when several time series are observed/sampled at different frequencies is proposed. It generalizes the common frequency analysis introduced by Engle and Kozicki. So doing we rely on the mixed frequency VAR representation investigated in Ghysels for stationary time series. However, for non-stationary time series in the level, one needs to account for the presence of two set of long-run relationships. The first set is implied by identities stemming from the fact that the difference of the high frequency regressors are stationary. The second set comes from additional long-run relationships between one of the high and the low frequency variables. Those transformed VECM representations generalize the results of Ghysels and determine the correct set of instruments to be used in a common cycle investigation. Empirical analyses with the quarterly US GDP and monthly IPIs illustrate our new approach.

CS29 Room 9 BAYESIAN EMPIRICAL MACROECONOMICS Chair: Deborah Gefang

C638: **Model switching in time-varying parameter regression models**
Presenter: **Miguel Belmonte**, Stratlhclyde, United Kingdom
Co-authors: Gary Koop
The usefulness of switching Gaussian state space models as a tool for implementing dynamic model selection in time-varying parameter regression models is investigated. Dynamic model selection methods allow for model switching, where a different model can be chosen at each point in time. Thus, they allow for the explanatory variables in the time-varying parameter regression model to change over time. We compare our exact approach to a popular existing procedure which relies on the use of forgetting factor approximations. In an application, we investigate which of several different forecasting procedures works best for inflation. We also investigate whether the best forecasting method changes over time.

C739: **Monetary policy transmission during financial crises: An empirical analysis**
Presenter: **Tatjana Dahlhaus**, Universitat Autonoma de Barcelona, Spain
The effects of a monetary policy expansion in the U.S. during times of financial crises are studied. The analysis is done introducing a Smooth Transition Factor Model where the transition between states ("normal" times and financial crises) depends on a financial condition index. The model is estimated using Bayesian MCMC methods. Employing a quarterly dataset over the period 1970Q1-2009Q2 containing 108 U.S. macroeconomic and financial time series it is found that a monetary expansion has stronger and more persistent effects on macroeconomic variables such as output, consumption and investment during financial crises than during normal times. Differences in effects among the regimes seem to originate from nonlinearities in the credit channel.

C851: **Industry productivity in Europe**
Presenter: **Deborah Gefang**, Lancaster University, United Kingdom
The industry productivity in Europe is investigated using Bayesian stochastic frontier methods. Making use of the translog system and a recently introduced nonparametric clustering approach, we consider the situation that an industry of a country falls into two groups: one group containing the same industries across all countries, and the other group containing all the industries in that country. We use EU KLEMS data to rank the productivities of 20 industries in 10 European countries according to industry productivity, country productivity, and both industry and country productivity, respectively.

CS34 Room 2 HIGH DIMENSIONAL BAYESIAN TIME SERIES ECONOMETRICS Chair: Richard Hahn

C236: **Large time-varying parameter VARs**
Presenter: **Dimitris Korobilis**, University of Glasgow, United Kingdom
Co-authors: Gary Koop
Methods for estimation and forecasting in large time-varying parameter vector autoregressive models (TVP-VARs) are developed. To overcome computational constraints with likelihood-based estimation of large systems we rely on Kalman filter estimation with forgetting factor. We also draw on ideas from the dynamic model averaging literature and extend the TVP-VAR so that its dimension can change over time. A final extension lies in the development of a new method for estimating, in a time-varying manner, the parameter(s) of the shrinkage priors commonly-used with large VARs. These extensions are operationalized through the use of forgetting factor methods and are, thus, computationally simple. An empirical application involving forecasting inflation, real output, and interest rates demonstrates the feasibility and usefulness of our approach.

C363: **Factor model shrinkage for linear instrumental variable analysis with many instruments**
Presenter: **Richard Hahn**, University of Chicago, United States
Co-authors: Hedibert Lopes
Sparse factor models are adapted to the instrumental variables linear model in order to mitigate the difficulty of causal estimation in the many-instruments context. An efficient and modular computational approach is described, which allows rapid comparison of posterior inference under a variety of different first-stage models.

C443: **Credit shocks, monetary policy, and business cycles: Evidence from a structural time varying Bayesian FAVAR**
Presenter: **Pooyan Amir Ahmadi**, Goethe University, Germany

A Bayesian factor-augmented vector autoregression model is estimated using a large panel of macroeconomic and credit spread data from the United States for the period 1926-2009. The model has time varying parameters and volatilities. A number of episodes with high volatility in the common component of credit spreads are identified. Often, though not always, these episodes coincide with (or lead) NBER recessions. During these episodes, credit spread shocks and monetary policy shocks are found to have much stronger effects on macroeconomic variables than on average. The degree of amplification of those responses reaches at its peak a factor of up to ten. These amplified responses tend to exhibit a larger persistence.

CS58 Room Multipurpose BOOSTRAPPING PANEL AND TIME SERIES DATA: THEORY AND APPLICATIONS Chair: Jean-Pierre Urbain

C441: **Wild bootstrap tests for autocorrelation in vector autoregressive models**
Presenter: **Niklas Ahlgren**, Hanken School of Economics, Finland
Co-authors: Paul Catani

Conditional heteroskedasticity is a common feature of many macroeconomic and financial time series. Standard tests for error autocorrelation are derived under the assumption that the errors are IID and are unreliable in the presence of conditional heteroskedasticity. We propose wild bootstrap tests for autocorrelation in vector autoregressive (VAR) models when the errors are conditionally heteroskedastic. We use a residual-based recursive wild bootstrap procedure. In particular, we investigate the properties of Lagrange multipliers (LM) and F-type tests. Monte Carlo simulations show that the wild bootstrap tests have satisfactory size properties in models with constant conditional correlation generalised autoregressive conditional heteroskedastic (CCC-GARCH) errors. In contrast, standard asymptotic and residual-based bootstrap tests are shown to be oversized. Some simulation evidence on the power of the tests is given. The tests are applied to credit default swap prices and Euribor interest rates. The results show that there are significant ARCH effects in the residuals from the estimated VAR models. The empirical examples demonstrate that wild bootstrap tests should be preferred over standard asymptotic and residual-based bootstrap tests.

C651: **Bootstrapping subset test statistics in IV regression**
Presenter: **Noud van Giersbergen**, Amsterdam, Netherlands

The finite-sample performance of various bootstrap procedures is studied by simulation in a linear regression model containing 2 endogenous regressors. Besides several residual-based bootstrap procedures, we also consider the GMM bootstrap. The test statistics include t-statistics based on k-class estimators and the robust subset quasi-LR (MQLR) statistic. In the simulations, the restricted fully efficient (RFE) bootstrap DGP based on Fuller estimates and the LIML t-statistic performs best of the Wald-type statistics. Unfortunately, the bootstrap only marginally reduces the conservativeness of the subset MQLR statistic. Finally, the GMM bootstrap does not seem to improve upon the asymptotic approximation. An empirical example illustrates the use of these procedures.

C842: **Unit root testing using modified wild bootstrap methods**
Presenter: **Jean-Pierre Urbain**, Maastricht University SBE, Netherlands
Co-authors: Stephan Smeekes

Several modified wild bootstrap methods are considered in order to test for unit roots. The wild bootstrap methods are modified in the sense that they can capture dependence over time as well as heteroskedasticity, making them attractive in the unit root testing framework. We apply the methods to panel unit root testing where the modified wild bootstrap methods provide a simple way to take heteroskedasticity, autocorrelation and cross-sectional dependence into account simultaneously. We derive the asymptotic properties of the methods and perform a simulation study to evaluate their small sample properties.

CS60 Room 5 INDIRECT ESTIMATION METHODS Chair: David Veredas

C168: **Which model to match**
Presenter: **Roxana Halbleib**, University of Konstanz, Germany
Co-authors: David Veredas, Matteo Barigozzi

The asymptotic efficiency of indirect estimation methods, such as the efficient method of moments and indirect inference, depends on the choice of the auxiliary model. To date, this choice is somehow ad hoc and based on an educated guess. We introduce information criteria that help the user to optimize the choice among nested and non–nested auxiliary models. They are the indirect analogues of the widely used Akaike–type criteria. A thorough Monte Carlo study based on two simple and illustrative models shows the usefulness of the criteria.

C263: **Indirect inference for time series observed with error**
Presenter: **Paolo Santucci de Magistris**, CREATES - Aarhus University, Denmark
Co-authors: Eduardo Rossi

The properties of the indirect inference estimator of a structural model when the observed series are contaminated by a measurement error are analyzed. The binding function is characterized when the signal is a stationary ARMA process and the error is an i.i.d. process. The outcomes of the estimation are not only the estimates of the structural parameters, but also the estimate of the measurement error variance. An extensive simulation analysis is carried out. The results highlight the bias reduction of the indirect estimates obtained when the measurement error is explicitly estimated with respect to the case when it is neglected. An application to the continuous time stochastic volatility model estimation shows the advantages of the proposed method.

C311: **Indirect inference versus data cloning for estimating multiple-membership random effects logit models**
Presenter: **Anna Gottard**, University of Firenze, Italy
Co-authors: Giorgio Calzolari

Multiple-membership logit models with random effects are models for data in which each statistical unit can belong to more than one group. For these models, the likelihood function is analytically intractable, while inference based on quasi-likelihood leads to biased estimates. Data cloning and indirect inference approaches are proposed for these models. A Monte Carlo experiment compares the two approaches on a set of simulated data. The indirect estimator seems to be almost as efficient as the maximum likelihood one based on data cloning, but has an enormous advantage in computational efficiency (seconds or minutes instead of hours). Data cloning is a novel approach producing maximum likelihood estimates, through the posterior distribution of an adequate Bayesian model fitted on cloned data. In fact, the posterior distribution collapses on the maximum likelihood estimate as the number of clones goes to infinity. A data clone algorithm specific for the case of multiple-membership data is proposed. Indirect inference is a non-likelihood-based method for estimating parameters. The idea is to use an auxiliary model, not necessarily correctly specified, to select a sensible estimate. An auxiliary model having the same dimension of parameters space as the target model is proposed.

CS93 Room 12 ECONOMETRIC MODELLING AND APPLICATIONS Chair: Edoardo Otranto

C903: **A generalized maximum entropy approach to small area estimation**
Presenter: **Rosa Bernardini-Papalia**, University of Bologna, Spain
Co-authors: Esteban Fernandez-Vazquez
Small area estimation techniques are becoming increasingly used in survey applications to provide estimates for local areas of interest. These techniques combine direct survey estimates with auxiliary data to produce estimates for areas with very few sampled units. Traditionally, two types of small area estimation are considered: direct and indirect. Direct small area estimation is based on survey design. On the other hand, indirect approaches of small area estimation include synthetic and composite estimators, which are applied to obtain the empirical best linear unbiased prediction (EBLUP) estimators. These methodologies in small area estimation are usually heavily dependent on the assumptions related to the model and/or to the linearity of the estimators of the small area means or totals. Alternatively, an information theoretic (IT) approach can be used to produce estimates for small domains by combining different macro and micro data sources. In this paper, we suggest applying IT based methods to estimate small area economic indicators by taking advantage of sample and non-sample (aggregate) data. More specifically, we develop a Generalized Maximum Entropy (GME) estimator to be applied to direct and indirect small area estimation. The performance of the GME estimator is evaluated by means of numerical simulations.

C913: **Maximum likelihood estimation for linear mixed model with unobserved common correlated effects via EM algorithm**
Presenter: **Shou-Yung Yin**, National Central University, Taiwan
Co-authors: Chih-Chiang Hsu, Chang-Ching Lin
Classical linear mixed models often impose the assumptions that the errors are normally distributed and are cross-sectionally independent. However, they are unrealistic in many applications with panel data. A flexible linear mixed model is provided in which errors cannot only be characterized by a multivariate skew normal distribution, but also be cross-sectionally dependent by imposing common correlated effects. The EM algorithm is implemented after taking a suitable transformation. Our Monte Carlo simulation shows that the proposed method is quite accurate in the presence of common correlated effects, while conventional models without taking these effects into account can result in severely biased parameter estimates.

C1030: **Consistent estimation of pseudo panels in presence of selection bias**
Presenter: **Jhon James Mora**, Icesi, Colombia
Co-authors: Juan Muro
In the presence of selection bias the traditional estimator of pseudo panel data is inconsistent. We show appropriate conditions under which the consistency is achieved in pseudo panel and propose a simple test of selection bias. A Wald test under the null hypothesis of absence of sample selection bias is proposed. The test enables us to make a consistent estimation of the pseudo panel parameters when the null hypothesis is rejected.

C998: **Mathematical models in the Romanian annuity market and pension funds**
Presenter: **Mihaela Covrig**, The Bucharest University of Economic Studies, Romania
Co-authors: Iulian Mircea
Some models of mortality rates and models of longevity risk are discussed. In recent years, due to increased life expectancy and decrease in the number of employee taxpayers, major difficulties have occurred on the annuity market and pension funds. As a consequence, many companies have closed the retirement plans that they used to offer to their employees. In addition, the Romanian government, like other governments, increased the retirement age by two or five years to take into account longevity improvements, aging and the retirement funding. As members of some reference population live longer on average than anticipated, life insurance companies should revise or adjust their pricing and reserving calculations. We propose some mathematical models in forecasting mortality rates for the Romanian population, and also, we present some models for the securitization of longevity bonds or loans. We provide a numerical illustration of the above models.

CS95 Room 11 FACTOR MODELS Chair: Alain Kabundi

C293: **Using factor model recursive parameters for business cycle convergence analysis**
Presenter: **Sonia De Lucas-Santos**, Universidad Autonoma de Madrid, Spain
Co-authors: Maria Jesus Delgado-Rodriguez
Factor model offers a useful information to contribute to business cycle analysis by identifying countries sharing common factors, monitoring structural stability and convergence of the countries to the common cycle. Application to European Business Cycle will be shown to demonstrate the capabilities of the model.

C747: **A novel criterion in multivariate linear factor models selection**
Presenter: **Quang Phan**, University of Warwick, United Kingdom
Co-authors: Valentina Corradi
An alternative criterion for factor models selection, based on the degree of sparsity of the estimated errors covariance matrix, is introduced. This criterion can be used as an alternative for the well-known AIC and BIC but with some advantages. The idea is based on the fact that a good multivariate linear factor model will force the idiosyncratic errors to have small cross-section dependence, i.e. the errors covariance matrix must have many zero-entries. This type of many-zeros matrix is called sparse matrix. Therefore, in order to provide a specification test for a factor model, we will estimate the degree of sparsity of the errors covariance matrix. This degree of sparsity is obtained by specifying a threshold and then count the number of entries below the threshold in the matrix. The value of this thresholding value is discussed, alongside with some simulations and literature reviews.

C859: **Inference for stochastic dynamic factor models**
Presenter: **Geert Mesters**, VU University Amsterdam, Netherlands
Co-authors: Siem Jan Koopman
The dynamic factor model $y_{i,t} = \lambda_i' f_t + \omega_i \varepsilon_{i,t}$ is considered for the case where all its components are modeled by independent latent Gaussian processes. For example, we may model f_t and $\varepsilon_{i,t}$ by auto-regressive processes and λ_i and ω_i by diffuse Gaussian densities. For this stochastic formulation of the dynamic factor model we present new filtering and smoothing recursions. The derivation of the new recursions is similar compared to the derivation of the standard Kalman filter and smoothing recursions, but depends on a novel decomposition of the prediction error variance, which enables us to separate the updating for the factors, loadings, scalings and disturbances. The resulting smooth estimates are shown to be the minimum variance linear unbiased estimates. Further, the stochastic formulation of the model ensures that the number of underlying parameters is low and independent from the number of time series considered. These parameters are estimated by maximum likelihood either via; the exact marginal likelihood which is computed from the prediction error decomposition provided by the filtering recursions, or by the EM algorithm, which depends on the expected complete likelihood that is evaluated by the smoothing recursions. The estimation methodology is easily extended to allow for missing values, regression effects, lagged dependent variables and time-varying factor loadings. We illustrate the methods by a large Monte Carlo study and two empirical applications, for interest rates and economic growth, respectively.

C935: **Efficient estimation of high dimensional factor models under cross sectional dependence**
Presenter: **Rachida Ouysse**, University of New South Wales, Australia

Efficient estimation of factor models is attracting considerable attention because recent empirical evidence suggests the estimates are adversely affected by the inability to account for the cross sectional dynamics. Let X_t be a N-dimensional vector of stationary variables observed at time $t = 1, \cdots, T$ such that $X_t = \Lambda F_t + \varepsilon_t$, where the common factors F_t and their loadings Λ are unobserved. A factor structure is approximate when the idiosyncratic errors ε_t are weakly correlated across the variables. Principal components analysis (PCA) provides consistent estimation of the factor structure and efficiency can be achieved using robust econometric tools such as generalized PCA and quasi maximum likelihood. However when $N > T$, the sample covariance matrix is singular and accounting for cross-sectional dynamics is challenging without imposing a structure on these dynamics. Instead we use the approximate structure assumption of bounded $\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}|E(\varepsilon_{it}\varepsilon_{jt})|$, as a constraint in the PCA framework. Our penalized PCA can be interpreted as a shrinkage regression, where the off-diagonal elements of the covariance matrix are shrunk towards zero as N grows large. We show that our estimators are consistent and more efficient thanPCA. Furthermore, simulation experiments show that our approach compares well with other alternatives that make use of a known covariance structure.

CS86 Room 10 VOLATILITY MODELS Chair: Esther Ruiz

C602: **Encompassing asymmetric stochastic volatility models in just one model**
Presenter: **Xiuping Mao** , Universidad Carlos III, Spain
Co-authors: Esther Ruiz Ortega, Maria Helena Veiga

A new specification that encompasses three previous asymmetric stochastic volatility models is proposed. We derive and compare the properties of each of the particular cases of the new model. As the model based on adding a noise to the EGARCH specification encompasses the model based on adding lagged returns to the log-volatility equation, both models share similar properties with the former having the possibility of generating returns distributed with larger kurtosis and stronger dependencies in power-transformed absolute returns. We also show that the constant term in the Threshold SV model is crucial for representing the leverage effect. Consequently, we derive its statistical properties when the persistence and the variance of the log-volatility noise are fixed and the constant changes according to the sign of one-lagged past returns. For the parameter values often encountered in practice, the Threshold SV model is able to capture higher leverage effect than the stochastic EGARCH model, while the latter is more flexible in representing volatility clustering. Consequently, the new model proposed encompasses the EGARCH and Threshold SV specifications. All the results are illustrated with simulated and real time series of financial returns.

C694: **Time varying volatility, default and the sovereign risk premium**
Presenter: **Hernan Seoane**, Universidad Carlos III de Madrid, Spain

The aim is to study the effects of changes in the macroeconomic volatility on the sovereign risk premium, default probability and external financing conditions. A model of endogenous sovereign risk premium for a small open economy is proposed. Domestic output is exogenously determined and follows a first order autoregressive process with stochastic volatility. Using Particle Filter the output process for Spain for the period 1970 to 2011 is estimated. The estimated parameters are fed to the model and the remaining parameters are calibrated to standard values in the literature. Changes in the macroeconomic volatility are found to have a large impact on the saving and default decisions as well as on sovereign risk premium, price of debt and default probability. An increase in volatility might induce an increase in savings. However, when there is a substantial default probability and a low cost of default, an increase in volatility might generate a decrease in savings, suggesting that in this scenario the default option is more attractive. An increase in volatility is then associated to an increase in default probability and hence, induce a lower price of debt and a higher risk premium. This mechanism rationalizes the positive correlation between risk premium and volatility, observed in the data.

C713: **Stochastic and deterministic filtering for stochastic volatility models**
Presenter: **Constantin Weiser**, Johannes Gutenberg University Mainz, Germany
Co-authors: Florian Heiss

Stochastic Volatility models (SVM) have become a widely used tool for analyzing financial time series. They are able to capture flexible heteroscedasticity patterns inherent in those data. The estimation of SVM by maximum likelihood or similar methods is difficult because they do not have an analytically tractable likelihood function. Approximations using stochastic filters based on particle filters have been suggested and successfully applied. They suffer from two potential problems: While the likelihood function can be approximated arbitrarily accurately, a close approximation can be computationally very demanding. And the approximated likelihood function is not smooth in the estimated parameters, impeding its numerical maximization. As an alternative, we suggest specific deterministic filters based on numerical integration methods to solve these problems and facilitate the use of SVM in practice. Monte Carlo simulations and empirical examples indeed show an impressive performance of the suggested algorithm which turns out to be more accurate than particle filters by orders of magnitude and gives a smooth approximation of the likelihood function.

C715: **On the need of intra-daily data to forecast daily volatility**
Presenter: **Denisa Georgiana Banulescu**, University of Orleans, France
Co-authors: Bertrand Candelon, Christophe Hurlin, Sebastien Laurent

Considering mixed data sampling (MIDAS) regressions, we analyze the influence of the sampling frequency of intra-daily predictors on the accuracy of the volatility forecasts. We propose various in-sample and out-of-sample comparisons of daily, weekly and bi-weekly volatility forecasts issued from MIDAS regressions based on intra-daily regressors sampled at different frequencies. First, we show that increasing the frequency of the regressors improves the forecasting abilities of the MIDAS model. In other words, using regressors sampled at 5 minutes gives more accurate forecasts than using regressors sampled at 10 minutes, etc. These results are robust to the choice of the loss function (MSE, Qlike, etc.) and to the choice of the forecasting horizon. Third, the MIDAS regressions with high-frequency regressors (sampled between 5 minutes and 30 minutes) provide more accurate in-sample forecasts than a GARCH model based on daily data. However, except the one-period-ahead forecasts of the calm period, the out-of-sample forecasts of MIDAS models during the crisis period are not significantly different from the GARCH forecasts, whatever the sampling frequency used, confirming that the direct use of high-frequency data does not necessarily improve volatility predictions.

CS71 Room 6 SOLVING INTERNATIONAL PORTFOLIO MODELS Chair: Michel Juillard

C170: **Computing the risky steady state in DSGE models**
Presenter: **Oliver de Groot**, University of Cambridge, United Kingdom

A simple procedure for solving the risky steady state in medium-scale macroeconomic models is described. The risky steady state is the "point where agents choose to stay at a given date if they expect future risk and if the realization of shocks is 0 at this date". This new procedure is a direct method which makes use of a second-order approximation of the macroeconomic model around its deterministic steady state, thus avoiding the need to employ an iterative algorithm to solve a fixed point problem. The paper uses an example to show the computation time and accuracy of the approximation.

C549: **Solving portfolio models with a perturbation approach**
Presenter: **Michel Juillard**, Bank of France, France

A previously proposed method to handle the singularity of portfolio models is discussed in the framework of the perturbation approach. We compare this method with an approximation around the risky steady state and discuss the consequences of various methodological choices in the implementation of the method.

C385: **Comparison of solutions to dynamic general equilibrium model with portfolio choice**
Presenter: **Ludovic Gauvin**, EconomiX-University of Paris West and Banque de France, France

In recent years, a vast literature concerning general equilibrium models with dynamic portfolio choice has emerged. In general, these studies have employed three main methodologies for their resolution. On the one hand, perturbation methods are particularly fast, yet suffer increasing approximation errors as they move away from the steady-state. Furthermore, the nonstochastic steady-state and the first-order approximation of portfolio components may be difficult to determine with this technique. On the other hand, projection methods are much more accurate but may suffer the so-called curse of dimensionality. Thirdly, hybrid methods, which are a combination of projection and perturbation methods, have also been proposed. The aim is to compare the accuracy of some existing solutions for general equilibrium models with dynamic portfolio choice and clarify the advantages and drawbacks of each method for this specific type of model.

CS28 Room 1 STRUCTURAL SYSTEMS: DIMENSIONALITY AND IDENTIFICATION Chair: John Galbraith

C1005: **Confidence sets for ratios in dynamic panels: persistence and identification**
Presenter: **Marcel Voia**, Carleton University, Canada
Co-authors: Lynda Khalaf, Jean-Thomas Bernard

The problem of estimating parameter ratios frequently arises in statistics and economics. Elasticities and/or curve tipping points are common examples of interest. We focus on possibly persistent dynamic panels and compare the Delta to the Fieller method, based on previous procedures. The Fieller method is expected to outperform the Delta method when the ratio is potentially weakly-identified. Whether the latter holds in dynamic panels or whether any of these methods works in this context is an open question, particularly with persistent data. An extensive simulation study is conducted to assess these questions, using an empirically relevant design, various combinations of the time series and cross-sectional dimensions, and mild to severe persistence schemes. Results confirm that the Fieller method based on one of the previous procedures works remarkably well even in near-unit root situations and relatively small samples.

C979: **Dynamic panels with MIDAS covariates: Estimation and fit**
Presenter: **Charles Saunders**, Carleton University, Canada
Co-authors: Lynda Khalaf, Maral Kichian, Marcel Voia

Higher frequency data has become more readily available, but not for all series of interest many of which remain at lower frequencies. The introduction of Mixed Data Sampling (MIDAS) time series methods has allowed researchers to efficiently exploit information at available frequencies. MIDAS methods so far have not formally been extended to the panel context. Because of a dynamic context by construction, available procedures in time series do not readily extend to panels. We introduce MIDAS to panel regression models suitable for analysis with GMM methods of previous forms. Because MIDAS specification tests are lacking even in time series contexts, we propose inference methods that statistically embed specification checks, as has recently been popularized in the weak-IV literature. We focus on confidence set estimation for the MIDAS parameters for which empty outcomes signal lack of fit. Proposed confidence sets invert a model specification test which we carefully design to adequately track the lag structure underlying GMM despite mixing time frequencies. The underlying statistics are (asymptotically, under standard regularity conditions) pivotal regardless of the considered data frequency. A simulation study illustrates promising size and power properties for our procedures.

C183: **Forecasting financial volatility with QML and LAD-ARCH estimators of the GARCH model**
Presenter: **John Galbraith**, McGill University, Canada
Co-authors: Liam Cheung

GARCH models and their variants are usually estimated using quasi-Maximum Likelihood. Recent work has shown that by using estimates of quadratic variation, for example from the daily realized volatility, it is possible to estimate these models in a different way which incorporates the additional information. Theory suggests that as the precision of estimates of daily quadratic variation improves, such estimates (via LAD- ARCH approximation) should come to equal and eventually dominate the QML estimators. The present paper investigates this using a five-year sample of data on returns from all 466 S&P 500 stocks which were present in the index continuously throughout the period. The results suggest that LAD-ARCH estimates, using realized volatility on five-minute returns over the trading day, yield measures of 1-step forecast accuracy comparable to those obtained from QML estimates. Combining the two estimators appears to produce a clear improvement in forecast accuracy relative to either of the two different forecasting methods alone.

Saturday 01.12.2012 17:40 - 19:20 Parallel Session H – ERCIM

ES04 Room 9 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING I **Chair: Igor Pruenster**

E408: **Semiparametric posterior limits under local asymptotic exponentiality**
Presenter: **Bartek Knapik**, CREST - Universite Paris Dauphine, France
Co-authors: Bas Kleijn

The asymptotic behavior of the marginal posterior for the parameter of interest in nonregular semiparametric estimation problems is investigated. It is known that there are few nonregular cases for which the posterior distributions converge. A crucial tool in studying the asymptotic behavior of the posterior distribution is likelihood expansion. We focus on nonregular models that exhibit a so-called local asymptotic exponentiality (LAE). This setting covers important examples such as densities with jumps, uniform distributions, etc. We show that under certain straightforward and interpretable conditions, the marginal posterior for the parameter of interest converges in total variation to a negative exponential distribution, with parameters present also in the LAE expansion. In addition to the local asymptotic exponentiality, the model is required to satisfy metric entropy conditions, while the nuisance prior must assign non-zero mass to certain Kullback-Leibler neighborhoods. In addition, the marginal posterior is required to converge at parametric rate (i.e. when the nuisance is fixed). The results are applied to estimation of location in a model, where underlying densities possess a jump at zero, with a Gaussian prior for the nuisance.

E397: **Finite sample posterior concentration in high-dimensional regression**
Presenter: **Nathaniel Strawn**, Duke University, United States
Co-authors: Artin Armagan, David Dunson, Rayan Saab, Lawrence Carin

Linear models with Gaussian noise and the number of samples much less than the number of predictors ($n \ll p$) are considered. In this regime, frequentist estimators have been studied extensively, but the posterior is often an afterthought. Leveraging metric geometry and nonparametric Bayesian techniques, we exhibit universal finite undersampling concentration bounds of Bayesian posteriors for linear regression. The form of these bounds makes it clear that only priors with sufficient concentration on the compressible vectors enjoy strong contraction around the true parameter in this regime. Based upon these bounds, we also obtain asymptotic contraction rates for well-structured priors.

E152: **Bayesian multivariate mixed-scale density estimation**
Presenter: **Antonio Canale**, University of Turin, Italy
Co-authors: David Dunson, Pierpaolo De Blasi

Although univariate continuous density estimation has received abundant attention in the Bayesian nonparametrics literature, there is essentially no theory on multivariate mixed scale density estimation. We consider a general framework to jointly model continuous, count and categorical variables, under a nonparametric prior, which is induced through rounding latent variables having an unknown density with respect to Lesbesgue measure. For the proposed class of priors, we provide sufficient conditions for large support, strong consistency and rates of posterior contraction. These conditions are primarily related to the prior on the latent variable density and heaviness of the tails for the observed continuous variables. They allow one to convert sufficient conditions obtained in the setting of multivariate continuous density estimation to the mixed scale case. We provide new results in the multivariate continuous density estimation case, showing the Kullback-Leibler property and strong consistency for different mixture priors including priors that parsimoniously model the covariance in a multivariate Gaussian mixture via a sparse factor model. In particular, the results hold for Dirichlet process location and location-scale mixtures of multivariate Gaussians with various prior specifications on the covariance matrix.

E240: **Species sampling models: Consistency for the number of species**
Presenter: **Pier Giovanni Bissiri**, Universita degli Studi di Milano-Bicocca, Italy
Co-authors: Andrea Ongaro, Stephen Walker

Recent focus has been on Bayesian nonparametric models for approaching species sampling problems and in particular on Gibbs-type priors. A discrete random measure can have a countable infinite number of atoms and, for the class of model considered, is associated with a parameter $\sigma \in (0,1)$. On the other hand, we can also consider a random measure generated with K atoms and allow K to be arbitrarily large. This is associated with the same model as the countable infinite case but with a key difference that $\sigma < 0$. It is the latter class of model we consider here, due to the existence and interpretation of K as the number of species. In particular, we provide a result which demonstrates the consistency of the posterior distribution of K as the sample size increases.

ES05 Room 5 OPTIMIZATION HEURISTICS IN ESTIMATION AND MODELLING **Chair: Manfred Gilli**

E178: **Optimised U-type designs on flexible regions: An implementation for higher dimensions**
Presenter: **Christopher Sharpe**, Justus Liebig University Giessen, Germany
Co-authors: Peter Winker

It has been demonstrated that optimised U-type designs with a low discrepancy can be generated for adjustable, symmetrical input spaces named flexible regions. These are comprised of convex and non-convex sets, a departure from the standard rectangular input space. In addition, the Centered Composite Discrepancy was shown to be a suitable discrepancy measure for such regions. A first implementation was limited to the calculation of regions in 2-dimensions only, which prevents it from being applied to create designs for real world experiments. A new implementation allows the Centered Composite Discrepancy to be calculated efficiently for flexible regions in higher dimensions. Results for optimised U-type designs are presented for 3-dimensional regions, with the heuristic optimisation technique threshold accepting used for design optimisation.

E233: **Vector STAR model: In search of reliable starting values**
Presenter: **Frauke Schleer**, Centre for European Economic Research, Germany

Within the framework of multivariate non-linear Vector STAR models, optimization problems related to flat or non-convex, thus containing multiple local minima, likelihood functions may arise. Hence, in empirical applications derivative-based optimization algorithms may either converge slowly or to a local optimum which is not globally optimal. The outcomes of derivative-based methods crucially rely on the starting values. Initializing the algorithm with starting values close to the global optimum helps the algorithm to converge to the nearest, global optimum. Before applying a Vector STAR model to economic problems aiming at interpreting economic processes properly, one has to ensure that the respective numerical algorithm is provided with appropriate starting values. The problem of finding good starting values is considered. Different procedures for finding starting values are evaluated and their performance is assessed with respect to model fit and computational effort. We employ i) a grid search algorithm, ii) heuristic optimization procedures and iii) Monte Carlo Markov Chain methods. It is expected that the inherent stochastics and hill-climbing features of heuristic optimization procedures may deliver advantages in terms of efficiency, speed and extent to which the surface area can be explored, resulting in superior starting values.

E617: **Robust estimation of shape constrained state price density surfaces**
Presenter: **Markus Ludwig**, University of Zurich, Switzerland

In order to better capture empirical phenomena, research on option price and implied volatility modeling increasingly advocates the use of non-parametric methods over simple functional forms. This, however, comes at a price, since they require dense observations to yield sensible results. Calibration is therefore typically performed using aggregate data. Ironically, the use of time-series data in turn limits the accuracy with which current observations can be modeled. We propose a novel approach that enables the use of flexible functional forms using daily data alone. The resulting estimators yield excellent fits and generalize well beyond available data, all the while respecting theory imposed shape constraints. We demonstrate the numerical stability and the pricing performance of our method by approximating arbitrage-free implied volatility, price and state price density surfaces from S&P 500 options over a period of 12 years.

E324: **A simple method for constructing joint confidence bands for impulse response functions**
Presenter: **Peter Winker**, University of Giessen, Germany
Co-authors: Helmut Luetkepohl, Anna Staszewska-Bystrova

In vector autoregressive analysis typically confidence intervals for individual impulse responses are reported to indicate the sampling uncertainty in the estimation results. In this note a simple method is proposed for constructing joint confidence bands, given a prespecified coverage level for the impulse responses at all horizons considered. Although the method may not produce bands of optimal size, it is demonstrated in a simulation comparison that it is competitive with other methods that have been proposed for this purpose including a method based on heuristic optimization.

ES11 Room 12 SPARSE DIMENSION REDUCTION Chair: Nickolay T. Trendafilov

E414: **Penalized likelihood factor analysis via non-convex penalties**
Presenter: **Kei Hirose**, Osaka University, Japan
Co-authors: Michio Yamamoto

In exploratory factor analysis, the model is usually estimated by the maximum likelihood method, and then a rotation technique such as the varimax method is utilized to find a meaningful relationship between the observable variables and the common factors. However, it is well known that the maximum likelihood method often yields unstable estimates because of overparametrization when the number of variables is large. In such cases, a penalized likelihood procedure that can achieve sparse models may produce more stable estimates and also can be applied to high dimensional data. We introduce a penalized likelihood factor analysis via non-convex penalties. In order to compute the entire solution path, a new efficient algorithm via the EM algorithm along with coordinate descent is proposed. Furthermore, the proposed procedure allows us to analyze high-dimensional data such as gene expression data. We introduce a new graphical tool that outputs path diagram, goodness-of-fit indices, and model selection criteria simultaneously, which is helpful in finding a suitable value of the regularization parameter.

E391: **A generalisation of sparse PCA to multiple correspondence analysis**
Presenter: **Gilbert Saporta**, CNAM, France
Co-authors: Anne Bernard, Christiane Guinot

Principal components analysis (PCA) for numerical variables and multiple correspondence analysis (MCA) for categorical variables are well-known dimension reduction techniques. PCA and MCA provide a small number of informative dimensions: the components. However, these components are a combination of all original variables, hence some difficulties in the interpretation. Factor rotation (varimax, quartimax etc.) has a long history in factor analysis for obtaining simple structure, ie looking for combinations with a large number of coefficients either close to zero or to 1 or -1. Only recently, rotations have been used in Multiple Correspondence Analysis. Sparse PCA and group sparse PCA are new techniques providing components which are combinations of few original variables: rewriting PCA as a regression problem, null loadings are obtained by imposing the lasso (or similar) constraint on the regression coefficients. When the data matrix has a natural block structure, group sparse PCA give zero coefficients to entire blocks of variables. Since MCA is a special kind of PCA with blocks of indicator variables, we define sparse MCA as an extension of group sparse PCA. We present an application of sparse MCA to genetic data (640 SNP's with 3 categories measured on 502 women) and a comparison between sparse and rotated components.

E332: **Orthogonal factor analysis subject to direct sparseness constraint on loadings**
Presenter: **Kohei Adachi**, Osaka University, Japan
Co-authors: Nickolay Trendafilov

A sparse factor analysis (FA) procedure is proposed for obtaining the optimal orthogonal factor solutions subject to constraints which directly requires a specified number of factor loadings to be exactly zeros with no restriction on their locations. The proposed sparse FA considerably differs in two ways from the existing sparse principal component analysis (PCA) procedures developed in the last decade. First, in FA, unique variances are estimated that are not considered in PCA. Second, the FA sparseness constraint is integrated into the convergence procedure of the FA solution, which is obtained by minimizing the standard FA loss function using an alternating least squares algorithm. This is in contrast to the available PCA procedures where sparseness is achieved by adding penalty functions to the original PCA loss function. Further, a sparseness selection procedure is presented based on information criteria. The proposed procedure is evaluated in a simulation study, and illustrated with well known data sets.

E164: **From simple structure to sparse components: a comparative introduction**
Presenter: **Nickolay Trendafilov**, Open University, United Kingdom

First, the main approaches to interpret the results from principal component analysis (PCA) during the last 50–60 years are briefly reviewed. The simple structure approach is compared to the modern approach of sparse PCA where interpretable solutions are directly obtained. It is shown that their goals are identical but they differ by the way they are realized. Then, the most popular and influential methods for sparse PCA are listed. They produce either only orthonormal loadings, or only uncorrelated components, or, most frequently, neither of them. To overcome this weakness, a new approach to define sparse PCA is introduced. In contrast to the existing methods, it makes it possible to achieve simultaneously nearly uncorrelated sparse components with nearly orthonormal loadings. Several alternative definitions are considered and illustrated on a well-known data set. Finally, it is demonstrated, how one of these possible versions of sparse PCA can be used as a sparse alternative to the classical rotation methods.

ES29 Room 13 ADVANCES IN ROBUST DATA ANALYSIS I Chair: Alfonso Gordaliza

E228: **Choice of the tuning constant in robust inference for generalized linear models and generalized additive models**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain

The usual trade-off between robustness and power is reflected in the tuning constant c if we make robust inference based on Huber M-estimators. The choice of c could be very arbitrary and not reflecting the reality, in the case of Generalized Linear Models or Generalized Additive Models based on Huber M-estimators.We apply the ideas of a previous work on a way to choose from different test results based on the $p-$value line, to choose the value of c in this type of model.

E389: **Robust segmentation through Hotelling T^2 test**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Andrea Cerioli, Anthony Atkinson
A form of robust segmentation using the forward search that allows the data to determine the number of clusters is described. Similarly to other existing robust methods of cluster analysis we also allow for outliers and so do not force all units to be clustered. Given that our analysis relies on forward plots of particular statistics, in order to provide sensitive inferences about the existence of clusters, it is necessary to augment such graphs with envelopes of the distributions of the statistics being plotted. We focus on a sort of Hotelling t-test where the two samples are made up respectively of the units which are inside and outside the subset at step m of the forward search. This allows for a robust split of each tentative group in two separate subgroups which could be useful also in hierarchical divisive clustering.

E527: **Constrained EM algorithms for Gaussian mixtures of factor analyzers**
Presenter: **Salvatore Ingrassia**, University of Catania, Italy
Co-authors: Francesca Greselin
Mixtures of factor analyzers are becoming more and more popular in the area of model based clustering of high-dimensional data. In data modeling, according to the likelihood approach, it is well known that the loglikelihood function may present spurious maxima and singularities and this is due to specific patterns of the estimated covariance structure. To reduce such drawbacks, we propose a constrained procedure which maintains the eigenvalues of the covariance structure into suitable ranges. Applications of this approach in robust clustering is then outlined.

E563: **Robust conditional tests for elliptical symmetry**
Presenter: **Isabel Rodrigues**, Technical University of Lisbon and CEMAT, Portugal
Co-authors: Ana Bianco, Graciela Boente
A robust procedure for conditional testing the hypothesis of elliptical symmetry is presented. That also provides a specialized test of multivariate normality. We derive the asymptotic behaviour of the robust conditional testing procedure under the null hypothesis of multivariate normality and under an arbitrary elliptically symmetric distribution. Their influence function is also studied. A simulation study allows us to compare the behaviour of the classical and robust test, under different contamination schemes.

ES01 Room 2 MATRIX COMPUTATIONS AND STATISTICS Chair: Ahmed Sameh

E1060: **On some matrix based methods for analyzing networks and text**
Presenter: **E. Gallopoulos**, University of Patras, Greece
Co-authors: V. Kalantzis, E.-M. Kontopoulou, C. Bekas
Matrix based methods in areas such as sociometrics and text mining has been established for some time but in recent years the field has been growing with great strides in order to cope with the vast increase in the scale of social and other networks and the massive amounts of text available over the Web by attempting the best possible use of the capabilities of modern computer systems. We will present some of our work in network and text mining where matrix iterative methods and statistics meet. Specifically, we will show methods for estimating graph centrality suitable for large networks. We will also show some new developments on a software environment, the Text to Matrix Generator MATLAB Toolbox (TMG), for dimensionality reduction and its application in text mining.

E865: **Window estimation of the general linear model with singular dispersion matrix**
Presenter: **Stella Hadjiantoni**, Queen Mary, University of London, United Kingdom
Co-authors: Erricos John Kontoghiorghes
The re-estimation of the least squares solution after observations are deleted, known as downdating, is often required. The problem of deleting observations from the general linear model (GLM) with a possible singular dispersion matrix is considered. A new method to estimate the downdated GLM, updates the original model with the imaginary deleted observations. This results in a non-positive definite dispersion matrix which comprises complex covariance values. For the case of a non singular dispersion matrix, this updated GLM has been proved to give the same normal equations than the downdated model. The proposed updated GLM is formulated as a generalized linear least squares problem (GLLSP). The main computational tool is the generalized QR decomposition (GQRD) which is employed based on hyperbolic Householder transformations; however, no complex arithmetic is used in practice. The special structure of the matrices and previous computations are efficiently utilized.

E1046: **Computational strategies for non-negativity model selection**
Presenter: **Cristian Gatu**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Erricos John Kontoghiorghes
The problem of regression subset selection under the condition of non-negative coefficients is considered. The straightforward solution would be to estimate the corresponding non-negative least squares of all possible submodels and select the best one. A new computationally efficient procedure which computes only unconstrained least squares is proposed. It is based on an alternative approach to quadratic programming that derives the non-negative least squares by solving the normal equations for a number of unrestricted least squares subproblems. The algorithm generates a combinatorial tree structure that embeds all possible submodels. This innovative approach is computationally superior to the straightforward method. Specifically, it reduces the double exponential complexity to a a single traversal of a tree structure. The computational efficiency of the new selection strategy is further improved by adopting a branch-and-bound device that prunes non-optimal subtrees while searching for the best submodels. The branch-and-bound algorithm is illustrated with a real dataset. Experimental results on artificial random datasets confirm the computational efficacy of the new strategy and demonstrates its ability to solve large model selection problems that are subject to non-negativity constrains.

E1047: **A GSVD strategy for estimating the SEM and SURE models**
Presenter: **Mircea Ioan Cosbuc**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Cristian Gatu, Erricos John Kontoghiorghes
Computational strategies for estimating the Simultaneous Equations Model and the Seemingly Unrelated Regression Equations are proposed. The methods use the Generalized Singular Value Decomposition as the main numerical tool. They exploit the block diagonal and banded structure of the matrices involved in the factorization and do not presume that the covariance matrices are nonsingular. An efficient implementation of these methods is described and its performance is compared to classical methods that treat the models as Generalized Linear Models. Not only are the procedures faster for large systems, but they also preserve a banded structure of the data matrices, allowing the efficient storage of the sparse matrices involved. The algorithms make use of the LAPACK and BLAS functions through their C interfaces.

ES31 Room Multipurpose STATISTICS IN FUNCTIONAL AND HILBERT SPACES I Chair: Gil Gonzalez-Rodriguez

E374: **Introducing a roughness penalty in the functional PLS regression model**
Presenter: **Ana M. Aguilera**, University of Granada, Spain
Co-authors: M. Carmen Aguilera-Morillo, Cristian Preda
Partial least squares (PLS) regression was recently extended to the functional domain for estimating a scalar response variable in terms of a functional predictor variable having as realizations square integrable curves observed at a finite set of points. When the sample curves are smooth but the observed records are noisy, the estimated functional parameter is affected by high variability, and some type of smoothing is necessary. The functional partial least squares components are generalized linear combinations of the functional predictor that have maximum covariance with the response variable. A new inner product that takes into account the roughness of the PLS weight functions is considered in the model formulation. By using basis expansions of the sample curves and the weight functions, this smoothed version of the functional PLS is proved to be equivalent to the multivariate PLS regression of the response on a transformation of the vector of basis coefficients associated with the sample curves. A simulation study is developed to test the good performance of the proposed method with respect to non penalized functional PLS.

E478: **Estimating surfaces and spatial fields via regression models with differential regularization**
Presenter: **Laura M. Sangalli**, Politecnico di Milano, Italy
Co-authors: James Ramsay
Interfacing statistical methodology and numerical analysis techniques, we propose regression models with partial differential regularization, that accurately estimate surfaces and spatial fields. In particular, the proposed models are able to deal with data scattered over complex bi-dimensional domains, including domains with irregular shapes and holes; they allow for spatially distributed covariate information and can impose various conditions over the boundaries of the domain. Accurate surface estimation is achieved resorting to finite elements and full uncertainty quantification is provided via inferential tools. Important extensions of the proposed models include the possibility to deal with data distributed over non-planar domains and to incorporate a priori information about the spatial structure of the phenomenon under study.

E649: **Time-varying additive model for longitudinal data**
Presenter: **Jane-Ling Wang**, University of California Davis, United States
Co-authors: Xiaoke Zhang, Byeong Park
Additive model is an effective dimension reduction approach that provides flexibility to model the relation between a response variable and key covariates. The literature is largely developed for scalar response and vector covariates. More complex data is of interest, where both the response and covariates may be functions. A functional additive model is proposed together with a new smooth backfitting algorithm to estimate the unknown regression functions, whose components are time-dependent additive functions of the covariates. Due to the sampling plan, such functional data may not be completely observed as measurements may only be collected intermittently at discrete time points. We develop a unified platform and efficient approach that can cover both dense and sparse functional data and the needed theory for statistical inference. The oracle properties of the component functions are also established.

E880: **Tools for checking bootstrap consistency in separable Hilbert spaces**
Presenter: **Gil Gonzalez-Rodriguez**, University of Oviedo, Spain
In inferential studies, and particularly for high dimensional or Hilbert valued random elements, the asymptotic distributions are not always useful in practice, as they are either partially unknown or not close enough to the sample ones. In these cases, bootstrap is usually a good option to approximate the corresponding sampling distributions. A way of checking the bootstrap consistency in the framework of separable Hilbert spaces through its connection with bootstrap for general empirical process will be analyzed. Such a connection will allow us to transfer many bootstrap approaches studied in the general empirical process context to the case of Hilbert space valued data. As an example, Efron's bootstrap and the exchangeable weighted bootstrap are considered. The usefulness of this technique will be illustrated in several situations including the ANOVA test, the developing of confidence bands or test for linearity.

ES40 Room 4 ADVANCED TIME SERIES ANALYSIS Chair: Alessandra Luati

E902: **Spectral filtering for trend estimation**
Presenter: **Andrea Martinelli**, University of Insubria, Italy
Co-authors: Marco Donatelli, Alessandra Luati
Trend estimation at the boundaries of a time series by means of smoothing methods is considered. After deriving the asymptotic properties of sequences of matrices associated with linear smoothers, two classes of asymmetric filters that approximate a given symmetric estimator are introduced: the reflective filters and antireflective filters. The associated smoothing matrices, though non-symmetric, have analytically known spectral decomposition. The paper analyses the properties of the new filters and considers reflective and antireflective algebras for approximating the eigensystems of time series smoothing matrices. Based on this, a thresholding strategy for a spectral filter design is discussed.

E605: **Locally stationary latent factors**
Presenter: **Giovanni Motta**, Columbia University, United States
Co-authors: Michael Eichler
Current approaches for fitting dynamic factor models to multivariate time series are based on the principal components of the spectral matrix. These approaches rely on the assumption that the underlying process is temporally stationary which appears to be restrictive because, over long time periods, the parameters are highly unlikely to remain constant. A more general model recently introduced allows the spectral matrix to be smoothly time-varying, which imposes very little structure on the moments of the underlying process. However, the estimation becomes fully non-parametric and delivers time-varying filters that are high-dimensional and two-sided. Moreover, the estimation of the spectral matrix strongly depends on the chosen bandwidths for smoothing over frequency and time. As an alternative, we propose a semi-parametric approach in which only part of the model is allowed to be time-varying. More precisely, the latent factors admit a dynamic representation with time-varying autoregressive coefficients while the loadings are constant over time. Estimation of the model parameters is accomplished by application of the EM algorithm and the Kalman filter. The time-varying parameters are modeled locally by polynomials and estimated by maximizing the likelihood locally. Simulation results show that compared to estimation by principal components, our approach produces superior results in particular for small cross-sectional dimensions. We illustrate the performance of our approach through applications to real data.

E909: **Filtering with heavy tails**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Andrew Harvey
An unobserved components model in which the signal is buried in noise that is non-Gaussian may throw up observations that, when judged by the Gaussian yardstick, are outliers. We develop an observation driven model, based on a conditional Student $t-$distribution, that is tractable and retains some of the desirable features of the linear Gaussian model. Letting the dynamics be driven by the score of the conditional distribution leads

to a specification that is not only easy to implement, but which also facilitates the development of a comprehensive and relatively straightforward theory for the asymptotic distribution of the maximum likelihood estimator.

ES47 Room 6 SMALL AREA ESTIMATION **Chair: Domingo Morales**

E221: **An application of small domain estimation in surveys of SMEs**
Presenter: **Tomasz Jurkiewicz**, University of Gdansk, Poland

Small domain estimation techniques give the researcher various opportunities of improving the quality of inference. However, applications of the vast majority of those techniques, because of their complexity, require good statistical knowledge and an adequate computer software. Another practical challenge is the necessity of gathering and selecting auxiliary information that is incorporated in the process of estimation. Market research institutes tend to exhibit a shortage of time and suffer from the lack of expertise in the area of statistics in order to use small area estimation efficiently. The purpose of the paper is to evaluate methods and techniques which under the constraints mentioned above can be used in practice, and indicate possible sources of auxiliary data. As an example discussed in the paper, a survey aimed at assessing various financial aspects of the activity of SMEs in Poland is presented.

E416: **Small area estimates of counts using M-quantile Poisson regression models**
Presenter: **Maria Giovanna Ranalli**, University of Perugia, Italy
Co-authors: Nicola Salvati, Nikos Tzavidis, Emanuela Dreassi

M-quantile regression models have recently gained much attention in small area estimation representing a robust and flexible alternative to the widespread use of random effects models. However, since quantiles, and more generally M-quantiles, are only uniquely defined for continuous variables, M-quantile models have been applied up to now only when the variable of interest is continuous. We propose a new approach to small area estimation for count data based on a Poisson-like M-quantile model. Parameter estimates are obtained by suitably extending a robust version of estimating equations for generalized linear models to handle M-quantile regression. A measure of accuracy of the proposed small area estimators is provided that uses nonparametric bootstrap. The proposed methodology is compared with classical small area estimation methods based on generalized linear mixed models via simulation studies. The study is motivated by the need to obtain estimates from the Italian National survey on Health Conditions and Appeal to Medicare for the average number of visits to physicians. Small area estimation techniques are needed because the survey is designed to provide reliable estimates at the level of Administrative Regions - NUTS2 level – while estimates are sought at the level of Health Districts (partition of the Administrative Regions).

E502: **Impact of design on quality of small area estimators**
Presenter: **Thomas Zimmermann**, University of Trier, Germany
Co-authors: Jan Pablo Burgard, Ralf Muennich

It is now well understood that the interplay of statistical modelling and survey weighting should be taken into account when selecting small area estimators. Additionally, special attention has to be paid on the interaction between sampling designs and quality measures of small area estimates. Understanding these interactions furnishes a better selection of estimation methods, especially for data producers such as National Statistical Institutes. For business statistics, some central quality requirements to be achieved are stated in the EU Structural Business Survey Regulation 58/97. The present study focuses on the impact of different sampling designs on quality measures of small area estimators. The evaluation problem is tackled from a decision-theoretic perspective allowing for a balanced consideration between the needs of reliable small area estimates and efficient estimates on national level. A design-based simulation study illustrates the impact of design-effects and the variability of design weights on the quality of model versus design-based estimation methods. This study is based on a close to reality dataset generated from Italian business data.

ES49 Room 1 STATISTICAL INFERENCE BASED ON DEPTH **Chair: Davy Paindaveine**

E257: **DDalpha-classification with the random Tukey depth**
Presenter: **Tatjana Lange**, University of Merseburg , Germany
Co-authors: Karl Mosler, Pavlo Mozharovskyi

The DDα-procedure is presented to solve the problem of classifying d-dimensional objects into $k \geq 2$ classes. The procedure is completely nonparametric: First the data is transformed to a k-variate depth-depth (DD) plot, and then the α-procedure is employed to classify the objects in the DD-plot. The α-procedure proves to be particularly efficient as k is not large. As a d-variate data depth function we use the random Tukey depth. In case of more than two classes several binary classifications are performed and a majority rule is used. The DDα-procedure is applied to simulated as well as real data, and the results are compared with those of similar procedures that have been recently proposed. In particular, the performance on asymmetric and fat-tailed distributions is investigated and the number of randomly chosen directions in approximating the depth is optimized.

E331: **Multivariate functional halfspace depth**
Presenter: **Mia Hubert**, KU Leuven, Belgium
Co-authors: Gerda Claeskens, Leen Slaets, Kaveh Vakili

In recent years several depth functions have been proposed for functional data. The observations typically consist of one set of curves, measured at a discrete set of time points. The depth functions then provide an ordering of the observed curves from the center outwards. As a generalisation, multivariate functional data are considered, where the measurements consist of several sets of curves, measured at the same time points. A new depth function is proposed for this type of data, called multivariate functional halfspace depth. It acknowledges some important characteristics of functional data, namely differences in the amount of local amplitude variation, shape variation and phase variation. A population version of this depth function is defined and its properties are studied. On real data, it is shown how this depth function allows us to estimate the central tendency of the curves, and to estimate the variability among the curves. Also its usefulness for the classification of new curves is illustrated.

E522: **Depth based classification of functional data**
Presenter: **Daniel Hlubinka**, Charles University in Prague, Czech Republic

Data depth has become a popular nonparametric tool for random vectors. Hence, the depth of functional data is a natural extension of the concept. Unfortunately, such a generalisation cannot be done by simple modification of multivariate data depth and slightly different approach is needed. There are many possible definitions of functional data depth; we propose to use not only the value but also the derivative(s) of the observed functional data. The idea is illustrated on classification problems where it is shown that derivatives may improve the link between the depth function and the position of function with respect to the distribution of functional random variables. In particular, the quality of classification may be substantially improved by use of better functional depth.

E1071: **Boxplots for functions based on halfgraph orders**
Presenter: **Juan Romo**, Universidad Carlos III de Madrid, Spain
Co-authors: Belen Martin-Barragan, Rosa Lillo
A new functional boxplot that extends the classical univariate boxplot is proposed. This approach makes use of new orders for functional data, instead of using functional depths. These orderings are based on the functions epigraphs and hypographs and allow us to define the functional quartiles. The method is computationally efficient, requiring a computational time of $o(n^2)$, in contrast to $o(n^3)$ for already existing methods. Simulated and real examples show that this new boxplot is a convenient visualization technique with a great potential for analyzing functional data.

ES51 Room 10 STATISTICAL ANALYSIS OF GENOME-WIDE ASSOCIATION STUDIES Chair: Taesung Park

E196: **Model selection approaches to GWAS analysis**
Presenter: **Florian Frommlet**, Medical University Vienna, Austria
Genome wide association studies (GWAS) are usually performed to learn about genetic causation of complex traits. It appears to be natural in that context to approach association with multiple regression analysis, where the problem of detecting associated markers can be formulated in terms of model selection. Although there are several software packages available to analyze GWAS in that way, it seems that marginal tests of SNPs combined with multiple testing correction is still considered to be the gold standard for GWAS analysis. We will present MOSGWA, a model selection approach based on some modification of the Bayesian Information Criterion (mBIC2) which controls the false discovery rate of detected regressors. A comprehensive simulation study based on real SNP data shows some surprising deficiencies of single marker tests, and to which extent model selection can help to overcome these difficulties. We will compare MOSGWA with two other packages, HLASSO which has implemented a Bayesian approach with shrinkage priors, and GWASelect which is combining iterated sure independence screening, LASSO and stability selection. Our simulation studies show that MOSGWA has only slightly smaller power than HLASSO and GWASelect, while it is detecting much less false positives.

E291: **A powerful approach for association analysis incorporating imprinting effects**
Presenter: **Wing K. Fung**, University of Hong Kong, China
Co-authors: Fan Xia, Jiyuan Zhou
For a diallelic marker locus, the transmission disequilibrium test (TDT) is a simple and powerful design for genetic studies. The TDT was originally proposed for use in families with both parents available (complete nuclear families) and has further been extended to 1-TDT for use in families with only one of the parents available (incomplete nuclear families). Currently, the increasing interest of the influence of parental imprinting on heritability indicates the importance of incorporating imprinting effects into the mapping of association variants. We shall extend the TDT-type statistics to incorporate imprinting effects and develop a series of new test statistics in a general two-stage framework for association studies. Our test statistics enjoy the nature of family-based designs that need no assumption of Hardy-Weinberg equilibrium. Also, the proposed methods accommodate complete and incomplete nuclear families with one or more affected children. In the simulation study, we verify the validity of the proposed test statistics under various scenarios, and compare the powers of the proposed statistics with some existing test statistics. It is shown that our methods greatly improve the power for detecting association in the presence of imprinting effects. We further demonstrate the advantage of our methods by the application of the proposed test statistics to a rheumatoid arthritis data set.

E783: **Detecting gene-gene interactions associated with the survival phenotype**
Presenter: **Seungyeoun Lee**, Sejong University, Korea, South
Co-authors: Jinseok Oh, Min-Seok Kwon, Taesung Park
While most of statistical methods in genome-wide association studies (GWAS) have been developed to identify SNP-SNP interactions based on a binary phenotype for case-control studies, there is an interest in identifying genetic variants associated with the survival phenotype in cohort studies. A novel method, called Surv-MDR, has been proposed for detecting gene-gene interactions associated with the survival time by modifying the multifactor dimensionality reduction (MDR) method based on the log-rank test. However, Surv-MDR has some limitations: it cannot adjust the covariates and requires rather intensive computations. An alternative approach, called Cox-MDR, has been recently proposed for detecting gene-gene interactions based on a Cox model by extending generalized multifactor dimensionality reduction (GMDR) to the survival phenotype. The main idea of Cox-MDR is to use a martingale residual as a score to classify multi-level genotypes as high and low risk groups. The Cox-MDR method allows the effects of discrete and quantitative covariates to be easily adjusted in the Cox model and requires much less computation than Surv-MDR. We propose to generalize the idea of a Cox-MDR to the parametric regression model for identifying gene-gene interactions associated with the survival phenotype. In the parametric regression model, a standardized residual can be used as a score to classify multi-level genotypes as high and low risk groups while keeping the rest of the MDR procedure unchanged. We compare the proposed method with Cox-MDR as well as Surv-MDR in detecting two-way interactions with the survival phenotype with adjusting the covariates versus without adjusting the covariates throughout the simulation studies as well as analysis of a real dataset.

E1068: **Genome-wide gene-gene interaction analysis based on ultrahigh-dimensional variable selection**
Presenter: **Masao Ueki**, Yamagata University, Japan
Co-authors: Gen Tamiya
Genome-wide gene-gene interaction analysis using single nucleotide polymorphisms (SNPs) is an attractive way for identification of genetic components that confers susceptibility of human complex diseases. Individual hypothesis testing for SNP-SNP pairs as in common genome-wide association study (GWAS) however involves difficulty in setting overall p-values due to complicated correlation structure, namely, the multiple testing problem that causes unacceptable false negative results. A large number of SNP-SNP pairs than sample size, so-called the large p small n problem, precludes simultaneous analysis using multiple regression. A method that overcomes the above issues is thus needed. We adopt an up-to-date method for ultrahigh-dimensional variable selection termed the sure independence screening (SIS) for appropriate handling of numerous number of SNP-SNP interactions by including them as predictor variables in logistic regression. We propose novel ranking strategies using dummy coding methods and following variable selection procedure in the SIS method suitably modified for gene-gene interaction analysis. We also implemented the procedures in a software program, EPISIS, using the cost-effective GPGPU (General-purpose computing on graphics processing units) technology. EPISIS can complete exhaustive scan for SNP-SNP interactions in standard large scale GWAS dataset within several hours. The proposed method works successfully in simulation experiments and in application to real WTCCC (Wellcome Trust Case-Control Consortium) data.

ES56 Room 3 COPULAS I **Chair: Wolfgang Trutschnig**

E155: **Multivariate exponential distributions constructed from certain infinitely divisible processes**
Presenter: **Jan-Frederik Mai**, Technische Universitaet Muenchen, Germany
Co-authors: Matthias Scherer
A stochastic representation for multivariate extendible distributions with exponential minimums (exEM), whose components are conditionally iid in the sense of de Finetti's theorem, is presented. It is shown that the "exponential minimum property" is in one-to-one correspondence with the conditional cumulative hazard rate process being infinitely divisible and satisfying a certain time consistency. The Laplace exponents of such processes are characterized in terms of a Bernstein function applied to the state variable and are linear in time. Examples comprise killed Levy subordinators, monomials whose slope is randomized by a stable random variable, and several combinations thereof. Finally, it is shown that when a certain subfamily of these processes is used in the construction leading to exEMs, the result is the proper subclass of extendible min-stable multivariate exponential (MSMVE) distributions. Since MSMVEs coincide with the family of multivariate distributions whose margins are exponential and whose copula is an extreme-value copula, as an immediate corollary a canonical construction for arbitrary extendible extreme-value copulas is provided.

E161: **H-extendible copulas**
Presenter: **Matthias Scherer**, Technische Universitaet Muenchen, Germany
Co-authors: Jan-Frederik Mai
Adequately modeling the dependence structure of high-dimensional random vectors is challenging. One typically faces a tradeoff between models that are rather simple but computationally efficient on the one hand, and very flexible dependence structures that become unhandy as the dimension of the problem increases on the other hand. Several popular families of copulas, especially when based on a factor-model construction, are extendible. Even though such structures are very convenient in large dimensions (due to the factor model / conditional i.i.d. structure), the assumption of conditional i.i.d. may be over-simplistic for real situations. One possibility to overcome extendibility without giving up the general structure is to consider hierarchical (or nested) extensions of the dependence structure under consideration. Heuristically speaking, the dependence structure of hierarchical copulas is induced by some global stochastic factor affecting i.i.d. components and by additional group-specific factors that only affect certain sub-vectors. We present a survey of recent developments on hierarchical models, such as hierarchical Archimedean and Marshall-Olkin type dependence structures, and unify the literature by introducing the notion of h-extendibility. This definition generalizes extendible models in a natural way to hierarchical structures.

E185: **Invariant dependence structures and L_1 Dirichlet random vectors**
Presenter: **Piotr Jaworski**, University of Warsaw, Poland
The correspondence between two families of bivariate distributions is introduced. The first family consists of irreducible copulas invariant with respect to lower conditioning of the first variable. Such copulas describe the limit behavior of a random vector when one of the components is taking on extreme values. The second one consists of distributions of L_1 Dirichlet random vectors called also l_1-norm symmetric or simplicially contoured distributions. Such distributions are considered as an alternative of spherical distributions in risk modelling. The Euclidean L_2 norm is replaced by L_1 one and the survival copulas become Archimedean. Basing on the above mentioned correspondence new characterizations of both families are provided.

E200: **A new measure of dependence invariant under bijective transformations**
Presenter: **Songkiat Sumetkijakan**, Chulalongkorn University, Thailand
Co-authors: Tippawan Santiwipanont, Pongpol Ruankong
Based on the modified Sobolev norm, we define a new norm for bivariate copulas which is invariant under shuffling of continuous random variables. It turns out that copulas with the maximum new norm not only include all complete dependence copulas, i.e. $C_{X,f(X)}$ and $C_{g(Y),Y}$, but also all copulas of the form $C_{X,f(X)} * C_{g(Y),Y}$ where $*$ is a given product of copulas. As a consequence, the corresponding measure of dependence satisfies most of Renyi's postulates and that continuous random variables whose copula is $C_{X,f(X)} * C_{g(Y),Y}$ have maximum measure of dependence. It remains to find a probabilistic interpretation of such copulas.

ES62 Room 8 IMPRECISE PROBABILITY THEORIES FOR STATISTICAL PROBLEMS **Chair: Ines Couso**

E379: **Continuity of the imprecise expectation**
Presenter: **Kevin Loquin**, LIRMM, France
The continuity of an application suggests stability between the application domain and its image. For instance, the application which links a probability to its expectation operator is continuous for the total variation distance. Therefore, a probability model Q closer to the probability P than a probability R will lead to an expectation E_Q closer to E_P than E_R. The aim is to study the continuity of the application which links a coherent imprecise probability model to its natural extension: its imprecise expectation. The continuity of this application is proved for the Hausdorff distance between coherent lower previsions models (and thus between coherent lower probability models). This Hausdorff distance between two lower previsions (probabilities) is constructed from the total variation distance between any two elements of their respective credal sets. More than establishing stability, what is particularly interesting is that we can relate the imprecision of the elicitation of an imprecise probability model to the imprecision of its associated imprecise expectation. This is due to the choice of the Hausdorff distance which can be considered as an imprecision marker of an imprecise prevision model.

E870: **Predicting structured data: Ideas and first results on ranking**
Presenter: **Sebastien Destercke**, Centre National de la Recherche Scientifique, France
Imprecise probabilistic methods are particularly interesting when data are incomplete or scarce and when some cautiousness in the prediction is needed. We present some first arguments about why imprecise probabilistic approaches may be interesting to deal with structured data. We then show how this can be done on the specific example of label ranking problem.

E1009: **A comparison of conventional and imprecise probability approaches to statistics**
Presenter: **Marco Cattaneo**, LMU Munich, Germany
In recent years, theories of imprecise probability have been suggested as alternative approaches to statistical problems. These approaches will be compared with the conventional ones (based on precise probability theory), both from a theoretical perspective, and from the pragmatical perspective of their application to the so-called "fundamental problem of practical statistics".

E1002: **Imprecise regression: A comparative study**
Presenter: **Andrea Wiencierz**, LMU Munich, Germany
Imprecise regression methods are regression methods that do not aim at a single estimated regression function but allow for a set-valued result. Sets of regression functions can be obtained from regression methods that are based on imprecise probability models, like the recently introduced

Likelihood-based Imprecise Regression, but they can also be derived from classical regression methods, e.g., in the form of confidence bands for Least Squares regression. Hence, the question arises, which approach leads to better results. To investigate this question, we will compare a selection of imprecise methods in the setting of simple linear regression. On the one hand the comparison will be based on general properties of the regression methods, like the coverage probability of the result or the robustness of the method. On the other hand, we will compare the performance in a practical setting, where we will consider the case of precise data as well as the case where the variables are only imprecisely observed.

ES74 Room 11 MULTIVARIATE EXTREMES AND REGRESSION ANALYSIS Chair: Armelle Guillou

E122: **Kernel regression with Weibull-type tails**
Presenter: **Yuri Goegebeur**, University of Southern Denmark, Denmark
Co-authors: Tertius de Wet, Armelle Guillou
The estimation of the tail index of a Weibull-type distribution in the presence of random covariates is considered. The approach followed is non-parametric and consists of locally weighted estimation in narrow neighbourhoods in the covariate space. We introduce two flexible families of estimators and study their asymptotic behavior under some conditions on the conditional response distribution, the weight function, the density of the independent variables, and for appropriately chosen bandwidth and threshold parameters. We also introduce a Weissman-type estimator for estimating extreme conditional quantiles. The finite sample behavior of the proposed estimators is examined with a small simulation experiment.

E153: **Testing for a δ-neighborhood of a generalized Pareto copula**
Presenter: **Michael Falk**, University of Wuerzburg, Germany
Co-authors: Stefan Aulbach
A multivariate distribution function F is in the max-domain of attraction of an extreme value distribution if and only if this is true for the copula corresponding to F and its univariate margins. It is known that a copula satisfies the extreme value condition if and only if the copula is tail equivalent to a generalized Pareto copula (GPC). We propose a χ^2-goodness-of-fit test in arbitrary dimension for testing whether a copula is in a certain neighborhood of a GPC.

E289: **Bernstein polynomial-based inference for multivariate extremes**
Presenter: **Miguel de Carvalho**, Pontificia Universidad Catolica de Chile, Chile
Co-authors: Timothy E. Hanson, Anthony C. Davison
The statistical modelling of multivariate extremes plays a key role in the analysis of the joint occurrence of infrequent but disastrous events. To model such events we need to consider the dependence structure within the joint tail, and this introduces complex issues into the analysis. In particular, this modelling cannot be based on a direct dimensional extension of univariate techniques, as there is no parametric family able to represent the class of multivariate extreme value distributions. One of such modelling difficulties arises in the so-called spectral measure which characterizes the interaction between joint extremes. The spectral measure is constrained to satisfy certain marginal moment constraints, and it is thus complex to design appropriate models and develop estimators that satisfy the constraints. We discuss how to model the spectral measure of a multivariate extreme value distribution using multivariate Bernstein polynomials on the simplex. Our multivariate Bernstein mean-constrained approach works in a similar way for any number of dimensions, and posterior inference is conducted using an adaptive componentwise random-walk Metropolis-Hastings algorithm.

E340: **Short-range temporal dependence within extremes of stock market indices**
Presenter: **Alexandra Ramos**, Universidade do Porto, Portugal
To characterize the extremal behaviour of stationary time series, attention has been given to the within-cluster behaviour of the extremes of a series, which is determined by the short-range temporal dependence. Most of its characterization has been done based on the assumption of Markovianity of the time series, as the class of dth-order Markov chains is sufficiently general and tractable. A simple quantifier of the extremal within-cluster dependence for any stationary time series satisfying the Leadbetter D condition is the threshold dependence extremal index, $\theta(u)$, which measures the extent of clustering of exceedances of the process above the high threshold u. We study and estimate this function for asymptotically independent Markov chains and model temporal dependence within these series extremes. Applying these techniques to daily log-returns of stock indices, we examine large consecutive changes associated with large financial gains or losses.

ES28 Room 7 ADVANCES IN HAMILTONIAN MONTE CARLO Chair: YeeWhye Teh

E436: **Quasi-Newton Markov chain Monte Carlo**
Presenter: **Charles Sutton**, University of Edinburgh, United Kingdom
The performance of Markov chain Monte Carlo methods is often sensitive to the scaling and correlations between the random variables of interest. An important source of information about the local correlation and scale is given by the Hessian matrix of the target distribution, but this is often either computationally expensive or infeasible. We propose MCMC samplers that make use of quasi-Newton approximations, which approximate the Hessian of the target distribution from previous samples and gradients generated by the sampler. A key issue is that MCMC samplers that depend on the history of previous states are in general not valid. We address this problem by using limited memory quasi-Newton methods, which depend only on a fixed window of previous samples. On several real world datasets, we show that the quasi-Newton sampler is more effective than standard Hamiltonian Monte Carlo at a fraction of the cost of MCMC methods that require higher-order derivatives.

E444: **Split Hamiltonian Monte Carlo**
Presenter: **Shiwei Lan**, UC Irvine, United States
Co-authors: Babak Shahbaba, Wesley Johnson, Radford Neal
It is shown how the Hamiltonian Monte Carlo algorithm can sometimes be sped up by "splitting" the Hamiltonian in a way that allows much of the movement around the state space to be done at low computational cost. One context where this is possible is when the log density of the distribution of interest (the potential energy function) can be written as the log of a Gaussian density, which is a quadratic function, plus a slowly varying function. Hamiltonian dynamics for quadratic energy functions can be analytically solved. With the splitting technique, only the slowly-varying part of the energy needs to be handled numerically, and this can be done with a larger stepsize (and hence fewer steps) than would be necessary with a direct simulation of the dynamics. Another context where splitting helps is when the most important terms of the potential energy function and its gradient can be evaluated quickly, with only a slowly-varying part requiring costly computations. With splitting, the quick portion can be handled with a small stepsize, while the costly portion uses a larger stepsize. We show that both of these splitting approaches can reduce the computational cost of sampling from the posterior distribution for a logistic regression model, using either a Gaussian approximation centered on the posterior mode, or a Hamiltonian split into a term that depends on only a small number of critical cases, and another term that involves the larger number of cases whose influence on the posterior distribution is small. Supplemental materials for this paper are available online.

E565: **Geometric methods for MCMC in statistics and machine learning**

Presenter: **Simon Byrne**, University College London, United Kingdom
Co-authors: Mark Girolami

Using tools from differential geometry, the Riemannian manifold Langevin and Hamiltonian Monte Carlo algorithms have been shown to produce efficient MCMC samplers by adapting to the geometric structure of the statistical model. However these methods are fundamentally "local" in that they only exploit information such as gradients and metrics that can be obtained from the neighbourhood of the current point. However, many models that are widely used in statistics and machine learning often have a rich "global" geometric structure that can also be exploited. It will be demonstrated how such geometry might be used to sample from difficult models, such as high-dimensional probability simplices and hierarchical models.

E1048: **Large scale Bayesian topic modelling of text corpora**

Presenter: **Yee Whye Teh**, Oxford University, United Kingdom
Co-authors: Sam Patterson

Topic models like latent Dirichlet allocation (LDA) have found widespread use in machine learning and natural language processing problems, where they are used to uncover the semantic structures underlying document corpora. Recent years has seen the easy availability of very large scale document corpora with which to fit LDA. We explore the use of Hamiltonian Monte Carlo methods for inference in such models on large scale data. Different parameterizations of the model and Riemannian metric structures used significantly affect the speed of convergence of Hamiltonian Monte Carlo methods, and we discuss the various pitfalls, lessons learnt, and choices made along the way to arriving at an algorithm that produces state-of-the-art performance.

Sunday 02.12.2012 08:45 - 10:25 Parallel Session I – CFE

CSI01 Room Multipurpose ADVANCES IN FINANCIAL TIME SERIES Chair: Monica Billio

C824: **Time-varying parameters and changing association**
Presenter: **Andrew Harvey**, University of Cambridge, United Kingdom
Co-authors: Stephen Thiele
Multivariate Gaussian time series models in which correlations change over time can be reformulated as regression models with time-varying parameters. However, the multivariate approach yields models with more satisfactory properties and generalization is more straightforward. Although models can be set up in an unobserved components framework, the recently developed dynamic conditional score approach appears to be more tractable. Simulation experiments on estimating such models are reported and the methods are applied to the Phillips curve.

C1050: **On the univariate representation of BEKK models with common factors**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Sebastien Laurent, Franz C. Palm
The minimal order univariate representation of some well known $n-$dimensional conditional volatility models is investigated. Even simple low order systems (e.g. a multivariate GARCH(0,1)) for the joint behavior of several variables imply individual processes with a lot of persistence in the form of high order lags. However, we show that in the presence of common GARCH factors, parsimonious univariate representations (e.g. GARCH(1,1)) can result from large multivariate models generating the conditional variances and conditional covariances/correlations. The trivial diagonal model without any contagion effects in conditional volatilities gives rise to the same conclusions though. Consequently, we then propose an approach to detect the presence of these commonalities in multivariate GARCH process. We compare the small sample performances of two strategies. First, we use reduced rank regressions in a multivariate system for squared returns and cross-returns. Second, we investigate a likelihood ratio approach, where under the null the matrix parameters of the BEKK have a reduced rank structure. It emerged that the latter approach has quite good properties enabling us to discriminate between a system with seemingly unrelated assets (e.g. a diagonal model) and a model with few common sources of volatility.

C1065: **Mixtures models, jumps and option pricing**
Presenter: **Jeroen Rombouts**, HEC Montreal, Canada
Asymmetric heteroskedastic normal mixture models are used to fit return data and to price options. The models can be estimated straightforwardly by maximum likelihood, have high statistical fit when used on S&P 500 index return data, and allow for substantial negative skewness and time varying higher order moments of the risk neutral distribution. When forecasting out-of-sample a large set of index options between 1996 and 2011, substantial improvements are found compared to several benchmark models in terms of dollar losses and the ability to explain the smirk in implied volatilities.

CS89 Room 4 FINANCIAL MARKETS I Chair: Sylvia Kaufmann

C968: **The effectiveness of changes in settlement procedures**
Presenter: **Emily Lin**, SJU, Taiwan
Taiwan futures market experienced 6 changes in the settlement procedure within the past 13 years. This makes the market a paragon for testing effectiveness of changes in settlement procedures. This study summarizes the settlement procedures conducted in the world-wide futures markets, explores for expiration-day effect on trading activity unexplored in expiration-day literature, e.g. illiquidity, market depth, order imbalance and quoted spread, and sophisticates the measurement of expiration-day phenomena in the cash market. The study proposes a model to verify the unresolved hypothesis that the expiration-day trading may split between market open and close and explores the settlement procedure affecting the trading activity at the expiration of the index derivatives. Expiration-day trading exhibits an intraday seasonality in the stock index market. The aim is also to offer criteria for selecting the settlement timing, that is, whether the settlement day should be same as the expiration day or different from the expiration day. Although it is found that the Taiwan stock market trading in terms of volatility and share volume on settlement day is higher than normal when TAIEX futures and MSCI TW futures settle, price reversal is not statistically higher, but the underlying index reverses stronger when MSCI TW futures settle than when TAIEX futures settle.

C994: **Building synthetic OIS curves in emerging markets currencies**
Presenter: **Jaroslav Baran**, University of Economics, Czech Republic
Possible proxies for OIS rates for currencies with no active or limited OIS market (Czech market in particular) which are used for discounting collateralized trades and their respective discount factors are constructed from the set of available marketable instruments with the emphasis on practicality and applicability. It is examined if it is suitable to use liquid proxies in other currencies and adjusting them with cross-currency basis when building synthetic OIS curves in emerging markets currencies. Individual components that affect cross-currency basis swap spreads and their respective implied discount curves are also analyzed.

C996: **Multivariate analysis of abnormal returns: The case of corporate social responsibility indices revisions**
Presenter: **Bartosz Zawadzki**, Wroclaw University of Economics, Poland
Co-authors: Karolina Daszynska-Zygadlo, Tomasz Slonski
Since 2008 total European SRI (Sustainable and Responsible Investing) assets under management almost doubled, exceeding approximately 5 trillion EUR. According to the semi-strong efficiency theory, capital markets need time to adapt to new information flow. Event studies measure the impact of a particular event on the rates of return generated by the companies. A novel method of testing the statistical significance of abnormal returns is being introduced. The procedure is based on testing the statistical significance of the average differences between the expected (estimated) return and the realized return in the days preceding and following the day of the inclusion/exclusion from corporate social responsibility (CSR) index. This research is based on a sample of companies included in European Sustainability indexes. When considering particular measurements of each abnormal return as variables, and taking into account their correlation, using single point approach becomes faulty. A proposition to avoid the error is to construct a $p-$variate confidence region based on the measurement of the distance between two multivariate means - the Mahalanobis distance and compare it with the overall similarity region (with assumed confidence level alpha).

C999: **Reliability and stability of different performance measures**
Presenter: **Pilar Grau Carles**, Universidad Rey Juan Carlos, Spain
Co-authors: Luis Miguel Doncel, Jorge Sainz
Some of the most popular measures of a portfolio performance are the reward-to-risk ratios, such as Sharpe ratio, Reward-to-VaR, Reward-to-modified VaR, Reward-to-Expected Shortfall and RAP (Risk Adjusted Performance) that differ in the way they measure the down-side risk. An alternative approach is the use of bootstrap methodology to take into account risk estimation by calculating the Double Sharpe ratio. The objective of this research is to ascertain that the choice of a particular performance measure has an impact on the ranking of alternative investments. Also,

the measures are not stable in time, especially in the case of a bear market. We show, using data from UK mutual funds, that different measures produce different rankings and that the performance measures are not stable over time. Also, using bootstrap we obtain that the Spearman rank correlation coefficient and Kendall tau show different rankings for different measures. The disagreement between the rankings is greater when there are major differences in the higher moments of the distribution of returns; and in the case of bear markets, some measures provide incorrect ranking classification.

C1006: **Earnout agreements in corporate acquisitions and bidders' gains**
Presenter: **Leonidas Barbopoulos**, University of St Andrews, United Kingdom
The gains to US bidders in M&A deals in which the payment is contingent upon post-merger performance (earnouts) are examined. Earnout is posited as a device to mitigate the adverse selection and moral hazard problems in corporate acquisitions more effectively than other payment methods. A logistic model that allows for the effects of transaction-, firm- and country-specific factors is constructed to predict the 'correct' use of earnouts in acquisitions. Our analysis shows that the use of earnout adds value and bidders making the 'correct' use of earnouts enjoy significantly higher returns from foreign target acquisitions than from domestic deals. Based on the propensity score matching method which addresses sample selection biases, the Rosenbaum Bounds, and on the identification of the correct use of earnout, we find that foreign bids when financed with the correct use of earnouts persistently outperform other deals. Overall, the findings suggest that the correct use of earnouts provides an effective mechanism for mitigating the risks of adverse selection and moral hazard in corporate acquisitions, in particular, cross-border acquisitions thereby generating significantly more value for acquirer shareholders.

CS48 Room 2 BUSINESS CYCLE ANALYSIS Chair: Francesco Ravazzolo

C344: **Short-term forecasting: Norges Bank's density combination approach**
Presenter: **Anne Sofie Jore**, Norges Bank, Norway
Co-authors: Knut Are Aastveit, Karsten R. Gerdrup, Francesco Ravazzolo
Norwegian real-time vintage data are used to produce combined density short-term forecasts for quarterly Norwegian Mainland GDP and inflation. We combine the forecasts using different pooling approaches from three main types of models typically used at central banks: Vector autoregressive models, leading indicator models and factor models. The combined forecasts are updated several times during the quarter to highlight the importance of new data releases, and results show that the performance in terms of point and density forecasting improves steadily as new information arrives. We apply the combined density forecasts for GDP to predict turning points of the business cycle in real time. We find the combined forecast is accurate in forecasting turning points in real time and superior to strategies based on selecting the ex-ante best model at the middle of the first forecasting quarter. The uncertainty in the definition of the cycle related to data revision play, however, a crucial role in the evaluation.

C492: **Money, credit, housing and the great recession: demand and supply channels**
Presenter: **Antonio Conti**, Universite Libre de Bruxelles, Belgium
The role of money, credit and housing in the business cycle to deepen the analysis on the Great Recession is evaluated. We estimate a Bayesian VAR with sign restrictions on US quarterly data up to 2011:Q4 to disentangle the dynamic effects of demand and supply disturbances in each of the above financial markets. In particular, we are interested in identifying the effects of monetary and credit aggregates and housing on output and prices, beyond highlighting the linkages among these variables. We find that housing and credit are non—negligible sources of the US business cycle. Notably, a combination of adverse credit supply and housing demand shocks seems to have played a major role in the origin and propagation of the Great Recession. Furthermore, results do not point at clear evidence of changes in monetary transmission to loans or house prices.

C656: **Pooling versus model selection for nowcasting with many and selected predictors**
Presenter: **Ard den Reijer**, Sveriges Riksbank, Sweden
Pooling versus model selection for now- and forecasting Swedish GDP growth in the presence of model and data uncertainty is discussed. Large macroeconomic data sets are typically unbalanced due to different sampling frequencies and publication delays. We employ a large monthly data set of about hundred monthly indicators with the aim of forecasting quarterly GDP growth rates. We compare the forecasting performance of different types of models based on the complete data set with the corresponding ones based on an optimal subset of selected predictors. The data selection procedure is designed to avoid oversampling and only includes those predictors that reduce the conditional in-sample variance. The different types of models consist of factor models that are designed to handle large data sets and mixed-data sampling (MIDAS) regressions relating monthly factors to quarterly GDP. Apart from predictor selection, model uncertainty relates to factor estimation method, number of factors and lag length. As there are many potential sources of misspecification, one strategy is to employ pooling over a large set of models with different specifications. We finally compare the results of pooled versus selected models based on the full sample versus the selected predictors.

C1054: **Global and regional business cycles: Shocks and propagations**
Presenter: **Leif Anders Thorsrud**, Norwegian Business School, Norway
The aim is to study the synchronization of real and nominal variables across four different regions of the world; Asia, Europe, North and South America, covering 32 different countries. Within a FAVAR framework we distinguish between global and regional demand and supply shocks, and document the relative contribution of these shocks in explaining macroeconomic fluctuations and synchronization. Our results support the decoupling hypothesis put forward in other recent business cycle studies, although the period 2007 to 2009 has had a large impact on the synchronization patterns. Importantly, the transmission mechanism of global demand shocks is different in Asia and South America than in Europe and North America, and the role played by demand shocks is larger than found in related studies. Finally, only innovations to the Asian activity and price factors have significant spillover effects on the common world factors, pointing out the growing importance of Asia in the global economy.

CS56 Room 3 MULTIVARIATE VOLATILITY MODELS Chair: Luc Bauwens

C403: **Data-driven estimation of smooth correlation changes in a semiparametric dynamic conditional correlation model**
Presenter: **Yuanhua Feng**, University of Paderborn, Germany
The purpose is to study the estimation of smooth volatility and correlation changes of financial returns under semiparametric extensions of the constant conditional correlation (CCC) and the dynamic conditional correlation (DCC) models. A multi-step estimation procedure is employed, which has some clear advantages compared to a uni-step estimation procedure. Different nonparametric approaches are proposed and compared. The CCC and DCC parameters are estimated after removing the nonparametric components properly. Detailed asymptotic results are obtained. Effect of the errors in the nonparametric estimates on the resulting parametric estimates is also discussed. It is in particular shown that, under regularity conditions, the parametric estimation is still efficient. Suitable iterative plug-in algorithms are developed for selecting the bandwidths. The proposed models are then applied to decompose well known financial market risk measures. Furthermore, nonparametric detection of possible structural breaks in volatility and correlations are investigated. It is shown that these approaches work well in practice, which also provide us with a quantitative method for defining financial crises.

C409: **Modeling the dependence of conditional correlations on volatility**
Presenter: **Edoardo Otranto**, University of Messina, Italy
Co-authors: Luc Bauwens
Several studies have tried to uncover the determinants of correlations between time series of financial returns, but few of them are based on econometric models that relate explicitly the correlations to their determinants. This is an important topic, because the increase in correlations is linked to the phenomenon of contagion, so the possibility to forecast the correlations is crucial in the study of spillover effects between financial markets, as well as of portfolio choice, asset allocation, etc. The most common empirical finding is that volatility is a major determinant of correlation, in particular during turmoil periods. Several models are developed to capture explicitly the dependence of the conditional correlation matrix of a set of financial time series on the volatility of the market of interest. The models include market volatility as determinant of correlations, either through time-varying probabilities in a Markov-switching framework, or as an additive effect in a classical dynamic correlation model. The models are applied to the 30 assets composing the Dow Jones industrial index.

C479: **Computationally efficient inference procedures for vast dimensional realized covariance models**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Luc Bauwens
Some computationally efficient estimation procedures for the estimation of vast dimensional realized covariance models are illustrated. In particular, we derive a Composite Maximum Likelihood (CML) estimator for the parameters of a Conditionally Autoregressive Wishart (CAW) model incorporating scalar system matrices and covariance targeting. The finite sample statistical properties of this estimator are investigated by means of a Monte Carlo simulation study in which the data generating process is assumed to be given by a scalar CAW model. The performance of the CML estimator is satisfactory in all the settings considered although a relevant finding of our study is that the efficiency of the CML estimator is critically dependent on the implementation settings chosen by modeller and, more specifically, on the dimension of the marginal log-likelihoods used to build the composite likelihood functions.

C593: **Common dynamics in volatility: An additive common component vMEM approach**
Presenter: **Giampiero Gallo**, Universita di Firenze, Italy
Co-authors: Fabrizio Cipollini
Measures of financial volatility exhibit clustering and persistence and can be jointly modeled as the element by element product of a vector of conditionally autoregressive scale factors and a multivariate i.i.d. innovation process (vector Multiplicative Error Model – vMEM). Since similar profiles are shared across measures, a restricted vMEM decomposes the conditional expected volatility into the sum of a common (persistent) component and a vector of measure specific components. With data on absolute returns, realized kernel volatility and daily range for five stock market indices, we show that indeed such a common component exists with the desired properties. The transitory components happen to have different features across indices.

C457: **Multivariate rotated ARCH models**
Presenter: **Diaa Noureldin**, Oxford, United Kingdom
Co-authors: Neil Shephard, Kevin Sheppard
This paper introduces a new class of multivariate volatility models which is easy to estimate using covariance targeting, even with rich dynamics. We call them rotated ARCH (RARCH) models. The basic structure is to rotate the returns and then to fit them using a BEKK-type parameterization of the time-varying covariance whose long-run covariance is the identity matrix. This yields the rotated BEKK (RBEKK) model. The extension to DCC-type parameterizations is given, introducing the rotated DCC (RDCC) model. Inference for these models is computationally attractive, and the asymptotics are standard. The techniques are illustrated using data on some DJIA stocks.

CS61 Room 1 TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

C269: **Principal components and model averaging**
Presenter: **Martin Wagner**, Institute for Advanced Studies, Austria
Co-authors: Stefan Zeugner
Model averaging has proven popular for inference with many potential predictors in small samples. However, it is frequently criticized for a lack of robustness with respect to prediction and inference. The aim is to explore the reasons for such robustness problems and to propose to address them by transforming the subset of potential 'control' predictors into principal components in suitable datasets. A simulation analysis shows that this approach yields robustness advantages vs. both standard model averaging and principal component-augmented regression. Moreover, we devise a prior framework that extends model averaging to uncertainty over the set of principal components and show that it offers considerable improvements with respect to the robustness of estimates and inference about the importance of covariates. Finally, we empirically benchmark our approach with popular model averaging and PC-based techniques in evaluating financial indicators as alternatives to established macroeconomic predictors of real economic activity.

C295: **Jittered phase diagrams for seasonal patterns in time series**
Presenter: **Robert Kunst**, Institute for Advanced Studies, Austria
In the analysis of seasonal time series, an issue of main concern is the discrimination of deterministic patterns (shape reversion) and seasonal unit roots (permanent shape transition). We suggest complementing the customary hypothesis tests with jittered phase diagrams on a discretized set of basic seasonal shapes that represent the dynamic transition between the patterns. These jittered phase diagrams provide a convenient visualization tool that supports the discrimination among the main classes of potential seasonal data-generating processes. For several classes of interest, characteristic patterns are determined. We also consider a nonparametric hypothesis test constructed from the charts. Whereas the visual tool mainly targets the case of quarterly data, an extension to the monthly case is also considered. Exemplary applications to economics data and to some other variables are presented.

C559: **Cointegrating relationship with spatial lags**
Presenter: **Jan Mutl**, European Business School, Austria
Co-authors: Leopold Soegner
Dynamic panel data models that exhibit a cointegration relationship that includes a spatial lag are studied. Such a relationship postulates that the long run outcome for a particular cross-section is affected by a weighted average of the outcomes in the other cross-sections. This model produces outcomes that are non-linear in the coefficients to be estimated. It also renders the existing estimation methods that typically assume cross-sectional independence inapplicable. Assuming that the weights are exogenously given, we extend the dynamic ordinary least squares (DOLS) methodology and provide a dynamic instrumental variable estimator. We derive the large sample properties of our proposed estimator under a set of low-level assumptions and investigate its small sample distribution in a simulation study. We illustrate our estimation methodology by investigating the structure of spatial correlation of firm-level implied credit risk. Credit default swaps, firm specific data and industry data from the US market are used to estimate spatial correlation models. We construct the economic space based on a 'closeness' measure for firms based on input-output

matrices. Our estimates show that the spatial correlation of credit default spreads is substantial and highly significant. We also demonstrate how these models can be used to price newly issued credit default swaps and show that they lead to better performance compared to standard panel regression settings.

C249: **Forecasting in nonlinear panel data regression by stepwise updating of product kernel weights**
Presenter: **Joachim Schnurbus**, Bielefeld University, Germany
Co-authors: Harry Haupt

Forecasting of $Y_{i,T+p}$ using a general class of nonlinear panel data models $Y_{i,t} = g(\mathbf{X}_{i,t}, \mathbf{Z}_{i,t}) + U_{i,t}$ with error process $\{U_{i,t}\}$ and unknown smooth regression function $g(\cdot)$, cross-section $i = 1, ..., N$ and time dimension $t = 1, ..., T$, continuous and categorical predictor variables $\mathbf{X}_{i,t}$ and $\mathbf{Z}_{i,t}$, where the former may contain exogenous and lagged endogenous variables and the latter typically contains deterministic trend and calendar effects t and s, respectively, is considered. In the framework of multiple nonparametric mixed kernel regression, observations are weighted by a product kernel function $W(\mathbf{h})$ based on the simultaneously estimated bandwidths in the vector $\mathbf{h}$ and the values of continuous and categorical predictor variables in $\mathbf{X}$ and $\mathbf{Z}$. The aim is to show how mixed kernel regression can be used for panel forecasting and to demonstrate how the initially estimated bandwidths vector $\mathbf{h}_T = (\mathbf{h}_{T,X}, \mathbf{h}_{T,Z})$ can be updated for p-step forecasts for each new cross-section of observations $T + p, p = 1, 2, \ldots$ We find that fully nonparametric forecasts tend to dominate various parametric benchmarks for all forecast horizons. In addition, using updated kernel weights $W(\widehat{\mathbf{h}}_{T+p})$ based on forecasted bandwidths obtained from nonlinear innovation state-space models often leads to further improvements of p-step forecasts.

CS90 Room 5 TIME SERIES ECONOMETRICS I Chair: Gianluca Cubadda

C468: **Forecasting with unobserved heterogeneity**
Presenter: **Matteo Guido Richiardi**, University of Torino, Italy

Forecasting based on random intercepts models requires imputation of the individual permanent effects to the simulated individuals. When these individuals enter the simulation with a history of past outcomes this involves sampling from conditional distributions, which might be unfeasible. We present a method for drawing individual permanent effects from a conditional distribution which only requires to invert the corresponding estimated unconditional distribution. While the algorithms currently available in the literature require polynomial time, the proposed method only requires matching two ranks and works therefore in $N \log N$ time.

C914: **Forecasting integer autoregressive process of order 1: Analyzing whether INAR models are really better than AR**
Presenter: **Margherita Gerolimetto**, Ca' Foscari, Venice, Italy
Co-authors: Luisa Bisaglia, Silvano Bordignon

Recently there has been a considerable interest in nonnegative integer-valued time series, particularly time series of counts. In some cases, the discrete values of the time series are large numbers and may be analyzed using continuous-valued models such as ARMA with Gaussian errors. However, when the values of the time series are small, as in the case of counting processes, the usual linear ARMA processes are of limited use for modeling and, especially, forecasting purposes, since they lead to non-integer values. While the theoretical properties of INAR models have been discussed extensively in the literature, there are relatively few contributions on the development of coherent, forecasting methods, i.e. such that only integer forecasts of the count variable are obtained. These methods are not easy to implement and it is interesting to investigate whether the more simple ARMA models may be a useful description of the real process in terms of providing good forecasts, even if the DGP is a count process. To provide evidence on this issue, an extensive Monte Carlo experiment is here carried out. INAR DGP's are considered and the performance of the competing class of models (in particular ARMA) is expressed in terms of their forecast accuracy.

C806: **Forecasting with estimated multi-frequency temporally aggregated linear processes**
Presenter: **Lyudmila Grigoryeva**, Universite de Franche-Comte, France
Co-authors: Juan-Pablo Ortega

The aim is to propose a finite sample based predictor for estimated linear one dimensional time series models and compute the associated total forecasting error. The expression for the error that we present takes into account the estimation error. Unlike existing solutions, our formulas require neither assumptions on the second order stationarity of the sample nor Monte Carlo simulations for their evaluation. This result is used to prove the pertinence of a new hybrid scheme that we put forward for the forecast of linear temporal aggregates. This novel strategy consists of carrying out the parameter estimation based on disaggregated data and the prediction based on the corresponding aggregated model and data. We show that in some instances this scheme has a better performance than the all-disaggregated approach presented as optimal in the literature.

C515: **Testing for trends and trend breaks in nonstationary long memory processes**
Presenter: **Petra Brand**, Julius-Maximilians-Universitaet Wuerzburg, Germany
Co-authors: Andreas Dechert

The decision about the existence of a deterministic trend in a time series is very important for practitioners, because it can severely affect statistical inferences. Hence, several different ways to deal with this decision arised in the last few years. Since most of empirical time series seem to be nonstationary it is interesting how to test for a linear trend, when there is evidence of nonstationarity. The aim is to generalize the approach proposed by Ventos-Santaularia and Gomez, who allowed the existence of a unit root only. Assuming that time series don't have to be integrated of an integer digit, but rather could by a fractional order, one knows that they are nonstationary for $d > 0.5$. Hence, this work simulates critical values for a range of nonstationary integration parameters d and gives response regression results for tests of the presence of a deterministic trend. Furthermore it takes into account that the trend component can underly a structural break and gives critical values according to d and the break location, also. As an empirical example the approach is applied on logs of industrial production data of several developed and emerging countries to distinguish between a deterministic and stochastic trend.

CS73 Room 6 FINANCIAL MODELLING Chair: Gianluca Fusai

C829: **FFT-network for bivariate Levy option pricing**
Presenter: **Weiyin Wang**, The Chinese University of Hong Kong, China
Co-authors: Hoi Ying Wong

A two-dimensional network to retrieve the price of two-asset option under Levy processes by using the fast Fourier transform (FFT) is proposed. It can be applied to different multivariate Levy constructions such as subordination and linear combination provided that the joint characteristic function is obtainable. With the prevalent implementation of FFT, the network approach results in significant computational time reduction, while maintaining satisfactory accuracy. In general, multi-dimensional option pricing problems are also solvable by extending this network. Furthermore, we investigate option pricing on a single asset where the asset return and its volatility are driven by a pair of dependent Levy processes. Such a model is also called the random time-changed Levy process. Numerical examples are given to demonstrate the efficiency and accuracy of FFT-network applied to exotic and American-style options.

C858: **CEV asymptotics of American options**
Presenter: **Chi Seng Pun**, The Chinese University of Hong Kong, China
Co-authors: Hoi Ying Wong
The constant elasticity of variance (CEV) model is a practical approach to option pricing by fitting to the implied volatility smile. Its application to American-style derivatives, however, poses analytical and numerical challenges. By taking the Laplace-Carson transform (LCT) to the free-boundary value problem characterizing the option value function and the early exercise boundary, the analytical result involves confluent hypergeometric functions. Thus, the numerical computation could be unstable and inefficient for a certain set of parameter values. We solve this problem by an asymptotic approach to the American option pricing problem under the CEV model. We demonstrate the use of the proposed approach using perpetual and finite-time American puts.

C947: **Non-stationary volatility with highly anti-persistent increments: An alternative paradigm in volatility modeling**
Presenter: **Ladislav Kristoufek**, Charles University in Prague, Czech Republic
An alternative paradigm to volatility modeling of financial securities is introduced. On the example of three stocks of highly capitalized companies, we show that the volatility process is non-stationary and its logarithmic transformation together with logarithmic increments are approximately normally distributed. Further, the increments are shown to be highly anti-persistent. Together with the assertion that logarithmic returns are normally distributed, and uncorrelated with time-varying volatility, we propose the new returns-generating process. The whole procedure is based on empirical observations without any limiting assumptions. We construct the returns series which remarkably mimic the real-world series and possess the standard stylized facts – uncorrelated returns with heavy tails, strongly autocorrelated absolute returns and volatility clustering. Therefore, the proposed methodology opens an alternative approach to research of financial volatility.

C951: **Estimation of high-frequency volatility in duration-based approach**
Presenter: **Sing Fan Chan**, Chinese University of Hong Kong, China
Co-authors: Kai Pui Lam
Estimating the intraday volatility of stock price by integrating the instantaneous conditional return variance (ICV) per unit time obtained from autoregressive conditional duration(ACD) model, called ACD-ICV method, has been proposed in 2011. Empirical results show that ACD-ICV method is a useful tool for estimating high-frequency volatility. Using ACD-ICV method to estimate high-frequency volatility, we find that the empirical dynamics of the duration of price-change events behaves asymmetric. To improve the modelling of duration of price-change events, we consider the price-change events as marked duration process and employ asymmetric ACD model to capture the asymmetric dynamics of price-change events. We compare the daily volatilities estimated using asymmetric ACD model and other volatility estimators such as bi-power variation estimator and duration-based estimate of RV. The asymmetric ACD volatility estimate obtains high correlation with RV estimates. A potential research direction of ICV class estimates is intraday volatility prediction.

C921: **Neglecting structural breaks in DCC models**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Co-authors: Andreea Halunga
Parameter regime changes in dynamic conditional correlation (DCC) models and corrected DCC (cDCC) are considered. It is shown that if these parameter regime changes are not accounted for in estimations, it causes the sum of the estimated short and long run persistence parameters to converge to one. This is usually taken as evidence of high persistence in conditional correlations, but in the case of neglected parameter changes, it leads to spurious inferences. Results suggest that not accounting for shifts in the unconditional correlations of the stochastic processes may bias upward short and long run estimates of persistence in correlation. Thus, they contaminate the usefulness of the DCC models in situations in which the degree of permanence is important.

Sunday 02.12.2012 08:45 - 10:25 Parallel Session I – ERCIM

ES10 Room 9 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING II **Chair: Jim Griffin**

E232: **Approximate Bayesian approaches in quantile regression for panel data**
Presenter: **Brunero Liseo**, Sapienza, Italy
Co-authors: Antonio Pulcini

Availability of repeated measurements on the same unit are nowadays common in many statistical applications and (generalized) linear mixed models represent the standard tool in statistical analysis. However it is very common the case that the covariates pattern helps to explain not only changes in location (i.e. mean or median) of the response variable; more generally it can modify the entire distributional shape and this can happen both at a given observation time and dynamically as well. In this situation a quantile regression (QR) approach may be more adequate. Standard parametric QR approaches are based on the asymmetric absolute loss function, which enables a maximum likelihood solution based on the use of the asymmetric Laplace distribution (ALD). We propose a novel nonparametric approach based on the so called Substitution Likelihood. We generalize previous results to the panel data case and we show how our method can easily deal with the non-crossing condition among different quantiles. Posterior computations are performed via an adaptive MCMC algorithm The performance of the proposed approach is evaluated with simulated data sets and compared with other existing methods.

E182: **Recursive partitioning and Bayesian inference on conditional distributions**
Presenter: **Li Ma**, Duke University, United States

A Bayesian framework is introduced for nonparametric inference on conditional distributions in the form of a prior called the conditional optional Polya tree. The prior is constructed based on a two-stage nested procedure, which in the first stage recursively partitions the predictor space, and then in the second generates the conditional distribution on those predictor blocks using a further recursive partitioning procedure on the response space. This design allows adaptive smoothing on both the predictor space and the response space. We show that this prior obtains desirable properties such as large support, posterior conjugacy and weak consistency. Moreover, the prior can be marginalized analytically producing a closed-form marginal likelihood, and based on the marginal likelihood, the corresponding posterior can also be computed analytically, allowing direct sampling without Markov Chain Monte Carlo. Our prior serves as a general tool for nonparametric inference on conditional distributions. We illustrate its work in density estimation, model selection, hypothesis testing, and regression through several numerical examples.

E248: **Bayesian nonparametric inference with continuous-state dependent processes**
Presenter: **Matteo Ruggiero**, University of Torino, Italy

Some recent advances on continuous-time continuous-state extensions of some widely used random probability measures are reviewed. Applications to Bayesian inference for discretely observed continuous-time quantities are considered. These include diffusive mixture models based on geometric stick-breaking diffusions and on a diffusive extension of the Dirichlet process.

E193: **Dependent Dirichlet processes and application to ecological data**
Presenter: **Julyan Arbel**, CREST - Universite Paris Dauphine, France
Co-authors: Kerrie Mengersen, Judith Rousseau

A new class of dependent Dirichlet processes is presented. The considered processes are an extension of Dirichlet processes that make use of covariates. It allows nonparametric density regression, i.e. regression varying with a covariate index. The dependent Dirichlet process is defined with varying weights and clusters (two infinite dimensional parameters of a Dirichlet process). It is novel in the way we define the weights, and it allows a sensible tuning of the dependence of the resulting predictor-dependent probability distributions. This construction results marginally in a Dirichlet process prior for the random measure at any specific predictor value. A blocked Gibbs sampler is developed for posterior computation. The method is illustrated with the study of abundance and diversity of Antarctic ice microbes with varying environmental covariates.

ES20 Room 10 ADAPTIVE MONTE CARLO **Chair: Nicolas Chopin**

E512: **Locally adaptive Markov chain Monte Carlo**
Presenter: **Anthony Lee**, University of Warwick, United Kingdom
Co-authors: Christophe Andrieu, Arnaud Doucet

The use of auxiliary variables in various Monte Carlo methods has proliferated both explicitly and implicitly over the last two decades, as our understanding of how to devise effective algorithms has grown. In addition, massively parallel 'many-core' processors have become the focus of the high performance computing community, for a variety of physical reasons, providing a strong incentive for algorithms in computational statistics to exhibit specific types of parallelism. Within the field of Monte Carlo methodology, population-based methods such as sequential Monte Carlo, parallel tempering and pseudo-marginal methods are promising auxiliary variable algorithms that can take advantage of available parallel resources while allowing the advantageous, principled interaction of simulated random variables. A perspective on auxiliary variables within reversible Markov chain Monte Carlo (MCMC) kernels that allows for the flexible construction of population-based MCMC kernels is described. One opportunity the methodology presents is the "local adaptation" of MCMC kernels with a given equilibrium distribution, in that some aspects of the kernels are automatically adapted to the target distribution of interest. Such kernels hold some promise in a variety of applications.

E533: **On the convergence properties of the Wang-Landau algorithm**
Presenter: **Robin Ryder**, Universite Paris-Dauphine, France
Co-authors: Pierre Jacob

The Wang-Landau algorithm is an adaptive MCMC algorithm which generates a Markov chain designed to move efficiently in the state space, by constantly penalizing already-visited regions. It hence falls into the class of exploratory algorithms, especially when the chosen regions correspond to different levels of density values. We explore two novel aspects of the Wang-Landau algorithm. First, we show that the algorithm reaches the so-called Flat Histogram criterion in finite time, which ensures convergence properties. Second, we examine the effect of using multiple chains, interacting through a common component. We shall present a theoretical study of the effect of parallelization using Feynman-Kac semigroups.

E557: **Optimal implementation of SMC methods for general Metropolis proposals**
Presenter: **Michael Pitt**, University of Warwick, United Kingdom
Co-authors: Robert Kohn, Arnaud Doucet

The optimal number of samples used in the construction of the unbiased likelihood estimator is first considered. Essentially, a trade-off is needed. Many samples for the estimator will result in a MCMC scheme which has similar properties to the case where the likelihood is exactly known but will be expensive. Few samples for the construction of the estimator will result in a faster estimation but at the expense of slower mixing in the resulting Markov chain. We consider general proposals in a Metropolis scheme which would result in a finite variance (for a given function of the parameters) if the likelihood were known exactly. We find that the performance of the unbiased likelihood MCMC method is governed by the

standard deviation of the log of the simulated likelihood (at a particular parameter ordinate) and that the optimal value of this standard deviation is around 1.

E595: **Adaptive projection Monte Carlo to estimate a Bayesian semiparametric multiplicative error model for daily realized volatility**
Presenter: **Antonietta Mira**, University of Lugano, Switzerland
Co-authors: Reza Solgi
A Markov chain Monte Carlo augmentation scheme involving adaptive projection on the space of identifiable models is proposed to estimate a semi parametric multiplicative error model (MEM) for daily realized volatility. In traditional MEM, the innovations are typically assumed to be Gamma distributed (with one free parameter that ensures unit mean of the innovations and identifiability of the model), however empirical investigations unveil the inappropriateness of this choice. In the proposed approach, the conditional mean of the time series is modeled parametrically, while we model its conditional distribution nonparametrically by Dirichlet process mixture with the Gamma distribution as its kernel. This model is applied to the time series of daily realized volatility of some indices, and is compared to similar parametric models available. Our simulations and empirical studies show better predictive performance, flexibility and robustness to misspecification of our Bayesian semi parametric approach.

ES35 Room 8 MODEL CHOICE AND FUNCTION SPECIFICATION TESTS IN SEMIPARAMETRIC REGRESSION Chair: Thomas Kneib

E123: **Restricted likelihood ratio testing in linear mixed models with general error covariance structure**
Presenter: **Sonja Greven**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Andrea Wiencierz, Helmut Kuechenhoff
The problem of testing for zero variance components in linear mixed models with correlated or heteroscedastic errors is considered. In the case of independent and identically distributed errors, a valid test exists, which is based on the exact finite sample distribution of the restricted likelihood ratio test statistic under the null hypothesis. We propose to make use of a transformation to derive the (approximate) null distribution for the restricted likelihood ratio test statistic in the case of a general error covariance structure. The method can also be applied in the case of testing for a random effect in linear mixed models with several random effects by writing the model as one with a single random effect and a more complex covariance structure. The proposed test proves its value in simulations. It is finally applied to the question of whether subjective well-being as a function of age is different from a quadratic or cubic polynomial as is typically used in the research of well-being economics.

E148: **Practical variable selection for generalized models**
Presenter: **Giampiero Marra**, University College London, United Kingdom
The problem of variable selection is considered within the class of generalized models, when there are many covariates to choose from but the number of predictors is still somewhat smaller than the number of observations. Two simple but effective shrinkage methods and an extension of the nonnegative garrote estimator are introduced. The proposals allow component selection to be carried out in one single step as opposed to many selection procedures which involve a search among many possible models. The empirical performance of the proposed methods is compared to that of some available techniques via a simulation study. The results show under which conditions one method can be preferred over another, hence providing applied researchers with some practical guidelines.

E235: **A volume-of-tube based specification test for penalized splines estimators**
Presenter: **Manuel Wiesenfarth**, University of Mannheim, Germany
Co-authors: Tatyana Krivobokova, Stephan Klasen, Stefan Sperlich
A simple and fast approach is proposed to testing polynomial regression versus a quite general nonparametric alternative modeled by penalized splines. For the construction of the test we exploit novel results on simultaneous confidence bands using the approximation to the tail probability of maxima of Gaussian processes by the volume-of-tube formula in combination with the mixed models representation of penalized splines. In simulations we show that the proposed test not only performs competitively compared to restricted likelihood ratio tests, but also allows the incorporation of smooth curves that enter an additive model, are spatially heterogeneous and are estimated from heteroscedastic data. Further advantages include very good small sample properties and the analytical availability, i.e. no computationally intensive procedures such as bootstrapping are necessary. We apply the method to analyze determinants of child undernutrition in Kenya with special interest in the highly nonlinear age pattern as well as in possible nonlinearities in the effects of the mother's body mass index and height. The method is implemented in the R package AdaptFitOS.

E501: **Gradient boosting with kernel-based regularized least squares base-learners**
Presenter: **Matthias Schmid**, University of Erlangen-Nuremberg, Germany
Co-authors: Robert Hable
Regularized kernel methods have become an important tool to model the functional relationships between the outcome variable and a set of predictor variables in regression and classification problems. Specifically, kernel methods constitute an efficient way to incorporate complex interaction terms into the prediction function. We investigate the properties of regularized kernel methods when used as base-learners in gradient boosting algorithms. Because gradient boosting results in an additive combination of the base-learners, the proposed approach can be used to derive structured additive prediction functions containing higher-order interaction effects.

ES41 Room 13 ADVANCES IN ROBUST DATA ANALYSIS II Chair: Luis Angel Garcia-Escudero

E405: **Robust text mining**
Presenter: **Domenico Perrotta**, EC Joint Research Centre, Italy
Co-authors: Marco Turchi, Marco Riani, Andrea Cerioli
The Forward Search approach to robust data analysis is extended to address problems in text mining. In this domain datasets are collections of an arbitrary number of documents represented as vectors of thousands of elements according to the so called Vector Space Model. Since the number of variables is much larger than the dataset size, the traditional Mahalanobis metric cannot be used as a similarity distance between documents. Instead, the cosine (dis)similarity (the angle between two vectors) is used. We show that by monitoring this measure with the Forward Search it is possible to perform robust estimation for a document collection and order the documents so that the most dissimilar (possibly outliers, for that collection) are left at the end. We show how to use appropriate confidence bands for this measure to detect outliers and reveal the presence of more groups of documents in the collection.

E927: **Sparse and robust partial least squares regression**
Presenter: **Valentin Todorov**, UNIDO, Austria
Co-authors: Peter Filzmoser
A suitable alternative to ordinary least squares regression for analyzing high dimensional data is the partial least squares (PLS) regression. A disadvantage of PLS, similar to PCA, is that the obtained latent variables will be linear combinations of all original predictors, which would severely hamper the interpretability of the resulting model. To cope with this problem a sparse version of PLS (SPLS) can be used which simultaneously

achieves good prediction performance and improved interpretability due to the obtained sparse linear combinations of the original variables. Since practical data often contains outliers the results of both standard and SPLS can be misleading. A solution to this problem in the case of standard PLS already exists. Here we will address the robustness of SPLS and will propose to replace the dimension reduction step by sparse and robust PCA carried out on the sample cross-covariance matrix between the response variables and the predictors. We discuss the selection of the tuning parameters and show the robustness of the proposed approach on simulation and real data sets. The proposed methods are implemented in the R package **rrcovHD** available at CRAN.

E1034: **New directions in TCLUST methodology**
Presenter: **Agustin Mayo-Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Carlos Matran
The aim is to enlarge the TCLUST methodology in order to include multivariate models with skew-symmetric populations or common principal components. Theoretical properties and the robustness of estimators will be analyzed. We will show empirical evidence of the appropriateness of the methodology. These modifications will be incorporated into TCLUST R package.

E1038: **Estimating hospital expected costs with censored data and outliers**
Presenter: **Alfio Marazzi**, Centre Hospitalier Universitaire Lausanne, Switzerland
Co-authors: Isabella Locatelli
The problem of estimating the expected hospital cost of stays of a class of patients (e.g., a diagnosis related group) as a function of the patient characteristics is considered. The statistical analysis is complicated by the asymmetry of the cost distribution, the possibility of censoring on the cost variable, and the occurrence of outliers. These problems have often been treated separately, and a method offering a joint solution to all of them is still missing. Indirect procedures have been proposed, which consistently treat censoring on the cost scale by combining an estimate of the duration distribution with an estimate of the conditional cost for a given duration. We propose a parametric version of this approach, allowing for asymmetry and censoring in the cost distribution and providing a mean cost estimate which is robust in the presence of extreme values. In addition, the new method takes covariate information into account.

ES46 Room 11 RECENT ADVANCES IN SMALL AREA ESTIMATION Chair: Isabel Molina

E271: **Studying the penetration of information technology in establishments**
Presenter: **Lola Ugarte**, Universidad Publica de Navarra, Spain
Co-authors: Ana F. Militino, Tomas Goicoa
Knowledge of the current state of the art in information technology (IT) of businesses in small areas is very important for Central and Local Governments, markets, and policy-makers because information technology allows us to collect information, to improve the access to information, and to be more competitive. Information about IT is obtained through the Information Technology Survey which is a common survey in Western countries. We focus on developing small area estimators based on a double logistic model with categorical explanatory variables to obtain information about the penetration of IT in the Basque Country establishments in 2010. Auxiliary information for population totals is taken from a Business Register. A model-based bootstrap procedure is also given to provide the prediction MSE.

E503: **Small area estimation of poverty indicators under spatio-temporal area-level models**
Presenter: **Domingo Morales**, Universidad Miguel Hernandez de Elche, Spain
Co-authors: Isabel Molina, Yolanda Marhuenda
Small area estimation is studied under a spatio-temporal area-level models. Model fitting based on restricted maximum likelihood is described and empirical best linear unbiased predictors are derived under the model. A parametric bootstrap procedure is proposed for the estimation of the mean squared error of the small area estimators. The spatio-temporal model is compared with simpler models through simulation experiments, analyzing the gain in efficiency achieved by the use of the more complex model. The performance of the parametric bootstrap estimator of the mean squared error is also assessed. An application to the Spanish Income and Living Conditions survey data is carried out to obtain estimates of poverty indicators for Spanish provinces at year 2008, making use of survey data from years 2004-2008.

E523: **Semiparametric Bayesian small area estimation based on the Dirichlet process prior**
Presenter: **Silvia Polettini**, Sapienza Universita di Roma, Italy
Small area estimation relies heavily on mixed effects regression models. The default normality assumption for the random effects is difficult to check. Their distribution may be multimodal due to missing covariates, or skewed. Accurate estimation of the random effects is crucial for small area prediction. Although point prediction is robust from deviations from normality, precision may be affected; estimation of nonlinear functionals may also be affected. Consequently, it would be useful to rely on a model having a flexible specification of the random components. A semiparametric Bayesian area level linear mixed effects model extending the Fay-Herriot model is considered. The random effects are modelled through a Dirichlet process with precision parameter M and base measure ϕ, assumed here to be a normal distribution. It is important to assess the estimator with respect to measures of variability and coverage. Those are the features where the effect of a flexible specification of the random effects is expected. The model is compared with the EBLUP and the corresponding standard hierarchical Bayesian (HB) model. Viability of the approach wll be investigated, in particular with the aim of understanding under which conditions higher precision of small area predictions may be achieved.

E684: **Estimation of labour force indicators in counties of Galicia using a multinomial mixed model with correlated time and area effects**
Presenter: **Esther Lopez-Vizcaino**, Instituto Galego de Estatistica, Spain
Co-authors: M^{a} Jose Lombardia, Domingo Morales
The unemployment rate in Spain in the fourth quarter of 2011 reached 22.9%, 13 points higher than in 2008. In Galicia (Spain) the labour market situation is not very different, the unemployment rate is 18.3% almost 10 points higher than in 2008 and unemployed people has increased in 132 000 people. The Spanish government in 2012 predicts an unemployment rate of 24.3% with a destruction of 630 000 jobs. In this situation, besides having global measures to measure unemployment, it is necessary to have indicators, like totals of employed, unemployed and inactive people or unemployment rates, to assess the impact of employment policy at local level. The objective is to estimate these indicators in counties of Galicia using small area estimation techniques. We propose a multinomial logit mixed model with correlated time and area effects to obtain small area estimates of labour force totals and unemployment rates in Galician counties. Model-based estimates tend to be biased, so the estimates' accuracy is a fundamental issue. For this reason we calculate the mean squared error and we propose estimation procedures by using analytical expressions and parametric bootstrap.

ES81 Room 7 STATISTICS IN FUNCTIONAL AND HILBERT SPACES II Chair: Gil Gonzalez-Rodriguez

E616: **Frequency domain analysis of stationary functional data**
Presenter: **Victor Panaretos**, EPFL, Switzerland

The problem of drawing statistical inferences on the second-order structure of weakly dependent functional time series is considered. Much of the research in functional time series has focused on inference for stationary time series that are linear. We consider the problem of inferring the complete second-order structure of stationary functional time series without any structural modeling assumptions. Our approach is to formulate a frequency domain framework for weakly dependent functional data, employing suitable generalisations of finite-dimensional notions. We introduce the basic ingredients of such a framework, propose estimators, and study their asymptotics under functional cumulant-type mixing conditions.

E698: **Bayesian bandwidth estimation for a nonparametric functional regression model with unknown error density**
Presenter: **Han Lin Shang**, Monash University, Australia

Error density estimation in a nonparametric functional regression model with functional predictor and scalar response is considered. The unknown error density is approximated by a mixture of Gaussian densities with means being the individual residuals, and variance as a constant parameter. This proposed mixture error density has a form of a kernel density estimator of residuals, where the regression function is estimated by the functional Nadaraya-Watson estimator. A Bayesian bandwidth estimation procedure that can simultaneously estimate the bandwidths in the kernel-form error density and the functional Nadaraya-Watson estimator is put forward. We derive a kernel likelihood and posterior for the bandwidth parameters under the kernel-form error density. A series of simulation studies show that the proposed Bayesian estimation method performs on par with the least-squares cross validation for estimating the regression mean function, but it performs better than the likelihood cross validation for estimating the regression error density. We apply the proposed Bayesian procedure to a nonparametric functional regression model, where the functional predictors are spectroscopy wavelengths and the scalar responses are fat/protein/moisture content, respectively. Because the proposed approach allows the estimation of error density, it provides a nonparametric means of constructing prediction interval for the response variable.

E991: **Penalized scalar on function regression with interaction term**
Presenter: **Karen Fuchs**, LMU Munich - Siemens AG Munich, Germany
Co-authors: Fabian Scheipl, Sonja Greven, Evamaria Stuetz

A generalized regression model for scalar responses with functional covariates, including linear functional terms as well as a functional interaction term, is presented. We extend a previous work, which includes main effects, but no interactions, of functional covariates. Coefficient functions are estimated using basis expansions, supposing a regular grid for the observation points of the covariates, and maximization of a log-likelihood, which is penalized to impose smoothness. The respective smoothing parameters for the penalties are estimated from the data, e.g. via generalized cross-validation. Open source software of the R-package mgcv can be used to do this. Further functional or scalar terms as well as functional interactions of higher order can be added within the same framework. We test the goodness of estimation for our approach in simulations. Our work is motivated by and applied to cell chip sensor data, where three functional signals are measured over time. The main aim is to predict the concentration of a chemical substance in the cell chip medium.

E504: **Bivariate functional depth: Application to biomechanics**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: Alvaro F. Page

The aim is to introduce and apply the bivariate functional depth, where the data consists of pairs of functions. The proposed depth is simple, computationally effective and well-behaved, satisfying the desirable properties for functional depths. Furthermore, it can be easily extended to compute the depth of a finite number of functions simultaneously. Specifically, we use it to compute the depth of the pairs formed by a function and its derivative. This is an important topic in Biomechanics where the use of functional data is getting common as they allow for a complete description of the human movements. We focus on the phase portraits data due to their interest in the analysis of movements as they represent the relationship between the movement variables and their derivatives. Particularly we apply the proposed depth to cervical motion data.

ES59 Room 12 NEW DEVELOPMENTS IN QUANTILE REGRESSION Chair: Melanie Birke

E204: **Significance testing in quantile regression**
Presenter: **Melanie Birke**, University of Bayreuth, Germany
Co-authors: Stanislav Volgushev, Holger Dette, Natalie Neumeyer

The problem of testing significance of predictors in multivariate nonparametric quantile regression is considered. A stochastic process is proposed, which is based on a comparison of the responses with a nonparametric quantile regression estimate under the null hypothesis. It is demonstrated that under the null hypothesis this process converges weakly to a centered Gaussian process and the asymptotic properties of the test under fixed and local alternatives are also discussed. In particular, we show, that - in contrast to the nonparametric approach based on estimation of L^2-distances - the new test is able to detect local alternatives which converge to the null hypothesis with any rate $a_n \to 0$ such that $a_n\sqrt{n} \to \infty$ (here n denotes the sample size). We also present a small simulation study illustrating the finite sample properties of a bootstrap version of the corresponding Kolmogorov-Smirnov test.

E322: **Model selection via Bayesian information criterion for quantile regression models**
Presenter: **Hohsuk Noh**, Universite catholique de Louvain, Belgium
Co-authors: Eun Ryung Lee

Bayesian Information Criterion (BIC) is well known to identify the true model consistently as long as the predictor dimension is finite. Although this property of BIC has been well studied in many kinds of regression models, most works have been done in mean and robust regression, not in general quantile regression. In this paper we study model selection consistency of BIC in quantile regression models including parametric and nonparametric ones. Further, we show how it can be used for a selection of a regularization parameter of penalized approach which is designed to obtain results of variable selection and shrinkage estimators simultaneously. Finally, we illustrate the established theoretical results via some simulated examples and a real data analysis on economic growth.

E490: **Quantile regression in spatially varying coefficient models for neuroimaging data**
Presenter: **Linglong Kong**, University of Alberta, Canada
Co-authors: Hongtu Zhu

Neuroimaging studies aim to analyze imaging data with complex spatial patterns in a large number of locations (called voxels) on a two-dimensional (2D) surface or in 3D volume. We propose a multiscale adaptive composite quantile regression model (MACQRM) that has four attractive features: being robustness, being spatial, being hierarchical, and being adaptive. MACQRM utilizes imaging observations from the neighboring voxels of the current voxel and borrows strength from the nearby quantile regressions of the current regression to adaptively calculate parameter estimates and test statistics. Theoretically, we establish consistency and asymptotic normality of the adaptive estimates and the asymptotic distribution of

the adaptive test statistics. Our simulation studies and real data analysis on ADHD confirm that MACQRM significantly outperforms MARM and conventional analyses of imaging data.

E690: **Adaptive semiparametric M-quantile regression**

Presenter: **Fabian Sobotka**, University Goettingen, Germany

Co-authors: Nicola Salvati, Giovanna Ranalli, Thomas Kneib

M-Quantiles combine the robustness and interpretability of quantiles with the flexibility and intuitive estimation of expectiles. They allow for an iteratively weighted least squares estimation including quadratic penalties to incorporate a semiparametric model. The inclusion of p-splines and spatial effects, like Markov random fields, is possible. And by definition their estimate is still robust against outliers. However, this is only true for homoscedastic scenarios. In heteroscedastic cases the distinction between outliers and "trustable" observations is likely to fail. Here, we introduce adaptive M-Quantile regression models to overcome this problem by replacing the tuning constant of the M-Quantile estimation with a robustness curve. The latter is constructed from the scale part of a location-scale model. Our findings will be analysed in a simulation study and made available as R-package "Mreg".

Sunday 02.12.2012 10:55 - 13:00 Parallel Session J – CFE

CS14 Room 1 MODELLING WITH HEAVY TAILS: APPLICATIONS **Chair: Svetlana Makarova**

C171: **On the European sovereign CDS prices**
Presenter: **David Veredas**, ECARES - University of Brussels, Belgium
Credit Default Swaps should reflect the probabilities of tail events in the underlying security. The higher its price the higher the probability of a tail event. We investigate if the yields of European Sovereign bonds CDS do explain the probability of tail events in the bonds. If not, then CDS prices are driven by other factors, such as speculation and irrational behavior. To measure these relations we rely on TailCoR, a new measure of tail association introduced recently.

C320: **Tempered Ornstein-Uhlenbeck processes: A practical view**
Presenter: **Michele Leonardo Bianchi**, Bank of Italy, Italy
Co-authors: Svetlozar T. Rachev
One dimensinal Ornstein-Uhlenbeck (OU) processes with marginal law given by tempered stable (TS) and tempered infinitely divisible (TID) distributions are studied. In general, the use of non Gaussian OU processes is obstructed by the difficulty of calibrating and simulating them. For this reason, we investigate transition law between consecutive observations of tempered OU processes and we evaluate the characteristic function of integrated tempered OU processes in three particular cases - classical tempered stable (CTS), rapidly decreasing tempered stable (RDTS) and variance gamma (VG) - with a view toward practical applications. Then, we analyze how one can draw random samples from this class of processes by taking into consideration both the classical inverse transform algorithm and an acceptance-rejection method based on the simulation of stable random sample. Finally, by considering a maximum likelihood estimation method based on the classical fast Fourier transform, we conduct an empirical investigation to assess the simulation algorithm performance.

C356: **Generating tempered stable random variates from mixture representation**
Presenter: **Piotr Jelonek**, University of Leicester, United Kingdom
A new method of random number generation for tempered stable distribution is presented. This method is easy to implement, faster than other available approaches when tempering is moderate and more accurate than the benchmark. All the results are given as parametric formulas that may be directly used by practitioners.

C362: **Simulated minimum distance estimators in macroeconomics**
Presenter: **Wojciech Charemza**, University of Leicester, United Kingdom
Co-authors: Svetlana Makarova, Yuanyuan Wang
The problem of parameters estimation and distributional selection is tackled for the case where the density function either does not exist in a closed form or is awkward to compute. It is suggested that in such cases the density function should be simulated and the parameters estimated through minimization of a distance criteria through an iterative grid search. Such an approach, although computationally expensive, guarantees finding a minimum for a unimodal monotonous function. Two types of distributions are analysed: tempered stable and skewed normal. It is shown through numerical experiments that the estimator proposed is relatively precise, in terms of average bias and root mean square error, and the application of the minimum distance criteria does not lead to distributional misspecification.

C364: **Testing exponentiality versus Pareto via likelihood ratio**
Presenter: **Anna Panorska**, University of Nevada, United States
Co-authors: Tomasz Kozubowski, Fares Qeadan, Alexander Gershunov, Debra Rominger
The problem of maximum likelihood estimation of the parameters of the Pareto Type II (Lomax) distribution is considered. We show that in certain parametrization and after modification of the parameter space to include exponential distribution as a special case, the MLEs of parameters always exist. Moreover, the MLEs have a non-standard asymptotic distribution in the exponential case due to the lack of regularity. Further, we develop a likelihood ratio test for exponentiality versus Pareto II distribution. We emphasize that this problem is non-standard, and the limiting null distribution of the deviance statistic in not chi-square. We derive relevant asymptotic theory as well as a convenient computational formula for the critical values for the test. An empirical power study and power comparisons with other tests are also provided.

CS49 Room 2 MULTIVARIATE TIME SERIES **Chair: Marco Reale**

C426: **On the approximation of long-memory process by ARMA models**
Presenter: **Davide Delle Monache**, Queen Mary University London, United Kingdom
A long-memory process can be well approximated by ARIMA models. We show how this is true only asymptotically and the sample properties show those specifications are not very attactive. We propose an autoregressive approximation based on state-space methods. In particular, we modify the truncated autoregressive method by an optimal truncated approximation. We show the forecasting performances and the estimation properties of our approach in small samples as well as asymptotically. Our approach can be related to the Durbin-Levinson algorithm and to the Haslett-Raftery method, however the advantages of this approach are twofold: a lower degree of numerical complexity and the use of state-space method allows us to deal with signal extraction and missing observations. Finally, we show how we can perform the L_q-likelihood approach, based on iterative re-weighted schema, in order to be robust to different types of contaminations.

C514: **The combined use of CVAR and BEKK models for portfolio selection of Italian stock-market**
Presenter: **Andrea Pierini**, Roma Tre, Italy
Co-authors: Marco Reale, Alessia Naccarato
The use of bivariate cointegrated vector autoregressive models (CVAR) and Baba-Engle-Kraft-Kroner (BEKK) models is suggested for the selection of a stock portfolio (Markowitz type portfolio) based on estimates of average returns on shares and volatility of share prices. In other words, we address the problem of estimating average returns and the associated risk on the basis of the prices of a certain number of shares over time. The estimate is then used to identify the assets offering the best performance and hence constituting the best investments. The use of VAR(1) model is a common practice; we suggest instead VEC models, which make it possible to take into account any cointegration between the series employed and the market trend as measured by means of the Equity Italy Index. The model put forward is applied to a series of data regarding the prices of 150 best capitalized shares traded on the Italian stock market (BIT) between 1 January 1975 and 31 August 2011; it takes into account the intrinsic value of the stocks to select the best performing ones. Moreover, we find the efficient portfolio minimizing the risk, that is to say analyzing the efficient frontier.

C950: **Modeling cointegrated time series with vector ARIMA models**
Presenter: **Jose Gallego**, Universidad de Cantabria, Spain
Co-authors: Carlos Diaz
The analysis of multiple time series with vector ARIMA models is extended to deal with cointegration. The role of the AR and MA matrix polynomials in the cointegration analysis is discussed and used to propose modeling procedures consisting of data simulation, estimation and inference in cointegrated vector ARIMA models. A multi-platform software has been developed to build this class of models.

C700: **Testing the constancy of conditional correlations in multivariate type-GARCH models**
Presenter: **Bilel Sanhaji**, Aix Marseille University, France
Co-authors: Anne Peguin Feissolle
Two multivariate constant conditional correlation tests that require little knowledge of the functional relationship determining the conditional correlations are introduced. The first test is based on artificial neural networks, which are universal approximators. The second one is based on a Taylor expansion of each conditional correlation around a given point in a sample space. These new tests are easy to implement and can be seen as general misspecification tests of a large set of GARCH-type multivariate models. We investigate the size and the power of these tests by Monte Carlo experiments by comparing them to other constant conditional correlation tests.

C545: **Sparse structural VAR's and multiple testing**
Presenter: **Marco Reale**, University of Canterbury, New Zealand
Co-authors: Shaodan Lin, Patrick Wongsaart
Graphical modeling can be used to build a sparse structural VAR model (SVAR). The presence of each variable in the SVAR is initially assessed by testing the significance of its partial correlation. However given the many partial correlations that need to be tested this leads to a multiple testing problem. The problem will be illustrated as well as possible alternative strategies. A comparison of the different strategies will be illustrated through a simulation study.

CS52 Room 5 CREDIT RISK Chair: Simona Sanfelici

C984: **Dynamic model of losses on a large mortgage portfolio**
Presenter: **Petr Gapko**, Charles University in Prague, Czech Republic
Co-authors: Martin Smid, Jan Vorisek
A new credit risk model based on the standard approach by Merton and Vasicek is introduced, which we generalize the latter in several ways. First, additionally to default rates (DR), it involves the losses given default (LGD). Second, the model is dynamic, allowing a general dependency structure of the factors driving DR and LGD. Finally, our model is multi-generational in the sense that it keeps track of individual generations of debtors according to the origination of their debt. A likelihood function of the model is formulated; as the model is only partially tractable, Monte Carlo simulation techniques are used to compute the likelihood function and transformations of the input data into the underlying factors. We apply the model using a nationwide portfolio of US mortgages assuming that the factors follow a VECM model with several macroeconomic indicators as explanatory variables. We demonstrate that our model is able to describe the evolution of credit risk more accurately than standard techniques and can therefore save partly the capital held against the credit risk. Moreover, as our model relates the credit variables to the macroeconomic environment, it is also able to capture the negative effect of the credit crisis in 2008-2009. Moreover, as the transformation between DR and LGD and the factors is nonlinear, the model is able to explain both an unexpected strength and the speed of the crisis.

C258: **A comprehensive framework for bilateral collateralized CVA and funding costs**
Presenter: **Andrea Pallavicini**, Banca IMI, Italy
Co-authors: Daniele Perini, Damiano Brigo
A bilateral collateralized counterparty valuation adjusted pricing equation by resorting to a risk-neutral approach is presented. It allows us to price a deal while taking into account credit and debt valuation adjustments (CVA, DVA) along with margining and funding costs, all in a consistent way. We find that the equation has a recursive form, making the introduction of a purely additive funding valuation adjustment (FVA) difficult. Yet, we can cast the pricing equation into a set of iterative relationships which can be solved by means of standard least-square Monte Carlo techniques. We consider realistic settings and include in our models the common market practices suggested by ISDA documentation, without assuming restrictive constraints on margining procedures and close-out netting rules. In particular, we allow for asymmetric collateral and funding rates, and exogenous liquidity policies and hedging strategies. Re-hypothecation liquidity risk and close-out amount evaluation issues are also covered. Finally, relevant examples of non-trivial settings illustrate how to derive known facts about discounting curves from a robust general framework and without resorting to ad hoc hypotheses.

C404: **Pricing credit derivatives in a Wiener-Hopf framework**
Presenter: **Gianluca Fusai**, Cass Business School and Universita del Piemonte Orientale, United Kingdom
Co-authors: Guido Germano, Daniele Marazzina
Fast and accurate pricing techniques for credit derivative contracts are presented when discrete monitoring is applied and the underlying evolves according to an exponential Levy process. Our pricing approaches are based on the Wiener-Hopf factorization. Numerical results are shown for different Levy models in order to validate the pricing algorithm. Moreover, an analysis on the sensitivity of the probability of default and the credit spread term structures with respect to the process parameters is considered. An empirical analysis on market data is also performed

C882: **Risky coupon bond option pricing: An intensity approach**
Presenter: **Marek Kolman**, University of Economics Prague, Czech Republic
A closed-form bond option pricing formula incorporating features of interest-rate models was firstly discovered in 1989. That formula could only price options on risk-free bonds. Later on, in a framework of intensity models an option pricing model on defaultable zero-bond was developed. The unsolved issue has been an option on defaultable coupon bond in an intensity framework. The proposed model fills this gap and moreover brings an intensity term-structure into the model as a new feature improving its flexibility. It is also shown how the model can be fitted not only to default-driving factors, but also to market factors represented by a choice of interest-rate model. Last but not least, it shall be noted that the model works in an analytical environment and thus, there is no need for simulations which results in faster runtime and unique solutions.

C229: **Default probability estimation under microstructure effects**
Presenter: **Simona Sanfelici**, University of Parma, Italy
Co-authors: Flavia Barsotti
Equity returns and firm's default probability are strictly interrelated financial measures capturing the credit risk profile of a firm. In structural credit risk models, a firm's equity and debt are viewed as contingent claims that partition the asset value of the firm. Firm's asset value and volatility are not directly observed. The idea is then to use the information content of high-frequency equity prices and then back out the firm's asset volatility. Nevertheless, the market microstructure literature strongly suggests that trading noises can affect equity prices. In the specific context of structural

credit risk models, ignoring microstructure effects could non-trivially inflate one's estimates for the "true" asset volatility, and this would produce misleading estimates for default probabilities and credit spreads. We propose a Monte Carlo simulation study, trying to compare the performance of different non parametric equity volatility estimators in their capability of i) filtering out the market microstructure noise, ii) extracting the (unobservable) true underlying asset volatility level, iii) predicting default probabilities deriving from calibrating the Merton structural model. Our analysis shows that the choice of the volatility estimator can largely affect calibrated default probabilities and hence risk evaluation, leading to a significant underestimation of default probabilities.

CS62 Room 4 SPARSE BAYESIAN MODEL CHOICE Chair: Helga Wagner

C278: **Model-based clustering based on sparse finite Gaussian mixtures**
Presenter: **Gertraud Malsiner-Walli**, University Linz, Austria
Co-authors: Sylvia Fruehwirth-Schnatter, Bettina Gruen

In model-based clustering selecting a suitable number of components for a finite mixture distribution is a challenging problem. We want to contribute to this issue and propose a new Bayesian approach. We propose to deliberately overfit the mixture model by specifying K^{max} components, K^{max} being a reasonable upper bound on the number of components. Simultaneously, we specify a sparse hierarchical prior on the component weights which has the effect of emptying superfluous components during MCMC sampling. A straightforward criterion for estimating the true number of components is given by the most frequent number of nonempty groups visited during MCMC sampling. In addition, we also examine the effect of specifying a sparse hierarchical prior on the component means, namely the normal gamma prior. This reduces the MSE of estimated parameters in the presence of noise variables. We perform Bayesian estimation of finite Gaussian mixture models using MCMC methods based on data augmentation and Gibbs sampling. To obtain an identified mixture model, in a post-processing step the MCMC output is relabeled using k-centroid cluster analysis based on the Mahalanobis distance. We evaluate our proposed method in a simulation setup with artificial data and by applying it to benchmark data sets.

C496: **Time-varying sparsity in dynamic regression models**
Presenter: **Maria Kalli**, Canterbury Christ Church University, United Kingdom
Co-authors: Jim Griffin

A novel Bayesian method for dynamic regression models is proposed where both the value of the regression coefficients and the variables selected change over time. We focus on the parsimony and forecasting performance of these models and develop a prior which allows the shrinkage of the regression coefficients to change over time. An efficient MCMC method for computation is described. The new method is then applied to two forecasting problems in econometrics: equity premium prediction and inflation forecasting. The results show that this method outperforms current competing methods.

C550: **Objective Bayesian search of Gaussian DAG models with non-local priors**
Presenter: **Guido Consonni**, Universita Cattolica del Sacro Cuore, Italy
Co-authors: Davide Altomare, Luca La Rocca

Directed Acyclic Graphical (DAG) models are increasingly employed in the study of physical and biological systems. The aim is to identify the structure of the graph from data, assuming that the variables satisfy a given ordering. We propose an objective Bayesian method for searching the space of Gaussian DAG models, using *non-local* parameter priors, which are especially suited for learning sparse graphs, because in any pairwise comparison of nested models they allow a faster learning rate, relative to ordinary local parameter priors, when the true unknown sampling distribution belongs to the simpler model. We implement an efficient stochastic search algorithm, which deals effectively with data sets having sample size smaller than the number of variables, and apply our method to a variety of simulated and real data sets. Our approach compares favorably to current state-of-the-art frequentist methods relying on the assumption of ordered variables. Importantly, with reference to learning the skeleton of the DAG for unordered variables, we find that our method is still competitive, provided the ordering of the variables is only moderately misspecified.

C587: **Decoupled shrinkage and selection in linear models**
Presenter: **Carlos Carvalho**, The University of Texas at Austin, United States
Co-authors: Richard Hahn

A new variable selection approach from a fully Bayesian decision theory viewpoint is proposed. By drawing an explicit distinction between actions and inferences our method is able to effectively manage the the trade-off associated with the competing goals of predictive generalization and interpretability. By decoupling posterior learning from model reporting, our approach creates a flexible framework where sparse solutions can be obtained even when using continuous shrinkage priors.

CS65 Room 3 ADVANCES IN ASYMMETRIC AND NONLINEAR FINANCIAL ECONOMETRICS Chair: Yasuhiro Omori

C131: **Asymmetric causal relationships between the stock markets of Europe and East Asian countries**
Presenter: **Feng Yao**, Kagawa University, Japan
Co-authors: Yirong Ying, Shugeng Dai

Empirical results are shown related to the asymmetric causal characteristics of the spotlighted stock market composite indices of Europe and East Asian countries in view of the previous results concerning the Wald test of one-way effect causal measure. Based on the error correction model for non-stationary daily open and close market data, we investigate the stock market causal relationships between the United Kingdom, France, Germany in Europe, and Japan, China, Hong Kong in East Asia in time domain and frequency domain. The composite indices we discussed in the models are FTSE 100 for London Stock Exchange, CAC 40 for Euronext Paris, DAX for Frankfurt Stock Exchange, Nikkei 225 for Tokyo Stock Exchange, Shanghai Composite for Shanghai Stock Exchange, and HANG SENG Index for Hong Kong Stock Exchange. Furthermore, the long-run and short-run causal relations are also discussed by the Wald test of local one-way effect.

C197: **Stepwise local influence in GARCH models**
Presenter: **Lei Shi**, Yunnan University of Finance and Economics, China
Co-authors: Rahman Md. Mostafizur, Wen Gan

Outlier detection is important for achieving the right statistical inference and has been an intriguing topic of numerous studies. We propose a new procedure for detecting a patch of outliers or influential observations in GARCH models using stepwise local influence method. We first studied the performance of local influence analysis under innovative perturbation scheme, additive perturbation scheme and data perturbation scheme. From simulations we found that an innovative perturbation scheme gives better results than the other two perturbation schemes although this perturbation scheme suffers from masking effects. Then we used stepwise local influence method to detect group of outliers. The simulated results confirm that the new technique can successfully detect all of the outliers. At our empirical study we consider monthly returns of Dow Jones industrial average index. Real data also shows the efficiency of the stepwise local influence method to detect multiple outliers in case of GARCH model under innovative perturbation scheme.

C407: **Analysis of the dynamics of the seasonality in a time series**
Presenter: **Ming-Yen Cheng**, National Taiwan University, Taiwan
Co-authors: Yu-Chun Chen, Hau-Tieng Wu
Seasonality is a feature aimed to describe the oscillatory pattern inside an observed signal, for example, the incidence time series. Identifying seasonality from an observed noisy signal is an important issue in many scientific fields, like public health and economics. Several methods, for example, seasonal autoregressive integrated moving average (SARIMA), frequency analysis (FFT), seasonal index, singular spectrum analysis (SSA), etc, are available to identify the seasonality. However, it is not an easy task to analyze the dynamical behavior of the seasonality by these methods. The purpose is to provide a general model describing the dynamics of the seasonality and introduce an adaptive method, referred to as the Synchrosqueezing transform, to accurately extract the seasonality from a given time varying signal. The identifiability problem of the model and the robustness property of Synchrosqueezing transform to time-dependent noise is theoretically analyzed.

C548: **Tail dependence analysis for high-frequency transaction data**
Presenter: **Alexander Malinowski**, Georg-August University Goettingen, Germany
Co-authors: Martin Schlather, Zhengjun Zhang
Marked point processes (MPPs) represent a widely used tool in financial econometrics to handle irregularly spaced events, such as single transactions in high-frequency financial data. At the same time, rare extreme events and the tail behavior of financial time series are of increasing interest, especially in the context of risk measurement and risk management. Extreme value theory provides a solid basis for the statistical modeling of these events. We combine MPP theory with extreme value analysis in order to detect and to quantify interaction effects between extremes of the marks and the pattern of point locations. With new measures of interaction based on second order moment measures of MPPs, we show that the extreme value parameter of inter-transaction returns in the German stock index is locally increased by the existence of nearby transactions, independently of traded price and volume. This local conditional tail index and related measures may thus serve as additional characteristics to judge the instantaneous risk of assets and may also have an impact on trading and hedging strategies. The results are confirmed by a simulation study.

CS74 Room 6 COMMODITY MARKETS Chair: Ivan Petrella

C174: **The recent behavior of commodity prices: Fundamentals, speculative bubbles and relation to the global economic environment**
Presenter: **Isabel Figuerola-Ferretti**, Business Department, Universidad Carlos III Madrid, Spain
Co-authors: Roderick McCrorie, Chris Gilbert
The aim is to test for the existence of speculative bubbles in five key commodities – crude oil, gold, silver, aluminum and copper – over the last decade or so, focusing on possible bubble characteristics around the time of the financial crisis. We utilize a new methodology that allows for date-stamping the origination and collapse of bubbles, but in a form where critical values are made robust to allow for possibly different data spans and sampling frequencies. We find evidence of bubble behavior in the copper, gold and silver markets in the first half of 2006. Results are less conclusive for the aluminum market. We fail to find convincing evidence for a 2007-08 crude oil bubble.

C296: **Spikes and stochastic volatility in commodity prices: Evidence from crude oil futures prices using conditional moments of integrated volatility**
Presenter: **Paola Zerilli**, University of York, United Kingdom
Co-authors: Christopher Baum
The main goal is to test various possible specifications of commodity prices using intraday tick by tick crude oil prices from 2003 to 2010. We build a simple GMM estimator using the information provided by high frequency intraday futures prices by matching sample moments of the realized volatility to the corresponding population moments of the integrated volatility. As expected, we find strong evidence for jumps in futures prices and in volatility particularly around the beginning of the subprime crisis (July 2007) and around the Greek crisis (May 2010). We use the best performing models in order to produce short-run predictions. Our findings show the importance of unconstrained jumps and stochastic volatility in constructing reliable forecasts of oil price movements.

C350: **Speculation in the oil market**
Presenter: **Ivan Petrella**, Birkbeck University of London, United Kingdom
Co-authors: Luciana Juvenal
The run-up in oil prices after 2004 coincided with a growing flow of investment to commodity markets and an increased price comovement between different commodities. We analyze whether speculation in the oil market played a key role in driving this salient empirical pattern. We identify oil shocks from a large dataset using a factor-augmented autoregressive (FAVAR) model. We analyze the role of speculation in comparison to supply and demand forces as drivers of oil prices. The main results are as follows: (i) While global demand shocks account for the largest share of oil price fluctuations, financial speculative demand shocks are the second most important driver. (ii) The comovement between oil prices and the price of other commodities is explained by global demand and financial speculative demand shocks. (iii) The increase in oil prices in the last decade is mainly explained by the strength of global demand. However, financial speculation played a significant role in the oil price increase between 2004 and 2008, and its subsequent collapse. Our results support the view that the financialization process of commodity markets explains part of the recent increase in oil prices.

C848: **Modeling commodity markets: The challenge of leptokurtic return distributions**
Presenter: **Arnab Laha**, Indian Institute of Management Ahmedabad, India
Co-authors: Pravida Raja
The aim is to model leptokurtic daily log-return distributions of three commodities: gold, silver and crude oil. Three modeling approaches are tried out namely (a) a two component mixture of normal distributions model, (b) Variance Gamma (VG) distribution model and (c) Generalized Secant Hyperbolic (GSH) distribution model. The two component mixture of normal distributions model is found to be a reasonable model for daily log-returns of gold and crude oil. The VG distribution model and the GSH distribution model are found not to be suitable for modeling log-returns for any of the three commodities considered.

C936: **Estimating two-factor Gaussian term-structure model with flexible variance of measurement errors**
Presenter: **Hiroaki Suenaga**, Curtin University, Australia
A Gaussian factor model has been widely considered for analyzing term-structure of commodity prices. In calibrating model to empirical data, a measurement error is added to the estimation equation so as to allow dispersions between observed and model's implied values of financial assets. The model is characterized as a latent factor model, which is usually estimated by a filtering method. Recently, models with additional factors and/or stipulating increasingly complex stochastic process of each latent factor have been considered. However, these models still impose a simple distributional structure on the measurement error. This study estimates a two-factor Gaussian term-structure model with flexible variance structure of the measurement error and compares it with the model imposing a simple variance structure of the measurement error. Empirical applications of the model to four different commodities indicate that estimates of model parameters are affected substantially by the distributional assumption of

the measurement error. In particular, the model with simple variance structure yields smaller estimates of diffusion rate parameters than the model allowing flexible variance of the measurement error. Model's implications for variance and cross-covariance of financial assets are substantially affected by the specification of the variance structure of the measurement error.

CP01 Room Hall POSTER SESSION Chair: Maria Jesus Garcia-Ligero

C645: **Using (1+1) evolutionary algorithms to formalize the resistant to innovation behavior**
Presenter: **Elena Druica**, University of Bucharest, Romania

A large and consistent body of macroeconomic literature shows innovation as one of the main engines of economic growth. Standard economic theory pays little attention to the effects of innovation on the economic decisions of individuals. Behavioral economics, however, describes the main triggers that drive this puzzling psychology of resistance against innovation, and explains its effects on the behavior of consumers. The aim is to use (1+1) Evolutionary Algorithms in order to formally explain some of the results obtained in the area of resistant behavior. Based on the main empirical findings in the area of resistance against innovation, this research makes the link between behavioral economics and the characteristics of the algorithm. Several theoretical results are proved in order to show how a computational approach to describing real economic behavior leads to a particular context in the process of searching for an optimum. Several practical applications of the model and a few directions for further research are discussed.

C759: **Solving norm constrained portfolio optimizations via coordinate-wise descent algorithms**
Presenter: **Yu-min Yen**, Academia Sinica, Taiwan
Co-authors: Tso-jung Yen

Fast and easy-to-implement coordinate-wise descent type algorithms are developed to solve portfolio selection problems in which asset weights are constrained by L_q norms for $1 \leq q \leq 2$. A special case is when $q = 1$, in that the L_1 norm constraint promotes zero values for the weight vector, leading to an automatic asset selection for the portfolio. Two benchmark data sets are used to examine performances of the norm constrained portfolio. When the sample size is not large in comparison with the number of assets, the minimum variance portfolio with the norm constraint tends to have a lower out-of-sample portfolio variance, lower turnover rate, fewer numbers of active assets and shortsales positions, but higher Sharpe ratio than the one without such norm constraints. We then show some extensions; in particular, we derive an algorithm for solving an mvp problem in which assets are allowed to be chosen grouply.

C861: **Nonparametric betas for conditional factor models**
Presenter: **Susan Orbe**, University of the Basque Country, Spain
Co-authors: Maria Victoria Esteban

A generalization of the market model that includes time-varying beta coefficients allows us to estimate the proposed model using nonparametric techniques. Time-varying factor loadings are nonparametrically estimated by using smoothing splines and taking into account the returns conditional variance structure. The evaluation of the ability of the nonparametric approach to capture the dynamics of state-varying beta coefficients is assessed by a Monte Carlo simulation study. In particular, the CAPM model and the three-factor model of Fama and French have been simulated under several alternative specifications for the beta coefficients and the error term. More precisely, beta coefficients have been generated as linear or nonlinear functions of some state variables, and the error term from a left-skewed normal distribution and a GARCH process. Simulation results indicate that the proposed two-stage estimator is more accurate than the other alternative estimators considered here. Finally, the nonparametric method is applied to estimate the traditional CAPM and the three-factor model of Fama and French using the 25 portfolios by size and book-to-market. Results indicate that the beta coefficients corresponding to the lowest and highest book-to-market portfolios, for each size, are more volatile.

C933: **Dynamic correlations in exchange rate time series: A copula approach**
Presenter: **Krenar Avdulaj**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik

Comovement of financial markets is an important topic of international portfolio modeling with direct implications for risk diversification. Thus understanding dynamics underlying the financial market dependencies is crucial for financial practitioners. The methodology we employ has been developed recently and it extends the theory of copulas allowing for conditioning variables. It measures the dependence by the use of a two-step procedure: first modeling the marginal distributions and then utilizing the dynamic copulas. We study the dynamic dependence in exchange rates of British Pound, Euro and Swiss Franc. We evaluate the estimated dependence paths by various bivariate copula types. The approach confirms that dependence between the studied exchange rates is not constant and it evolves over time. Moreover, we use the non-parametric measure of realized correlation for comparison. Results reveal interesting tail dependence in the data accounting for the large part of the dynamics. When modeled correctly, it can improve the forecasting as well as risk management.

C987: **Realized wavelet jump-GARCH model: On the wavelet decomposition of volatility to improve forecasts**
Presenter: **Jozef Barunik**, Academy of Sciences of the CR, Czech Republic
Co-authors: Lukas Vacha

A forecasting model for volatility based on its decomposition to several investment horizons and jumps is proposed. As a forecasting tool, we utilize the Realized GARCH framework which models jointly returns and realized measures of volatility. For the decomposition, we use the jump wavelet two scale realized volatility estimator (JWTSRV). While the main advantage of our time-frequency estimator is that it provides us with realized volatility measures robust to noise, as well as with consistent estimates of jumps, it also allows us decomposing the volatility into the several investment horizons. On currency futures data covering the period of the recent financial crisis, we compare forecasts from the Realized GARCH(1,1) model using several measures. Namely, we use the realized volatility, the bipower variation, the two- scale realized volatility, the realized kernel and ourJWTSRV. We find that in-sample as well as out-of-sample performance of the model significantly differs based on the realized measure used. When the JWTSRV estimator is used, the model produces significantly best forecasts. We also utilize jumps and build a Realized Jump-GARCH model. We conclude that realized volatility measurements in the time-frequency domain and the inclusion of jumps improves the volatility forecasting considerably.

C988: **Wavelet-band least squares estimation of fractional cointegration**
Presenter: **Lukas Vacha**, UTIA AV CR, Czech Republic
Co-authors: Jozef Barunik

The aim is to develop a new tool for the estimation of fractional cointegrating relation based on wavelets, a wavelet band least squares (WBLS). The main advantage of WBLS in comparison to other frequency domain methods is that it allows us to work conveniently with potentially non-stationary time series due to the properties of wavelets. In comparison to the frequency domain least squares methods based on the Fourier transform, our methodology allows us to study the dynamics of the relationship in the time-frequency domain. In the simulation study, we compare our estimator to the narrow-band least squares (NBLS) and fully modified narrow-band least squares (FMNBLS) based on the Fourier frequencies.

C1067: **Effects of systematic risk modelling in conditional version: Empirical evidence from the Madrid stock market**

Presenter: **Maria Dolores Lagoa-Varela**, Corunna, Spain

Co-authors: Susana Iglesias Antelo

From the success of models autoregressive conditional heteroskedasticity (ARCH) in its application to capital markets, some authors have proposed the replacement of static models of asset pricing by other dynamic modeling incorporating varying risk time. In order to verify the relevance of this measure, this work is carried out in a practical application that systematic risk is estimated in two versions, classical (static) and conditional (time varying or dynamic) as a GARCH bivariate, and contrasts the CAPM with both approaches. We present an empirical application to the Madrid stock market during the period 2002-2004. The results achieved with a dynamic structure to generate returns on assets are so similar to those obtained by the conventional method that we believe may be interpreted as an indication of inefficiency ARCH modeling in this context, understanding efficiency in the economic sense, i.e. in terms of the effort required to implement a computational level is not offset by a substantial improvement in the results.

Sunday 02.12.2012 10:55 - 13:00 Parallel Session J – ERCIM

ESI02 Room Multipurpose BAYESIAN NONPARAMERICS Chair: Igor Pruenster

E399: **Adaptive-modal Bayesian nonparametric regression**
Presenter: **George Karabatsos**, University of Illinois-Chicago, United States
Co-authors: Stephen G. Walker

A Bayesian nonparametric regression model defined by a covariate-dependent infinite-mixture of unimodal densities is proposed and investigated. The mixture weights are designed to specifically account for the number of unimodal densities required to account for the variability at each point in the covariate space. Multimodal models at a covariate value lead to meaningless prediction, we argue, and hence we anticipate that a good collection of covariates would provide roughly a single unimodal density as explaining all the variability at each point in the covariate space. The anticipation is then needed to be matched by a model which specifically models the number of unimodal densities. The model is illustrated through the analysis of real and simulated data sets. The real data applications reflect a wide range of substantive areas, including meta-analysis. We also consider and illustrate a spatio-temporal extension of the regression model, having mixture weights specified by space- and time-varying coefficients.

E631: **Species sampling prior for the analysis of array CGH data**
Presenter: **Michele Guindani**, MD Anderson Cancer Center, United States
Co-authors: Thiago Costa, Federico Bassetti, Fabrizio Leisen, Edoardo Airoldi

Many popular Bayesian Nonparametric priors can be characterized in terms of exchangeable species sampling sequences. However, in some applications, exchangeability may not be appropriate. We introduce non exchangeable generalized species sampling sequences characterized by a tractable predictive probability function with weights driven by a sequence of independent Beta random variables. We compare their clustering properties with those of the Dirichlet Process and the two parameters Poisson-Dirichlet process. We propose the use of such sequences as prior distributions in a hierarchical Bayes modeling framework. We detail on Markov Chain Monte Carlo posterior sampling and discuss the resulting inference in a simulation study, comparing their performance with that of popular Dirichlet Processes mixtures and Hidden Markov Models. Finally, we discuss an application to the detection of chromosomal aberrations in breast cancer using array CGH data.

E1064: **A nonparametric approach to adaptive sampling**
Presenter: **Robert Kohn**, University of New South Wales, Australia
Co-authors: Michael Pitt, Minh Ngoc Tran

Interesting and efficient general-purpose adaptive samplers based on nonparametric fitting of multivariate mixtures of t distributions and mixtures of t analyzers are constructed. The mixtures are fitted by an efficient variational Bayes algorithm in which the number of components is determined automatically. The sampler consists of various steps including independent, correlated and random walk Metropolis Hastings steps in such a way that the adaptive chain can explore the state space efficiently both globally and locally. The adaptive sampler is designed in such a way that ergodic convergence is ensured. Performance of the proposed sampler is evaluated using simulated and real data examples both when the likelihood is available and when only an unbiased estimate is available.

ES12 Room 9 HIGH DIMENSIONAL DATA ANALYSIS AND STATISTICAL LEARNING Chair: Ejaz Ahmed

E610: **Nonlinear shrinkage estimation of large-dimensional covariance matrices**
Presenter: **Michael Wolf**, University of Zurich, Switzerland
Co-authors: Olivier Ledoit

Many statistical applications require an estimate of a covariance matrix and/or its inverse. When the matrix dimension is large compared to the sample size, which happens frequently, the sample covariance matrix is known to perform poorly and may suffer from ill-conditioning. There already exists an extensive literature concerning improved estimators in such situations. In the absence of further knowledge about the structure of the true covariance matrix, the most successful approach so far, arguably, has been shrinkage estimation. Shrinking the sample covariance matrix to a multiple of the identity, by taking a weighted average of the two, turns out to be equivalent to linearly shrinking the sample eigenvalues to their grand mean, while retaining the sample eigenvectors. This approach is extended by considering nonlinear transformations of the sample eigenvalues. We show how to construct an estimator that is asymptotically equivalent to an oracle estimator suggested in previous work. As demonstrated in extensive Monte Carlo simulations, the obtained bona fide estimator can result in sizeable improvements over the sample covariance matrix and also over linear shrinkage.

E594: **A lasso-penalized BIC for parsimonious Gaussian mixture models**
Presenter: **Sakyajit Bhattacharya** , University of Guelph, Canada

In the context of Gaussian mixture model-based clustering analysis, we propose a penalized likelihood approach with a penalty function akin to the Lasso. We use a modified AECM algorithm to fit our proposed model with some pre-defined priors. A Lasso-Penalized BIC (LPBIC) is proffered for selection of the number of components and the penalization parameter. Our approach is especially suitable for high dimension, low sample size settings. We show that this LPBIC has some useful asymptotic properties, i.e. under some regularity assumptions the criterion can correctly choose the true set of parameters. Data analyses are carried out applying the Parsimonious Gaussian Mixture Models (PGMM); BIC and LPBIC are compared for model selection.

E597: **Pooled estimation in CAR models**
Presenter: **Abdulkadir Hussein**, University of Windsor, Canada
Co-authors: Ejaz Ahmed, Marwan Al-Momani

Conditionally autoregressive (CAR) models are common in many areas of research where spatial correlations are present. In this talk, we consider CAR regression models and construct shrinkage-type estimators for the regression coefficients. We show that such estimators are more efficient in terms of asymptotic quadratic risks than those based on reduced as well as full CAR models. We illustrate the usefulness of the proposed estimators by using a real data example and by Monte Carlo comparisons with estimators obtained through the LASSO family of methods.

E614: **On random network inference using bootstrap**
Presenter: **Lilia Leticia Ramirez Ramirez**, Instituto Tecnologico Autonomo de Mexico, Mexico
Co-authors: Yulia Gel, Mary Thompson

Models based on networks have been widely employed to describe characteristics and dynamics of various phenomena, in contexts ranging from the natural and social sciences to technological systems. Examples include the study of susceptibility to failure of power grid networks and modeling the spread of infectious outbreaks. However, most of the currently developed network methodology assumes the complete knowledge of the graph or its characteristics. Alternatively, the nature of the network can be described, for example, in terms of the degree distribution of a

random network, and it has been observed that dynamics, such as resilience of the network and spread of infections, are heavily affected by this hypothesized structure. However, in most practical studies this distributional structure is inferred using only partial data from the network, which leads to additional (sampling) errors and instability of estimates. Hence, it is of key importance to develop a realistic and data-driven model for the network structure. We present a semi-parametric approach to network inference under limited information that may be viewed as a fusion of modern re-sampling procedures for time series with classical network methodology. We illustrate our idea by application to estimation of general characteristics of a degree distribution, e.g. mean degree and goodness-of-fit testing.

E297: **Perspectives on human bias versus machine bias: Weibull censored regression model**
Presenter: **Ejaz Ahmed**, Brock, Canada

The problem of estimating a vector of regression parameters in the Weibull censored regression model will be addressed. The main objective is to provide natural adaptive estimators that significantly improve upon the classical procedures in the situation where some of the predictors may or may not be associated with the response. In the context of two competing Weibull censored regression models (full model and candidate submodel), we consider an adaptive shrinkage estimation strategy that shrinks the full model maximum likelihood estimate in the direction of the submodel maximum likelihood estimate. The shrinkage estimators are shown to have higher efficiency than the classical estimators for a wide class of models. Further, we consider a LASSO type estimation strategy and compare the relative performance with the shrinkage estimators. Monte Carlo simulations reveal that when the true model is close to the candidate submodel, the shrinkage strategy performs better than the LASSO strategy when, and only when, there are many inactive predictors in the model. Shrinkage and LASSO strategies are applied to a real data set from Veteran's administration (VA) lung cancer study to illustrate the usefulness of the procedures in practice.

ES23 Room 7 PARTICLE FILTERING Chair: Veronika Czellar

E599: **SMC^2: An efficient algorithm for sequential analysis of state-space models**
Presenter: **Nicolas Chopin**, CREST-ENSAE, France
Co-authors: Pierre Jacob, Omiros Papaspiliopoulos

The generic problem of performing sequential Bayesian inference is considered in a state-space model with observation process y, state process x and fixed parameter theta. An idealized approach would be to apply an iterated batch importance sampling (IBIS) algorithm previously developed. This is a sequential Monte Carlo algorithm in the theta-dimension, that samples values of theta, reweights iteratively these values using the likelihood increments $p(y_t|y_1:t-1,\theta)$, and rejuvenates the theta-particles through a resampling step and a MCMC update step. In state-space models these likelihood increments are intractable in most cases, but they may be unbiasedly estimated by a particle filter in the $x-$dimension, for any fixed theta. This motivates the SMC^2 algorithm that will be proposed: a sequential Monte Carlo algorithm, defined in the theta-dimension, which propagates and resamples many particle filters in the $x-$dimension. The filters in the $x-$dimension are an example of the random weight particle filter. On the other hand, a particle Markov chain Monte Carlo (PMCMC) framework previosuly developed allows us to design appropriate MCMC rejuvenation steps. Thus, the theta-particles target the correct posterior distribution at each iteration t, despite the intractability of the likelihood increments. We explore the applicability of our algorithm in both sequential and non-sequential applications and consider various degrees of freedom, as for example increasing dynamically the number of $x-$particles. We contrast our approach to various competing methods, both conceptually and empirically through a detailed simulation study, included here and in a supplement, and based on particularly challenging examples.

E702: **Information content of S&P500 and VIX derivatives markets**
Presenter: **Elise Gourier**, University of Zurich - Swiss Finance Institute, Switzerland
Co-authors: Markus Leippold, Chris Bardgett

The S&P 500 and VIX markets are closely intertwined, both being strongly impacted by volatility movements and the combined speculative and hedging possibilities that they offer. A large dataset of options and index prices are used to analyze the empirical performance of affine jump-diffusion processes for the S&P 500 returns to accurately reproduce the patterns of both indices as well as consistently price S&P 500 and VIX derivatives. Based on the affine relationship of the VIX squared with respect to the latent factors, we extend the Fourier Cosine Expansion to efficiently price VIX derivatives. These tools allow us to build a fast auxiliary particle filter and investigate the usefulness of jumps and of each added factor depending on the market considered. We also analyze the out-of-sample forecasting performance of the models depending on which products and markets are considered in the in-sample estimation procedure. We find that jumps in volatility are necessary to reproduce the positive skewness observed in the VIX option market. Additionally, including the financial crisis in our sample reveals that modeling the long-term mean reversion parameter in Heston-like models as a stochastic process significantly improves the fit of the model to the time series of spot and derivatives market prices.

E963: **Fixed-parameters estimation in piecewise deterministic processes via particle MCMC**
Presenter: **Axel Finke**, University of Warwick, United Kingdom
Co-authors: Adam Johansen, Dario Spano

Filtering and fixed-parameter estimation for discretely-observed piecewise deterministic processes (PDPs) via sequential Monte Carlo methods are discussed. PDPs are stochastic processes that jump randomly at a countable number of stopping times but otherwise evolve deterministically in continuous time. For instance, PDPs can be used to model the trajectories of fighter aircraft. Here, a pilot accelerates at random times which correspond to jumps in the acceleration dynamics. Between such jumps, the trajectory of the aircraft is just a deterministic function of the aircraft's position, speed, and acceleration at the most recent jump. To this model, we apply an existing sequential Monte Carlo filter for PDPs, based on the SMC-sampler framework. In addition, we develop particle MCMC methods for fixed-parameter estimation in this context.

E820: **Multiperiod corporate default prediction with the partially-conditioned forward intensity**
Presenter: **Andras Fulop**, ESSEC Business School, France
Co-authors: Jin-Chuan Duan

A recently introduced forward intensity model has proved to be a consistent and parsimonious way to model defaults at multiple periods. However, it has a noticeable shortcoming; it does not allow for default correlation among the individual obligors. We propose a new forward-intensity approach that allows for correlation of intensities among individual obligors by conditioning the forward default intensities on common risk factors (observed and/or latent). The resulting model is richly parameterized, so we turn to a pseudo-Bayesian numerical device for parameter estimation. In particular, we use the pseudo-likelihoods through time to define a sequence of pseudo-posteriors from which one can sample by sequential Monte Carlo. The pseudo-posterior means provide consistent and asymptotic normal parameter estimates. Inference is conducted by self-normalized asymptotics using these recursive estimates. The model is implemented on a large sample of US industrial and financial firms spanning the period 1991-2009 on a monthly basis. We find that our new model is able to correct the structural biases at longer horizons previously observed. Further, allowing for default correlation is found to be crucial to describe the joint default behavior.

E818: **Robust filtering**
Presenter: **Veronika Czellar**, HEC Paris, France
Co-authors: Laurent Calvet, Elvezio Ronchetti
Filtering methods are powerful tools to estimate the hidden state of a state-space model from observations available in real time. However, they are known to be highly sensitive to the presence of small misspecifications of the underlying model and to outliers in the observation process. We show that the methodology of robust statistics can be adapted to the framework of sequential filtering. We introduce an impact function that quantifies the sensitivity of the state distribution with respect to new data. Since the impact function of a standard filter is unbounded even in the simplest cases, we propose a filter with a bounded impact function which provides accurate state and parameter inference in the presence of model misspecifications. We illustrate its good properties in several examples including linear models and nonlinear financial models of stochastic volatility.

ES25 Room 12 STOCHASTICS ORDERS AND THEIR APPLICATIONS Chair: Jose Maria Fernandez-Ponce

E133: **Multivariate convex transform orders**
Presenter: **Julio Mulero**, Universidad de Alicante, Spain
Co-authors: Felix Belzunce, Alfonso Suarez-Llorens, Jose Maria Ruiz
The study of random variables or vectors usually involves the analysis of skewness. That can be done by using single coefficients, but sometimes they are not very informative. The convex transform order is a useful tool to compare the skewness of two random variables. Different extensions to compare random vectors in these terms have been proposed. Three new generalizations to the multivariate case of the convex transform order are defined, namely, the multivariate upper orthant convex, the convex transform order and the conditional convex transform orders. Properties and relationships between these new multivariate orders and some applications are discussed too.

E146: **An application of an aging notion in a breast tumor growth study**
Presenter: **Rosario Rodriguez-Grinolo**, Universidad Pablo de Olavide, Spain
Co-authors: Jose-Maria Fernandez-Ponce, Franco Pellerey
Different aging notions have been defined and studied in reliability theory. In particular, the New Better than Used in Expectation (NBUE) concept has been defined to describe the property satisfied by an item whose expected residual lifetime is decreasing. NBUE distributions can be characterized by different stochastic orderings. We defined a multivariate version of the mean residual life function and multivariate NBUE version to be applied in a breast tumor growth study. It is well-known that a breast cancer screening has been recommended every two years in women between the ages of 50 and 74 years. Our results give clear evidence for the need to consider breast tumor screening annually in 42-year-old women.

E175: **On the Lp-metric between a probability distribution and its distortion**
Presenter: **Miguel A. Sordo**, University of Cadiz, Spain
Co-authors: Alfonso Suarez-Llorens, Miguel Lopez-Diaz
In actuarial theory, the Lp-metric is used to evaluate how well a probability distribution approximates another one. In the context of the distorted expectation hypothesis, the actuary replaces the actual probability distribution by a distorted probability, so it makes sense to interpret the Lp metric between them as a characteristic of the underlying random variable. We show that this is a characteristic of the variability of the random variable, study its properties and give some applications.

E189: **A new dispersion order based on the comparison of the tail probabilities**
Presenter: **Alfonso Suarez-Llorens**, Universidad de Cadiz, Spain
Co-authors: Miguel Angel Sordo, Marilia Souza
A new univariate dispersion order which can be interpreted in terms of the comparison of the tail probabilities is introduced. We study the relationship with other classical dispersion orders in the literature and provide some characterizations that lead to many possible comparisons. Finally, an interesting application in Risk Theory is given.

E838: **On suffcient conditions for the mean residual life and right spread orders**
Presenter: **Felix Belzunce**, Universidad de Murcia, Spain
Co-authors: Carolina Martinez-Riquelme, Jose M. Ruiz
The purpose is to show that the non-monotonicity of the ratio of survival functions or of the difference of quantile functions are still sufficient conditions for the mean residual life and right spread orders, under some additional mild conditions. Applications to compare some parametric models of distributions and generalized order statistics are provided.

ES36 Room 13 ROBUSTNESS ISSUES IN COMPLEX DATA ANALYSIS Chair: Giancarlo Ragozini

E370: **Robust archetypal analysis in R**
Presenter: **Friedrich Leisch**, Universitaet fuer Bodenkultur Vienna, Austria
Co-authors: Manuel Eugster
Archetypal analysis has the aim to represent observations in a multivariate data set as convex combinations of a few, not necessarily observed, extremal points (archetypes). The archetypes themselves are restricted to being convex combinations of the individuals in the data set and lie on the boundary of the data set. Archetypal analysis approximates the convex hull of the data set – this suggests itself that outliers have a great influence on the solution. In fact, the farther a data point is from the center of the data set, the more influence it has on the solution. Although archetypal analysis is about the data set boundary, practice has shown that in many cases one primarily is interested in the archetypes of the large majority than of the totality. The original archetypes estimator has been adapted to be a robust M-estimator using an iteratively reweighted least squares fitting algorithm. This robust archetypal analysis algorithm is based on weighting the residuals and observations, respectively. As a byproduct an algorithm for weighted archetypal analysis is available.

E387: **A forward approach for supervised classification with model selection**
Presenter: **Aldo Corbellini**, University of Parma, Italy
Co-authors: Gianluca Morelli, Fabrizio Laurini
Supervised methods of classification naturally exploit linear and non linear relationships between explanatory variables and a response. However, the presence of clusters may lead to a different pattern within each group. For instance, data can naturally be grouped in several linear structures and so, simple linear regression models can be used for classification. Estimation of linear models can be severely biased by influential observations or outliers. A practical problem arises when the groups identifying the different relationships are unknown, and the number of "relevant" variables is high. In such a context, supervised classification problems can become cumbersome. As a solution, within the general framework of generalized linear models, a new robust approach is to exploit the sequential ordering of the data provided by the forward search algorithm. Such an algorithm

will be used two-fold to address the problems of variable selection for model fit, while grouping the data naturally "around" the model. The influence of outliers, if any is inside the dataset, will be monitored at each step of the sequential procedure. Preliminary results on simulated data have highlighted the benefit of adopting the forward search algorithm, which can reveal masked outliers, influential observations and show hidden structures.

E413: **Finite sample efficiency of the RDELA location and scatter estimator**
Presenter: **Steffen Liebscher**, Martin-Luther-University Halle-Wittenberg, Germany
Co-authors: Thomas Kirschstein

A common strategy to obtain robust estimates of location and scatter (as e.g. followed by the minimum covariance determinant estimator, MCD) is to identify an outlier-free subset of the data which is used for estimation instead of the whole data. The size of the subset is crucial: choosing a subset of minimal size yields highly robust estimators, but leads to a huge loss in efficiency. Therefore the initial subset is usually enlarged in a second step (called reweighting), which increases the efficiency (while conserving the robustness properties). For the decision as to how much the initial subset should be enlarged critical values are required. For the RDELA estimator a theoretical framework to derive such values does not exist so far. It is shown that Chebyshev's inequality can be used for that purpose. A simulation study is conducted in order to investigate the finite-sample efficiency of the RDELA estimator (using Chebyshev's inequality) compared with the MCD and S estimator.

E431: **On concurvity in nonlinear and non parametric regression models**
Presenter: **Sonia Amodio**, Leiden University, Netherlands
Co-authors: Roberta Siciliano, Massimo Aria

When data are affected by multicollinearity in the linear regression framework, then concurvity will be present in fitting a Generalized additive model (GAM). The term concurvity describes nonlinear dependencies among the predictor variables. As collinearity results in inflated variance of the estimated regression coefficients in the linear regression model, the result of the presence of concurvity leads to instability of the estimated coefficients in GAM. Even if the backfitting algorithm will always converge to a solution, in the case of concurvity the final solution of the backfitting procedure in fitting a GAM is influenced by the starting functions. While exact concurvity is highly unlikely, approximate concurvity, the analogue of multicollinearity, is of practical concern as it can lead to upwardly biased estimates of the parameters and to underestimation of their standard errors, increasing the risk of committing type I error. We compare the existing approaches to detect concurvity, pointing out their advantages and drawbacks, using several simulated and real data sets. As a result, a general criterion to detect concurvity in nonlinear and non parametric regression models will be provided to ensure the robustness of the parameters estimation.

E763: **Hidden Markov structures for dynamic copulae**
Presenter: **Ostap Okhrin**, Humboldt-University Berlin, Germany
Co-authors: Weining Wang, Wolfgang Haerdle

Understanding the dynamics of a high dimensional non-normal dependency structure is a challenging task. A multivariate Gaussian or mixed normal time varying models are limited in capturing important types of data features such as heavy tails, asymmetry, and nonlinear dependencies. This research aims at tackling this problem by building up a hidden Markov model (HMM) for hierarchical Archimedean copulae (HAC). HMM applied to HAC flexibly models high dimensional non-Gaussian time series. We apply the expectation maximization (EM) algorithm for parameter estimation. Consistency results for both parameters and HAC structures are established in an HMM framework. The model is calibrated to exchange rate data with a VaR application. We compare the forecasting performance with other classical dynamic models.

ES75 Room 11 EXTREMES AND DEPENDENCE Chair: Michael Falk

E145: **Goodness-of-fit tests for spatial extremes models based on max-stable processes**
Presenter: **Ivan Kojadinovic**, University of Pau, France
Co-authors: Hongwei Shang, Jun Yan

Parametric max-stable processes are increasingly used to model spatial extremes. Starting from the fact that the dependence structure of a simple max-stable process is completely characterized by an extreme-value copula, a class of goodness-of-fit tests is proposed based on the comparison between a nonparametric and a parametric estimator of the corresponding unknown multivariate Pickands dependence function. Because of the high-dimensional setting under consideration, these functional estimators are only compared at a specific set of points at which they coincide with estimators of the extremal coefficients. The nonparametric estimators of the Pickands dependence function used are those recently studied by Gudendorf and Segers. The parametric estimators rely on the use of the *composite pseudo-likelihood* which extends the concept of composite likelihood to a rank-based context. Approximate p-values for the resulting margin-free tests are obtained by means of a *one- or two-level parametric bootstrap*. Conditions for the asymptotic validity of these resampling procedures are given based on the work of Genest and Rémillard. The finite-sample performance of the tests is investigated in dimensions 20 and 40. An application of the tests to Swiss extreme precipitation data is finally presented.

E167: **Modelling and estimation of extreme events observed in space and time**
Presenter: **Claudia Klueppelberg**, Technische Universitaet Muenchen, Germany
Co-authors: Richard A. Davis, Christina Steinkohl

When modelling complex systems as e.g. with meteorological data like precipitation and wind fields, statistical methodology can often be applied to reconcile the physical models with the data. For an adequate risk analysis, the statistical modelling of extreme events, such as heavy rain and floods, or severe wind gusts or storms, is essential. Historically, Gaussian random fields play a central role in modelling spatio-temporal data. When it comes to extremes and extremal dependence, Gaussian processes are not appropriate, since observations at two different locations and time points are in Gaussian models independent at high levels. A natural extension from uni- and multivariate extreme value theory is formed by so-called max-stable random fields. We suggest statistical models for extreme data measured in space and time. We present the basic aspects and challenges of simulation as well as methods for parametric and non-parametric estimation of max-stable spatio-temporal random fields.

E216: **Estimation of the marginal expected shortfall: the mean when a related variable is extreme**
Presenter: **Juan-Juan Cai**, Tilburg University, Netherlands
Co-authors: John H.J. Einmahl, Laurens de Haan, Chen Zhou

Denote the loss return on the equity of a financial institution as X and that of the entire market as Y. For a given very small value of $p > 0$, the marginal expected shortfall (MES) is defined as $E(X \,|\, Y > Q_Y(1-p))$, where $Q_Y(1-p)$ is the $(1-p)$-th quantile of the distribution of Y. The MES is an important factor when measuring the systemic risk of financial institutions. For a wide nonparametric class of bivariate distributions, we construct an estimator of MES and establish the asymptotic normality of the estimator when $p \downarrow 0$, as the sample size $n \to \infty$. Since we are in particular interested in the case $p = O(1/n)$, we use extreme value techniques for deriving the estimator and its asymptotic behavior. The finite sample performance of the estimator and the adequacy of the limit theorem are shown in a detailed simulation study. We also apply our method to estimate the MES of three large U.S. investment banks.

E309: **Conditional simulation of max-stable processes**
Presenter: **Mathieu Ribatet**, Montpellier II, France
Co-authors: Clement Dombry, Frederic Eyi-Minko
Since many environmental processes such as heat waves or precipitation are spatial in extent, it is likely that a single extreme event affects several locations and the areal modelling of extremes is therefore essential if the spatial dependence of extremes has to be appropriately taken into account. A framework for conditional simulations of max-stable processes is proposed and closed forms for Brown-Resnick and Schlather processes are given. We test the method on simulated data and give an application to extreme rainfall around Zurich and extreme temperature in Switzerland. Results show that the proposed framework provides accurate conditional simulations and can handle real-sized problems.

E440: **Multivariate extremal dependence: Estimation with bias correction**
Presenter: **Anne-Laure Fougeres**, Universite Claude Bernard Lyon 1, France
Co-authors: Laurens de Haan, Cecile Mercadier
Estimating extreme risks in a multivariate framework is highly connected with the estimation of the extremal dependence structure. This structure can be described via the stable tail dependence function L, for which several estimators have been introduced. Asymptotic normality is available for empirical estimates of L, with rate of convergence $k^{1/2}$, where k denotes the number of high order statistics used in the estimation. Choosing a higher k might be interesting for an improved accuracy of the estimation, but may lead to an increased asymptotic bias. We provide a bias correction procedure for the estimation of L. Combining estimators of L is done in such a way that the asymptotic bias term disappears. The new estimator of L is shown to allow more flexibility in the choice of k. Its asymptotic behavior is examined, and a simulation study is provided to assess its small sample behavior.

ES76 Room 10 MEDICAL STATISTICS Chair: Maria De Iorio

E437: **On the role of formal risk-benefit decision-making in the regulation of medicines**
Presenter: **Deborah Ashby**, Imperial College London, United Kingdom
The regulation of medicine requires evidence of the efficacy and safety of medicines, and methods are well-developed to deal with the latter and to a lesser extent the former. However, until recently, assessment of risk- benefit especially in relation to alternatives has been entirely informal. There is now growing interest in the possibilities of more formal approaches to risk-benefit decision-making. We review the basis of drug regulation, the statistical basis for decision-making under uncertainty, current initiatives in the area, and discuss possible approaches that could enhance the quality of regulatory decision-making.

E438: **Validation measures for risk prediction models for correlated binary and survival outcomes**
Presenter: **Rumana Omar**, University College London, United Kingdom
Co-authors: Shafiqur Rahman, Gareth Ambler
Risk prediction models are developed to guide patients' clinical management or monitor performances of health institutions. Data from multiple centres are usually used to develop risk models. Random effects/frailty models can be used to develop models for binary and survival outcomes correlated within centres. Some of the existing validation measures for use with these models are extended. An 'overall' and a 'pooled' approach are proposed. The 'overall' measures use comparisons of patients both within and between centres. The 'pooled' approach is based on a weighted average of the cluster specific validation measures. Each approach can produce different values for the validation measures depending on predictions: a) conditional predictions using the estimates of the random effects or setting these as zero and b) marginal predictions obtained by integrating out the random effects. The measures are illustrated with real data and their performances investigated using simulation studies. The 'overall' measures based on predictions including random effects generally performed well; those using predictions with random effects as zero or marginal predictions were biased for moderate within centre correlation. The 'pooled' measures are not recommended for small centres. The appropriateness of these measures for validation data with same or new centres are considered.

E556: **Ridge regression for genetic data: A semi-automatic method to guide the choice of ridge parameter**
Presenter: **Erika Cule**, Imperial College London, United Kingdom
Co-authors: Maria De Iorio
The challenge of predicting disease risk is addressed using data obtained in a genome-wide association study (GWAS). Because genotypes are fixed at birth, but disease symptoms may not appear until later in life, risk prediction using genetic data could suggest lifestyle or pharmaceutical interventions in higher risk individuals, which may reduce future disease susceptibility. Traditional SNP-by-SNP analyses of GWAS data do not take into account the potential combined effect of multiple SNPs, a plausible hypothesis for complex diseases. Standard multiple regression approaches cannot be used for the analysis of GWAS data due to the high dimensionality of the data and the correlation structure among SNPs. Ridge regression (RR) is a penalized regression method that gives rise to stable coefficient estimates even when the number of predictors exceeds the number of observations. One challenge in the use of RR is the choice of the ridge parameter, which affects the model fit. We present a semi-automatic method to choose the ridge parameter that offers improved predictive performance over other regression methods in a GWAS setting. We apply our method to out-of-sample prediction using two Bipolar Disorder GWAS. We present an R package, ridge, which implements the method for genome-wide SNP data.

E685: **An application of a random partition model with regression on covariates for inference in a longitudinal study**
Presenter: **Gary Rosner**, Johns Hopkins University, United States
Co-authors: Peter Mueller, Fernando Quintana
The aim is to consider inference for longitudinal data based on mixed-effects models with a nonparametric Bayesian prior on the treatment effect. The proposed nonparametric Bayesian prior is a random partition model with regression on patient-specific covariates. The main feature and motivation for the proposed model is the use of covariates with a mix of different data formats and possible high-order interactions in the regression. The regression is not explicitly parametric but is implied by the random clustering of subjects. The motivating application is a study of the hypertensive side effect of an anticancer drug. The study involves blood pressure measurements taken periodically over several 24-hour periods for 54 patients.

E1045: **Uncertainty in the statistical analysis of expression data**
Presenter: **William Astle**, McGill University, Canada
Co-authors: Ernest Turro
Gene expression is the central biological process by which information encoded in genes in DNA is used to synthesise functional biochemicals via intermediate RNA molecules. High throughput sequencing of RNA (RNA-seq) technology provides an unprecedented opportunity for the fine-grained genome-wide characterisation of gene expression in biological samples. Most methods for estimating the differential expression of genes (between treatment groups) from RNA-seq data rely on the availability of sequence read counts specific to each feature of interest, which can be a gene, isoform or haplotype-specific isoform. Unfortunately, reads cannot in general be uniquely mapped to features, as features often share significant stretches of sequence. The uncertainty associated with the read-feature mapping, a form of measurement error, can be large relative

to biological variation. In extreme cases, substantial sequence sharing coupled with insufficient read depth, can make it virtually impossible to distinguish the expression levels within certain groups of features. A procedure to deal with this feature identifiability problem will be described. A Bayesian model selection approach to differential expression analysis will be discussed. It can account for read count uncertainties in addition to the biological variability between tissue samples.

ES63 Room 8 COPULAS II **Chair: Wolfgang Trutschnig**

E201: **Approximations by patched bivariate copulas**
Presenter: **Tippawan Santiwipanont**, Chulalongkorn University, Thailand
Co-authors: Songkiat Sumetkijakan, Nattakarn Chaidee

Patched bivariate Fréchet copulas to approximate any copulas in the uniform and L^2 norms have been introduced. After dividing the domain of a given copula into equal rectangles, a patched copula is constructed by replacing conditional copulas on sub-rectangles by any copulas, while keeping the original probability masses and marginal distributions of sub-rectangles. We show that patched copulas also converge to the original copula in the Sobolev norm. However, it is not always possible to construct a copula from arbitrary probability masses and marginal distributions. We obtain a set of necessary and sufficient conditions under which a copula with such probability masses and marginal distributions exists. It turns out that one can always use uniform distributions on all sub-rectangles. We then investigate the approximations by patched copulas whose all marginal distributions are uniform. In particular, we illustrates that the Sobolev norm of the independence copula can be attained in the limit by a sequence of copulas which are conditionally mutually completely dependent.

E202: **Dynamics of dependence properties for random lifetimes**
Presenter: **Rachele Foschi**, IMT Institute for Advanced Studies Lucca, Italy

Tail dependence and default contagion emerge as two subjects of interest in multivariate survival models and are typically related with the analysis of non-negative random variables $T_1,\ldots,T_n$, that have the meaning of waiting times until top events for several different units. Such phenomena are equally relevant in several fields, such as e.g. Reliability, Risk theory and Finance. Tail dependence refers to those aspects of stochastic dependence (among units) that are related with a high frequency of simultaneous occurrence of extreme events for the different units. The term "default contagion" describes the cases when the default of a unit causes a decrease in the joint survival probability of residual lifetimes of the remaining units. Both tail dependence and default contagion are related to the evolution, at certain instants, of stochastic dependence. We describe stochastic dependence in terms of copulas. The evolution of dependence will be recast by means of the family of survival copulas $\{\hat{C}_t\}_{t\geq 0}$. We present a study of the evolution of such a family of copulas, both in the case characterized by absence of defaults and given the occurrence of non-simultaneous defaults. In particular, we will focus on analytical properties of the family and on implied conditions for preservation of some positive dependence properties.

E209: **A new methodology of construction of n-copulas**
Presenter: **Jose Maria Gonzalez-Barrios**, Universidad Nacional Autonoma de Mexico, Mexico
Co-authors: Maria Magdalena Hernandez-Cedillo

A new methodology for the construction of n-copulas for $n \geq 3$ is proposed. This construction generalizes both patchwork and ordinal sum construction, and it is based on a multivariate version of the orthogonal grids. This methodology allows us to modify an n-copula in a fixed n-box and it is quite useful when modelling multivariate dependencies.

E787: **Skew $t-$copula and tail dependence**
Presenter: **Tonu Kollo**, University of Tartu, Estonia
Co-authors: Gaida Pettere

Let X and Y be two (possibly dependent) random variables with marginal distribution functions $F(x)$ and $G(y)$. Then the joint distribution function $H(x;y)$ can be presented through a copula. Several classes of copulas have been constructed which have different mathematical properties. In applications an important property of a copula is tail dependence. Tail dependence refers to the degree of dependence in the corner of the lower-left quadrant or upper-right quadrant of a bivariate distribution. Whether a large value of X increases the probability that Y is also large is characterized mathematically by the limit of the conditional probability $P(Y|X)$ when X tends to infinity and Y is greater or equal to X. The limit is the characteristic of tail dependence. The Gaussian copula does not have the upper tail dependence as the described above limit equals to zero. This property makes Gaussian copula not suitable for many applications. The same time $t-$copula takes into account tail dependence when using it as a data model. In 2010 the skew $t-$copula was introduced. The construction is based on the multivariate skew $t-$distribution. First applications have shown that the distribution can successfully be used in practice. To examine the tail behavior of the skew $t-$copula and speed of convergence to the limit we have carried out a simulation experiment. Results of theoretical as well as empirical studies will be discussed.

E970: **EM algorithms for estimating the Bernstein copula**
Presenter: **Satoshi Kuriki**, The Institute of Statistical Mathematics, Japan
Co-authors: Xiaoling Dou, Gwo Dong Lin

The so-called Baker's family of multivariate distributions with fixed marginals have been proposed few years ago. In the bivariate case, its density is represented as $h(x,y)=\sum_{i,j=1}^{n} r_{ij} f_i(x) g_j(y)$, where $\sum_{i=1}^{n} r_{ij} = \sum_{j=1}^{n} r_{ij} = 1/n$, $r_{ij} \geq 0$, and $f_i(x)$ and $g_j(y)$ are density functions of the ith and jth smallest order statistics of X and Y with sample size n, respectively. This is a finite mixture of n^2 populations. On the other hand, since the density of the order statistic is described in terms of the Bernstein polynomials, Baker's distribution is constructed by the copula function with Bernstein polynomial basis (Bernstein copula). We propose EM algorithms for estimating Baker's distribution, or the Bernstein copula. We focus on two examples. The first one is for a particular form of Baker's distribution $h(x,y)=(1-q)f(x)g(y)+qh_n^+(x,y)$, $0\leq q\leq 1$, where $f(x)$, $g(y)$ are densities of X, Y, and $h_n^+(x,y)$ is the joint density with maximum positive correlation among Baker's distributions. The second EM algorithm is for the general form of Baker's distribution. Our proposed algorithm is proved to have the property of local convergence of first order. We discuss statistical inference based on these models, and finally give illustrative data analyses.

EP03 Room Hall POSTER SESSION III **Chair: Maria Jesus Garcia-Ligero**

E835: **Influence and explained variation in survival analysis**
Presenter: **Refah Alotaibi**, Newcastle University, United Kingdom
Co-authors: Robin Henderson, Malcolm Farrow
The amount of explained variation R^2 is an overall measure used to quantify the information in a model and especially how useful the model might be when predicting future observations. R^2 is useful in guiding model choice for all types of predictive regression models, including linear and generalized linear models and survival analysis. We consider how individual observations in a data set can influence the value of R^2. Influence of a subject is a measure of the effect on estimates of deleting him/her from the data set. Influence on regression coefficients has had much attention but there has not been work in influence for R^2 for survival data analysis or other measures of predictive accuracy. Generally in reasonable size data sets the deletion of a single subject has no effect on conclusions. However, examination of distance between estimates with and without the subject can be useful in distinguishing abnormal observations. Based on the influence function, we discuss methodologies for assessing influence on Graf *et al.*'s R^2_G measure, the C-index and Nagelkerke's R^2_G. The ideas are illustrated on data on 1391 patients diagnosed with Non-Hodgkin's Lymphoma (NHL).

E837: **Distributed fixed-point smoother from uncertain observations with Markovian random delays**
Presenter: **Maria Jesus Garcia-Ligero**, Universidad de Granada, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez
The signal estimation problem from measurements coming from multiple sensors has received considerable attention because of the complexity of current systems. The fusion of the information provided by different sensors is generally carried out by two different methods, centralized and distributed. The computational advantages provided by the distributed method lead us to consider this versus the centralized one. On the other hand, in communication systems, errors during the signal transmission that can lead to uncertainties and delays are unavoidable. These uncertainties and random delays have been usually modeled by independent Bernoulli random variables. However, in real communication systems, current time delays are usually correlated with previous ones; a reasonable way to model this dependence is to consider the random delay as a homogeneous Markov chain. Under this assumption the least-squares linear fixed-point smoothing problem for stochastic systems with uncertain observations which can be randomly delayed by one sampling time is addressed in a multisensor environment. Using the information provided by the covariance functions of the signal and noise processes, local fixed-point smoothers from each sensor are derived and the distributed smoother is obtained as linear combination of them by using least squares criterion.

E839: **Evaluation of the patient service quality at a university hospital by the service quality measurement scale**
Presenter: **Fezan Sahin Mutlu**, Eskisehir Osmangazi University, Turkey
Co-authors: Unal Ayranci, Ahmet Musmul, Ipek Aciksoz, Muzaffer Bilgin, Kazim Ozdamar
The aim was to evaluate the patient service quality(PSQ) at a university hospital by the service quality measurement scale (SERVQUAL). The study comprised of 267 patients applied to the polyclinics of a university hospital of Western Turkey between 1st June 2011– 29th Feb., 2012. The SERVQUAL was used to determine PSQ. Each of the expectation and perception scores consisted of 22 questions identical. The 5 Likert scale was applied in two sets of variables, with 5 dimensions: tangibles, reliability, responsiveness, assurance, empathy. The Cronbach Alpha Coefficient was obtained for each dimensions of the first 4 questions for tangibles, the 5th-9th questions for reliability, 10th-13th questions for responsiveness, 14th-17th questions for assurance, and 18th-22nd questions for empathy. To test the accuracy of the dimensions, the confirmatory factor analysis was performed. According to the index of Root Means Square Error of Approximation(RMSEA), 5 dimensions in the model were compatible (0.083 for expectation, 0.062 for perception). Thus, the SERVQUAL scores for the 5 dimensions were received by subtracting the expected dimensions from the perceived dimensions. The SERVQUAL scores indicated that the SERVQUAL gave a result of middle level. The hospital management should make arrangements to improve the SERVQUAL, and use this scale.

E855: **A new approach to estimate sources contribution in stable isotope mixing models**
Presenter: **Josep Carrasco**, Universitat de Barcelona, Spain
Co-authors: Josep Puig-Martinez, Lluis Jover
Stable isotope mixing models are used to quantify contributions of different sources within a mixture and thereby to infer diet composition. Several approaches, either fully parametric as non-parametric, have been proposed to estimate the proportions of the mixture. Furthermore all of these approaches involve solving a system of equations. Here we aim to carry out a simulation study to evaluate a novel non-parametric approach by comparing its performance with that of two existent approaches. The evaluation of estimation methods will be carried out in terms of coverage and length of intervals as well as accuracy of estimates.

E869: **A double-sigmoidal logistic diffusion process**
Presenter: **Francisco Torres-Ruiz**, University of Granada, Spain
Co-authors: Eva Poza-Cruz, Patricia Roman-Roman
In order to describe sigmoid growth patterns in many scientific fields, a wide variety of curves have been used. Among others, we can cite the logistic, Gompertz, Bertalanffy, Richards. All these curves have a point of inflection that is always at a fixed proportion of its asymptotic value. Nevertheless, there are problems corresponding to a double sigmoid behaviour. For example, the growth of some fruits (berries, guayabas, nectarines,...) presents more than one point of inflection showing two phases represented separately for sigmoid patterns. Also this kind of behaviour has been studied in other fields. For instance, it has been used for representing earth-states transitions in geological models. However, the models used in the above contexts are deterministic and do not take into account external influences. In order to take into account random environmental effects, a stochastic model related to a double sigmoidal curve, concretely a transformation of the well-known logistic model, is presented. This model is a diffusion stochastic process which mean is a curve of this type. Also, with the purpose of fitting this model to real data, the maximum likelihood estimation of the parameters of the model is studied. Finally, some simulated examples are presented.

E885: **New estimators for the population proportion using auxiliary information**
Presenter: **Antonio Arcos**, Universidad de Granada, Spain
Co-authors: Maria del Mar Rueda, Juan Francisco Munoz, Encanacion Alvarez
The problem of estimating a population proportion is an important topic that has numerous applications to many areas. For example, applications can be found in the health field with the problem of estimating prevalences. Recently, many studies about the estimation of a population proportion and based upon auxiliary information have been done, and the ratio, difference and regression techniques were used. The aim is to contribute by defining exponentiation type estimators for a population proportion. Using data extracted from the Spanish National Health Survey, the proposed estimator is numerically compared to alternative estimators. The aim in this study was to estimate prevalences for the diseases asthma and allergy. Results derived from this study reveal that the proposed exponentiation estimator can be more efficient than alternative estimators.

E894: **Globally optimal scoring weights and their application for test score equating in IRT model**

Presenter: **Sayaka Arai**, The National Center for University Entrance Examinations, Japan
Co-authors: Shin-ichi Mayekawa

In the area of educational assessment, test scores play an important role in making decisions about course placement, university admission, etc. Test scores can be broadly divided into two types: (1) the weighted total scores and (2) the scaled scores using item response theory (IRT). The weighted total scores are easy to calculate, however the statistical properties of the score depend on the weights. On the other hand, although IRT scaled scores are complicated for students to calculate, they have advantage over the weighted total scores: comparability of test scores from different test forms. The globally optimal scoring weights as a new way of weighting items to have the advantages of both of the scores were previously developed. Globally optimal scoring weights are a set of weights which maximize the expected test information under IRT models. The results showed that they reduced the posterior variance when evaluating the posterior distribution of latent ability, theta, given the weighted total scores. Globally optimal scoring weights are applied in a situation where test scores are equated between two test forms and problems and solutions are discussed.

E898: **The asymptotics of maximum likelihood for censored data**

Presenter: **Inmaculada Barranco-Chamorro**, University of Seville, Spain
Co-authors: Juan Luis Moreno-Rebollo, M. Dolores Jimenez-Gamero

In life testing experiments, modelled by continuous parametric distributions, quite often data are censored. In this context, inference is usually based on likelihood function. We carry out a comparison of properties of maximum likelihood estimators (MLEs) for different models under different censorship mechanisms. First, we note that different censorship mechanisms lead to similar likelihood functions. Second, we focus on the study of conditions under which asymptotic properties of ML estimation can be applied. At this point our objectives are twofold. We give a review of asymptotic results when sampling from regular models, and study the asymptotic properties when regularity conditions do not hold. This is the innovative part of our study. Simulations and examples are included to illustrate the performance of our results.

E901: **Case-deletion type diagnostics for penalized calibration estimators and blup under mixed linear models**

Presenter: **Juan Luis Moreno-Rebollo**, University of Seville, Spain
Co-authors: Inmaculada Barranco-Chamorro

Case-deletion diagnostics are commonly applied in most fields of statistics to detect influential observations. It is generally assumed that the observations are independent and identically distributed (iid). In the iid framework, deleting an observation involves a reduction of the sample size: an iid sample of size n is reduced by one unit, resulting in an iid sample of size $n-1$. However, in survey sampling under the design-based approach the meaning of deleting an observation from the sample is not clear. Putting a deleted unit at the same level as a nonresponse case and based on the use of calibration techniques as a way of handling nonresponse, a case-deletion type diagnostic is proposed for a wide class of estimators. The class includes penalized-calibration estimators and the best linear unbiased predictor under mixed linear models. The resulting diagnostic can be calculated from quantities related to the full data set, namely the weight, the residual and the leverage. It is therefore unnecessary to recalculate the weights and the estimates for each deleted case. This fact undoubtedly eases the practical application of the diagnostic.

Sunday 02.12.2012 14:30 - 16:10 Parallel Session K – CFE

CS11 Room 1 VOLATILITY MODELLING IN THE PRESENCE OF OUTLIERS **Chair: Angeles Carnero**

C128: **Instabilities and breaks on European credit default swaps: Effects over a crisis period**
Presenter: **Jose Olmo**, Centro Universitario de la Defensa de Zaragoza, Spain
Co-authors: Burcu Kapar, Ricardo Laborda
The determinants of major European corporate credit default swap contracts during the recent financial crisis are investigated. In particular, we find that the European iTraxx credit risk portfolio and the idiosyncratic stock price corresponding to each CDS contract are sufficient to describe the long-run dynamics of European corporate CDS spreads. The effects of the crisis are reflected in instabilities and structural breaks in the long and short run dynamics of the cointegrated variables. We find two different types of instabilities in these series: a first type that is due to structural breaks in the nonstationary factors driving the spreads and a second type that is intrinsic to the CDS market and that produces changes in the relationship between CDS spreads and the explanatory factors. Our results reveal the importance of considering aggregate credit risk measures and idiosyncratic market risk measures that incorporate the occurrence of structural breaks.

C163: **Effects of outliers on asymmetric GARCH models**
Presenter: **Angeles Carnero**, Universidad de Alicante, Spain
Co-authors: Ana Perez-Espartero, Esther Ruiz
It is well known that outliers affect the identification and estimation of the parameters and the underlying volatilities in symmetric GARCH models. However, nothing is known about their effects on asymmetric conditional heteroscedastic models when the volatility response is different depending on whether past returns are positive or negative. We first analyze the effect of additive outliers on the identification of asymmetries when this is based on the sample cross-correlations between returns and future squared returns. Second, we analyze whether positive and negative outliers of equal size have different impacts on the estimated model parameters and volatilities in the context of the Threshold GARCH (TGARCH) and Exponential GARCH (EGARCH) models. Different parameter estimators are compared: Gaussian maximum likelihood, QML based on maximizing the Student likelihood (QML-t), robust BM and variance-targeting estimators. Given that there are no closed-form expressions for these estimators, the comparison is based on Monte Carlo experiments. We also look at the effects of outliers on the estimates of the underlying volatilities and VaR obtained using the standard and robust filters. The results are illustrated with an empirical application.

C299: **Regime switches in volatility and correlation of financial institutions**
Presenter: **Kris Boudt**, KULeuven-VU Amsterdam, Belgium
Co-authors: Jon Danielsson, Siem Jan Koopman, Andre Lucas
A regime switching model is proposed to characterize the dynamics in the volatility and correlation of financial institutions. The within regime dynamics in the volatility and correlation depend on the shape of the Student t innovations. The across-regime dynamics in the transition probabilities are driven by systemic risk variables.

C912: **Testing for jumps in GARCH models, a robust approach**
Presenter: **Laurent Sebastien**, Maastricht University, Netherlands
Co-authors: Christelle Lecourt, Franz Palm
Financial series occasionally exhibit large changes. Assuming that the observed return series consist of a standard normal ARMA-GARCH component plus an additive jump component, we propose a new test for additive jumps in an ARMA-GARCH context. The test is based on standardised returns, where the first two conditional moments are estimated in a robust way. Simulation results indicate that the test has very good finite sample properties. We apply the test on about 50 US stocks and show the supremacy of our model compared to standard GARCH-type models but also the recent Generalized Autoregressive Score volatility model (GAS).

CS12 Room 4 BAYESIAN NONLINEAR ECONOMETRICS **Chair: Roberto Casarin**

C373: **Learning about long-run risk: Self-referential beliefs and equilibrium asset pricing**
Presenter: **Daniele Bianchi**, Bocconi University, Italy
The standard full information assumption is amended to allow for structural uncertainty in a consumption-based asset pricing framework. The expected growth rate of the economy (consumption growth) is assumed unobservable and sequentially learned by a representative agent. Yet, the agent's beliefs are affected by a feedback effect from lagged endogenous variables. The impact of feedback and structural uncertainty on the perceived law of motion is two-fold. First, incomplete information introduces high persistency in the dynamics of the perceived long-run expected growth rate, endogenously generating long-risk features which are not detectable in the data generating process. Second, feedback raises dynamic endogeneity in the learning scheme, making the sequential updating beliefs self-referential, then systematically affecting the actual law of motion. These effects have considerable asset pricing implications in equilibrium. Both an endogenous Kalman Filter and a Particle Learning algorithm are used to investigate (analytically and numerically) the asset pricing implications of self-referential sequential learning in a general equilibrium framework with recursive utilities.

C462: **Nowcasting business cycle turning points in an uncertain environment**
Presenter: **Francesco Ravazzolo**, Norges Bank, Norway
Co-authors: Knut Are Aastveit, Herman K. van Dijk
A combined Bayesian factor model to estimate and nowcast US business cycle turning points in real-time is constructed. Factor models are often used to construct business cycle indicators. Two approaches have mainly been used. The first approach is to identify turning points individually in a large number of series and then average (date then average). The second approach is to look for turning points in a few, or just one, aggregate (average then date). We study both approaches within a factor model framework. In the former approach we apply a nonlinear transformation to all the data series and then estimate factors. In the latter approach we estimate a single series, such as GDP, using factor models. We then apply a given rule to the estimated series. In both approaches we propose to use a combined Bayesian factor model. An advantage of using model averaging is that we account for several sources of uncertainty such as parameter uncertainty and data input and factor structure uncertainty.

C534: **Bayesian nonparametric copulas for multivariate time series**
Presenter: **Concepcion Ausin**, Universidad Carlos III de Madrid, Spain
Copula models have become a very common tool in financial time series as they make it possible to define separately the marginal distributions of assets and their dependence structure. The usual approach is to assume a parametric model for the copula function, which can be constant or time-varying. However, the choice of adequate parametric copula models which describe appropriately the time series dependence is complicated in practice. A Bayesian nonparametric approach is developed to approximate multivariate copulas in financial time series using Bernstein polynomials. This class of polynomials can be used to approximate arbitrarily any distribution and density function defined on the unit hypercube and, in

particular, a multivariate copula. A Bernstein-Dirichlet prior is assumed whose properties have been studied in previous works. Inference and prediction is carried out using Markov chain Monte Carlo methods. The procedure is illustrated using simulated and real data sets.

C608: **Beta-product dependent Pitman-Yor processes**
Presenter: **Roberto Casarin**, University Ca Foscari of Venice, Italy
Co-authors: Fabrizio Leisen, Federico Bassetti

Time series data may exhibit clustering over time and, in a multiple time series context, the clustering behavior may differ across the series. The motivation is the Bayesian non-parametric modeling of the dependence between the clustering structures and the distributions of different time series. We follow a Dirichlet process mixture approach and introduce a new class of multivariate dependent Dirichlet processes (DDP). The proposed DDP are represented in terms of vectors of stick-breaking processes with dependent weights. The weights are beta random vectors that determine different and dependent clustering effects along the dimension of the DDP vector. We discuss some theoretical properties and provide an efficient Monte Carlo Markov Chain algorithm for posterior computation. The effectiveness of the method is illustrated with a simulation study and an application to the United States and the European Union industrial production indexes.

CS27 Room 2 APPLICATIONS OF PENALIZED SPLINES TO ECONOMICS Chair: Goeran Kauermann

C134: **Smooth estimation of urban house price surfaces under conditional price and spatial heterogeneity**
Presenter: **Harry Haupt**, Bielefeld University, Germany
Co-authors: Pin T. Ng

The problem of incorporating nonlinearities in both the systematic and the spatial component of price regressions is addressed by applying a flexible semiparametric approach incorporating additive univariate and bivariate penalized spline structures. We demonstrate that such a strategy imposes only weak and empirically testable assumptions on the structure of the inherent data generating processes while avoiding the curse of dimensionality of fully nonparametric approaches. The resulting anisotropic spatial nature of the price surface allows the existence of spatial subcenters –i.e. horizontal market segmentation. Further, to simultaneously embed nonlinearities and spatial association in a framework allowing to estimate submarket specific price surfaces (and shadow prices) we incorporate the semiparametric approach into the quantile regression framework. The virtues of the latter are that different market mechanisms might be prevalent across different parts of the conditional price distribution –low price, middle price, and high price segments may exhibit different functional relationships (through different subsets of characteristics). This can be directly estimated via the respective conditional quantiles, allowing a vertical segmentation of a housing market. Further, the exact-fit-property of quantile regression can be used to estimate the effective dimension of a hedonic price surface even in highly nonlinear contexts. The latter not only constitutes a valuable information in the model selection process but also is essential for economic interpretation of hedonic shadow prices and conditional price predictions. An application to German rental data formally and graphically illustrates the virtues of the proposed modeling approach.

C138: **Bayesian nonparametric instrumental variable regression based on penalized splines and Dirichlet process mixtures**
Presenter: **Thomas Kneib**, University of Goettingen, Germany
Co-authors: Manuel Wiesenfarth

A Bayesian nonparametric instrumental variable approach is proposed that allows us to correct for endogeneity bias in regression models where the covariate effects enter with unknown functional form. Bias correction relies on a simultaneous equations specification with flexible modeling of the joint error distribution implemented via a Dirichlet process mixture prior. Both the structural and instrumental variable equation are specified in terms of additive predictors comprising penalized splines for nonlinear effects of continuous covariates. Inference is fully Bayesian, employing efficient Markov Chain Monte Carlo simulation techniques. The resulting posterior samples do not only provide us with point estimates, but allow us to construct simultaneous credible bands for the nonparametric effects, including data-driven smoothing parameter selection. In addition, improved robustness properties are achieved due to the flexible error distribution specification. Both these features are extremely challenging in the classical framework, making the Bayesian one advantageous. In simulations, we investigate small sample properties and an investigation of the effect of class size on student performance in Israel provides an illustration of the proposed approach which is implemented in an R package *bayesIV*.

C218: **Smoothing and forecasting seasonal time series with P-spline mixed models**
Presenter: **Maria Durban**, Universidad Carlos III de Madrid, Spain
Co-authors: Lee Dae-Jin

A method for smoothing and forecasting seasonal time series is proposed. It is based on the mixed model formulation of f trend and seasonal modulation models using varying coefficient terms. The approach allows us to decompose the fitted values as the sum of a time trend and seasonality. Under the mixed model framework the model can be used to forecast future observations and estimate missing observations; it also allows the incorporation of additional structures in the residuals, such as autorregressive components, and estimate the smooth term and the error structure simultaneously. The model will be illustrated with the analysis of economic time series.

C671: **Applications of penalized splines to economics**
Presenter: **Michael Wegener**, Deka Investment GmbH, Germany

Considering empirical surveys on financial markets or economic relationships one finds primarily parametric, often times linear, approaches to explain or forecast the data at hand. However, parametric modelling in financial and economic research has its shortcomings – most notably the need for a-priori specification of the functional form. Non- and semi-parametric approaches do not rely on a-priori specification and hence serve as viable alternatives to parametric modelling. In practice, however, one is faced with a multitude of different smoothing algorithms (e.g. kernel smoothers, spline methods, or wavelets) and it's not obvious which one is appropriate. To obtain wide acceptance in financial and economic research, a smoothing technique needs to be flexible, quick, and easy to implement. It should yield robust results even when small changes are applied to parameters like bandwiths or the number of knots. The method should be able to correctly handle various model complexities. Examples here include: additive models, generalized and semi-parametric models, as well as data suffering from autocorrelation and heteroscedasticity. Furthermore, having an easy and fast way at automatically determine the degree of smoothness is highly desirable. In recent years, penalized splines have become popular and have proven their adequacy in a multitude of simulations and practical applications. A fundamental advantage lies in the possibility that P-splines can be interpreted as a mixed model. As such the many instruments and widely available software for mixed models can be used to calculate P-splines quickly and efficiently. Hence, we present several economic and financial applications of penalized spine models which show that P-splines fulfill a great many requirements needed to establish itself as the smoothing method of choice in financial and economic research.

CS50 Room 5 CORRELATION BY VOLATILITY Chair: Jeroen Rombouts

C222: **The value of multivariate model sophistication: An application to pricing Dow Jones Industrial Average options**
Presenter: **Jeroen Rombouts**, HEC Montreal, Canada
Co-authors: Francesco Violante, Lars Stentoft
The aim is to assess the predictive accuracy of a large number of multivariate volatility models in terms of pricing options on the Dow Jones Industrial Average. We measure the value of model sophistication in terms of dollar losses by considering a set of 248 multivariate models that differ in their specification of the conditional variance, conditional correlation, and innovation distribution. All models belong to the dynamic conditional correlation class which is particularly suited because it allows us to consistently estimate the risk neutral dynamics with a manageable computational effort in relatively large scale problems. It turns out that the most important gain in pricing accuracy comes from increasing the sophistication in the marginal variance processes (i.e. nonlinearity, asymmetry and component structure). Enriching the model with more complex correlation models, and relaxing a Gaussian innovation for a Laplace innovation assumption improves the pricing in a smaller way. Apart from investigating directly the value of model sophistication in terms of dollar losses, we also use the model confidence set approach to statistically infer the set of models that delivers the best pricing performance.

C272: **Option pricing with GARCH and flexible conditional distributions**
Presenter: **Lars Stentoft**, HEC Montreal, Canada
The GARCH framework has by now been used for option pricing with quite some success. While the initial work assumed conditional Gaussian distributions recent contribution relaxes this assumption and allows for more flexible parametric specifications of the underlying distribution. However, even within this generalized framework it may be difficult to accommodate e.g. the amount of skewness and leptokurtosis needed to match actual asset return data. In this paper we provide a new method which uses the actual historical innovations as the underlying distribution and we show how to obtain the appropriate risk neutral dynamics in this framework. We compare the model to existing parametric GARCH models in a Monte Carlo study and show empirically the value of the distributional flexibility.

C474: **Infinite-states Markov-switching for dynamic volatility and covariance models**
Presenter: **Arnaud Dufays**, Universite Catholique de Louvain, Belgium
Dynamic volatility models with fixed parameters such as GARCH models are too restrictive. One interesting way of making these models more flexible is enriching them with a dynamic discrete latent state Markov process in such a way that the parameters can abruptly switch from one value to another. These models are called Markov-switching when the chain is recurrent. We propose a new Markov-switching model based on the sticky infinite hidden Markov-chain framework that deals with models exhibiting path dependence such as the GARCH one. Our MCMC sampler endogenously determines the number of regimes. It uses a Metropolis-Hastings update that allows for decreasing the computational time compared to existing algorithms. Moreover its building block is very close to the standard forward-backward algorithm and they share the same computational order. A brief study on the mixing properties of the algorithm is provided. We illustrate our method on the GARCH and the *RM*1994 models. We first provide detail results on simulated data and on the daily index S&P 500 from 1999 to 2011 (3000 observations). We end by comparing our Markov-switching *RM*1994 to the *RM*1994 with fixed parameter.

C620: **Volatility forecasting and explanatory variables: A tractable Bayesian approach to stochastic volatility**
Presenter: **Christian Dorion**, HEC Montreal, Canada
Co-authors: Nicolas Chapados
A formulation of stochastic volatility (SV) based on Gaussian processes, a flexible framework for Bayesian nonlinear regression, is provided. The advantage of using Gaussian processes (GP) in this context is to place volatility forecasting within a regression framework; this allows a large number of explanatory variables to be used for forecasting, a task difficult with standard volatility-forecasting formulations. Our approach builds upon a range-based estimator to provide much greater accuracy than traditional close-to-close estimators using daily data. The GP framework allows for a richer characterization of the volatility dynamics which, even without explanatory variables, entails significant improvements over SV and GARCH benchmarks in terms of out-of-sample one-week- to one-year-ahead forecasts. Augmenting the model with explanatory variables leads to even greater improvements. The augmented GP volatility model significantly outperforms all benchmarks at forecasting volatility one day to one year ahead, out of sample.

CS55 Room 6 CONTINUOUS TIME FINANCIAL MODELS Chair: Leopold Soegner

C214: **Method of moments estimation and affine term structure models**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Jaroslava Hlouskova
Parameter estimation for affine term structure models by means of the generalized method of moments is investigated. First we apply some results obtained by mathematical finance literature on m-polynomial processes to derive the moments of the affine latent process driving the term structure. After specifying the properties of the micro-structure noise, we derive the moments of the yields. Equipped with these moments we estimate the model parameters by means of the generalized method of moments. The estimation procedure is applied to simulated and empirical interest rate data.

C292: **Filter based volatility and regime switching models in continuous time**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany
A continuous time regime switching model, where the observation process is a diffusion whose drift and volatility coefficients jump governed by a continuous time Markov chain, can explain some of the stylized facts of asset returns, even in this simple linear and non-autoregressive form. But due to the switching volatility, in continuous time the underlying Markov chain could be observed and no filtering is needed (in theory). Therefore, even if e.g. in Finance theoretical results could be obtained, they often don't provide a good approximation for the discretely observed model in which we have to filter. On the other hand, a continuous time hidden Markov model (HMM), where only the drift jumps and the volatility is constant, allows for explicit calculations but has not such good econometric properties. To combine useful aspects of both models, we look at a HMM where the volatility depends on the filter for the underlying Markov chain. We analyze its relation to Markov switching models and show that it has better econometric properties than the simple HMM. 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.

C569: **Doubly robust estimation of causal effects with multivalued treatments**
Presenter: **Derya Uysal**, Institute for Advanced Studies, Austria
A simple doubly robust estimation method for multivalued treatments is proposed and applied to an interesting data set. The method combines weighting and regression approaches and requires two model specifications. The advantage over single methods is that it stays consistent even if one of the models is misspecified. The asymptotic distribution of the estimator is also provided and a Monte Carlo study to evaluate the small sample performance of the proposed estimator. In the second part, the returns to schooling are estimated using the rich data set of British Cohort Study

where schooling is used as a multivalued treatment variable. The analysis is carried out for female and male samples separately to capture possible gender differences. Average returns are estimated for entire population, as well as conditional on having a specific educational achievement.

C1063: **Analyzing whether imperfection is better: Evidence from predicting stock and bond returns**
Presenter: **Katarina Kvasnakova**, Vienna Graduate School of Finance, Austria

The short-horizon stock and bond return predictability in a predictive regression and a predictive system is analyzed by using a Bayesian framework. In contrast to the predictive regression where the expected returns are modeled as a linear function of predictors, in the predictive system this assumption is relaxed and predictors do not explain all variance in the expected returns. We argue that a fair comparison of the models has not been drawn yet. We propose an approach that allows for a comparison of the corresponding models. We allow for various levels of optimism about predictability by different priors on the R^2. Comparing the models, we look at out-of-sample performance of the asset allocation problem and show that relaxing the assumption of perfect predictors does not pay off out-of-sample.

CS72 Room 3 ADVANCES IN DSGE MODELING Chair: Michel Juillard

C169: **Accelerating the resolution of sovereign debt methods with an endogenous grid method**
Presenter: **Sebastien Villemot**, CEPREMAP, France

Existing algorithms for solving sovereign debt models with endogenous default rely on value function iterations, because the default decision involves the comparison between two value functions. These algorithms are therefore very slow, and become almost intractable even for a state space of dimension as low as 3. It is shown how to apply the endogenous grid method to such models, leading to a dramatic speed gain. This improvement opens the possibility of solving larger such models or of estimating these with Bayesian techniques. A second contribution is to quantify and compare the accuracy of the computed solutions for both algorithms.

C173: **A note on particle filters applied to DSGE models**
Presenter: **Angelo M Fasolo**, Central Bank of Brazil, Brazil

The aim is to compare the properties of two particle filters – the Bootstrap Filter and the Auxiliary Particle Filter – applied to the computation of the likelihood of artificial data simulated from a basic DSGE model with nominal and real rigidities. Particle filters are compared in terms of speed, quality of the approximation of the probability density function of data and tracking of state variables. Results show that there is a case for the use of the Auxiliary Particle Filter only when the researcher uses a large number of observable variables and the number of particles used to characterize the likelihood is relatively low. Simulations also show that the largest gains in tracking state variables in the model are found when the number of particles is between 20,000 and 30,000, suggesting a boundary for this number.

C191: **Moments and stable simulations of higher order DSGE models via a nonlinear moving average**
Presenter: **Alexander Meyer-Gohde**, Humboldt University Berlin, Germany
Co-authors: Hong Lan

We apply a nonlinear moving average as an alternative to the standard state-space policy function for the analysis of a nonlinear DSGE model. The ensuing policy function is a direct mapping from shocks to endogenous variables up to the order of approximation and provides a natural framework for approximative studies of moments and simulations. In comparison with state-space policy functions, the nonlinear moving average eliminates (1) the need for 'pruning' in simulations as higher-order approximations inherit the stability from the first-order solution and (2) the infinite recursion in moments of the approximative policy function.

C199: **Generalizing the Taylor principle: New comments**
Presenter: **Magali Marx**, Banque de France, France
Co-authors: Jean Barthelemy

A controversy between two previous works about determinacy conditions in Markov switching rational expectations models is clarified. We show that one of them is valid in a set of bounded solutions, that we call Markovian, which only depend on a finite number of past regimes. Enlarging the scope of the analysis to all type of bounded solutions, we derive determinacy conditions, i.e. conditions ensuring existence and uniqueness of a bounded solution. If the determinacy conditions are satisfied, the solution is always the Markovian solution of such an approach. We finally illustrate the results in the standard new Keynesian model studied in those works and depict the determinacy region in this case.

Sunday 02.12.2012 14:30 - 16:10 Parallel Session K – ERCIM

ESI04 Room Multipurpose SPATIAL STATISTICS Chair: Nial Friel

E354: **Bayesian parameter estimation for latent Markov random fields and social networks**
Presenter: **Richard Everitt**, University of Reading, United Kingdom
In a range of applications, including population genetics, epidemic modelling and social network analysis, the data from which we wish to estimate parameters of interest consists of noisy or incomplete observations of an unobserved process. Bayesian statistics offers a framework in which to tackle this problem, accurately accounting for the uncertainty present due to the missing data. However, standard Markov chain Monte Carlo (MCMC) methods that are used to implement the Bayesian approach can perform poorly in this situation. We describe two alternatives to standard MCMC approaches: approximate Bayesian computation (ABC) and particle MCMC. Both methods are applied to parameter estimation of a hidden Markov random field, and are compared to the standard data augmentation approach.

E1039: **Approximate Bayesian computation methods for model choice, application to latent Gibbs random fields**
Presenter: **Jean-Michel Marin**, University Montpellier, France
The Approximate Bayesian computation (ABC) techniques have become an essential tool for the analysis of complex stochastic models. It has been advocated the use of ABC for Bayesian model choice in the specific case of Gibbs random fields, relying on a inter-model sufficiency property to show that the approximation was legitimate. We will first recall the ABC model choice methodology. We will notably emphasize on the fact that the choice of the summary statistics is paramount to produce a valid outcome and we will introduce necessary and sufficient conditions on those statistics for the corresponding Bayes factor to be convergent, namely to asymptotically select the true model. Then, we will consider the specific case latent Gibbs random fields.

E662: **Approximate MCMC with noisy acceptance probabilities**
Presenter: **Nial Friel**, University College Dublin, Ireland
It turns out that the likelihood function for many spatial statistics models is very difficult to compute. There have been many different approaches to overcome this obstacle, for example, using tractable approximations to the likelihood, such as composite likelihoods. An alternative strategy is to use simulation based methods where the central notion is that it is possible to simulate from the complicated likelihood function, even though it may be impossible to actually evaluate the likelihood function. The aim is to explore how such approximations can be used instead of the true likelihood in a Markov chain Monte Carlo setting.

ES42 Room 13 ROBUST STATISTICAL MODELLING Chair: Anthony C. Atkinson

E383: **A generalized weighted likelihood estimator**
Presenter: **Michael Amiguet**, University of Lausanne, Switzerland
Co-authors: Alfio Marazzi
A new robust estimation method is proposed for regression in a wide framework including the GLM. Let $P_{\alpha,\mu}(Y \leq z)$ be a family of distributions depending on two parameters α and μ, where $\mu = E(Y)$ and α is a shape, a scale, or a dispersion parameter. Consider a general regression model where the response Y has cdf $P_{\alpha_0,\mu_0(x)}(Y \leq z)$, x is a covariate vector, $E(Y|x) = \mu_0(x) = h(\beta_0^{\mathrm{T}} x)$, and h is a given link function. A specific example is the negative binomial (NB) regression model, where $Y \sim NB(\alpha_0, \mu_0(x))$ and $\mathrm{Var}(y|x) = \mu_0(x) + \alpha_0 \mu_0(x)^2$. Let $(x_1, y_1), ..., (x_n, y_n)$ be a random sample and consider estimation of α_0 and β_0. If Y is continuous, the tail probabilities $P_{\alpha_0,\mu_0(x_i)}(Y \leq y_i)$ are a sample of a uniform distribution. If Y is discrete, generate m random samples $(u_{11}, ..., u_{n1})$, ..., $(u_{1m}, ..., u_{nm})$ from the uniform distribution on $[0,1]$ and consider the "randomized tail probabilities" $p_{\alpha_0,\mu_0(x_i)}(z_i, u_{ij}) = P_{\alpha_0,\mu_0(x_i)}(Y \leq y_i) - u_{ij} P_{\alpha_0,\mu_0(x_i)}(Y = y_i)$, $i = 1, ...n$, $j = 1, ..., m$. For a given j, $p_{\alpha_0,\mu_0(x_i)}(z_1, u_{1j}), ..., p_{\alpha_0,\mu_0(x_i)}(z_n, u_{nj})$ is a sample from a uniform distribution. A set of weights is then derived from a measure of disparity between the empirical distribution of the (randomized or not) tail probabilities (or of a transformation of them) and their theoretical distribution under the model. These weights are used in a weighted likelihood procedure to estimate α_0 and β_0.

E574: **New LD estimators for length of stay in hospitals**
Presenter: **Nataliya Horbenko**, Fraunhofer ITWM/ University of Kaiserslautern, Germany
Co-authors: Matthias Kohl, Peter Ruckdeschel
Length of stay (LOS) is an important factor for hospital management as hospitals are paid by cases. Atypical cases (outliers) are dangerous for overestimation of LOS, therefore robust procedures are needed for this purpose. We model LOS with Weibull, Gamma, GPD and GEV distributions and apply Location-Dispersion (LD) estimators, based on median and Qn, Sn or kMAD, to estimate LOS on a real data set supplied by the Center for Sepsis Control and Care of the University Hospital Jena. We also determine the impact of outliers on LOS by means of influence functions and describe possible disastrous scenarios.

E448: **The minimum trimmed deviances estimator for the logistic regression model**
Presenter: **Tsung-Chi Cheng**, National Chengchi University, Taiwan
A minimum trimmed deviances (MTD) estimator to deal with multiple outliers for the logistic regression model is proposed. An adaptation of MTD, called the minimum trimmed maximum-dual deviances (MTMdD) estimator, is applied to estimate the common odds regression model and hence to identify outlying tables when some possible outliers exist among several 2×2 contingency tables. We will illustrate the proposed methods by analyzing real data examples in the literature.

E603: **A framework for the comparison of robust regression estimators**
Presenter: **Anthony Atkinson**, London School of Economics, United Kingdom
Co-authors: Marco Riani, Domenico Perrotta
There are several methods for obtaining very robust estimates of regression parameters that asymptotically resist 50% of outliers in the data. Differences in the behaviour of these algorithms depend on the distance between the regression data and the outliers. We introduce a parameter λ that defines a parametric path in the space of models and enables us to study, in a systematic way, the properties of estimators as the groups of data move from being far apart to close together. We examine, as a function of λ, the variance and squared bias of five estimators and we also consider their power when used in the detection of outliers. This systematic approach provides tools for gaining knowledge and better understanding of the properties of robust estimators.

ES48 Room 10 DISEASE MAPPING Chair: Lola Ugarte

E243: **CAR and P-spline models for detecting high risk areas**
Presenter: **Tomas Goicoa**, Universidad Publica de Navarra, Spain
Co-authors: Lola Ugarte, Jaione Etxeberria, Ana F. Militino

There has been a considerable amount of research in disease mapping in the last twenty years. Smoothing risks is one of the main goals in this area to overcome the high variability of classical measures such as the standardized mortality ratio. However, smoothing risks might hinder the detection of high risk areas, since these two objectives are somewhat contradictory. Most of the work on smoothing risks and detection of high risk areas has been derived using conditional autoregressive (CAR) models. Very recently, P-splines models have been incorporated into the disease mapping toolkit and hence it is very interesting to study their performance. Confidence intervals for the log-relative risk predictor will be derived as a tool to detect high-risk areas. P-spline and CAR models will be compared in terms of smoothing (relative bias), sensitivity (ability to detect high risk areas), and specificity (ability to discard false patterns created by noise) through a simulation study based on the well-known Scottish lip cancer data.

E313: **Assessing local spatial clustering in disease from relative risk models**
Presenter: **Peter Congdon**, QMUL, United Kingdom

Spatial analyses of health outcomes at a small-area scale are often used for detecting areas with elevated risk. In spatial analysis of area health outcomes under a Bayesian perspective, the outcomes, namely relative disease risks, are unknowns, and inferences about them are based on hierarchical models. Such models generally focus on the posterior probabilities of elevated risks in each area separately, without necessary regard to the broader local clustering pattern around each area. For example, a high risk area may be an outlier (in local terms) if it is surrounded by low risk areas, whereas a high risk cluster would occur when both the area and its surrounding locality of nearby areas demonstrate common elevated risk. The present analysis discusses the detection and measurement of different forms of local clustering in hierarchical models for small area disease counts. It focusses especially on local join-count statistics as these enable discrimination between different types of clustering. This strategy enables a typology of areas (e.g. areas that are centres of high risk clusters as against areas that are high risk but outliers in terms of the wider locality of areas, etc), and partitioning of the total variation in relative risk between such area types. One may then assess how area predictors (e.g. deprivation) reduce variation in relative risks according to area type. A Bayesian estimation strategy is adopted with prior densities specified on unknown parameters, and posterior inferences based on an MCMC estimation. An application involves hospitalisations for intentional self-harm (ICD10 codes X60 to X84) over a 5-year period in 113 small areas (Middle Super Output Areas) in four London boroughs.

E367: **A smoothed ANOVA model for multivariate ecological regression**
Presenter: **Miguel Martinez-Beneito**, Centro Superior de Investigacion en Salud Publica, Spain
Co-authors: Marc Mari Dell'Olmo, Merce Gotsens, Laia Palencia

Smoothed ANOVA has recently been proposed as a new method to carry out multivariate disease mapping. One of the main features of Smoothed ANOVA is that one can define specific contrasts or comparisons, which may be of interest for the problem at hand. These contrasts are incorporated into the model and conclusions can be drawn from them. This work generalizes Smoothed ANOVA to the case of multivariate ecological regression, that is, when we are interested in determining the relationship between one (or more) covariate(s) and more than one geographically referenced outcome variable. This proposal allows us to decompose the variance on the outcome variables (usually the Relative Risks for a set of diseases) into several orthogonal components explained either by the covariate(s) or by other unobserved factors. Moreover, within this proposal it is also possible to determine and measure the proposed contrasts, which better explain the differences between diseases, and whether these are related to the covariate(s) under study or due to other factors.

E571: **A flexible spatio-temporal functional model for the analysis of mortality risks**
Presenter: **Maria Pilar Frias Bustamante**, University of Jaen, Spain
Co-authors: Maria Dolores Ruiz-Medina, Lola Ugarte, Ana F. Militino

One of the more popular models to analyze the temporal evolution of the geographical patterns of mortality risks in disease mapping is the CAR model. However, the temporal and spatial correlation has a particular and fixed structure. A flexible spatio-temporal functional parametric model, allowing for strong dependence in time and /or space, is used instead. The model will be fitted to the data in the spectral domain, in terms of log-wavelet regression parameter estimators. For smoothing purposes, a spatiotemporal mixed-effect linear approach is considered in the logarithmic scale. Results will be illustrated using mortality data due to prostate cancer in Spain in the period 1975-2008.

ES58 Room 11 RECENT ADVANCES IN NON- AND SEMIPARAMETRIC REGRESSION WITH CENSORED DATA Chair: Anouar El Ghouch

E247: **Single-index model under censoring: A comparative study**
Presenter: **Ewa Strzalkowska-Kominiak**, Universidade da Coruna, Spain
Co-authors: Ricardo Cao

The single-index model is a useful tool to incorporate a d-dimensional vector of covariates X into a regression model avoiding the so called "curse of dimensionality". By assuming that there exists a vector of parameters θ_0 so that the response variable depends only on the projection $\theta_0' X$, one avoids a multivariate regression. This assumption is a reasonable compromise between a fully parametric and a fully nonparametric model. Additionally, in medical or economic studies, the large number of explanatory variables is not the only problem. Very often the response variable is only partly observed and so censored from the right. The goal is to estimate the vector θ_0 using two different estimators of the conditional distribution function and density under censoring. The first one is based on Kaplan-Meier integrals in the presence of the covariates. The second one is based on the Beran estimator for the conditional distribution function. In both cases, the estimation of θ_0 and the choice of bandwidth base on maximum likelihood method. The two methods will be compared in the simulation study and applied to a real data example.

E339: **Nonparametric regression with right-censored and generalized selection biased data**
Presenter: **Cedric Heuchenne**, University of Liege, Belgium
Co-authors: Geraldine Laurent

Suppose the random vector (X,Y) satisfies the nonparametric regression model $Y = m(X) + \varepsilon$, where $m(x) = E[Y|X = x]$ and $\sigma^2(x) = Var[\varepsilon|X = x]$ are unknown smooth functions and the error ε has zero mean and finite variance conditionally on $X = x$. The pair (X,Y) is subject to generalized selection bias while the response is possibly right-censored. We construct new estimators for $m(\cdot)$ and $\sigma^2(\cdot)$ by extending the conditional distribution estimation methods for classical length bias and fixed difference between censoring time and sampling time. Asymptotic properties of the resulting estimators are obtained and the proposed procedures are studied via extended simulations. Finally, a data set on the mortality of diabetics is analyzed.

E358: **Improved semiparametric inference of competing risks data with additive risks and missing indicators**
Presenter: **Jean-Yves Dauxois**, INSA-IMT Toulouse, France
Co-authors: Laurent Bordes, Pierre Joly
The main purpose is to show that, in some situations, the estimators in semiparametric models for competing risks data with missing indicators can be improved in order to reach an asymptotic optimality criterion. More precisely, we assume an additive hazards model on each cause-specific hazard rate function. We also suppose that a random right censoring occurs and that the failure cause is missing for a subset of individuals. Our goal is to estimate the regression parameters as well as functional parameters like the baseline cumulative hazard rate functions or the cumulative incidence functions. We first introduce preliminary estimators of the unknown (Euclidean and functional) parameters when cause of death indicators are missing completely at random (MCAR). They are obtained using the observations with known cause of failure. The advantage of considering such a simple MCAR model appears when we show that, in this case, the information given by the observed lifetimes with unknown failure cause can be used to improve our previous estimates. The large sample behavior of our estimators is obtained and their performance on finite sample sizes illustrated through a simulation study.

E517: **Estimation of varying coefficient models with randomly censored data**
Presenter: **Seong Jun Yang**, Universite catholique de Louvain, Belgium
Co-authors: Anouar El Ghouch, Ingrid Van Keilegom
The varying coefficient model is a useful alternative to the classical linear model, since the former model is much richer and more flexible than the latter. We propose estimators of the coefficient functions for the varying coefficient model in the case where different coefficient functions depend on different covariates and the response is subject to random right censoring. We employ a smooth backfitting technique, that is known to be an effective estimation in structured nonparametric models, since our model has an additive structure and requires multivariate smoothing. The estimators are based on synthetic data obtained by an unbiased transformation. The asymptotic normality of the estimators is established and a simulation study is shown to illustrate the reliability of our estimators.

ES66 Room 7 RECENT ADVANCES IN FUNCTIONAL DATA ANALYSIS Chair: German Aneiros

E119: **Investigating functional datasets with the BAGIDIS semimetric**
Presenter: **Catherine Timmermans**, Universite catholique de Louvain, Belgium
A new method for investigating functional datasets is highlighted. The method is centered on the definition of a new functional, data-driven and highly adaptive semimetric for measuring dissimilarities between curves. It is based upon the expansion of each curve of a dataset into a different wavelet basis, one that is particularly suited for its description. The expansions remain however comparable as they rely on a common notion of hierarchy in describing the curve. Measuring dissimilarities in such a way implies comparing not only the projections of the curves onto the bases but also the bases themselves. Therefore, the name of the method stands for BAses GIving DIStances. Due to its above-mentioned properties, the BAGIDIS semimetric is really powerful when dealing with curves with sharp local features that might be affected simultaneously by horizontal shifts and vertical amplification. As the method overcomes the limitation of expanding all the curves of a dataset in the same basis, it provides for a new paradigm for curve comparison, which opens attractive prospects. The performances of BAGIDIS in applied problems are assessed through the investigation of a spectrometric dataset issued from pharmaceutical research.

E205: **Modeling the probabilistic index in the presence of a functional covariate**
Presenter: **Vanda Inacio de Carvalho**, Pontificia Universidad Catolica de Chile, Chile
Co-authors: Wenceslao Gonzalez-Manteiga, Miguel de Carvalho
Recently, there is an increasing interest in the modeling of the so-called probabilistic index $P(Y_1 < Y_2)$, where Y_1 and Y_2 are two independent continuous random variables. Although there exist some approaches in the literature adjusting this index for covariates, most of the research in this topic concentrates on simply comparing two outcomes, with no covariate adjustment. We propose a nonparametric method for modeling $P(Y_1 < Y_2 \mid X)$ in the presence of a functional covariate X. The corresponding estimator can be viewed as an extension of the Mann-Whitney statistic to the functional context. The method is illustrated with real data and its finite sample performance is evaluated through an extensive simulation study.

E211: **Nonlinear functional data analysis with reproducing kernels**
Presenter: **Hachem Kadri**, INRIA Lille, France
Recent statistical and machine learning studies have revealed the potential benefit of adopting a functional data analysis (FDA) point of view to improve learning when data are objects in infinite dimensional Hilbert spaces. However, nonlinear modeling of such data (aka functional data) is a topic that has not been sufficiently investigated, especially when response data are functions. Reproducing kernel methods provide powerful tools for nonlinear learning problems, but to date they have been used more to learn scalar or vector-valued functions than function-valued functions. Consequently, reproducing kernels for functional data and their associated function-valued RKHS have remained mostly unknown and poorly studied. A learning methodology for nonlinear FDA is described based on extending the widely used scalar-valued RKHS framework to the functional response setting. It introduces a set of rigorously defined reproducing operator-valued kernels suitable for functional response data, that can be valuably applied to take into account relationships between samples and the functional nature of data. Finally, it is experimentally shown that the nonlinear FDA framework is particularly relevant for speech and audio processing applications where attributes are really functions and dependent of each other.

E465: **Functional data analysis for neural synchrony**
Presenter: **Aldana Gonzalez Montoro**, Universidade da Coruna, Spain
Co-authors: Ricardo Cao
A two way functional analysis of variance model is proposed to test hypothesis on a functional outcome variable in a neuroscience context. Synchrony dynamics in time between pairs of neurons is the functional response studied. Data come from a complex experimental setting where neural activity is recorded from the primary visual cortex (V1) of an anesthetized cat. Neurons in V1 are selective to orientation and, during the experiment, the preferred orientation of each cell is determined. Also, two other areas of the cortex are electrically stimulated affecting the connection between neurons in V1. The nature of the raw data makes the functional synchrony indexes exhibit some kind of dependence. The biological problem of interest is whether the levels of the two factors: 'area of stimulation' and 'difference in orientation selectivity', have differential effects in the synchrony between pairs of neurons. A recent functional method based on random projections is used to test for differences among the levels of these two factors along time, succeeding in finding intervals of time where the differences are statistically significant.

ES68 Room 8 COPULAS III **Chair: Wolfgang Trutschnig**

E268: **Detecting and measuring spatial contagion**
Presenter: **Enrico Foscolo**, Free University of Bozen-Bolzano, Italy
Co-authors: Fabrizio Durante
Contagion is usually referred to as the process that describes the spread of financial difficulties from one economy to others in the same region and beyond. In practice spatial contagion between two financial markets X and Y arises when significant increases in comovements of prices and quantities across markets appear, conditional on a crisis occurring in one market or group of markets. A nonparametric test and a related index to detect and measure the contagion effects are introduced. These tools are grounded on the concept of copula and on the use of related conditional Spearman's correlation coefficients. In order to distinguish between normal comovements, due to simple interdependence among markets, and excessive comovements, the proposed approach is based on the determination of a suitable threshold associated with a tail and a central set. As an empirical application, the proposed test is exploited in order to detect contagion in the recent crisis and the related index is used for obtaining a contagion–based hierarchical clustering of European stock market indices. The whole procedure is expected to be useful for portfolio management in crisis periods.

E401: **Bootstrapping copulas for dependence structure testing**
Presenter: **Dominik Sznajder**, KU Leuven, Belgium
Co-authors: Irene Gijbels
Very often a specific dependence structure can be seen as a feature of the underlying copula function. Different features identify the subclasses in the set of the copulas depending on their shape properties. Examples of such dependence structures can be quadrant dependence, tail monotonicity or stochastic monotonicity, and also symmetry properties or stochastic orderings. On the one hand, these kinds of relations between random variables are often encountered in financial, insurance and econometric theoretical considerations and are assumed in modeling. Therefore, it is desirable to test if the assumptions hold for the particular data at hand. On the other hand, there are many parametric copulas which, when used for modeling, impose these features and then a specific dependence structure test can justify if the actual data reveal such characteristics. We propose methodology which is generic and flexible and can be applied to many dependence structures, which can be expressed as shape constraints on a copula function.

E578: **Domain extendibility and limit laws of the multivariate geometric distribution**
Presenter: **Natalia Shenkman**, Technische Universitaet Muenchen, Germany
Co-authors: Jan-Frederik Mai, Matthias Scherer
The discrete multivariate lack-of-memory (LM) property is extended to random vectors with rational components. A set of necessary and sufficient conditions subject to which the domain of a d-variate geometric distribution satisfying the local LM property can be extended to $(k^{-1})^d$, for some $k \in \mathbb{N}$, without violating the LM structure, is derived. The resulting distribution is proved to be unique and is referred to as the k-domain extension. On the one hand, the narrow-sense geometric distribution yields to be k-domain extendible for all $k \in \mathbb{N}$. On the other hand, a multivariate geometric distribution which is k_n-domain extendible for some increasing sequence $\{k_n\}_{n\in\mathbb{N}}$ is geometric in the narrow sense. The k-domain extension of a multivariate geometric distribution is shown to converge weakly to a Marshall-Olkin distribution as $k \to \infty$. Extending the domain of a geometric distribution is relevant for practical applications when the time grid has to be refined without violating the initial dependence structure. Examples are multi-period models in portfolio credit risk.

E381: **Copulas from the Markov kernel point of view**
Presenter: **Wolfgang Trutschnig**, ECSC, Spain
Using the one-to-one correspondence between copulas and special Markov kernels (regular conditional distributions) allows the study of various well-known copula-related concepts in terms of operations on Markov kernels. First of all we will introduce the strong metric D_1 on the space of two-dimensional copulas, state some of its main properties, and show how the metric can be used to define a dependence measure with the seemingly natural property of assigning maximum dependence to (and only to) the class of completely dependent copulas. After that the well-known star product of copulas will be expressed in terms of the standard composition of Markov kernels. This interrelation allows, firstly, the construction of copulas with fractal support (Hausdorff dimension $dim_H \in (1,2)$) which are at the same time idempotent with respect to the star product and, secondly, to characterize the asymptotic behavior of iterates of generalized shuffles of copulas in terms of mixing properties of the corresponding λ-preserving transformations.

ES15 Room 9 BAYESIAN SEMI- AND NONPARAMETRIC MODELLING III **Chair: George Karabatsos**

E149: **Enriched stick breaking processes for functional data**
Presenter: **Bruno Scarpa**, University of Padua, Italy
Co-authors: David Dunson
In many applications involving functional data, prior information is available about the proportion of curves having different attributes. It is not straightforward to include such information in existing procedures for functional data analysis. Generalizing the functional Dirichlet process (FDP), we propose a class of stick-breaking priors for distributions of functions. These priors incorporate functional atoms drawn from a Gaussian process. The stick-breaking weights are specified to allow user-specified prior probabilities for curve attributes, with hyperpriors accommodating uncertainty. Compared with the FDP, the random distribution is enriched for curves having attributes known to be common. Theoretical properties are considered, methods are developed for posterior computation, and the approach is illustrated using data on temperature curves in menstrual cycles.

E186: **ABC-MCMC algorithms with quasi-likelihoods**
Presenter: **Stefano Cabras**, University Carlos III of Madrid, Spain
Co-authors: Maria Eugenia Castellanos, Erlis Ruli
Approximate Bayesian computation (ABC) are computational techniques for Bayesian inference useful when the likelihood function is analytically or computationally intractable. In the framework of ABC based on Markov chain Monte Carlo algorithm (ABC-MCMC) we discuss an automatic proposal distribution, for a scalar parameter θ, based on the theory of quasi-likelihood functions $L_Q(\theta)$. Differently from the ABC framework, in that of the quasi-likelihood the sampling model $\pi(y|\theta)$ is unknown while $L_Q(\theta)$ can be obtained from an unbiased estimating equation $\psi(y;\theta)$, under the assumption that there exists some $\pi(y|\theta)$ which matches $L_Q(\theta)$. We combine the two frameworks by discussing a proposal density for θ. Essentially, given a summary statistic s, we reparametrize θ through the regression function $E(s|\theta)$, whose parameters can be estimated, with the desired precision, by sampling from $\pi(y|\theta)$. Possible extensions are discussed when θ is a vector. We illustrate the method trough some examples.

E497: **Sequential Monte Carlo methods for Bayesian nonparametric models**
Presenter: **Jim Griffin**, University of Kent, United Kingdom

The use of Bayesian nonparametric priors for statistical modelling has been greatly enhanced by the availability of efficient Markov chain Monte Carlo methods. Increasingly structured nonparametric priors (with time-dependence, with covariate-dependence, etc.) are being proposed in the literature, which can lead to posteriors with well-separated modes. Markov chain Monte Carlo methods are not suitable for these types of posterior (unless, specially tailored moves can be developed). Sequential Monte Carlo methods are one class of computational algorithm that can work well on multi-modal posteriors. This talk will consider their application to nonparametric mixture models. The different algorithms will be compared on a range of data sets.

E156: **Bayesian conditional tensor factorizations for high-dimensional classification**
Presenter: **Yun Yang**, Duke University, United States
Co-authors: David Dunson

In many application areas, data of a categorical response and high-dimensional categorical predictors are collected, with the goal of building a parsimonious model for classification while doing inferences on the important predictors. In settings such as genomics, there can be complex interactions among the predictors. By using a carefully-structured Tucker factorization, we define a model that can characterize any conditional probability, while facilitating variable selection and modeling of higher-order interactions. Following a Bayesian approach, we propose a Markov chain Monte Carlo algorithm for posterior computation accommodating uncertainty in the predictors to be included. Under near sparsity assumptions, the posterior distribution for the conditional probability is shown to achieve close to the parametric rate of contraction even in ultra high-dimensional settings in which the number of candidate predictors increases exponentially with sample size. The methods are illustrated using simulation examples and several biomedical applications.

ES02 Room 12 PARALLEL MATRIX COMPUTATIONS AND APPLICATIONS Chair: Stratis Gallopoulos

E1055: **Scalable parallel sparse matrix computations**
Presenter: **Ahmed Sameh**, Purdue University, United States
Co-authors: F. Saied, O. Schenk, M. Manguoglu

Large-scale sparse matrix computations arise in numerous computational statistics and data analysis applications. Such computations often require the use of parallel computing platforms for reducing time-to-solution. Unfortunately, minor modifications of current uniprocessor sparse matrix kernels and algorithms are not sufficient to ensure realizing high performance on parallel computing platforms. Often, one needs to design, afresh, sparse matrix primitives and algorithms capable of minimizing memory references and interprocessor communications. Only then one can achieve high parallel scalability on computing platforms consisting of a single multicore node or many such nodes. We address: (i) designing sparse matrix primitives such as sparse matrix-vector multiplication, sparse matrix reordering (large graph manipulations) that exhibit high degree of parallel scalability, and (ii) designing a hybrid sparse linear system solver that is more scalable than current parallel direct sparse solvers, and more robust than preconditioned iterative solvers. Numerical experiments that demonstrate the robustness and parallel scalability of both our primitives and solver are presented and compared with their counterparts in the Intel Math Kernel Library (MKL), and with the well-known Trilinos library of Sandia's National Laboratories in the U.S.A.

E1051: **Parallel block Chebyshev subspace iteration algorithm optimized for sequences of correlated dense eigenproblems**
Presenter: **Edoardo Di Napoli**, Forschungszentrum Juelich, Germany
Co-authors: Mario Berljafa

In many material science applications simulations are made of dozens of sequences, where each sequence groups together eigenproblems with increasing self-consistent cycle outer-iteration index. Successive eigenproblems in a sequence possess a high degree of correlation. In particular, it has been demonstrated that eigenvectors of adjacent eigenproblems become progressively more collinear to each other as the outer-iteration index increases. This result suggests one could use eigenvectors, computed at a certain outer-iteration, as approximate solutions to improve the performance of the eigensolver at the next one. In order to optimally exploit the approximate solution, we developed a block iterative eigensolver augmented with a Chebyshev polynomial accelerator (BChFSI). Numerical tests show that, when the sequential version of the solver is fed approximate solutions instead of random vectors, it achieves up to a 5X speedup. Moreover the parallel shared memory implementation of the algorithm obtains a high level of efficiency up to 80 % of the theoretical peak performance. Despite the eigenproblems in the sequence being relatively large and dense, the parallel BChFSI fed with approximate solutions performs substantially better than the corresponding direct eigensolver, even for a significant portion of the sought-after spectrum.

E1049: **A new multi-threaded and recursive direct algorithm for parallel solution of sparse linear systems**
Presenter: **Ercan Selcuk Bolukbasi**, Middle East Technical University, Turkey
Co-authors: Murat Manguoglu

Many of the science and engineering applications need to solve linear systems to model a real problem. Usually these linear systems have sparse coefficient matrices and thus demand an effective solution of sparse linear systems which is usually the most time consuming operation. Circuit simulation, material science, power network analysis and computational fluid dynamics can be given as examples of these problems. With the introduction of multi-core processors, it becomes important to solve sparse linear systems effectively in parallel. To take advantage of these developments, a new direct multi-threaded and recursive algorithm based on the generalization of the banded Spike factorization to solve sparse linear systems will be presented. The algorithmic challenges of this approach will be studied on matrices from different application domains. The advantages and disadvantages of variations of the algorithm on different matrices will be discussed. Specifically, we study the effects of changing the number of threads, degree of diagonal dominance, the usage of sparse right hand side feature of Pardiso direct solver and alternative approaches for finding the exact solution. Furthermore, parallel scalability and numerical accuracy comparisons will be made against a well-known direct parallel sparse solver.

E908: **Using parallel computing and data structures on disk to analyze genomic data with R**
Presenter: **Ramon Diaz-Uriarte**, Universidad Autonoma de Madrid, Spain
Co-authors: Oscar Rueda, Daniel Rico

The study of alterations in copy number changes in genomic DNA uses technological platforms that, nowadays, can contain over five million probes per patient, and some of these studies involve hundreds or thousands of patients. Thus, the size of these data sets can create computational challenges. The vast majority of the state-of-the-art algorithms for the statistical analysis of these data are implemented in R, but they often work sequentially, and with data structures that are loaded in memory. Here we describe our work on parallelizing some of the more widely used algorithms while also using data structures that are stored on disk (via the ff package). The combination of these two features makes it possible to analyze very large data sets that will not fit in RAM, and make use of both multicore machines and clusters of workstations to distribute the work,

potentially leading to considerable speed ups. We compare several of our implementations (including some that use local, in-RAM, objects), and discuss some of the main challenges and opportunities of these types of approaches with the continuously increasing genomic data sets.

Sunday 02.12.2012 16:40 - 18:45 Parallel Session L – CFE

CSI02 Room Multipurpose NEW METHODS IN DYNAMIC MODELING AND ECONOMETRICS Chair: Willi Semmler

C1017: **Using nonlinear model predictive control to solve dynamic decision problems in economics**
Presenter: **Willi Semmler**, New School for Social Research, United States
Co-authors: Lars Grune, Marleen Stieler

A new approach to solve dynamic decision models in economics and finance is proposed. The procedure is called Nonlinear Model Predictive Control (NMPC). Similarly to dynamic programming it can solve dynamic decision problems globally. In DP a global solution to the optimal control problem is found by first computing an approximation to the optimal value V and then the optimal control is computed from V. Yet, it has the disadvantage that its numerical effort typically grows exponentially with the dimension of the state variable. As other numerical procedures, applied to solve dynamic decision models, such as provided by some numerical methods to solve DSGE models, DP makes strong assumption on the knowledge and information that agents should have in the context of an infinite horizon model. A remedy to these problems can be obtained by using nonlinear model predictive control (NMPC). Instead of computing the optimal value function for all possible initial states, NMPC only computes single (approximate) optimal trajectories. NMPC has the advantage to compute state variable trajectories and decision variables globally, not using terminal conditions and thus not working with linearizations or local approximations about the steady state. NMPC also allows us to study decision making in a shorter time horizon and agents making decisions over different time horizons. It also permits to study models with parameter uncertainties and learning, models with multiple equilibria, and jumps and regime switches in the dynamics. It does not require to know the terminal or steady state conditions. NMPC works for both known and unknown steady state conditions. Moreover, as compared to the infinite horizon models, there is much less requirement of information that agents need when making decisions. The solution procedure operates on a finite time horizon, but when the time horizon becomes very large it approximates the DP solution. Several examples from economics and finance are computed to illustrate the effectiveness of NMPC to solve those issues.

C1032: **Functional representation, approximation, bases, and wavelets**
Presenter: **James Ramsey**, New York University, United States of America

After stressing the importance of analyzing the various basis spaces, the exposition evaluates the alternative bases available to wavelet researchers. The next step is to demonstrate the impact of choice of basis for the representation or projection of the regressand. The similarity of formulating a basis is explored across a variety of alternative representations. This development is followed by a very brief overview of some articles using wavelet tools. The comparative advantage of wavelets relative to the alternatives considered is stressed.

C1069: **Stochastic extended path**
Presenter: **Michel Juillard**, Bank of France, France
Co-authors: Stephane Adjemian

The Extended Path (EP) approach is known to provide a simple and fairly accurate solution to large scaled nonlinear models. The main drawback of the EP approach is that the Jensen inequality is neglected, because future shocks are (deterministically) set to their expected value of zero. In a previous contribution we showed that the cost of this approximation is small compared to the cost of neglecting the deterministic nonlinearities (as those induced by the presence of a Zero Lower Bound on the nominal interest rate). Now we propose a simple extension to the EP approach by considering that the structural innovations in $t+1$ are non zero and keeping the innovations in $t+s$ ($s>1$) equal to their expected value of zero. We use a quadrature approach to compute the expectations under this assumption. We evaluate the the accuracy of the Stochastic Extended Path approach on a Real Businnes Cycle model. The computing time of this approach is polynomial in the number of endogenous variables but exponential with respect to the number of structural innovations.

CS67 Room 4 NON LINEAR MODELS FOR MACROECONOMICS Chair: Monica Billio

C223: **Testing common nonlinear features in nonlinear vector autoregressive models**
Presenter: **Dao Li**, Dalarna University, Sweden

A special class of vector smooth-transition autoregressive (VSTAR) models containing common nonlinear features (CNFs) is studied. To test the existence of CNFs in a VSTAR model, a triangular representation for such a system containing CNFs is proposed. A procedure of testing CNFs in a VSTAR model is consisting of two steps: first, test unit root in a STAR model against a stable STAR process for each individual time series; second, examine if nonlinear features are common in the system by a Lagrange Multiplier (LM) test when the null of unit root is rejected in the first step. The asymptotic distribution of the LM test is derived. Simulation studies of both unit root test and LM test have been carried out to investigate the finite sample properties. In the empirical application, the procedure of testing CNFs is illustrated by analyzing the monthly growth of consumption and income data of United States (1985:1 to 2011:11). The consumption and income system contains CNFs, and an estimated common nonlinear factor in VSTAR model is suggested.

C210: **An alternative methodology for turning-point detection in business cycle: A wavelet approach**
Presenter: **Peter Martey Addo**, University-of-Paris and EDEEM, France
Co-authors: Monica Billio, Dominique Guegan

A signal modality analysis is provided to characterise and detect non-linearity schemes in the US Industrial Production Index time series. The analysis is achieved by using the recently proposed 'delay vector variance' (DVV) method, which examines local predictability of a signal in the phase space to detect the presence of determinism and non-linearity in a time series. Optimal embedding parameters used in the DVV analysis are obtained via a differential entropy based method using wavelet-based surrogates. A complex Morlet wavelet is employed to detect and characterise the US business cycle. A comprehensive analysis of the feasibility of this approach is provided. Our results coincide with the business cycles peaks and troughs dates published by the National Bureau of Economic Research (NBER).

C778: **Comovement of the Chinese provincial business cycles**
Presenter: **Jamel Gatfaoui**, Aix-Marseille University, France
Co-authors: Eric Girardin

Fluctuations in Chinese growth have become a major concern not simply for the government but increasingly as a driving force for cycles in East Asian as well as world economies. However given the continental character of the Chinese economy, it is quite natural to inquire about the regional dispersion of the national business cycle, even if the existing work has exclusively focused on the latter. We use monthly industrial-output data for 28 provinces over 20 years and a variety of techniques to examine the nature and degree of comovement among Chinese provincial deviation-cycles. Provinces can be classified among five major clusters as a function of standard measures of cyclical characteristics. The main coastal provinces (especially Zhejiang and Guangdong) are by far more in concordance with the national cycle than any other region. Furthermore, they experience expansion phases much longer than recessions. In terms of their concordance with the national business cycle, the national recessions are less reflective of western, southern and northeastern provinces.

C665: **Generalizing smooth transition autoregressions**
Presenter: **Emilio Zanetti Chini**, University of Rome Tor Vergata, Italy

A variant of the smooth transition autoregression (STAR) is introduced. The proposed model is able to parametrize the asymmetry in the tails of the transition equation by using a particular generalization of the logistic function. The null hypothesis of symmetric adjustment toward a new regime is tested by building two different LM-type tests. The first one maintains the original parametrization, while the second one is based on a third-order expanded auxiliary regression. Three diagnostic tests for no error autocorrelation, no additive asymmetry and parameter constancy are also discussed. The empirical size and power of the new symmetry as well as diagnostic tests are investigated by an extensive Monte Carlo experiment. An empirical application of the so generalized STAR (GSTAR) model to four economic time series reveals that the asymmetry in the transition between two regimes is a feature to take into account for economic analysis.

C805: **A Smolyak method with an adaptive grid**
Presenter: **Rafael Valero**, University of Alicante, Spain
Co-authors: Kenneth L. Judd, Lilia Maliar, Serguei Maliar

The Smolyak (sparse) grid is a numerical technique for representing, integrating and interpolating functions on a multidimensional hypercube. Sparse grids are tractable in problems with high dimensionality and can be used to solve large-scale economic models. We show that the accuracy of numerical solutions produced by the Smolyak method can significantly depend on the way in which the state space of a given economic model is mapped into a multidimensional hypercube. We construct an adaptive Smolyak grid using a minimal hypercube that encloses a high-probability area of the ergodic set of a given economic model. Our construction relies on a principle-component transformation of the state variables. By making the Smolyak grid adaptive, we reduce the approximating errors by more than 5 times in one- and multi-agent examples.

CS10 Room 6 JUMPS IN PRICES AND VOLATILITIES Chair: Eduardo Rossi

C553: **Relevance of jumps in returns and jumps in volatility for VaR forecasting**
Presenter: **Juan-Angel Jimenez-Martin**, Complutense University, Spain
Co-authors: Alfonso Novales Cinca

The impact for VaR estimation of modelling jumps in returns and jumps in volatility is analyzed using international stock market index data as well as individual stock data. Our analysis is motivated by the overwhelming evidence available on the need to depart from standard stochastic processes specifications in order to explain the high levels of curtosis present in daily data on financial returns. Evaluation of VaR estimates is made using the Daily Capital Charges measure recently introduced. Incorporating jumps into the stochastic process for the underlying asset implies occasionally large isolated VaR values, which allows for a reduction in the volatility of the regular component of the stochastic noise. The implication is that for a similar accuracy of the VaR estimates, the Daily Capital Charges are lower.

C782: **Averaging tests for jumps**
Presenter: **Ana-Maria Dumitru**, University of Surrey, United Kingdom

A new procedure to detect jumps in prices in the presence of microstructure noise is proposed. The procedure averages the results from existing tests across sampling frequencies. This approach overcomes the sub-sampling and low-power problems that plague existing tests for jumps. We use a modified version of Fisher's method to combine p-values for previous tests applied at different sampling frequencies. We propose a double bootstrap procedure to obtain an empirical distribution for Fisher's test statistic. As an additional contribution, we prove the bootstrap consistency of those previous tests.

C962: **Potential over-detection of jumps with conventional sampling frequencies**
Presenter: **Taesuk Lee**, University of Auckland, New Zealand

Because of market micro structure noise, it is common in the empirical finance literature to choose somewhat moderate sampling frequency such as 5 minutes or 15 minutes. To focus on the potential jump over-detection with conventional choice of sampling frequency, this paper ignores the noise and introduces examples of hump-shaped size distortion of tests. If the underlying volatility follows the square root process with a fast mean-reverting drift and a large diffusion term, the bi-power variation may show the non-monotonic bias, which brings hump-shaped jump detection probabilities. The bias of bi-power variation is approximated by the Euler scheme under asymptotic assumptions on sampling frequency and diffusion parameters. Through Monte Carlo experiments, the hump-shaped rejection probabilities are illustrated with the approximated and simulated bias of bi-power variations. The approximated bias as well as the simulated bias can be used to check the depth of the bias of bi-power variations and size distortion of tests for jumps.

C538: **On the consistendy of classical option pricing models with observed option second order moment characteristics**
Presenter: **Matthias Fengler**, University of Sankt Gallen, Switzerland
Co-authors: Francesco Audrino

Ex post intra-day variability in options markets and the associated underlying markets are studied. We then ask whether the measures of realized variance are compatible with each other under the assumptions of classical option pricing models traditionally used. We study single diffusion and one-factor stochastic volatility models as well as the role of jumps in the underlying asset. We find that one-dimensional diffusions are not compatible with options realized variance.

C302: **Multiplicative error model with jumps**
Presenter: **Eduardo Rossi**, University of Pavia, Italy
Co-authors: Massimiliano Caporin, Paolo Santucci de Magistris

The volatility of financial returns is characterized by rapid and large increments. An extension of the Multiplicative Error Model (MEM) for estimating the presence of jumps in volatility is proposed (MEM-J), using a realized volatility measure as volatility proxy. The chosen specification for the volatility is given by a combination of a continuous volatility component combined with a discrete compound Poisson process for the jumps; the two elements are affecting the total volatility in a multiplicative framework. The moments and the likelihood function are obtained in closed form. A Monte Carlo simulation experiment shows the properties of the model and the finite-sample features of maximum likelihood estimation. The MEM-J model is able to capture jumps when those are present in the volatility process. The maximum likelihood estimates of the jump component seem to be reliable, even when jumps are rare events. The empirical application focuses on a set of 36 NYSE stocks, the estimation results highlights a positive probability of jumps in volatility, which is consistent with the findings of previous studies on the topic.

CS19 Room 5 MODELLING AND FORECASTING FINANCIAL MARKETS Chair: Dudley Gilder

C328: **Dynamic models of exchange rate dependence using option prices and historical returns**
Presenter: **Leonidas Tsiaras**, Aston University, United Kingdom
Models for the conditional joint distribution of exchange rate pairs are evaluated and compared. The conditional dependency is allowed to vary across time, as a function of either historical returns or a combination of past return data and option-implied dependence estimates. Using prices of currency options that are available in the public domain, risk-neutral dependency expectations are extracted through a copula representation of the bivariate risk-neutral density. Within this framework, one can extract forward-looking information regarding the overall degree of covariation, as well as, the level and direction of asymmetric dependence. Time-series models that include option-based measures in their information set are found to outperform, in-sample and out-of-sample, models that rely solely on historical returns.

C579: **Covariance forecasting using high-frequency data and a single factor model**
Presenter: **Dudley Gilder**, Aston Business School, United Kingdom
Co-authors: Mark Shackleton, Stephen Taylor
All models used to forecast covariance matrices are misspecified. This is due to a trade-off between being able to replicate the time series dynamics of positive-definite covariance matrices and the curse of dimensionality. We use high-frequency data to generate covariance matrix forecasts using a single factor model and compare the forecasts to those obtained from alternative high-frequency time series models. We use a proxy of the market portfolio as the factor and construct covariance matrix forecasts for the constituents of the Dow Jones Industrial Average. Due to the simplicity of the single factor model, we are able to examine whether incorporating forward-looking information from option prices improves our covariance matrix forecasts.

C419: **Sovereign credit ratings, market volatility information and financial gains**
Presenter: **Pedro Gomes**, Universidad Carlos III , Spain
Co-authors: Antonio Afonso, Abderrahim Taamouti
The reaction of bond and equity market volatilities in the EU to sovereign rating announcements (Standard & Poor's, Moody's, and Fitch) is investigated using panel regressions with daily stock market and sovereign bond returns. The parametric volatilities are defined using EGARCH specifications. Alternatively we use the absolute value and the squared returns as proxies of volatilities. We find that upgrades do not have significant effects on volatility, but downgrades increase stock and bond market volatility. Contagion is present, with a downgrade increasing the volatility of all other countries. Markets respond more to S&P and Moody's, with higher stock and bond returns' volatility when sovereign downgrades occur. Finally, we find that there is a financial gain for portfolio returns when taking into account sovereign credit ratings information for volatility modelling, with financial gains decreasing with higher risk aversion.

C791: **Hedge fund return predictability**
Presenter: **Spyridon Vrontos**, University of Piraeus, Greece
Co-authors: Ekaterini Panopoulou
While the majority of the predictability literature has been devoted to the predictability of traditional asset classes, the literature on the predictability of hedge fund returns is quite scanty. We focus on assessing in- and out-of-sample predictability of hedge fund strategies. We employ an extensive list of predictors. Aiming at reducing uncertainty risk associated with a single predictor model, we first engage into combining the individual forecasts. We consider various combining methods ranging from simple averaging schemes to more sophisticated ones, such as discounting forecast errors, cluster combining and principal components combining. Our second approach applies bootstrap aggregating (bagging) to a general-to-specific procedure based on a general dynamic linear regression framework containing our set of predictors. The statistical significance of our forecasts is assessed against the benchmark constant return model by employing suitable tests for equal predictive ability and forecast encompassing. To evaluate the economic significance of our approaches, we construct portfolios in the context of a component-rotation investment strategy.

C949: **Hedge fund return dynamics: Long memory and regime switching**
Presenter: **Mohamed-Ali Limam**, University of Montpellier, France
Co-authors: Virginie Terraza, Michel Terraza
The dynamics of hedge fund returns and their behavior of persistence are investigated in a unified framework through a Markov Switching ARFIMA model. Major results, based on the CSFB/Tremont hedge fund indexes monthly data during the period 1994-2011, highlight the importance of the long memory parameter magnitude, i.e shocks in shaping hedge fund return dynamics. They also show that the hedge fund dynamics are characterized by two levels of persistence. In the first one, associated to low-volatility regime, hedge fund returns are a stationary long memory process. In the second one, associated to high-volatility regime, returns exhibit higher parameter of fractional integration. More precisely, in high volatility regime the process tends to be non-stationary and exhibits a mean-reverting behavior in periods of turmoil. The findings are interesting and enable us to establish a relationship between hedge fund return states and memory phenomenon.

CS30 Room 1 TIME-VARYING PARAMETER MODELS Chair: Pawel Janus

C883: **Financial conditions, crises, and uncertainty: A dynamic factor perspective**
Presenter: **Scott Brave**, Federal Reserve Bank of Chicago, United States
Co-authors: Andrew Butters
The mixed frequency latent dynamic factor model behind the Federal Reserve Bank of Chicago's National Financial Conditions Index (NFCI) is used to investigate U.S. financial crises and uncertainty shocks. Episodes of financial stress in recent U.S. history are shown to be identified with shifts in the mean, variance, and kurtosis of shocks to the NFCI according to both 2 and 3-state Markov processes. Crises are characterized by a combination of high mean, high variance, and leptokurtic shocks which are only slowly unwound over time. In addition, the conditional variance of these shocks is shown to be highly correlated at contemporaneous and leading horizons with second-moment shocks derived from existing measures of economic and policy uncertainty. This is demonstrated nonparametrically using receiver operating characteristic (ROC) analysis as well as by mixed frequency vector auto-regression (VAR) techniques.

C572: **Model averaging in dynamic factor models: An application to electricity prices forecasting**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Andres M. Alonso, Guadalupe Bastos
In the context of the liberalization of electricity markets, forecasting of electricity prices is an essential task for all the agents involved in electricity markets' operation. With this aim, research has evolved in order to be able to model the particularities of electricity prices, which make their forecasting complex. In particular, dynamic factor models have been quite successful in the task. However, specifying a unique model for the factors, which are unobserved, might result in a difficult task, and there are no sound grounds to guarantee that such a unique model exists. Model averaging is employed with the objective of overcoming this difficulty, and as many research papers claim, with the expectation that electricity

prices would be better forecast by a combination of models rather than by a single one. Numerical results are provided both for the Iberian and the German markets, MIBEL and EEX, respectively.

C577: **Collapsed dynamic factor model: Application for the eurozone**
Presenter: **Irma Hindrayanto**, De Nederlandsche Bank, Netherlands
Co-authors: Jasper de Winter, Siem Jan Koopman
In the classic dynamic factor model (DFM), a small number of unobserved common dynamic factors drive the observed movements of a large number of economic time series. The technique has been proven a powerful tool for short-run forecasting and policy analysis. However choosing the appropriate dimension in DFM specifications is still a topic of ongoing research. We investigate the empirical performance of the so-called collapsed factor model, where an autoregressive (AR) term of the target variable is added into the framework and where instead of the large vector of indicator variables, a small vector of estimated common factors of these indicators is included in the equation. We use this parsimonious model to forecast both quarterly and yearly GDP growth of the eurozone and its five largest countries. We compare our results with the classic DFM in detail, including an in depth review of the relative forecasting performance during recession, expansions and the recent financial crisis. We are fully accounting for the unbalancedness of the data by performing an out-of-sample pseudo real-time forecasting exercise. We conclude that the inclusion of AR term dramatically improves the forecasting quality of yearly GDP growth, especially during the recent financial crisis.

C353: **Estimating daily variation in aggregate expectations from daily survey responses**
Presenter: **Frieder Mokinski**, ZEW Mannheim, Germany
An approach to estimate the day-to-day variation in aggregate expectations based on daily survey data is presented. The approach requires information about the day on which individual expectations have been formed. This kind of data is sometimes gathered in surveys. Understanding individual responses as noisy signals about latent aggregate expectations, a state space model is set up, in which the average response of a day is the measurement and aggregate expectations are the latent state. Contrary to standard applications of state-space models, since the measurement is an average of individual measurements, its precision varies exogenously with the number of responses on a day. By filtering and smoothing algorithms it is nevertheless a routine task to estimate the daily variation in the latent aggregate expectations. It is shown how to apply this approach both to point and to directional survey expectations. Moreover, it is shown how additional variables with daily variation can be used to improve the estimate of latent aggregate expectations, especially on days when no survey responses are available. In two empirical applications it is shown why estimates of daily variation in aggregate expectations are useful. First of all, they deliver more precise estimates of aggregate expectations in realtime. Secondly, they facilitate ex-post estimation of the impact of events on aggregate expectations.

C555: **Modeling daily financial covariance matrix by combining multiple realized measures**
Presenter: **Pawel Janus**, UBS, Switzerland
Co-authors: Siem Jan Koopman
A novel approach is proposed for modeling high dimensional daily covariance matrix of financial assets. The development of the model starts from joint modeling of daily asset returns and daily realized covariance measured from high-frequency intraday data. The key novelty is the dynamic model formulation for covariance matrix based on model-implied weighted combination of past realized measures of daily covariance matrices and past outer-products of daily returns. In principle, the proposed model is well-suited for the use of multiple noisy and frequency-varying signals of daily covariance matrix. We demonstrate that in our model specification each individual innovation driving the update of covariances exploits full likelihood information. Consequently, even a parsimonious model formulation allows for cross-asset effects implying that a number of model static parameters can be kept low in high dimensions without loss of flexibility. We illustrate our novel model to a set of NYSE equities, and we show that our model outperforms some alternative frameworks.

CS37 Room 2 NUMERICAL ANALYSIS AND COMPUTATION OF UNIT ROOT DISTRIBUTIONS Chair: Roderick McCrorie

C566: **The exact asymptotic first-order bias in least squares estimation of the AR(1) model under a unit root**
Presenter: **Roderick McCrorie**, University of St Andrews, United Kingdom
The still-open classical problem of fully characterizing the exact asymptotic bias in least squares estimation of the first-order autoregressive model is addressed. Under the unit root hypothesis, the asymptotic first-order bias is governed by a constant that is, by construction, the mean of the Dickey-Fuller distribution. In spite of the prevalence of this distribution, little is known about its mean when viewed as a mathematical constant. We show that characterizing the constant is essentially a Basler-type problem and offer an approach based on generalized hypergeometric series. The constant is shown to have the same essential form as Catalan's constant, many of whose number-theoretic properties remain unresolved even to this day. Using a group-theoretic approach, we classify a group of twelve generalized hypergeometric series allied to the constant, and provide closed forms for these series, none having been summed in the mathematical literature to date.

C516: **Closed-form approximations to moment values of an AR(1) model in unit-root and stationary cases**
Presenter: **Huihua Xie**, National University of Singapore, Singapore
Co-authors: Albert K. Tsui
It is well-known that the ordinary least-squares (OLS) estimator of first-order autoregressive (AR(1)) models is biased in finite samples. Unfortunately the finite sample distributions of such estimators remain unknown. To study its distributional properties, the lower order moments have been explored extensively in both analytical and computational approaches. However, most of such studies in the analytical approach focus on the first two moments and obtain only a few terms which are expressed in terms of the sample size and the autoregressive parameter. To the best of our knowledge, the closed-form expressions for mean, variance, skewness and kurtosis remain unknown. We revisit these classic problems, and are able to obtain analytically simple but accurate closed-form approximations to the first four moments of the OLS estimator of AR(1) model without intercept. It includes the stationary case where the autoregressive parameter is less than 1 and the unit-root case where the autoregressive parameter is equal to 1. Accuracies of the moment values computed by the closed-form approximations are compared with those exact values computed by numerical methods. It is found that our approximations are reasonably accurate for a wide range of sample size and autoregressive parameter, with errors less than 0.1% in most of the cases.

C701: **Simulating the asymptotic distributions and densities of the OLS estimator in autoregressive models with a unit root**
Presenter: **Liang Cao**, University of St Andrews, United Kingdom
The problem of simulating the distributions and densities of the OLS estimator in autoregressive models with a unit root is considered. This problem is non-trivial owing to the transcendental nature of the mathematical formulae describing the relevant distributions and densities that have hitherto been derived. It has almost always been attacked by first simulating data from the given autoregressive specification. We show that this approach makes it difficult to accurately approximate the tail behavior of the distributions. We outline two new approaches to the problem: first, we use and adapt previous analytical expressions for the various joint densities as the basis for simulating them in a numerically robust way; and secondly we simulate the same densities via the numerical inversion of the two-dimensional Laplace transforms that underlie them. Various methods of numerical inversion are compared and contrasted. Our results indicate that the method based on the numerical inversion of the Laplace transform

is the most robust for simulating the relevant densities, and using multi-precision arithmetic (i.e. using more than double-precision arithmetic) is essential to capture the tail behavior accurately.

C310: **Overlapping sub-sampling and invariance to initial conditions**
Presenter: **Maria Kyriacou**, University of Southampton, United Kingdom
The use of the overlapping blocking scheme in unit root autoregression is studied. When the underlying process is that of a random walk, the blocks' initial conditions are not fixed, but are equal to the accumulated sum of all the previous observations' error terms. Under the non-overlapping sub-sampling scheme, these initial conditions do not disappear asymptotically. It will be shown that a simple way of overcoming this issue is to use overlapping blocks. If the block length grows at the same rate as the sample size, the effect of initial conditions vanishes asymptotically. This implies that for each block, the same expansion can be used to describe the bias of the associated least squares estimator of the autoregressive parameter. This result motivates an application to jackknife estimators. A jackknife estimator based on overlapping blocks is shown to provide obvious reductions to both bias and mean square error.

C874: **Numerical distribution functions for seasonal unit root tests**
Presenter: **Ignacio Diaz-Emparanza**, Universidad del Pais Vasco UPV/EHU, Spain
When working with time series data observed at intervals of less than a year, it is often necessary to test for the presence of seasonal unit roots. One of the most widely used methods for testing for seasonal unit roots is based on regressing the seasonal difference of the series over the transformations of the series by applying specific filters for each seasonal frequency, which provides test statistics with non-standard distributions. A generalisation of this method for any periodicity is used and a response surface regressions approach is presented for calculating the $P-$values of the statistics whatever the periodicity, autorregressive order and sample size of the data. The algorithms are prepared with the Gretl open source econometrics package and two empirical examples are presented.

CS22 Room 3 WAVELETS AND MACROECONOMIC ANALYSIS Chair: Marco Gallegati

C316: **Oil shocks and the euro as an optimum currency area**
Presenter: **Luis Aguiar-Conraria**, Universidade do Minho, Portugal
Co-authors: Maria Joana Soares, Teresa Rodrigues
Wavelet analysis is used to study the impact of the Euro adoption on the oil price macroeconomy relation in the Euroland. We uncover evidence that the oil-macroeconomy relation changed in the past decades. We show that after the Euro adoption some countries became more similar with respect to how their macroeconomies react to oil shocks. However, we also conclude that the adoption of the common currency did not contribute to a higher degree of synchronization between Portugal, Ireland and Belgium and the rest of the countries in the Euroland. On the contrary, in these countries the macroeconomic reaction to an oil shock became more asymmetric after adopting the Euro.

C325: **A framework for a nondyadic wavelets analysis**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom
The conventional dyadic multiresolution analysis constructs a succession of frequency intervals in the form of $(pi/2j, \pi/2^{j--1}); j = 1,2,\ldots,n$ of which the bandwidths are halved repeatedly in the descent from high frequencies to low frequencies. Whereas this scheme provides an excellent framework for encoding and transmitting signals with a high degree of data compression, it is less appropriate to statistical data analysis. A wavelet analysis which allows the wave bands to be defined more flexibly than in the case of a conventional dyadic analysis is described. The wavelets that form the basis vectors for the wave bands are derived from the Fourier transforms of a variety of functions that specify the frequency responses of the filters corresponding to the sequences of wavelet coefficients.

C773: **Essay on wavelet analysis and the European term structure of interest rates**
Presenter: **Michaela Kiermeier**, University of Applied Sciences Darmstadt, Germany
Approaches to model and forecast term structure of European interest rates are summarized. For forecasting purposes the generalized, dynamic Nelson Siegel approach is examined. Five factors are used as explanatory risk factors in an arbitrage free model setting. In our empirical work we identify the number of factors within a factor analysis to find the number of risk factors evaluated by the market. Furthermore, due to the last financial market crisis we do not want to impose the assumption of parameter constancy so that we apply the Fama-McBeth approach to test for significant risk factors in a series of cross section regressions. We find that four factors are evaluated by the market for the determination of the term structure. Since price movements in financial markets are driven by news, we apply wavelet analysis to decompose the identified risk factors in their expected and unexpected components. In out of sample forecasts we assume that market participants form their expectations according to the most important time scales of the risk factors. We find good forecasting abilities of this approach, the one month forecast remains high even during times of the last financial market crisis.

C777: **Time-scale and the S-Curve: A wavelet analysis of trade dynamics**
Presenter: **Jonathan Thong**, Unversity of Sydney Business School, Australia
Co-authors: James Ramsey
The dynamic structural relationship between the trade balance and the terms of trade is studied by means of a time-scale decomposition. Our analysis is carried out using data obtained from ten developed economies; Australia, Austria, Canada, France, Italy, Japan, Norway, Switzerland, the United Kingdom and United States. In doing so, we make two substantive contributions. First, we perform Granger Causality tests to analyze the Harberger-Laursen-Metzler effect upon the international trade of goods and services for our economies of interest across different time scales. The results of these tests motivate a second, deeper empirical exploration into the underlying structural dynamic relationship of these two variables. Using wavelet analysis, we show that by employing a time scale decomposition of the data we are able to verify the presence of the *S*-curve previously predicted by uncovering a far more precise depiction of the shape of the cross correlation function between the trade balance and the terms of trade. Through our findings we demonstrate that time scale decompositions can uncover features of economic phenomena that may be undetected by standard techniques

C993: **A wavelet-based index of financial distress**
Presenter: **Marco Gallegati**, Polytechnic University of Marche, Italy
Co-authors: James B. Ramsey, Willi Semmler
As a result of the increased occurrence of crises events over the past decades several composite financial stress indexes (FSI) have been proposed to assess the presence of potential instability in the financial system. Variable selection and appropriate weighting represent potential problems of this methodology. In contrast to standard approach we consider that the aspect of financial stress captured by each of these variables can change with the time frame adopted and thus expected to vary across scales. We propose a new method to construct composite financial stress indexes by taking advantage of the time scale decomposition property of wavelet analysis, and of the frequency dependence of economic and financial relationships. Specifically, after decomposing all the variables used in the IMF's FSI into their time-scale components through the maximum overlap discrete

wavelet transform (MODWT), we construct a composite indicator for the US by simply adding different time scale components of several variables, i.e. corporate bond yield spread, inverted term spread and monthly stock returns. The performance of such a wavelet-based index is compared to well known episodes of financial stress identified in the literature and contrasted with a measure of output growth with interesting results.

Sunday 02.12.2012 16:40 - 18:45 Parallel Session L – ERCIM

ES06 Room 11 STATISTICS OF EXTREMES AND APPLICATIONS I **Chair: Juerg Huesler**

E208: **An asymptotically unbiased minimum density power divergence estimator of the Pareto-tail index**
Presenter: **Armelle Guillou**, Strasbourg, France
Co-authors: Goedele Dierckx, Yuri Goegebeur

The density power divergence between two densities functions f and g, denoted $\Delta_\alpha(f,g)$, was introduced in 1998 in order to develop a robust estimation criterion. It depends on a parameter $\alpha \geq 0$. Assume that the density function g depends on a parameter vector θ, and let $Y_1,\ldots,Y_n$ be a sample of independent and identically distributed random variables according to density function f. The minimum density power divergence estimator (MDPDE) is the point θ minimizing the empirical density power divergence. In case $\alpha = 0$, the empirical density power divergence corresponds with minus the log-likelihood function. The parameter α controls the trade-off between efficiency and robustness of the MDPDE: the estimator becomes more efficient but less robust against outliers as α gets closer to zero, whereas for increasing α the robustness increases and the efficiency decreases. In this talk we use the density power divergence criterion with the objective to obtain a robust unbiased estimator for the tail parameter of a Pareto-type distribution. Under a second order condition, we establish the asymptotic normality of this estimator and illustrate its efficiency on a small simulation study.

E220: **Periodicity and clustering of rare events on dynamical systems**
Presenter: **Jorge Freitas**, Universidade do Porto, Portugal
Co-authors: Ana Freitas, Mike Todd

Rare events for chaotic dynamical systems are studied. We will address this issue by two approaches. One regards the existence of Extreme Value Laws (EVL) for stochastic processes obtained from dynamical systems, by evaluating a fixed random variable along the orbits of the system. The other has to do with the phenomenon of recurrence to arbitrarily small (hence rare) events, which is commonly known as Hitting Time Statistics (HTS) and Return Time Statistics (RTS). We will show the connection between the two approaches both in the absence and presence of clustering. The strength of the clustering is quantified by the Extremal Index (EI), that takes values between 0 and 1. Using the connection between EVL and HTS/RTS, we will interpret the existence of an EI less than 1 as due to the presence of underlying periodic phenomena.

E415: **The extreme value Birnbaum-Saunders model in athletics and environment**
Presenter: **Ivette Gomes**, CEAUL and DEIO, Faculty of Science, University of Lisbon, Portugal
Co-authors: Marta Ferreira, Victor Leiva

The Birnbaum-Saunders (BS) model is a life distribution that has been largely applied in recent decades. A random variable following the BS distribution can be stochastically represented by another random variable, a standard normal, used as a basis. Consequently, the BS model can be generalized by switching the standard normal distribution of the basis random variable, allowing the construction of more general classes of models. Among those models, the extreme value Birnbaum-Saunders models, based on results from extreme value theory, have been recently introduced in the literature. The way these models can link with traditional extreme value analysis is now provided, through applications to athletics and environmental data.

E826: **Spatial analysis of extremes in environmental time series**
Presenter: **Manuela Neves**, Technical University of Lisbon, Portugal
Co-authors: Clara Cordeiro, Dora Prata Gomes

Spatial statistics deals with statistical methods in which spatial locations play an explicit role in the analysis of data. Classical geostatistics, widely used for spatial data, is mostly based on multivariate normal distributions that are inappropriate for modelling tail behaviour. Classical statistics of extremes is very well developed in the univariate context, for modelling and estimating parameters of rare events. However, such rare events often present a temporal aspect, a spatial aspect or both. The adequate spatial analysis of the extremes of a natural process such as precipitation or temperature is important in environmental time series, for adequate risk management. Max-stable processes arise from an infinite-dimensional generalisation of extreme value theory. They form a natural class of processes when sample maxima are observed at each site of a spatial process. Results for Spatial Analysis of Extremes recently developed are presented. The objectives are to show the recent developments of extremal spatial approaches, to show the main packages developed in R environment devoted to Extremal Spatial Statistics and to perform an application of the aforementioned procedures to an annual maximum series of daily precipitation levels in Portugal at different sites.

E862: **Bayesian Dirichlet mixture model for multivariate extremes**
Presenter: **Anne Sabourin**, Universite de Lyon, France
Co-authors: Philippe Naveau

The probabilistic framework of extreme value theory is well-known. The dependence among large events is characterized by an angular measure on the positive quadrant of the unit sphere. The family of these so-called "spectral measures" is non-parametric by nature. A Dirichlet mixture model for spectral measures was already studied. This semi-parametric model was theoretically valid in arbitrary dimension, but Bayesian inference with their proposed parametrization was very challenging in dimension greater than three. We propose a new parametrization allowing for a natural prior specification. A new reversible-jump algorithm is also built to approximate the posterior, and tested up to dimension five. We prove the prior's weak consistency and the ergodicity of the Markov chain. Finally, we propose a convergence assessment technique based on integration against Dirichlet test functions.

E932: **A Bayesian hierarchical model for spatial extremes with multiple durations**
Presenter: **Yixin Wang**, Hong Kong University of Science and Technology, China
Co-authors: Mike K.P. So

Bayesian spatial modeling of extreme values has become increasingly popular for its usefulness in obtaining relevant uncertainty measures of estimates. We propose a Bayesian hierarchical model to describe the dependence of extreme return levels on spatial locations and duration periods. The peak over threshold approach is adopted to characterize extreme intensities: we pool the data from all duration periods and incorporate duration periods as one of the covariates characterizing the latent spatial process that drives the behavior of Generalized Pareto Distribution (GPD) parameters. This potentially avoids possibly nonsensical return level estimates produced by modeling data from different duration periods separately. The models are fitted using a Markov Chain Monte Carlo (MCMC) algorithm to perform Bayesian hierarchical inference for the model parameters as well as uncertainty measures. The methodology is applied to simulated data and real precipitation data.

ES16 Room 13 ROBUST FUNCTIONAL DATA ANALYSIS Chair: Graciela Boente

E151: **Robust estimators under a functional common principal components model**
Presenter: **Juan Lucas Bali**, Universidad de Buenos Aires, Argentina
Co-authors: Graciela Boente
When dealing with several populations of functional data, equality of the covariance operators is often assumed. Usually, if this assumption does not hold, one estimates the covariance operator of each group separately, which leads to a large number of parameters. As in the multivariate setting, this is not satisfactory since the covariance operators may exhibit some common structure. We discuss the extension to the functional setting of projection–pursuit estimators for the common directions under a common principal component model that has been widely studied when dealing with multivariate observations. We present estimators of the unknown parameters combining robust projection–pursuit with different smoothing methods. We obtain consistency results under mild assumptions.

E172: **Quantiles for functional data**
Presenter: **Beatriz Pateiro-Lopez**, Universidad de Santiago de Compostela, Spain
Co-authors: Ricardo Fraiman
We propose a projection-based definition of quantiles in a multivariate and infinite-dimensional setting. The directional quantiles we define are shown to satisfy desirable properties of equivariance. Sample quantiles estimating the corresponding population quantiles are defined and consistency results are obtained. The new concept of principal quantile directions is closely related in some situations to principal component analysis. It is found specially attractive for reducing the dimensionality and visualizing important features of functional data. Asymptotic properties of the empirical version of principal quantile directions are also obtained. Based on these ideas, a simple definition of robust principal components for finite and infinite-dimensional spaces is also proposed.

E428: **Robust estimation of principal functions using M-estimation**
Presenter: **Nedret Billor**, Auburn University, United States
Co-authors: Seokho Lee, Hyejin Shin
A robust functional principal component analysis is proposed which is based on estimating principal functions by M-estimation with roughness penalty in the presence of outlying curves. This method is efficient since it requires to downweigh abnormally observed individual measurements that may exist in a single curve instead of removing or downweighing a whole curve so that the maximal information from whole observed curve is retained. The resulting principal functions by the proposed method are given as M-type smoothing splines, so that the associated computation using the properties of a natural cubic spline becomes fast even for long and dense functional data. We demonstrate the performance of the proposed method on simulated and lip movement data and compare it with classical and some existing robust functional component analysis techniques.

E471: **Detecting outlying curves based on functional depth**
Presenter: **Kaveh Vakili**, KU Leuven, Belgium
Co-authors: Mia Hubert, Gerda Claeskens
In functional analysis the observations typically consist of one set of curves, measured at a discrete set of time points. In the multivariate functional data setting, measurements consists of several sets of curves, measured at the same time points. A new depth function has recently been proposed for this type of data, called multivariate functional halfspace depth (MFHD). This depth function can for example also be applied to observed curves and their derivatives, or to warped curves and their warping functions. It is shown how our multivariate approach yields better estimates for the central tendency than existing depth functions for univariate curves. It is also illustrated how MFHD can be useful to detect globally outlying curves as well as curves that are only outlying on parts of their domain. Several graphical representations of the curves and their degree of outlyingness are presented.

E511: **S-estimators for functional principal component analysis**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Matias Salibian-Barrera
A well-known property of functional principal components is that they provide the best $q-$dimensional approximation to random elements over separable Hilbert spaces. Our approach to robust estimates of principal components for functional data is based on this property since we consider the problem of robustly estimating these finite-dimensional approximating linear spaces. We propose a new class of estimators for principal components based on robust scale functionals by finding the lower dimensional linear space that provides the best prediction for the data. In analogy to the linear regression case, we call this proposal S-estimators. This method can also be applied to sparse data sets when the underlying process satisfies a smoothness condition with respect to the functional associated with the scale defining the S-estimators. The motivation is a problem of outlier detection in atmospheric data collected by weather balloons launched into the atmosphere and stratosphere.

ES17 Room 8 STATISTICAL METHODS FOR NON-PRECISE DATA Chair: Angela Blanco-Fernandez

E510: **Further results on the confidence sets for the Aumann mean of a random closed set**
Presenter: **Raffaello Seri**, University of Insubria, Italy
Co-authors: Christine Choirat
A bootstrap method to build confidence sets for the Aumann mean of a random closed set as estimated through the Minkowski empirical mean has already been developed by the authors. The method relies on the characterization of the random closed set through its support function. The resulting confidence set is not exact but performs remarkably well in simulations. We further investigate some properties of this confidence set and, in particular, we propose an exact version of it. The performance of the new set is investigated through simulations.

E732: **Likelihood-based belief function induced by uncertain data**
Presenter: **Thierry Denoeux**, Universite de Technologie de Compiegne, France
As shown by several authors, the normalized likelihood function can be seen as the contour function (possibility distribution) of a consonant belief function in the parameter space. This approach is compatible with Bayesian inference, as combining this likelihood-based belief function with a Bayesian prior yields the Bayesian posterior. Other usual notions such as the likelihood ratio statistics or the profile likelihood receive interesting interpretations in this context. Here, we extend this formalism to the case where observations are uncertain and described by possibility distributions or, more generally, belief functions. We show that data uncertainty induces a loss of information in the likelihood-based belief function, which can be quantified. We give several applications to different statistical inference problems.

E225: **On possibilistic clustering with repulsion constraints for imprecise data**
Presenter: **Maria Brigida Ferraro**, Sapienza University of Rome, Italy
Co-authors: Paolo Giordani
In possibilistic clustering the objects are assigned to the clusters according to the so-called membership degrees taking values in the unit interval.

Differently from fuzzy clustering, it is not required that the sum of the membership degrees of an object in all the clusters is equal to one. This is very helpful in the presence of outliers, which are usually assigned to the clusters with membership degrees close to zero. Unfortunately, a drawback of the possibilistic approach is the tendency to produce coincident clusters. A remedy is represented by the use of a repulsion term among prototypes in the loss function forcing the prototypes to be 'enough' far from each other. Here, a possibilistic clustering model with repulsion constraints for imprecise data, managed in term of fuzzy sets, is introduced. Applications to synthetic and real fuzzy data are considered in order to analyze how the proposed clustering model works in practice.

E960: **Testing the inclusion of the Aumann expectation of random sets in a compact and convex set**
Presenter: **Ana Belen Ramos-Guajardo**, University of Oviedo, Spain
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez, Maria Angeles Gil
Random sets have been shown to be suitable for modeling real-life situations in which the experimental outcomes are nonempty compact and convex sets of $\mathbb{R}_p$ as, for instance, in the fields of spatial statistics and image analysis. The problem of solving one and k-sample tests for the Aumann expectation of random sets has been previously tackled. On the other hand, the null hypothesis assuming that the population mean of a normally distributed random variable belongs to a given interval has been tested in the literature by considering a multiple hypothesis test. The aim is to extend this previous test to the random sets framework by considering a measure of the inclusion of the set-valued expected value in a given set. The asymptotic distribution of the proposed test statistic based on such a measure is not suitable in practice. For this reason, a bootstrap procedure is developed and leads to convenient empirical results for moderate sample sizes. The study of a real-life situation is finally provided in order to illustrate the use of the test.

E896: **A linearity test for a set-arithmetic regression model between interval-valued variables**
Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain
Co-authors: Gil Gonzalez-Rodriguez, Ana Colubi
The problem of checking the adequacy of a linear regression model for modelling the relationship between two interval-valued variables is investigated. A linearity hypothesis test based on nonparametric regression is proposed. A bootstrap technique to solve the test is applied. If the linear model is valid, analytic expressions for the least squares estimators are obtained. The addition of some conditions to the estimation process in order to keep the coherency of the solutions with the interval arithmetic entails some differences with classical estimators. Under the alternative of non-linearity, a nonparametric model between the interval-valued variables is estimated by means of kernel regression techniques. Thus, the statistic of the test is defined to compare the variability of the residuals obtained from both the linear and the nonparametric regression, taking similar values in case of linearity. Both the practical applicability and the empirical performance of the test have been checked by simulation and some practical examples.

ES34 Room 9 MODEL VALIDATION Chair: M. Dolores Jimenez-Gamero

E140: **Testing departures from elliptical symmetry and specific elliptic distributions**
Presenter: **Apostolos Batsidis**, University of Ioannina, Greece
During the last decades, researchers have focused on the broad class of elliptic/spherical distributions, which includes among others as particular cases, the multivariate normal, t, Pearson type II and VII, symmetric Kotz type distribution. The importance of the elliptical distributions to formulate real phenomena, justifies the necessity of the existence of testing procedures which help to decide: a) whether a data set is coming from an elliptical (or spherical) distribution (hypothesis 1), and b) if the data are distributed according to a specific elliptical (spherical) distribution (hypothesis 2). Two procedures are presented to address the above problems. In particular, a new procedure, which is formulated on the basis of the power divergence test statistic and is in accordance with a characterization of the spherical distributions, is presented for testing hypothesis 1. In the sequel, based on an empirical estimator of Song's measure, we present a procedure for testing hypothesis 2. A Monte Carlo study is carried out to examine: a) the performance on controlling type I error rates of the proposed test statistics, b) the behavior of the test statistics, and c) the power of the tests. Finally, a well-known data set is used to illustrate the methods developed.

E288: **Assessing model adequacy in possibly misspecified quantile regression**
Presenter: **Anouar El Ghouch**, The University catholique de Louvain, Belgium
Possibly misspecified linear quantile regression models are considered and a measure for assessing the combined effect of several covariates on a certain conditional quantile function is proposed. The measure is based on an adaptation to quantile regression of the famous coefficient of determination originally proposed for mean regression, and compares a 'reduced' model to a 'full' model, both of which can be misspecified. We propose an estimator of this measure and investigate its asymptotic distribution both in the non-degenerate and the degenerate case. The finite sample performance of the estimator is studied through a number of simulation experiments. The proposed measure is also applied to a data set on body fat measures.

E290: **Boosted discrete choice models for product demand estimation**
Presenter: **Jianqiang Wang**, Hewlett-Packard Labs, United States
Estimating the aggregated market demand is intrinsically important to manufacturers and retailers. Motivated by the need for a demand model to facilitate pricing optimization at Hewlett-Packard, we have developed a Boosted Discrete Choice Model that takes product brand, features, price and sales volume as the input. In the proposed approach, the utility of a product is specified semiparametrically, either by a varying-coefficient linear model or a partially linear model. We formulate the multinomial likelihood and apply gradient boosting to maximize the likelihood. Several attraction functions like the multinomial logit (MNL), linear and constant elasticity of substitution (CES) attraction functions are compared empirically and the implications of the model estimates on pricing are discussed.

E368: **Graphical procedures and goodness-of-fit tests for the compound Poisson-exponential distribution**
Presenter: **Ioannis Koutrouvelis**, University of Patras, Greece
The compound Poisson-exponential distribution is a basic model in queueing theory, risk analysis and reliability theory. Graphical procedures for assessing this distribution are proposed which utilize the residuals from a regression model involving a standardized form of the cumulant generating function and its empirical counterpart. A method of mixed moments for estimating the unknown parameters is employed first. Plots furnished with a band of simultaneous confidence level are constructed. The critical points of the equivalent goodness-of-fit test are found by fitting equations involving the supremum of the standardized residualas, the sample size and the estimate of the Poisson mean. Simulation results are presented for the power of the test against other skewed distributions including the compound Poisson-gamma with shape parameter different from one and the compound Poisson-inverse Gaussian.

E606: **Model validation for multinomial logistic regression using the Cook's distance**
Presenter: **Nirian Martin**, Carlos III University of Madrid, Spain
The usual application of the Cook's distance is to provide influence measures for a model. We focus on the multinomial logistic regression model and show that it is possible to provide a measure of reliability of a specific prediction using the asymptotic distribution of the Cook's distance.

ES26 Room 7 BIOSTATNET: YOUNG RESEARCHERS FORUM II Chair: Carmen Cadarso-Suarez

E637: **Experimental designs for surgery with simultaneous equations models**
Presenter: **Victor Casero**, Universidad de Castilla-La Mancha, Spain
Co-authors: Jesus Lopez-Fidalgo
A simultaneous equations model is considered to obtain experimental designs for a real case of thoracic surgery in the treatment of lung cancer. The aim is to design an exercise test (riding a static bicycle) to obtain more information to predict complications of surgery. The model involves another variable that is unknown at the beginning of the experiment but it is known after the experiment is performed. Other authors have considered prior information about the unknown variable to obtain optimal experimental designs. We consider a simultaneous equations model to obtain D-optimal designs. A sensitivity analysis is performed to have an idea of the risk in choosing wrong nominal values of the parameters.

E719: **Choosing the primary endpoint in clinical trials using web-based application tools**
Presenter: **Moises Gomez-Mateu**, Universitat Politecnica de Catalunya, Spain
Co-authors: Guadalupe Gomez
The choice of the primary endpoint is a key issue at the design stage of randomised clinical trials. In many research areas it is common to use a composite endpoint consisting of the union of two or more outcomes to prove the efficacy of a given treatment. The Asymptotic Relative Efficiency (ARE) is used to decide between a composite endpoint and one of its components. The main goal is to make this method available to researchers by means of an interactive, flexible and friendly Web platform.

E733: **Randomization tests in clinical trials**
Presenter: **Arkaitz Galbete**, Universidad Publica de Navarra, Spain
Co-authors: Jose Antonio Moler, Fernando Plo
In the framework of clinical trials, legal and ethical restrictions make a population model unrealistic. In this case, randomization tests are a viable alternative to classical inference, since they do not require any condition on the data. When a response- adaptive design, as the Klein design, is applied, the allocation of treatments has a strong correlation structure. A recursive algorithm that allows us to obtain the exact distribution of a permutation test is presented. Besides, conditions that guarantee the asymptotic normality of this statistic are also obtained.

E751: **Markov models used in a 2-stage outcome cohort simulation for an economic evaluation**
Presenter: **Nuria Perez-Alvarez**, Universitat Politecnica de Catalunya, Spain
Co-authors: Guadalupe Gomez, Roger Paredes, Bonaventura Clotet
An economic evaluation model is proposed to assess the performance of 3 strategies for HIV infected patients. The building of the model is divided in several steps: design of the simulation model, summary of the data generated in terms of cost-effectiveness and model validation. The main contribution has been to simplify the system to simulate the cohort without losing truthfulness and to extend the model to reflect a different slope in the clinical outcome evolution.

E644: **Simultaneous inferences: Method of maximum combination applied to proportions**
Presenter: **Maria Alvarez Hernandez**, University of Granada, Spain
Co-authors: Antonio Martin Andres
Recently, the method of maximum combination has been presented as an alternative to the Scheffe procedure for simultaneous inferences. The method (based on the maximization of the test statistic in the coefficients of the combination) yields the same results as Scheffe's and allows in addition: (a) to determine the statistic which should be used for the test of global homogeneity; and (b) to determine the combination of maximum significance. The methodology of maximum combination is applied to the case of proportions. It allows us to get a homogeneity test more powerful than the classical chi-square test of Pearson. At www.ugr.es/local/bioest/Z LINEAR K.EXE a free program can be obtained to apply the method.

ES77 Room 10 CONTRIBUTIONS TO LONGITUDINAL/SPATIAL DATA ANALYSIS Chair: Nial Friel

E598: **Joint modeling longitudinal categorical responses and time-to-dropout**
Presenter: **Sara Viviani**, Sapienza University of Rome, Italy
Co-authors: Marco Alfo
Longitudinal studies often entail categorical outcomes; when dropout occurs and the response is Gaussian, non-ignorability is frequently accounted through joint models (JMs). We formulate an approach for non-Gaussian longitudinal outcomes in the framework of joint models. As an extension of shared parameter model, based on shared latent effects, we assume that the past history of the response up to the current time may influence the risk of dropout. Since the time a subject spends in the study is continuous, we parametrize the dropout process through a proportional hazard model. We give some specifications of this model for several longitudinal distributions (Poisson, Binomial among others). Moreover, some extensions on the random effects distribution, that is usually assumed to be Gaussian, are introduced. To estimate model parameters, we consider an EM-type algorithm. In this context, the maximization of the observed data log-likelihood requires numerical integration over the random effect posterior distribution, which is usually not straightforward. We investigate the behaviour of parameter estimates through a simulation study, and apply the proposed method to a benchmark dataset.

E633: **Missing data in longitudinal binary responses**
Presenter: **Maria Helena Goncalves**, CEAUL and FCT of UAlg, Portugal
Co-authors: Maria Salome Cabral
In the analysis of binary longitudinal data, and longitudinal data in general, a frequent problem is the presence of missing data since it is difficult to have complete records of all individuals. To overcome this problem likelihood inference is considered and serial dependence is modeled using Markov chains as the basic stochastic mechanism. Based on this approach missing values are allowed on the response provided that they are missing at random (MAR) with some restrictions for the presence of missing data when they occur in the middle of the profile. Random effects are also considered, in addition to the above form of serial dependence. A real dataset is analyzed using the R package bild where this methodology is implemented.

E761: **Developing longitudinal models for monitoring chronic diseases in computerised GP records**
Presenter: **Zalihe Yarkiner**, Kingston University, United Kingdom
Co-authors: Rosie O'Neil, Gordon Hunter
An emerging challenge in the field of statistical analysis is the development of methodologies to facilitate the analysis of large, complex longitudinal datasets. The application here is the natural history of Chronic Kidney Disease (CKD), a relatively recently recognised multi-stage, progressive disorder that affects up to 10% of the population and is associated with increased all cause mortality and co-morbidities. CKD is an important condition because interventions which can be delivered in primary care can slow or stop its progression but the current lack of knowledge is recognised to be an issue in the management of CKD. The investigation is based on ten years worth of GP records from 130 practices, and includes

almost 200 variables for more than one million patients. Some early results in identifying, developing and applying a longitudinal modelling framework to investigate the natural history of CKD are described. MIXED models and generalized linear mixed model approaches are used in this analysis. The research and particularly the methodologies developed should be applicable to many other progressive chronic disorders where long term management of a deteriorating condition is required. The methodologies could also be extended to investigate health outcomes in other longitudinal, routinely collected data.

E929: **Inference of gastrointestinal slow wave activity using statistical modeling**
Presenter: **Jiayi Lu**, The Chinese University of Hong Kong, China
Co-authors: Xiaodan Fan
Recently, pharmacologists who are interested in the discovery of drugs to prevent nausea and vomiting perform a study on the gastrointestinal tract. Stomach tissue of rat is taken out and is put on a microelectrode array (MEA), from which slow wave activity can be identified. In order to test whether the slow wave activity has some spatial pattern, the spatiotemporal data is modeled, assuming single electrophysiological source and one-dimensional propagating signals. Parameters characterizing the slow wave activity, such as the velocities, noise frequencies, and signal missing probabilities, are estimated using maximum likelihood method and hill climbing optimization. Performances of the approaches are demonstrated in a simulation study. The method provides a baseline for future studies on slow way activity that has multiple wavefronts propagating simultaneously or has two-dimensional propagating signals.

E938: **A spatial approach to individual health expenditures analysis**
Presenter: **Luca Grassetti**, University of Udine, Italy
Co-authors: Laura Rizzi
Determinants of individuals health expenditure in Region Friuli-Venezia Giulia (Italy) are studied considering individual costs for inpatient hospitalizations, prescribed drugs and outpatient services. Individual data are collected from the regional administrative databases along with contextual covariates: patients' age, sex and proxies of health status; general practitioners' classification and age, municipalities' altimetric area and population density. The reference year is 2010 and observations regard a study population of 1,000,000 patients. Evidences about spatial autocorrelation in health expenditure data have been highlighted in a set of well known researches. Typically, the analysis of spatial autocorrelation is developed considering area-level data. Here two-part random effects spatial stochastic cost function is adopted in order to deal with zero mass issue, cross-classified hierarchical data structure (patients belong to family doctor and municipality of residence) and spatial correlation. Univariate data analyses have been considered to explore health expenditures behaviour. Multivariate model generalization is finally considered in order to draw out the interdependence between different expenditures. Empirical results suggest that spatial correlation is not negligible and unobserved factors affect health expenditures at both family doctor and municipality level. These results can be interesting when health care costs mapping is considered in an evidence-based policy making process.

E944: **Source detection of food-borne disease outbreaks applied to the German E. coli outbreak 2011**
Presenter: **Juliane Manitz**, University of Göttingen, Germany
Co-authors: Martin Schlather, Thomas Kneib, Dirk Brockmann
As exemplified by the 2011 E. coli outbreak in Germany, one of the key challenges in designing efficient mitigation strategies is a fast and efficient source detection during food-borne disease outbreaks. Standard public health procedures typically use case-control studies and tracings along the food shipping chain. These methods are usually very time-consuming and suffer from biased data collected in patient interviews. We introduce a network theoretic method to detect the spatial source of such a food-borne disease outbreak that only requires infection reports regularly collected by the public health institutes and knowledge of the underlying food shipment network. Replacing ordinary geodesic distances, which are commonly used in conventional approaches, with effective distances obtained from the transportation network, our method efficiently identifies the epicentre of outbreak dynamics. We test our method in a spatial dynamic model for food-borne diseases and specifically validate our approach for the German E. coli outbreak in 2011.

ES95 Room 12 BAYESIAN STATISTICS Chair: Thomas Kneib

E668: **Bayesian generalized additive models for location scale and shape in zero-inflated count data**
Presenter: **Nadja Klein**, Georg-August-University Goettingen, Germany
Co-authors: Thomas Kneib, Stefan Lang
Bayesian generalized additive models for location scale and shape (GAMLSS) define a flexible, semi-parametric class of regression models. While ordinary regression analyses only the effects of covariates on the mean of a response, GAMLSS describes more complex parameters of the underlying distribution using semi-parametric additive predictors. Each non-linear effect on the predictors is written as a linear combination of B-spline basis functions. We focus on zero-inflated Poisson (ZIP) and negative binomial (ZINB) models, whereas the classical regression models for count data are extended in the way that excess of zeros can be accounted. In contrast to quantile regression, which avoids distributional assumptions, GAMLSS has the advantage that effects of covariates directly impact the parameters of the underlying distribution, and are therefore easier to interpret. In contrast, complex and numerically demanding likelihood-functions are a consequence. An alternative to likelihood-based estimations are efficient Markov chain Monte Carlo techniques (MCMC). Especially constructing adequate proposal densities, which automatically deliver approximations of the full conditionals, play a crucial role. The Bayesian approach will be compared to existing likelihood-based estimations, whereby the former one has the advantage that estimating the smoothing parameters is part of the procedure. Additionally, spatial and random effects can easily be included. An application to patent data will be presented.

E699: **Geometrically-inspired structured priors for Bayesian partition-based models**
Presenter: **Ludger Evers**, University of Glasgow, United Kingdom
Partition-based regression models, which include both recursive-partitioning methods such as CARTs as well as methods based on tessellations, have been used successfully for more than a decade. However, partition-based models typically lead to discontinuous predictions which increases the variance of the predictions. Bayesian approaches suffer in addition from rather poorly mixing MCMC algorithms. The aim is to show how geometrically inspired structured joint priors on the models inside each region can overcome both problems. The core idea behind the construction of the prior is to assume a linear, rather than constant, model inside each region and to penalise large differences between neighbouring regions in the partition implied by the method. This choice of prior reflects the geometry of the partition rather than the process used to generate it. This is the core difference to related methods suggested in the literature. The priors yield both "smooth-ish" predictive surfaces and better mixing MCMC chains and thus yield a powerful flexible regression method.

E731: **Tail area higher-order approximations in Bayesian inference**
Presenter: **Laura Ventura**, University of Padova, Italy
Co-authors: Erlis Ruli, Walter Racugno
Asymptotic tecniques are widely used in Bayesian inference, and in recent years there has been considerable development of so-called higher-order asymptotics. This theory provides very accurate approximations to marginal posterior distributions and to the corresponding tail area probabilities,

in a variety of parametric statistical problems even for small sample sizes. The aim is to discuss tail area higher-order approximations for practical use in Bayesian analysis, with a view to highlight the agreements between Bayesian and non-Bayesian approaches and their implementations and applications. From a theoretical point of view, we outline how the methodology proposed requires little more than standard likelihood (or pseudo-likelihood) quantities for its implementation, and hence it may be available at little additional computational cost over simple first-order approximations. From a practical point of view, we outline how modern asymptotic theory may routinely be applied in practice to compute posterior summaries and highest posterior density credible sets, and to perform model selection. Numerical illustrations are discussed and several interesting results emerge.

E796: **Bayesian analysis of misclassified polychotomous response data**
Presenter: **Lizbeth Naranjo**, University of Extremadura, Spain
Co-authors: Carlos J. Perez, Jacinto Martin

Two Bayesian approaches are proposed to model polychotomous response data that are subject to misclassification. Misclassification in the categorical response is considered through the development of Bayesian regression models for both ordered and unordered categories in the response variable. The computational difficulties have been avoided by using data augmentation frameworks. The idea of using data augmentation is exploited to derive efficient Markov chain Monte Carlo methods. A simulation-based example illustrates the model performance when comparing with standard methods that do not consider misclassification.

E959: **Bayesian evaluation of asymmetry in multidimensional scaling**
Presenter: **Kensuke Okada**, Senshu University, Japan

One of the major benefits of adopting an inferential/probabilistic approach in multidimensional scaling (MDS) is that it allows statistical model evaluation of asymmetry in data, which is difficult in classical approaches. The ratio of two marginal likelihoods is called the Bayes factor, and it represents how likely the data are under each one of the different models. Comparison of asymmetric and symmetric MDS models in terms of Bayes factors will provide important insights with implications for modeling and prediction. However, although it is conceptually appealing, computation of Bayes factor is often difficult because it involves integration over the entire parameter space. In order to deal with this problem, a version of the product space sampling approach is proposed. In this method, a mixture model is employed in which both asymmetric and symmetric model parameters are combined, and is used to generate samples from the joint posterior distribution of both model parameters and model indices. By repeating the bisection-based optimization of model priors, it becomes possible to efficiently estimate the required Bayes factor. Numerical examples with both artificial and real data are presented to illustrate the benefits provided by the proposed method.

E832: **Bayesian analysis of nonlinear, non-Gaussian state space models: The joint independent Metropolis-Hastings method**
Presenter: **Istvan Barra**, VU University Amsterdam, Netherlands
Co-authors: Lennart Hoogerheide, Siem-Jan Koopman, Andre Lucas

A new methodology is proposed for the Bayesian analysis of nonlinear non-Gaussian state space models where the signal is univariate and Gaussian. The joint independent Metropolis-Hastings procedure is introduced to sample from the joint posterior density of the parameters and states. The novelty in our approach lies in the development of a proposal density for the joint posterior distribution of parameters and states. This proposal density consists of two ingredients: a mixture of $t-$Student densities targeting the marginal posterior density of the parameters and a Gaussian density targeting the density of the states given observations and parameters. Examples of stochastic volatility and multivariate point processes show that our method is a highly efficient alternative to particle Markov chain Monte Carlo methods, and provides at least three advantages. First, our methodology is fully automated and there is no need for case by case fine tuning of the algorithm, whenever we change the observation density. Second, our methodology is easily parallelized. Hence we can utilize cutting edge graphics cards to speed up the calculations. Third, we show that our approach allows us to calculate the marginal likelihood using importance sampling with only a few additional computations.

EP05 Room Hall POSTER SESSION V Chair: Patricia Roman-Roman

E141: **Bayesian influence in a consensus value for radiocarbon dating**
Presenter: **Fatima Palacios-Rodriguez**, Universidad Pablo de Olavide, Spain
Co-authors: Jose-Maria Fernandez-Ponce, Rosario Rodriguez-Grinolo

Radiocarbon dating is vital in the establishment of time lines for many archaeological studies. The calibration curves necessary to map radiocarbon to calendar ages were originally estimated using only measurements on known age tree-rings. The types of records available for calibration have since diversified and a large group of scientists (known as the IntCal Working Group) from a wide range of backgrounds has come together to create internationally-agreed estimates of the calibration curves. An influence analysis on a finite population from a Bayesian perspective is developed to contribute further information to previous studies in radiocarbon dating. An expression for the Bayesian Conditional Bias of a statistic T that does not depend on the parameter is proved and a sufficient and necessary condition to obtain the Bayes estimator of the conditional bias of a statistic T is given. Some results of influence analysis from a Bayesian viewpoint are also shown.

E234: **Examining growth with statistical shape analysis and comparison of growth curve models**
Presenter: **Deniz Sigirli**, Uludag University, Turkey
Co-authors: Ilker Ercan

Growth curves have been widely used in the fields of biology, zoology and medicine, by taking a measurable trait of an organism, such as weight, height, volume or area, as a size measure. We compared the performances of the commonly used non-linear growth curves by taking centroid size as a size measure, which is a commonly used size measure in geometrical morphometry because of its feature of being independent from the shape. The performance of three parameter and four parameter logistic curves, the Gompertz curve and the Richards curve were compared, and their efficiencies in different sample sizes were examined using a simulation study. Furthermore, a practical example is given for examining the relationship between the centroid size of the cerebellum and the duration of multiple sclerosis (MS) disease in MS patients with the three- and four parameter logistic, Gompertz and Richards models. Mean square error decreased as the sample size increased, in all growth models. Richards model was not convenient for small samples in terms of both model performance and parameter estimates. The three parameter logistic and Gompertz models did not display explicit differences in parameter estimates by sample size. The three parameter logistic and Gompertz models were more preferable to the other two models, especially for small samples, for the proper situations.

E345: **Examining the type I error rate for the Hotelling T2 test for shape analysis**
Presenter: **Guven Ozkaya**, Uludag University, Turkey
Co-authors: Ilker Ercan, Deniz Sigirli

In many biological and biomedical investigations, the most effective way to analyze the forms of whole biological organs or organisms is by recording geometric locations of landmark points. If we want to compare shapes in geometric morphometry by shape analysis, individuals should be translated, rotated and scaled in such a way that all the individuals lay in the standard position and center. Bookstein carried out this by choosing two landmarks as reference landmarks. Each individual is translated, rotated and scaled according to these reference landmarks. It is aimed to examine if there are differences in the results of Hotelling T2 test, in case of choosing different reference landmarks. While variance level is equal

0.01 and 0.05, type I error rate is found between 0.049-0.051 for 3, 4, 5, 6 landmarks and 25, 50, 100, 250, 500 and 1000 sample sizes. While variance level is equal 737, type I error rate is found between 0.031-0.049. On the other hand, for 2949 variance level, type I error rate is found between 0.031-0.043. Simulations showed that, in the 0.01 and 0.05 variance level, the type I error rate of Hotelling T2 test which was performed for Bookstein coordinates obtained by taking different landmarks as reference, was quite consistent in the level of 0.05.

E351: **Type I error rate according to number of landmark for two-sample tests in statistical shape analysis**
Presenter: **Gokhan Ocakoglu**, Uludag University, Faculty of Medicine, Department of Biostatistics, Turkey
Co-authors: Ilker Ercan

In many biological and biomedical investigations, the most effective way to analyze the forms of whole biological organs or organisms is by recording geometric locations of landmark points. Therefore it is of great importance to compare shapes of the organs or organisms. Statistical shape analysis is a geometrical analysis of the statistics measured from sets of shapes that determines the features of similar shapes or of different groups comprising similar shapes. Newly developed methods utilize two-sample tests in statistical shape analysis which is a geometric morphometric concept. The Hotelling T2, Goodall F and James Fj tests as well as the lambda test statistic are used to compare the mean shapes of two samples from the statistical shape analysis literature according to type I error rates derived from various variance values in different sample sizes. Additionally, related tests performances are compared in terms of number of landmarks. The results indicate that tests perform better with large samples than with small samples. It is observed that tests performances increase with the decreasing number of landmarks.

E525: **Kernel principal components for economic time series poredictions**
Presenter: **Alessandro Giovannelli**, University of Tor Vergata - School of Economics, Italy

The main objective is to propose two nonlinear extensions for macroeconomic forecasting using large datasets. First we propose an alternative technique for factor estimation, i.e. kernel principal component, which allows the factors to be nonlinear related to the input variables. Second, we propose artificial neural networks as alternative to the factor augmented linear dynamic equation. These two extensions allow us to determine whether, in general, there is empirical evidence in favour of nonlinear methods and, in particular, to comprehend if the nonlinearity occurs in the estimation of the factors or in the functional form that links the target variable to the factors. To verify the empirical performances of the methods proposed, we conducted several pseudo forecasting exercises on industrial production index and consumer price index for Euro Area and US economies. These methods were used to construct the forecasts at 1-, 3-, 6-, and 12-month horizons using a large dataset containing 253 predictors for Euro Area and 131 predictors for US economy. The results obtained from the empirical study indicate that the construction of nonlinear factors significantly improves the quality of forecasts compared to the linear principal components. Moreover the results for artificial neural networks techniques have the same forecasting ability compared to the factor augmented linear dynamic equation.

E625: **Determining cut-off values for patients diagnosed with anxiety using receiver operation characteristics analysis**
Presenter: **Ilker Ercan**, Uludag University Faculty of Medicine, Turkey
Co-authors: Sengul Hafizoglu, Guven Ozkaya, Cengiz Akkaya, Elif Yalcintas, Selcuk Kirli

Receiver Operation Characteristics (ROC) analysis has widely been used in medical data analysis in measuring discriminatory ability of diagnostic or prognostic tests. This makes the ROC analysis one of the most active research areas in medical statistics. Anxiety is a physiological, behavioral, and psychological reaction. The State–Trait Anxiety Inventory (STAI) is one of the most widely used self-report measures of anxiety in clinical and research settings. Aim of this study is to determine cut off values for STAI scale. Seventynine patients diagnosed with anxiety and 56 subjects without any history of psychiatric diagnosis were included. There is no significant difference between groups according to age and gender ($p = 0.438$, $p = 0.728$). ROC analysis was applied to all subjects total score of STAI. Physician assessment for subjects were taken as a gold standard. Cut off value for STAI Form-I was found as 44 ($AUC = 0.737$, $p < 0.001$) and 45 for STAI Form-II ($AUC = 0.789$, $p < 0.001$). Subjects who had higher value than cut off values were classified as having anxiety and other were classified not. Determined cut off values is recommended using as a reference in screening studies.

E627: **Comparison of estimating approaches in the Poisson generalized linear mixed model: A simulation study**
Presenter: **Marti Casals**, Universitat de Barcelona, Spain
Co-authors: Klaus Langohr, Josep Lluis Carrasco, Lars Ronnegard

Generalized linear mixed models (GLMMs) are a flexible approach to fit non-normal data. GLMMs are known to be useful for accommodating overdispersion in Poisson regression models and modelling the cluster dependence structure in longitudinal or repeated measures designs. The main difficulty of GLMM is the parameter estimation because it is often not viable to obtain an analytic solution that allows the maximization of the marginal likelihood of data. Attention has been paid to this issue leading to several proposals that are able to estimate GLMM parameters. Hence, it is possible to find different approaches to fit GLMM implemented in the main statistical software packages. Among them we focus on the Gauss-Hermite quadrature (GHQ) estimation, hierarchical (h-likelihood) or Bayesian methods (Integrated nested Laplace approximation -INLA-) implemented in the R packages lme4, hglm and INLA, respectively. The purpose is to compare the performance of these GLMM estimation methods via a simulation study with different scenarios of overdispersion.

E137: **Identification of patients at high risk of metastatic breast cancer: A data mining approach**
Presenter: **Sayed Mohsen Hosseini**, Isfahan University of Medical Sciences, Iran
Co-authors: Razieh Hassannejad

Breast cancer is the most prevalent cancer among women and the second leading cause of cancer death. In this study, Data mining was carried out by utilizing association rules (AR) method and process often utilizes machine learning and statistical methods to analyze data. Data collection was conducted at Seido Shohada Hospital of Isfahan. Extracted information includes patient tumor and treatment factors. With combination of association rules mining and cumulative logit regression a new approach is introduced for analysis of categorical data. The Apriori algorithm was applied to the data set to extract association rules. The IMS Apriori algorithm created 79 rules of higher confidence using support difference (SD) approach as a minimum item support to prune rules and extracted interesting rules. The cumulative logit regression estimated parameters that represent each variable effect on the outcome. For instant, Rule1 represents 62.7% of patients with CA15-3 more than or equal to 30 U/ml had metastatic cancer. We can classify patients who are more likely to suffer metastatic breast cancer based on their characteristics. The most important results gained were to demonstrate data mining and association rules algorithms are useful to identify a group of patients at higher risk of metastatic breast cancer.

EP04 Room Hall POSTER SESSION IV **Chair: Patricia Roman-Roman**

E905: **A goodness-of-fit test for a family of leptokurtic symmetric distributions**
Presenter: **Virtudes Alba-Fernandez**, University of Jaen, Spain
Co-authors: M. Dolores Jimenez-Gamero, Inmaculada Barranco-Chamorro, J. Luis Moreno-Rebollo
A family of leptokurtic symmetric distributions has been introduced by means of the difference of two independent gamma variates. The classical Laplace and normal distributions are special cases of this family. While there is no general closed-form expression for the density function, the characteristic function has a closed, quite tractable expression. A Goodness-of-Fit (GoF) test for this family is proposed. The test statistic takes advantage of the convenient formula of the characteristic function of this distribution and compares it with the empirical characteristic function of the sample. Large sample properties of the proposed test are studied. In contrast to other previously proposed GoF tests, the one in this work is consistent against all alternatives. Some numerical results are reported which study the finite sample performance of the proposed test.

E945: **An approximation of the partial likelihood information in a follow-up study**
Presenter: **Maria Jesus Rivas-Lopez**, Universidad de Salamanca, Spain
Co-authors: Jesus Lopez-Fidalgo
In a follow-up study the time-to-event may be censored either because of dropouts or the end of study is earlier. When the model is to be fitted to n observed times, these are known, and for each of them, it is also known whether that time is censored or not. When the experiment is to be designed, neither the observed times nor the information about whether a particular unit will be censored are known. Thus, the design problem faces uncertainty about occurrence of unknown censored times. This situation is frequently modelled through a Cox-proportional hazards model and the Cox partial likelihood instead of the full likelihood is usually considered for these models. A partial information matrix should be built in this case. An application to a simple case with two possible treatments shows the direct way to approximate partial likelihood information depends on the unknown failure times. All this process is rather complex and we provide a more convenient approximation.

E931: **Interaction analysis of the HHP multi-dimensional model**
Presenter: **Yulia Malitskaia**, Stern New York University, United States
Co-authors: William Greene, Yuri Malitsky
Advancements in technology have made gathering large arrays of data faster and more efficient. Mining these data sets for useful information is a quickly growing field of research that has had impacts on many areas of application. This trend, for example, is represented in the Heritage Health Prize (HHP) competition, which aims to support the development of an efficient algorithm for predicting the number of days a patient will spend in the hospital given their medical history. An immediate and actual product of this competition will be the construction of a comprehensive multi-dimensional data model that captures the complex interaction effects among various attributes and features of patient histories, such as different types of diseases, procedures, places of services, and others. However, in the framework of the initial linear regression applications, the treatment of interaction effects represents a serious challenge associated with two contradictory issues: the principle of marginality and the implications of ultra-high scale. In our paper we apply a two-stage procedure based on the pairwise selection of the most influential interactions and subsequent joint analysis using the LIMDEP econometrics package. Our work provides a comprehensive overview of this approach and comparison of alternative techniques.

E946: **The determinants of passengers demand for the leisure air travel market: The case of Portugal**
Presenter: **Jose Manuel Vicente**, University of Evora, CEFAGE-UE, Portugal
Co-authors: Andreia Dionisio, Manuela Oliveira
The purpose is to analyse the determinants of the quarterly demand of passengers in Portuguese domestic leisure air travel market (between Lisbon, Oporto and Madeira, Azores Islands), from 2006 to 2010. The data refers to the routes under consideration as panel data, and the model was estimated by pooled OLS, fixed and random effects. The results were promising and provide important information, points to the following determinants: airfare, capacity, frequency and the 3rd quarter.

E953: **Change points analysis in regression models**
Presenter: **Nora M. Villanueva**, University of Vigo, Spain
Co-authors: Marta Sestelo, Javier Roca-Pardinas
Regression analyses are usually applied in order to obtain a fit of a regression curve. Sometimes a fitted curve might contain an isolated discontinuity or change point. In many cases interest focuses on the occurrence of such change points, however, identifying these points in a statistical framework can be a challenging task. We introduce a method based on regression models to detect changes both the location and the number of them. In order to asses this number, we have proposed a bootstrap-based test. Binning techniques were applied to speed up computation in the estimation and testing process.

E957: **Accelerated failure time model with left-truncated and right-censored reliability data**
Presenter: **Antonio Jesus Lopez-Montoya**, University of Granada, Spain
Co-authors: Irene Garcia-Garrido
The linear regression models are popularly considered to deal with filtered data, such as right-censored and left-truncated data in the context of reliability analysis. The motivation is to evaluate the risk of failure in water supply networks. To do it, a dataset consisting of breakdown data of the pipes in a network laid in a small city of the Mediterranean Sea is analyzed. Due to the nature of the dataset, it is thought appropriate to use a Cox model fit. However, this model is not appropriate since in this case the proportional hazards assumption does not hold. In consequence, an accelerated failure time model is proposed which allows estimating and make inferences about the parameters of the model, without assuming a particular distribution for the failure time random variable. Furthermore, the results are easily implemented and interpreted. An extended version of the Nelson estimator, modified in order to adapt the problem of right-censored alongside left-truncation has been used in estimation procedure of the model coefficients.

E967: **Optimal acceptance sampling plans for log-normal distributions using Bayesian risks**
Presenter: **Carlos J. Perez-Gonzalez**, Universidad de La Laguna, Spain
Co-authors: Arturo J. Fernandez
The design of reliability sampling plans is important to assess the acceptability of a lot of product, or its production process. Acceptance tests provide decision rules to either accept or reject the product based on the failure times of a random sample of items drawn from the lot or process. It is assumed that the probability of obtaining a defective unit, p, is not constant. A procedure for determining optimal failure-censored sampling plans is presented for log-normal lifetime distribution and a generalized beta model of p. These plans are designed using Bayesian producer and consumer risks. In many cases, the use of prior information can significantly reduce the amount of testing required. However, the main advantage of using a prior model for the fraction defective is not the reduced sample size but better assessment of the true sampling risks. Several examples are considered to illustrate the results developed for selected censoring levels and specifications according to the available prior information on p.

E975: **Comparison of the predictive values of binary tests in the presence of verification bias**

Presenter: **Ana Eugenia Marin**, University of Granada, Spain

Co-authors: Jose Antonio Roldan, Juan de Dios Luna

A diagnostic test is a medical procedure applied to detect whether a patient is free from a given disease or not. Some traditional parameters for evaluating the performance of a binary diagnostic test are the positive and negative predictive values. The positive predictive value is the probability that a patient with a positive result in the diagnostic procedure does in fact have the disease, and the negative predicative value is the probability that a patient undergoing the diagnostic procedure does not have the disease. The comparison of the predictive values of two or more binary diagnostic tests is a topic of particular relevance in the study of statistical methods applied to medical diagnosis, a global hypothesis test is studied in order to conduct a simultaneous comparison of the VPs of more than two binary diagnostic tests in the presence of missing data ignorable. The global hypothesis test deduced is based on the chi-squared distribution. Simulation experiments were carried out to study the type I error and the power of each global hypothesis test and of other alternative methods when comparing the predictive values of two and three binary diagnostic tests, respectively.

Monday 03.12.2012 08:30 - 10:10 Parallel Session M – CFE

CSI04 Room Auditorium ADVANCES IN ECONOMETRICS Chair: Jean-Michel Zakoian

C336: **Risk-parameter estimation in volatility models**
Presenter: **Christian Francq**, CREST and University Lille III, France
Co-authors: Jean-Michel Zakoian

The concept of risk parameter is introduced, for conditional volatility models of the form $\varepsilon_t = \sigma_t(\theta_0)\eta_t$, and statistical procedures are developed to estimate this parameter. For a given risk measure r, the risk parameter is expressed as a function of the volatility coefficients θ_0 and the risk, $r(\eta_t)$, of the innovation process. A two-step method is proposed to successively estimate these quantities. An alternative one-step approach, relying on a reparameterization of the model and the use of a non-Gaussian QML, is proposed. Asymptotic results are established for smooth risk measures as well as for the Value-at-Risk (VaR). Asymptotic comparisons of the two approaches for VaR estimation suggest a superiority of the one-step method when the innovations are heavy-tailed. For standard GARCH models, the comparison only depends on characteristics of the innovations distribution, not on the volatility parameters. Monte-Carlo experiments and an empirical study illustrate these findings.

C641: **Evaluation of multivariate count models. Trading activity in U.S. large banks**
Presenter: **Gloria Gonzalez-Rivera**, University of California Riverside, United States
Co-authors: Yingying Sun

The in-sample and out-of-sample evaluation of multivariate count models is rather thin. We contribute by generalizing an autocontour-based tests previously proposed so that they are applicable to univariate or multivariate systems with either continuous or discrete random processes. Our interest lies on the multivariate process of a vector of counts for which we specify the dynamics of the marginal distributions and a copula that ties up the marginals to produce their multivariate distribution. We work with a continued version of the probability integral transforms of the system that, under correct specification of the conditional model, should be i.i.d. U[0,1]. We construct hyper-cubes of different sizes within the maximum hyper-cube formed by a multidimensional uniform density $[0,1]^n$, and we assess the location of the empirical PITs within the corresponding population hyper-cubes. If the conditional model is correct, the volumes of the population hyper-cubes must be the same as those in their empirical counterparts. We construct a trivariate model of trade counts for three large U.S. banks. The contemporaneous dependence among institutions is asymmetric so that when liquidity drains in one institution, we should expect a concurrent effect among similar institutions, but when liquidity is plenty, trading of the institutions are not correlated.

C658: **A Bayesian nonparametric analysis of the relationship between returns and realized variance.**
Presenter: **John Maheu**, McMaster University, Canada
Co-authors: Mark Jensen, Thomas McCurdy

The relationship between risk and return is one of the most studied topics in finance. The majority of the literature is based on fixed parameter models. The relationship between returns and realized variances is nonparametrically modelled. An infinite mixture of bivariate normals is given a flexible Dirichlet process prior. From this, the nonparametric conditional distribution of returns on realized variance consists of an infinite mixture representation whose probabilities and arguments depend on the value of realized variance. This allows for a smooth nonlinear relationship between the conditional mean of returns and realized variance. The model is estimated with MCMC techniques based on slice sampling methods. Exact finite sample measures of the uncertainty of the relationship are available. We discuss methods to allow the nonlinear relationship to change over time and the role of volatility shocks.

CS04 Room 6 QUANTITATIVE EVALUATION OF FISCAL POLICIES Chair: Aurelien Eyquem

C570: **On tax policies in open economies**
Presenter: **Aurelien Eyquem**, ENS Lyon, France
Co-authors: Stephane Auray, Paul Gomme

Tax-based deficit reduction experiments for the U.S. and EMU-12 are conducted using an open economy model. In welfare terms, raising the consumption tax is the least costly, followed by the labor income tax, then the capital income tax. Use of an open economy model means that the incidence of the consumption tax is borne in part by foreign producers. Among revenue-neutral tax experiments, partially replacing the capital income tax is welfare-enhancing, although there are short term losses. Replacing labor income tax revenue with a consumption tax improves international competitiveness and is welfare-improving.

C646: **Housework and fiscal expansions**
Presenter: **Stefano Gnocchi**, UAB, Spain
Co-authors: Daniela Hauser, Evi Pappa

The data suggest that government spending crowds-in private consumption. Many authors tried to modify the standard neoclassical model to account for this fact by introducing nominal and financial frictions, by changing consumer preferences, or by assuming different rules for fiscal policy. We propose a New Keynesian model that includes an explicit household production sector to address this issue. The fiscal expansion induces a wealth effect that increases market hours, households substitute nonmarket for markets goods and market consumption increases in equilibrium. We estimate the model using Bayesian techniques. The data support our mechanism for the transmission of fiscal shocks.

C655: **Optimal fiscal and monetary rules in normal and abnormal times**
Presenter: **Paul Levine**, Surrey, United Kingdom
Co-authors: Cristiano Cantore, Giovanni Melina, Joseph Pearlman

Fiscal-monetary interactions in a NK DSGE model with deep habit, distortionary taxes and a sovereign risk premium for government debt are examined. Deep habit crucially affects the fiscal transmission mechanism in that it leads to a counter-cyclical mark-up, boosting the size of a demand-driven output expansion with important consequences for monetary and fiscal policy. We estimate the model by Bayesian methods and use it to compute optimal monetary and fiscal policy first in 'normal times' with debt at its steady state and then in a crisis period with a much higher initial debt-GDP ratio. Policy is conducted in terms of optimal commitment, time consistent and simple Taylor-type rules. Welfare calculations and impulse responses indicate that the ability of the simple rules to closely mimic the Ramsey optimal policy, observed in the literature with optimal monetary alone, is still a feature of optimal policy with fiscal instruments, but only with 'passive' fiscal policy. For crisis management we find there some support for slow consolidation with a more active role for tax increases rather than a decrease in government spending.

C551: **Smoothing shocks and balancing budgets in a currency area**
Presenter: **James Costain**, Bank of Spain, Spain
Co-authors: Beatriz de Blas

Simple fiscal rules for stabilizing the government debt in response to asymmetric demand shocks in a country that belongs to a currency union

are studied. We compare debt stabilization through tax rate adjustments with debt stabilization through expenditure changes. While rapid and flexible adjustment of public expenditure might seem institutionally or informationally infeasible, we discuss one concrete way in which this might be implemented: setting salaries of public employees, and social transfers, in an alternative unit of account, and delegating the valuation of this numeraire to an independent fiscal authority. Using a sticky-price DSGE matching model of a small open economy in a currency union, we compare the business cycle implications of several different fiscal rules that all achieve the same reduction in the standard deviation of the public debt. In our simulations, compared with rules that adjust tax rates, a rule that stabilizes the budget by adjusting public salaries and transfers reduces fluctuations in consumption and employment, thus bringing the market economy closer to the social planner's solution.

CS05 Room 3 BAYESIAN ECONOMETRICS WITH APPLICATIONS Chair: Nalan Basturk

C237: **Efficient Bayesian inference for multivariate factor stochastic volatility (SV) models**
Presenter: **Gregor Kastner**, WU Vienna University of Economics and Business, Austria
Co-authors: Sylvia Fruehwirth-Schnatter, Hedibert Lopes
Multivariate factor SV models are increasingly used for the analysis of multivariate financial and economic time series because they can capture the volatility dynamics by a small number of latent factors. The main advantage of such a model is its parsimony, where all variances and covariances of a time series vector are governed by a low-dimensional common factor with the components following independent SV models. For high dimensional problems of this kind, Bayesian MCMC estimation is a very efficient estimation method; however, it is associated with a considerable computational burden when the number of assets is moderate to large. To overcome this, we avoid the usual forward-filtering backward-sampling (FFBS) algorithm by sampling "all without a loop" (AWOL), consider various reparameterizations such as (partial) non-centering, and apply an ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation at an univariate level, which can be applied directly to heteroscedasticity estimation for latent variables such as factors. To show the effectiveness of our approach, we apply the model to a vector of daily exchange rate data.

C265: **Bayesian estimation of a DSGE model: A Monte Carlo experiment**
Presenter: **Alessia Paccagnini**, Bicocca University, Italy
The aim is to discuss the comparison of different Bayesian approaches to evaluate a New Keynesian macroeconomic model: the Dynamic Stochastic General Equilibrium Model (DSGE). From an econometric point of view, the performance of a DSGE model is often tested against an estimated VAR. This procedure requires a Data Generating Process (DGP) that is consistent with the theoretical economic model and it has a finite-order VAR representation. However, the statistical representation of a DSGE model is an exact VAR only when all the endogenous variables are observable; otherwise, a more complex Vector Autoregressive Moving Average model (VARMA) is needed. Hence, the main point in estimating a DSGE model is to consider the statistical representation. Monte Carlo experiments with different Data Generating Processes are conducted to verify the ability of different econometrics models (such as VAR, Bayesian VAR, DSGE-VAR) to test the hypothesis that a specific DSGE model is consistent with the data. The empirical strategy is completed with an exercise on the US economy. In our analysis, we focus on the marginal data density and the information criteria to determine lag-length in the implemented models.

C552: **Inflation regimes, technological change and weak identification in an NKPC model with forward looking price behavior**
Presenter: **Cem Cakmakli**, University of Amsterdam, Netherlands
Co-authors: Nalan Basturk, Herman Van Dijk, Pinar Ceyhan
Removing low frequency movements in inflation and marginal cost series, a possible weak identification of structural parameters in the New Keynesian Phillips curve model has been indicated in the literature. Using simulation based Bayesian inference we study the effect of considering the original series in a model where a structural time series approach for inflation and costs is combined with the Gali-Gertler model on inflation expectations. Our results indicate that the weak identification issue is reduced in the more general model. Inflation forecasts are obtained using our Bayesian approach on the generalized model.

C509: **Bayesian testing for multimodality using mixture distributions**
Presenter: **Nalan Basturk**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide, Peter De Knijf, Herman Van Dijk
In several applications the data comes from a non-standard, multimodal distribution. In such cases, standard exploratory data analysis can be misleading since possible multimodality is not taken into account. A Bayesian test for detecting the number of distinct modes in the data will be outlined. A mixture of shifted Poisson distributions is proposed to estimate the probability of multimodality. The method is applied to two datasets. First, we apply the proposed method on DNA tandem repeats' data from three groups of individuals with Asian, African and European origin. Second, we analyze possible multimodality in the economic growth performances of countries, measured by Gross Domestic Product per capita. The results are compared with those of standard frequentist tests.

CS20 Room 1 RISK MANAGEMENT IN ENERGY MARKETS: MARKET DESIGN AND TRADING STRATEGIES Chair: M. Dolores Furio

C398: **Oil price shocks and stock markets: A non-linear approach**
Presenter: **Rebeca Jimenez-Rodriguez**, University of Salamanca, Spain
There is no consensus among researchers regarding the effects of oil price shocks on stock returns. One possible explanation for this lack of consensus is that these studies have obviated that the link between oil prices and stock market prices may be non-linear. This paper tests for non-linearity in the relationship between oil prices and stock returns, and assesses the impact of oil price shocks on the stock markets of the main OECD countries.

C528: **Assessment of volatility in oil markets: An entropy analysis**
Presenter: **Silvia Serrano Calle**, UNED, Spain
The purpose is to investigate the uncertainty and volatility in oil markets. Using the concept of entropy as defined by Information Theory, trends in oil markets along past decades are explored. The time series analyzed are daily price differentials for spot and futures markets. Those time series are quantified in order to obtain sequences of discrete values. The discretization of the original time series is done to apply the model and concepts of entropy instead of differential entropy. The results show the utility and potentiality of the Information Theory approach to assess risk in energy markets.

C576: **Modelling price and volatility in the Iberian day-ahead electricity market**
Presenter: **Ainhoa Zarraga**, University of the Basque Country, Spain
Co-authors: Aitor Ciarreta
The interrelationships and volatility spillovers of prices in the integrated Iberian electricity market are modelled. It works as a single market whenever there are no congestions in the net, otherwise a market splitting mechanism is implemented and a different system marginal price is set in each country. The market began functioning in July 2007 and since then there has been a gradual increase in market interconnections. Thus,

using three different specifications of the dynamic conditional correlation (DCC) models we study up to what extent this process has favoured price convergence. Data consist of daily prices calculated as the arithmetic mean of the hourly prices and span from July 1st 2007 until February 29th 2012, thus overall we have 1705 daily observations. The DCC model in which the variances of the univariate processes are specified with a VARMA(1,1) excluding the lagged variance of the other series fits the data the best. Mean and variance spillovers are mostly significant indicating the strong interdependence between both markets.

C704: **Wind power forecast: Statistical analysis and economic benefits**
Presenter: **Pilar Munoz**, Universitat Politecnica de Catalunya, Spain
Co-authors: Josep Anton Sanchez, Maria Dolores Marquez, Sergio Baena, Patricia Tencaliec

The forecast of wind energy with accuracy is an issue that is not yet fully resolved because of the randomness of wind, which is the main input for models that forecast wind power. Therefore, the first thing that we must obtain is precise forecasts of wind speed. To do this, we have two sets of data, those obtained from the HIRLAM weather model and those which come from the wind farm. A model of data fusion is needed to merge these two sources of information, due to the fact that, in general, spatial locations as well as the temporal frequency are different. Wind power forecasting, in the short and medium term, is performed by predicting wind speed by means of nonparametric models as well as combining forecasts obtained from parametric and non-parametric models. Finally, the estimation of the potential economic benefits of using wind power is based on the electricity prices in the day-ahead market, which are obtained from OMEL, the Spanish electricity market operator. Forecast errors will be useful for generating different scenarios of wind power generation. Through simulations, it will be possible to calculate the benefits/loss of wind generation.

CS23 Room 4 COUNTERPARTY CREDIT RISK, A TRANSVERSAL FINANCIAL REGULATION Chair: Bertrand Hassani

C348: **Sovereign rating adjustment using market information**
Presenter: **Xin Zhao**, University Pantheon-Sorbonne, France
Co-authors: Dominique Guegan, Bertrand Hassani

An understandable and standardized rating system is highly demanded with the rapidly expanding international bond market. The fact puts the credit rating agencies to a more and more important position in financial markets. But the problems of their ratings have been unveiled by the East Asian Crisis and the European Sovereign Debt Crisis. To narrow the gap between the high market demands and the information insufficiency, we use market information to relieve the inconsistency between the agencies' ratings and the market practitioner's expectations. We propose a method to adjust the sovereign ratings issued by credit rating agencies using macroeconomic and market information. We adjust the agencies' ratings in terms of timeless and give the warning signals of default or downgrades from the market risk and liquidity risk besides the credit risk.

C461: **Analyzing whether the new counterparty risk regulation increase contagion**
Presenter: **Marius-Cristian Frunza**, Paris Sorbonne University/ Schwarzthal Kapital, France

The CVA-VAR is a metric of the Basel prudential standard meant to assess the potential credit cost for a trade and implies major changes in risk governance across major investment houses. CVA trading desks are mandated to optimize the exposure and the capital related to counterparty risk. Using few competitive models we design a framework for the computation of the CVA-VAR and we emphasize the heterogeneity of the metric depending on the nature of the model. Markets exhibiting tails and asymmetry, modelled with Generalized Hyperbolic distributions, consume more capital than traditional Gaussian models. Assuming a shock in the credit worthiness of a counterparty it appears clearly that the mandate of CVA traders to hedge the counterparty risk can transfer the shock to other underlings. The effect became significant if the number of counterparties exposed to the initial shock is important. We simulate a market with few banks hedging the CVA exposure and show the occurrence of contagion across various underlings as a result of an initial shock. We emphasize via Granger causality test the transmission of shocks for prices and volatilities and with structural break tests the drift in the correlation structure.

C659: **Asset pricing: Analyzing whether the counterparty credit risk is already taken into account**
Presenter: **Pierre-Andre Maugis**, Universite Paris 1, France
Co-authors: Bertrand Hassani

The aim is to determine whether (i) the counterparty credit risk is already taken into account in the General Model of Asset pricing, (ii) the credit value adjusted is a recognition of the risk free valuation failure, (iii) the counterparty credit risk is only a regulatory capital charge. Considering any asset, its value is characterised by the discounted sum of future cash flows. The discount rate should be the sum of the risk free rate, and a risk premium. This so-called risk premium should principally represent the counterparty credit risk. Consequently, on simple shares, we are able to evaluate the differences between the marked to market value and the price obtained from the risk free general model to obtain the market value of the target company credit risk. Consequently, we are able to reevaluate any derivative product, using a discount factor embbedding this risk premium (Binomial tree), and then compare it to the classical Black-Scholes Formula (risk free evaluated).

C506: **Impact of counterparty credit risk on risk appetite measures**
Presenter: **Bertrand Hassani**, Paris 1 Pantheon-Sorbonne, France
Co-authors: Cedric Naud

Risk appetite may be considered as the inverse function of the risk aversion. A financial institution has a portfolio of various types of risks. For example, with regards to Basel II accords these would be the market, the credit and the operational risks. Taking position on the market and contracting loans coupled with the pertaining business failures, result in a global risk position or GRP. Basel III comes up with several other requirements impacting this GRP, especially the counterparty credit risk. The GRP may be characterized by various risk measures, for instance, the VaR, the EPE, the Expected Shortfall, the Spectral and the Distorsion measures or even a combination of these. Based on the enumerated approaches and taking into account the dependencies, we propose to study the impact of the counterparty credit risk on the capital charge pertaining to the GPR and on a financial institution strategy which is informed by both risk measurement and understanding.

CS32 Room 5 FINANCIAL MARKETS LIQUIDITY Chair: Serge Darolles

C283: **Liquileaks**
Presenter: **Ting Wang**, VU University Amsterdam, Netherlands
Co-authors: Albert J. Menkveld

A security's liquidity properties have been studied in terms of mean and variance: liquidity level and liquidity risk, respectively. Tail events and liquidity disaster risk are explored. Liquidity might not be a worry to investors in normal market conditions, but it does become a first-order concern in case the security not only hits an illiquid state, but is trapped in it so that waiting a day will not restore liquidity. Empirically, such events, referred to as liquidity leaks or liquileaks, can be characterized with a Markov regime-switching model that alternates between a liquid and an illiquid state. Liquileak risk is naturally defined as the probability of finding the security in the illiquid state for more than a week. This probability is estimated based on an unbalanced sample of 2147 stocks from 1963 through 2008. Standard Fama-MacBeth regressions show that a one standard deviation increase in liquileak probability commands an annual premium of 1.33%. This premium has increased over time.

C337: **Identifying and analyzing hidden order placements**
Presenter: **Ruihong Huang**, Humboldt University of Berlin, Germany
Co-authors: Nikolaus Hautsch
Trading under limited pre-trade transparency becomes increasingly popular on financial markets. We provide first evidence on traders' use of (completely) hidden orders which might be placed even inside of the (displayed) bid-ask spread. Employing TotalView-ITCH data on order messages at NASDAQ, we propose a simple method to conduct statistical inference on the location of hidden depth and to test economic hypotheses. Analyzing a wide cross-section of stocks, we show that market conditions reflected by the (visible) bid-ask spread, (visible) depth, recent price movements and trading signals significantly affect the aggressiveness of 'dark' liquidity supply and thus the 'hidden spread'. Our evidence suggests that traders balance hidden order placements to (i) compete for the provision of (hidden) liquidity and (ii) protect themselves against adverse selection, front-running as well as 'hidden order detection strategies' used by high-frequency traders. Accordingly, our results show that hidden liquidity locations are predictable given the observable state of the market.

C382: **Liquidity in ETFs**
Presenter: **Fabrice Riva**, Universite Paris-Dauphine, France
Co-authors: Laurent Deville, Anna Calamia
The liquidity of ETFs and its determinants as the market develops and competition increases is studied. Our analysis is based on a long period database and a sample of equity ETFs trading on European markets. Instead of working with ETFs' holdings, we model the relationship existing between the liquidity of ETFs and that of the stocks composing their benchmark indices. The rationale for this link draws from the specific structure of ETFs. Among the factors affecting this relationship, we focus on issues related to the size of the market, the competition between ETFs, cross listing and replication style. We provide sound evidence that, as expected, the liquidity of ETFs, as measured by different low frequency spread proxies and trading volume, is related to the liquidity of the underlying index constituent stocks. However, our results also highlight the effects of competition on an index or an index class, of cross-listing and the importance of the way replication is achieved. Actually, swap-based ETFs present significantly higher liquidity than full replication ETFs and cross-listing of an ETF and higher competition on an index both result in improved liquidity.

C618: **MLiq a Meta Liquidity Measure**
Presenter: **Gaelle Le Fol**, CREST and Universite Paris - Dauphine, France
Co-authors: Serge Darolles, Jeremy Dudek
The 2008 financial crisis has shown the increasing importance of the liquidity risk, and its role in propagating the turmoil. However, seeking to control liquidity remains a harder task than controlling for any other risk. The first major obstacle comes from the difficulty to propose an accurate measure of liquidity as attested by the incredible number of market liquidity measures available. Liquidity is a multi-dimensional concept that not any of the numerous proxies are perfectly taking into account. Nevertheless, the choice of the liquidity measure is definitely not neutral to the analysis of the liquidity risk. Principal Component Analysis could be a solution to extract the liquidity factor from several liquidity measures. However, we show that it is inaccurate in some particular cases. We focus on serious liquidity problems: the only liquidity problems reported by all considered liquidity measures simultaneously. Our Meta-liquidity measure (MLiq) is defined as the probability to be in the state of high liquidity. MLiq is the probability to be in the state of high correlation between liquidity measures for one asset. In other words, the re-correlations of all dimensions of the liquidity measurement sheds light on the seriousness of the liquidity problem. Thus, we use a multivariate model allowing to measure correlations between liquidity measures jointly with a state-space model allowing to endogenously define the illiquid periods.

CS45 Room Multipurpose STATISTICAL SIGNAL EXTRACTION AND FILTERING Chair: Stephen Pollock

C529: **Credit default and macroeconomic factors: An empirical investigation**
Presenter: **Tatiana Cesaroni**, Bank of Italy, Italy
Over the past few years, a lot of attention has been devoted in analysing the determinants of Firms Probability Default. Given the strong link between economic activity conditions and credit risk, the use of default macro modelling is particularly indicated. To this end this paper analyses the linkages between Italian non financial firm's defaults and macroeconomic factors using a Vector Autoregressive model. In more detail we consider a model in which the effects of variables such as GDP, unemployment, interest rates etc on default rates behaviour are evaluated. To further inspect the issue we compare the results with those obtained from an analysis by sectors.

C648: **A note on pseudo-spectra and pseudo-covariance generating functions**
Presenter: **Marcos Bujosa**, Universidad Complutense de Madrid, Spain
Co-authors: Andres Bujosa, Antonio Garcia-Ferrer
Although the spectral analysis of stationary stochastic processes has solid mathematical foundations, this is not the case for non-stationary stochastic processes. The algebraic foundations of the spectral analysis of non-stationary ARIMA (or SARIMA) processes are established. For this purpose the Fourier Transform is extended to the field of fractions of polynomials. Then, the Extended Fourier Transform pair pseudo-covariance generating function / pseudo-spectrum, analogous to the Fourier Transform pair covariance generating function – spectrum, is defined. The new transform pair is well defined for stationary and non-stationary ARMA processes. This new approach can be viewed as an extension of the classical spectral analysis. It is shown that the frequency domain has some additional algebraic advantages over the time domain.

C678: **Reliability of the (new) tramo-seats automatic identification of reg-ARIMA models**
Presenter: **Agustin Maravall**, Bank of Spain, Spain
Insofar that –as Hawking and Mlodinow state– "there can be no model-independent test of reality," time series analysis applied to large sets of series needs an automatic model identification procedure. The presentation will discuss the results of the (default) automatic identification of regression-ARIMA models in the new versions of programs TRAMO and SEATS. The procedure identifies several types of outliers and calendar effects through regression, and identifies the ARIMA model for the stochastic series. Two sets of series are treated: one with 50,000 simulated series, generated by 50 different ARIMA models; the second with 14000 real economic series of different lengths. The first set shows the accuracy of the procedure in identifying the correct model when the series follows a reg-ARIMA model; the second set shows the ability of these models to capture real-world series dynamics, as evidenced by the validity of the n.i.i.d. assumption for the residuals and by the out-of-sample forecasts performance. A comparison with the automatic procedure in the present X12-ARIMA and DEMETRA+ programs (based on older versions of the TRAMO) is made.

C969: **SVD-based Kalman filtering and likelihood evaluation under linear constraints**
Presenter: **Christian Heinze**, University Bielefeld, Germany
Reformulations of the basic Kalman filtering equations abound in the literature on linear time series methods. Most algorithms either focus on state covariances or their inverses and, at the next level, consider these as is or in square root form. Covariance based approaches allow cheap time updates whereas information based approaches excel during the measurement step—a consequence of a correspondence of time and measurement updates between these classes. A more recent idea is filtering based on singular value decompositions (SVD). Since matrix inverses and square

roots are obtained in linear time once an SVD is available, the latter algorithms allow exploiting the just mentioned correspondence to combine some of the merits of both classes. The currently available formulation requires non-singular covariances. This requirement is removed here by utilizing the generalized SVD. Also, SVD-based Likelihood evaluation has received little attention, e.g., the only available R-package implementing SVD-based filtering increases the associated computation load by almost 50% via an additional SVD-step. It is shown how to avoid this by adapting available results in the singular and non-singular case. Finally, the SVD-approach is compared with a sequential processing covariance filtering in the presence of linear restrictions (benchmarking) in an application to German consumer price indices.

CS59 Room 2 FILTERING AND MODELLING FINANCIAL TIME SERIES Chair: Helena Veiga

C357: **Smooth filtering and likelihood inference in dynamic latent variables models**
Presenter: **Dennis Kristensen**, University College London, United Kingdom
Co-authors: Cristian Brownlees, Yongseok Shin

One of the main challenges in the implementation of dynamic latent variable models (DLVM's) is that their likelihood functions are in general not available in closed form. Novel Sequential Monte Carlo (SMC) methods for computing likelihood functions of DLVM's are proposed. We modify the standard SMC algorithm by introducing nonparametric smoothing in the resampling step. This delivers likelihood approximations that are smooth with respect to the underlying parameters, hence suitable for maximum likelihood estimation. The smoothing step introduces small additional noise which is beneficial when the state variable is highly persistent. The algorithm allows for general dependence between the errors entering the measurement and state equations and, in its more general form, does not require closed-form expressions of the measurement density. We establish rate results for the smooth filter and establish the asymptotic properties of the resulting estimators. The usefulness of our proposed method is demonstrated with Monte Carlo studies and an application to the estimation of continuous-time stochastic volatility models given first-step estimates of integrated volatility.

C477: **Maximum likelihood estimation of stochastic volatility models via iterated filtering**
Presenter: **Carles Breto**, Universidad Carlos III de Madrid, Spain

The likelihood of a variety of stochastic volatility models is maximized. We also give confidence intervals for the maximum likelihood point estimates and filtering intervals for the unobserved variables. To do this, we rely on the recently developed approach of iterated filtering, which we implement via particle filters. Iterated filtering only requires computer code that corresponds to the equations of the model at hand and no model-specific, analytical derivations or approximations are needed. This makes the approach easy to implement in a range of different stochastic volatility models. To illustrate the flexibility of the approach, we start by fitting basic stochastic volatility models to data. Then, we analyze the residuals looking for features that may be missing in the initial model. Finally, we fit modified models that incorporate such features, including time-varying leverage, regime switching and non-Gaussian noise distributions with heavy tails or asymmetries.

C487: **Regime switching GARCH-based clustering of financial time series**
Presenter: **Jose Dias**, Instituto Universitario de Lisboa ISCTE, Portugal
Co-authors: Sofia Ramos

A time series clustering methodology is introduced. It combines regime-switching models within time series and a heuristic clustering approach at the second level, exploring the distance between time series. Each time series is modeled as a regime switching model that captures for instance nonlinearities and breaks, i.e., a regime switching GARCH model is considered for each time series with different specifications of the error terms such as normal, skew-normal, t, and skew-t. The clustering of time series is, then, performed on the similarity between the posterior regime probabilities and subsequent application of a hierarchical clustering strategy. This methodology is first illustrated using synthetic data sets. Then, an empirical application with European stock markets shows the potential of the method in financial econometrics, in particular in the analysis of markets synchronization.

C159: **Risk premium, variance premium and the maturity structure of uncertainty**
Presenter: **Abderrahim Taamouti**, Universidad Carlos III de Madrid, Economics, Spain
Co-authors: Bruno Feunou, Jean-Sebastien Fontaine, Romeo Tedongap

Theoretical risk factors underlying time-variations of risk premium across asset classes are typically unobservable or hard to measure by construction. Important examples include risk factors in Long Run Risk [LRR] structural models as well as stochastic volatility or jump intensities in reduced-form affine representations of stock returns. Still, we show that both classes of models predict that the term structure of risk-neutral variance should reveal these risk factors. Empirically, we use model-free measures and construct the ex-ante variance term structure from option prices. This reveals (spans) two risk factors that predict the bond premium and the equity premium, jointly. Moreover, we find that the same risk factors also predict the variance premium. This important contribution is consistent with theory and confirms that a small number of factors underlies common time-variations in the bond premium, the equity premium and the variance premium. Theory predicts that the term structure of higher-order risks can reveal the same factors. This is confirmed in the data. Strikingly, combining the information from the variance, skewness and kurtosis term structure can be summarized by two risk factors and yields similar level of predictability. This bodes well for our ability to bridge the gap between the macro-finance literature, which uses very few state variables, and valuations in option markets.

Monday 03.12.2012 08:30 - 10:10 Parallel Session M – ERCIM

ES67 Room 9 HIGH DIMENSIONAL PROBLEMS IN APPLICATIONS TO GENETICS Chair: Florian Frommlet

E118: **Statistical methods for detecting single nucleotide polymorphisms using genome sequencers**
Presenter: **David Ramsey**, University of Limerick, Ireland
Co-authors: Andreas Futschik, Ali Sheikhi

The genome consists of sequences made of 4 nucleotides. At a majority of the sites along these sequences, each individual in a species will have the same nucleotide. A site where there is variation within a population is called a single nucleotide polymorphism (SNP). At virtually all such sites, just two of the four nucleotides appear. These variants are called alleles, the most common (rare) allele is termed the major allele (minor allele, respectively). Genome sequencers read these sequences by utilizing DNA from an individual (or pool) placed within a lane. Each lane gives a random number of reads for a particular site (can be modelled using the Poisson or negative binomial distribution). One could use the following simple test for the presence of a minor allele: accept that there is a minor allele if in any lane the number of reads for an allele that is not the major allele exceeds a given threshold. Due to the large number of sites on the genome it is necessary to adopt an appropriate multiple testing procedure (e.g. the Benjamini-Hochberg procedure). This simple test is compared to a likelihood ratio test. Adaptations of these tests to real data are also considered.

E287: **Learning causal models from interventions: Inference and active learning**
Presenter: **Alain Hauser**, ETH Zuerich, Switzerland

Causal relationships between random variables are commonly modeled using DAGs, directed acyclic graphs. DAGs can only be inferred up to Markov equivalence from observational data, that is, data produced by the undisturbed system. We show to which extent interventional data, that is, data produced under a perturbation of the system, improves the identifiability of causal models. For this aim, we extend the notion of Markov equivalence of DAGs to the interventional case and present a graph theoretic characterization of the corresponding equivalence classes. Furthermore, we consider Gaussian causal models and address the problem of calculating the maximum likelihood estimator from interventional data. We present a generalization of Chickering's Greedy Equivalence Search algorithm to interventional data that makes regularized maximum likelihood estimation computationally feasible. We demonstrate the performance of this algorithm in a simulation study. Finally, we address the question of active learning, that is, the question of finding interventions that lead to "optimal" identifiability of causal models. Both the greedy search algorithm and the active learning approaches are based on the graph theoretic characterization of interventional Markov equivalence classes.

E418: **Added predictive value of prediction models based on high-dimensional data**
Presenter: **Riccardo De Bin**, University of Munich, Germany
Co-authors: Anne-Laure Boulesteix

In the biomedical literature, numerous studies highlighted the relationship between some gene expressions and specific disease outcomes. Usually, they provide a signature useful for diagnosis purposes. Unfortunately, such studies often have a reduced impact in biomedical applications, partially because they focus only on the prediction ability of the proposed signature, without considering the effect of clinical predictors such as age, sex, tumor stage, etc. The latter are usually extensively investigated and validated, and they may provide the same information of molecular predictors, with the advantage of an easier and cheaper availability. For this reason, it is important to evaluate the added predictive value of a gene signature given the clinical predictors and, in a more extended way, to consider clinical predictors in procedures such as variable selection, model construction, etc. Here we briefly review some approaches for validating the added predictive value of high-throughput molecular data to clinical data, focusing on the case in which an independent validation dataset is given. Finally, an application in a study for the prediction of survival and time to treatment in chronic lymphocytic leukemia is presented.

E604: **Gene-gene interaction analysis for multiple phenotypes**
Presenter: **Taesung Park**, Seoul National University, Korea, South
Co-authors: Iksoo Huh, Min-seok Kwon

Most genome-wide association studies (GWAS) have been conducted by focusing on one trait of interest for identifying genetic variants associated with common complex traits. However, despite of many successful results from GWAS, only a small number of genetic variants tend to be identified and replicated given a stringent genome-wide significance criterion. Furthermore, in many GWAS, one binary trait is commonly used which is usually derived from multiple quantitative traits. For example, a binary trait representing hyperlipidemia status is defined by combining four quantitative traits: Total cholesterol (Tchl), High density lipoprotein (HDL) cholesterol, Low density lipoprotein (LDL), and cholesterol and Triglycerides (TG). The use of this summary phenotype may decrease power due to the loss of information about phenotypes. Therefore, we propose a multivariate approach which uses all information about multiple phenotypes. Especially, we overview an identification of gene-gene interaction effects on the multiple phenotypes. We propose a multivariate version of the multifactor dimensionality reduction (MDR) approach in order to identify gene-gene interaction for the multiple phenotypes. Our proposed multivariate MDR method uses multiple phenotypes simultaneously. We also consider an extension of the generalized estimating equations (GEE) approach. We applied the multivariate MDR approach to a GWA dataset of 8,842 Korean individuals genotyped by Affymetrix SNP array 5.0, and detected novel genetic variants associated with obesity-related phenotypes, which had been missed in an earlier single phenotype analysis.

ES39 Room 8 OPTIMAL EXPERIMENTAL DESIGN Chair: Jesus Lopez Fidalgo

E195: **Optimal experimental designs for AR(1) model**
Presenter: **Mariano Amo-Salas**, University of Castilla-La Mancha, Spain
Co-authors: Jesus Lopez-Fidalgo

Time series models are a sort of nonlinear models where the observations are not independent and the function which defines this correlation depends on the mean of the model. These models are studied within the framework of Optimal Experimental Designs in order to obtain the best moments to perform the experiments. The model and the covariance function are expressed in a suitable form to apply the usual techniques of Optimal Experimental Design. The expression of the Fisher information matrix is provided for estimating the parameters of the model when they appear in the covariance function. Optimal designs for the AR(1) model are computed for different nominal values of the nonlinear parameter and their efficiencies are compared.

E213: **Optimality in adaptive designs**
Presenter: **Jose Moler**, Universidad Publica de Navarra, Spain
Co-authors: Arkaitz Galbete, Fernando Plo

A generalized linear model is proposed to relate the response of patients and its allocation among several treatments in phase I, II or III of a clinical

trial. Then, optimal design techniques are taken into account in order to propose an optimal allocation depending on the inferential goal. This optimal allocation becomes the allocation target of the design. The inclusion of covariates in the model is also considered.

E242: **OAT designs for single and cross effects**
Presenter: **Joao Rendas**, CNRS-University of Nice, France
Co-authors: Jean-Marc Fedou, Gilles Menez, Luc Pronzato
Morris scheme for One At a Time (OAT) designs for sensitivity analysis is widely used for rapid identification of groups of important and linear inputs of a multivariate function $f(x), x \in A\mathbb{R}^k$. Given its higher efficiency, its clustered version that computes $m > 1$ samples of each derivative of f in the neighborhood of the same point x is especially appealing. We present a formal specification of a family of balanced clustered designs for arbitrary values of k and $m \leq 2^{k-1}$, extending the original presentation of Morris that was defined only for selected pairs (k,m). We also extend Morris concept of clustered designs to the sampling of second order derivatives. The correctness of our construction is demonstrated using a polynomial representation of sub-graphs of the unit cube Q_k, on which a convenient inner product is defined.

E392: **Formulating, via copula, and fitting latent variable models for paired comparisons and ranking studies: An application of optimal design theory**
Presenter: **Ben Torsney**, Glasgow, United Kingdom
In a paired comparisons experiment a subject has to indicate which of two 'treatments' T_i, T_j is preferred. We observe O_{ij}, the frequency with which T_i is preferred to T_j in n_{ij} comparisons. Examples naturally arise in Food Tasting and Market Research, while others can be extracted from studies in Health, Sport, Journal Citations. Under a class of models for such data, which include the Bradley Terry and Thurstone models, $P(T_i$ is preferred to $T_j) = F(lambda_i - lambda_j)$, where $F(.)$ is a symmetric distribution function and $lambda_i$ is a treatment index. For identifiability purposes constraints must be imposed on parameters. One is to assume that sum of $p_i = 1$, where $p_i = ln(lambda_i)$; an alternative is product of $p_i = 1$; i.e. sum of $lambda_i = 0$. Thus theorems identifying optimal design weights and algorithms for determining them carry over to the maximum likelihood estimation of these parameters. We will explore this fusion of topics, taking the opportunity to expand on the class of models, both for simple paired comparisons data and also for data consisting of orderings or rankings. In particular we will propose models based on copula for rankings and exploit multiplicative algorithms, developed in the design context, for maximum likelihood estimation.

ES92 Room 10 CLUSTERING AND CLASSIFICATION Chair: Jean-Michel Poggi

E274: **Distance-based approach to one-class classification**
Presenter: **Basilio Sierra**, University of the Basque Country, Spain
Co-authors: Concepcion Arenas, Itziar Irigoien
In a classification problem, usually units are available for all classes that can occur. Nevertheless, there are situations where only units of a class (target class) are available. Then, the main goal is to decide whether a new unit belongs to the target class. This kind of problem is called One-class classification. It is commonly tackled using methods based on density estimation, which are based on mixture of Gaussians or need Markov models. Other techniques are based on boundary approaches, but some of them, as the K-centers, require the estimation of some parameters. Other methods are based on support vector data description such as that implemented in libSVM . Although these methods show in general good results, they are only adequate for continuous variables, need a known distribution or require the estimation of some parameters. To avoid these restrictions a distance-based approach is presented. We evaluate the ability of our procedure using experimental datasets with any kind of data and presenting comparison results with other methods. All datasets considered in the evaluation have multiple classes. Each class in turn is considered as the target class and the units in the other classes are considered as new units to be classified.

E682: **Functional clustering of mouse ultrasonic vocalization data**
Presenter: **Xiaoling Dou**, The Institute of Statistical Mathematics, Japan
Co-authors: Shingo Shirahata, Hiroki Sugimoto, Tsuyoshi Koide
It has been found that male mice emit ultrasonic vocalizations (USVs) towards females during male-female interaction. Our purpose is to classify the waveforms of the mouse USV data. The data are transformed by FFT. Because the USV data are very noisy, it is difficult to analyze them by known software. In our data analysis, we first transform the data by a moving average method and decide the least intensity. If the intensity of the transformed data is less than the least intensity, then the data is regarded as noise. Then we specify the intervals of the USV calls and determine their lengths. According to the call shapes, we then smooth them with B-spline basis functions and count their number of breakpoints. The clustering is performed as follows. First, we classify the USV curves as the first level clusters by their number of breakpoints. Then, for the first level clusters, we construct clusters based on coefficient vectors of the curves. This analysis can help us to find some sequence patterns of USVs in mouse communication.

E779: **Combination of one-vs-one and one-vs-all using different base classifiers**
Presenter: **Inigo Mendialdua**, UPV/EHU, Spain
Co-authors: Basilio Sierra, Elena Lazkano, Ekaitz Jauregi
The binarization strategies decompose the original multiclass dataset into multiple two-class subsets, learning a different binary model for each new subset. One-vs-one and one-vs-all are two of the most well-known techniques: one-vs-one separates a pair of classes in each binary subproblem, ignoring the remaining ones; and one-vs-all distinguishes one class from all the other classes. A new method is proposed where both methods are combined and a different base classifier is applied in each binary subproblem. In order to validate the performance of the new proposal, an empirical study has been done where the new method has been compared with one-vs-one and one-vs-all. Experimental results show that the new method gets promising results.

E866: **Functional factorial $k-$means and the related methods**
Presenter: **Michio Yamamoto**, Osaka University, Japan
Co-authors: Yoshikazu Terada
A new procedure is developed to find optimal partitions of multivariate functional objects and also find an optimal subspace for partitioning, simultaneously. The method is based on the $k-$means criterion for functional data in the subspace and seeks the subspace that is maximally informative about the clustering structure. An efficient alternating least-squares algorithm is described, and the proposed method is extended to regularized method. Furthermore, the proposed method is theoretically and empirically compared with the related methods including functional reduced $k-$means, functional $k-$means and tandem analysis.

E772: **Clustering of financial time series in risky scenarios**
Presenter: **Roberta Pappada**, University of Padua, Italy
Co-authors: Fabrizio Durante, Nicola Torelli
A methodology is presented for clustering financial time series according to the association between their low values (defined according to a given threshold). The procedure is based on the calculation of suitable pairwise conditional Spearman's correlation coefficients extracted from the series.

The performance of the method has been tested via a simulation study. Moreover, an empirical analysis will be illustrated. The results could be applied to construct asset portfolios that reduce the effect of a simultaneous financial crisis in several markets.

ES72 Room 7 NONPARAMETRIC AND ROBUST MULTIVARIATE METHODS Chair: Marco Riani

E203: **On robust estimation of singular covariance matrices**
Presenter: **Pauliina Ilmonen**, Universite libre de Bruxelles, Belgium
Co-authors: Catherine Dehon

Minimum Covariance Determinant (MCD) method is one of the most well known methods for robustifying the estimation of location, and scatter of p variate observations. MCD method is based on considering all the subsets containing h sample points (usually 50%) of the original observations, and estimating the covariance matrix, and the mean vector, on the data of the subset associated with the smallest covariance matrix determinant. However, for singular square matrices, the determinant is always zero. Thus the original MCD method works only when the covariance matrix is nonsingular. The problem occurs also if the entire data (including outlying points) lies in $\mathbb{R}^p$, but majority of the observations lie in a smaller dimensional subspace. In many practical applications, for example in regression settings when some regressors are categorical, the covariance matrix may indeed be singular. We consider MCD type method for singular covariance matrices. The idea behind the new singular MCD method is simple. It is based on volume minimization of singular matrices. We also provide a method for choosing the number of included data points (h) wisely. The limiting properties of the new singular MCD method rely heavily on the provided method for choosing h. Based on simulated examples, the new singular MCD method works well also when the sample size is small.

E305: **Robust estimation in multivariate multiple regression using LAD-lasso**
Presenter: **Jyrki Mottonen**, University of Helsinki, Finland
Co-authors: Mikko Sillanpaa

In the univariate multiple regression case the regularization can be done, for example, by using lasso estimation method. The lasso estimates are found by minimizing residual sum of squares $\sum(y_i - \alpha - \beta_1 x_{i1} - \ldots - \beta_q x_{iq})^2$ subject to $\sum_{j=1}^{q} |\beta_j| \leq t$ where the tuning parameter $t \geq 0$ controls the amount of shrinkage. It is well known that the ordinary least squares estimates, and consequently the lasso estimates, are very sensitive to outliers. We consider the multivariate multiple regression case and propose a multivariate LAD-lasso based on spatial signs. The performance of the method is illustrated using multivariate simulated marker data examples.

E521: **On robustifying some second order blind source separation methods for nonstationary time series**
Presenter: **Klaus Nordhausen**, University of Tampere, Finland

Blind source separation (BSS) is an important analysis tool in various signal processing applications like image, speech or medical signal analysis. The most popular BSS solutions have been developed for independent component analysis (ICA) with identically and independently distributed (iid) observation vectors. In many BSS applications the assumption of iid observations is not realistic, however, as the data are often an observed time series with temporal correlation and even nonstationarity. In this talk, some BSS methods for time series with nonstationary variances are discussed. We also suggest ways to robustify these methods and illustrate their performance in a simulation study.

E856: **Robust extraction of relevant information from multivariate time series in real-time**
Presenter: **Matthias Borowski**, TU Dortmund University, Germany
Co-authors: Roland Fried

The real-time (also online or sequential) monitoring of multivariate data stream time series takes place in many fields. Since the time series are measured with high frequency, relevant information is corrupted by a changing amount of noise and outliers. The time series are often not stationary but show enduring and suddenly changing trends and level shifts. Moreover, cross-dependences between the univariate time series may change over time. The extraction of relevant information in real-time from such complex time series is an obvious but challenging task calling for fast, robust multivariate methods. New robust procedures are presented for multivariate signal extraction, for the online-estimation of the volatility, for the online-detection of signal changes (and hence for a data-adaptive choice of the window width) and for the online-surveillance of cross-dependences. All methods are based on robust regression in moving windows and therefore feasible for real-time application.

ES70 Room 13 ROBUST METHODS FOR COMPLEX DATA Chair: Matias Salibian-Barrera

E227: **Robust PARAFAC for incomplete data**
Presenter: **Tim Verdonck**, KU Leuven, Belgium
Co-authors: Mia Hubert, Johan Van Kerckhoven

Different methods exist to explore multiway data. We focus on the widely used PARAFAC (parallel factor analysis) model, which expresses multiway data in a more compact way without ignoring the underlying complex structure. An alternating least squares procedure is typically used to fit the PARAFAC model. It is, however, well known that least squares techniques are very sensitive to outliers, and hence, the PARAFAC model as a whole is a nonrobust method. Therefore a robust alternative, which can deal with fully observed data possibly contaminated by outlying samples, has already been proposed in literature. We present an approach to perform PARAFAC on data that contain both outlying cases and missing elements. A simulation study shows the good performance of our methodology. In particular, we can apply our method on a dataset in which scattering is detected and replaced with missing values. This is illustrated on a real data example.

E384: **On approximations of the finite sample distribution of support vector machines**
Presenter: **Andreas Christmann**, University of Bayreuth, Germany

The focus will be on the mainly open question how to draw statistical decisions based on SVMs. This question has not yet attracted much consideration in the literature. From an applied point of view, this might be considered as a serious gap, because knowledge of the finite sample or asymptotic distribution of a statistical method is the basis to make statistical decisions like confidence regions, prediction intervals, tolerance intervals or tests. Some recent results on theoretical properties of bootstrap approximations for the finite sample distribution of SVMs are given.

E432: **Minimum volume peeling: A multivariate mode estimator**
Presenter: **Giarncarlo Ragozini**, Federico II University of Naples, Italy
Co-authors: Giovanni C. Porzio, Steffen Liebscher, Thomas Kirschstein

Three concurrent measures of location are offered to students within standard statistical textbooks: the arithmetic mean, the median, and the mode. Amongst them, the mode is the most neglected parameter. However, the mode has many relevant properties: it is a location measure that is robust with respect to contaminated data; it is appropriate to locate skewed distributions; it is a convenient location parameter for truncated distributions. Finally, we should mention that the value that most likely occurs is an interesting parameter by itself. Estimating the mode has been extensively studied in the univariate case. On the other hand, a few contributions are available to estimate the mode of a multivariate distribution. For this reason, this paper introduces *minimum volume peeling*, a direct multivariate mode estimation procedure based on Sager's idea. We propose to direct estimate the multivariate mode through any point $\hat{\theta}$ that belongs either to the set with the minimum volume set containing a given fraction of points.

With this aim, and in analogy with the convex hull peeling procedure, we propose to adopt a peeling approach, i.e. we propose to iteratively select nested subsamples with a decreasing fraction of sample points, looking for the minimum volume subsample at each step. Properties, robustness, and implementing issues of our proposal are discussed.

E841: **S-estimators for heteroscedastic regression**
Presenter: **Peter Slock**, Ghent University, Belgium
Co-authors: Stefan Van Aelst, Matias Salibian-Barrera
Similarly to least squares estimators, regression S-estimators remain consistent under more general assumptions than standard linear regression. However, these estimators may become inefficient for specific model deviations, such as heteroscedastic errors. Therefore, S-estimators are extended to heteroscedastic regression models. The idea is to consider more complex linear models that account for heteroscedastic errors by modelling the variance as a parameterized (exponential) function of predictor variables x, regression parameters β, and heteroscedasticity parameters γ. For each choice of β and γ, a robust M-estimator of the corresponding residual scale $\hat{\sigma}(\beta,\gamma)$ is considered. The robust S-estimators β are then given by minimizing $\hat{\sigma}(\beta,\gamma)$ over β. First, it is shown how an optimal value of γ can be obtained corresponding to each choice of β. Based on this relation an efficient algorithm for these heteroscedastic S-estimators is constructed. The robustness and efficiency of these S-estimators are also studied and compared to the MLE for this model. Finally, robust inference procedures for these S-estimators are derived as well.

ES50 Room 11 CONTRIBUTIONS TO TIME SERIES MODELING AND COMPUTATION Chair: Manuel G. Scotto

E568: **Multivariate time series model with hierarchical structure for over-dispersed discrete outcomes**
Presenter: **Nobuhiko Terui**, Tohoku University, Japan
Co-authors: Masataka Ban
A multivariate time series model for over-dispersed discrete data is proposed to explore the market structure based on sales count dynamics. We first discuss the microstructure to show that over-dispersion is inherent in the modeling of market structure based on sales count data. The model is built on the likelihood function induced by decomposing sales count response variables according to products' competitiveness and conditioning on their sum of variables, and it augments them to higher levels by using Poisson-Multinomial relationship in a hierarchical way, represented as a tree structure for the market definition. State space priors are applied to the structured likelihood to develop dynamic generalized linear models for discrete outcomes. For over-dispersion problem, Gamma compound Poisson variables for product sales counts and Dirichlet compound multinomial variables for their shares are connected in a hierarchical fashion. Instead of the density function of compound distributions, we propose a data augmentation approach for more efficient posterior computations in terms of the generated augmented variables. We present an application using a weekly product sales time series in a store to compare the proposed model with one without over-dispersion by using several model selection criteria, including in-sample fit, out-of-sample forecasting errors, and information criterion, to show that the proposed model provides improved results.

E815: **Segmentation of time series via complexity of continuous functions**
Presenter: **Boris Darkhovsky**, Institute for Systems Analysis Russian Academy of Sciences, Russia
Co-authors: Alexandra Piryatinska
In many applications time series are sequences of connected, distinct segments which are generated by their own individual mechanisms. To analyze such series it is necessary to split them into these segments. If time series is generated by stochastic mechanisms, then the known change-point detection algorithms can achieve the segmentation. However it is not the case for deterministic or mixed mechanisms. A new approach to this problem based on the novel concept of the complexity of a continuous function is proposed. In this method, part of the continuous function values is discarded, while retaining the rest of the function values. Then the complexity of a continuous function is defined as the fraction of the retained values necessary to recover the original function via a certain fixed family of approximation methods without exceeding a given error. It is shown that the dependence of the complexity of a function on the reconstruction error can be well approximated in logarithmic coordinates by an affine function. Its parameters, calculated dynamically, are then used as diagnostic sequences to find the change-points of the original time series. The effectiveness of this procedure is verified in the case of several simulated time series.

E900: **A dynamic copula model for time series**
Presenter: **Wai Kit Jacky Fung**, Hong Kong University of Science and Technology, China
Co-authors: Mike K.P. So
A new dynamic copula model is introduced. The proposed model is useful in describing the dependence structure between time series variables. Existing dynamic copula models assume a particular parametric form of the copula function. If the copula function is wrongly specified, it will lead to unsatisfactory inference or poor prediction in time series. It is proposed to estimate dynamic copula function nonparametrically via kernel functions. Using a kernel based approach to estimate time-dependent copula functions can give a smooth and continuous copula while allowing the dependence measures to evolve over time. The estimation of parameters is done by a two-step approach. The simulations show that the estimation method is reliable. We apply the model to multiple financial time series to understand the time series properties of the dependence among financial returns.

E1010: **Bootstrap log-averaged sample spectral estimator: An application to long-memory models**
Presenter: **Valderio Anselmo Reisen**, DEST-CCE-UFES, Brazil
Co-authors: Glaura Franco, Giovani Comarela, Helene Boistard
A bootstrap procedure is employed to obtain replications of the periodogram in order to calculate a bootstrap log-averaged sample spectral estimator. Among the asymptotic properties discussed here, the consistency of the bootstrap spectral estimator is established for short and long-memory stationary processes. The inference based on this procedure is considered on the estimation of the memory parameter d of the stationary ARFIMA(p,d,q) process using a semiparametric regression approach. The proposed methodology is also used to obtain confidence intervals for d. The finite sample size properties are investigated through Monte Carlo experiments and, as expected, the empirical results evidence that the consistency property of the bootstrap spectral estimator is also translated into the finite sample properties of d estimates. Although, even asymptotically, the independence of the periodogram ordinates are not guaranteed for long-memory processes, the empirical investigation shows that the averaged bootstrap periodogram estimator, based on bootstrap replications of the periodogram, gives precise estimates of the memory parameter. In addition, the use of the averaged bootstrap periodogram estimator improves the accuracy of the interval estimates of parameter d by presenting better coverage rates with smaller width compared with the usual log-periodogram regression estimator. An application is presented to show the usefulness of the proposed methodology.

E1059: **Comparison of goodness-of-fit tests for GARCH models**
Presenter: **Kilani Ghoudi**, United Arab Emirates University, United Arab Emirates
Co-authors: Bruno Remillard
Using simulations, we compare goodness-of-fit tests for the null hypothesis of a Gaussian distribution for the innovations of GARCH models. More precisely, Cramér-von Mises and Kolmogorov-Smirnov type statistics are computed for empirical processes based on the standardized residuals and

their squares. For computing *P*-values we use the parametric bootstrap method and the multipliers method. We also use the so-called Khmaladze's transform to obtain an approximated Brownian motion under the null hypothesis, for which we compute also Cramér-von Mises and Kolmogorov-Smirnov type statistics, using both the standardized residuals and their squares.

ES79 Room 12 STATISTICS OF EXTREMES AND APPLICATIONS II Chair: Claudia Klueppelberg

E966: **Augmented cumulative distribution networks for multivariate extreme value modelling**
Presenter: **Gildas Mazo**, Inria, France
Co-authors: Florence Forbes, Stephane Girard

Max-stable distribution functions are theoretically grounded models for modelling multivariate extreme values. However they suffer from some striking limitations when applied to real data analysis due to the intractability of the likelihood when the number of variables becomes high. Cumulative Distribution Networks (CDN's) have been introduced recently in the machine learning community and allow the construction of max-stable distribution functions for which the density can be computed. Unfortunately, we show that the dependence structure expected in the data may not be accurately reflected by max-stable CDN's. To face this limitation, we therefore propose to augment max-stable CDN's with the more standard Gumbel max-stable distribution function in order to enrich the dependence structure.

E981: **Multifractal characteristics of spatial threshold exeedances**
Presenter: **Francisco Javier Esquivel**, University of Granada, Spain
Co-authors: Ana Esther Madrid, Jose Miguel Angulo

Structural characteristics of spatial threshold exceedance sets have been analyzed in a previous recent work in terms of the point patterns defined by centroids of connected components, in relation to the fragmented nature of such sets, as well as based on the point patterns defined by *A*-exit points, which reflect regularity of the boundary. In particular, correlation range and local variability properties of the underlying model are inherited by threshold exceedance sets, with a certain asymptotic behavior as the threshold increases. Multifractal characteristics displayed for both types of point patterns under certain general model formulations are studied here in terms of the associated generalized dimensions and multifractal spectrum, showing significant differences derived from the model parameter specifications as well as the increasing value of the threshold. A dependence coefficient for dimensional interaction of coordinates is also calculated based on entropy dimensions to discriminate behavior under different scenarios. Simulation results show the potential of these tools to help understanding the complexity inherent to extremal behavior defined by threshold exceedances in spatial processes.

E1023: **On heuristic choices of tuning parameters in the estimation of the extreme value index**
Presenter: **Manjunath Bangalore Govinda Raju**, University of Lisbon, Portugal
Co-authors: M. Isabel Fraga Alves, M. Ivette Gomes, Ligia Henriques-Rodrigues

An adequate estimation of the second-order parameters (ρ, β), where $\beta \neq 0$ and $\rho < 0$, in the class of peaks over random threshold (PORT) minimum variance reduced bias(MVRB) extreme value index (EVI) estimators, which are both location and scale invariant, depends on heuristic choices of some tuning parameters. We explore a new data-driven heuristic procedure in estimating the scale second order parameter β and shape second order parameter ρ and its stability in the choice of tuning parameters. Nevertheless, one can also use such type of estimators in estimating high-quantiles, probabilities of exceedances of high levels and so on. To motivate the use of all developed computational algorithms and illustrate their performance, we present their implementation in the R software.

E915: **Query by committee and extreme values**
Presenter: **Theo Rietsch**, CNRS, France
Co-authors: Nicolas Gilardi, Philippe Naveau, Armelle Guillou

The query by committee (QBC) is an application of neural networks used in order to build designs of experiments. Neural networks can be used to model functions when only a few values of the function are known. The QBC becomes very useful when the observations are scarce and very expensive (either in time or in money) to obtain. This method consists in building several models of our function called experts and to determine, via a certain cost function Δ, the points where the experts disagree the most (i.e the maximas of Δ), which will be the points where the new observations are to be made. The optimisation of the neural networks is usually done using least squares and the function Δ is the variance. Our goal is to model the behaviour of extreme rainfall on the territory, having only observations on a certain number of stations. This problem has already been tackled for the means, but it is more difficult to deal with extremes since they behave completely differently. The aim is to try to apply the QBC to solve this problem.

E924: **Estimation of max-stable processes by simulated maximum likelihood**
Presenter: **Erwan Koch**, CREST and ISFA, France

Max-stable processes are very appropriate for the statistical modeling of spatial extremes. Nevertheless their inference is tedious. Indeed the multivariate density function is not available and thus standard likelihood-based estimation methods cannot be applied. The commonly used method - based on composite likelihood - leads to non efficient estimators. In this study an approach based on nonparametric simulated maximum likelihood is developed. The multivariate density (in space) is approximated by kernel methods. Our approach can be applied for many subclasses of max-stable processes and can provide better results than methods already introduced. This is true in particular when only a few temporal observations of the process are available and when the spatial dependence is high. The estimator is efficient when both the temporal dimension and the number of simulations go to infinity. The methodology is examined on simulated data and applied to real data. Applications to insurance are discussed.

Monday 03.12.2012 10:40 - 12:20 Parallel Session N – CFE

CS07 Room Multipurpose MEASURING SYSTEMIC RISK Chair: Lorenzo Frattarolo

C130: **International transmission of equity volatilities with application to financial crises**
Presenter: **Alain Kabundi**, University of Johannesburg, South Africa
Co-authors: Andrew Duncan
A parsimonious empirical model of cross-market volatility transmission is developed. The model, a time-varying version of the factor-augmented vector autoregression (FAVAR) approach, is used to assess the effects on the world equity market of volatility shocks coinciding with financial crises in Asia (1997-8) and the United States (US, 2007-8). Our analysis is based on monthly data from 1994 to 2008 for a panel of 25 developed and 20 emerging markets, which together constitute over 95 percent of world market capitalisation. Of the two crises, the results indicate that the US shock leads to a larger increase in global volatility. However, the effect of the Asian crisis is more pronounced in the case of most emerging markets.

C1042: **Determining the number of regimes in Markov-switching VARMA models**
Presenter: **Maddalena Cavicchioli**, Ca Foscari University of Venice, Italy
Stable finite order VARMA$(p*;q*)$ representations for M-state Markov switching VARMA$(p;q)$ processes are given, where upper bounds for $p*$ and $q*$ are elementary functions of the dimension K of the process, the number M of regimes, and the orders p and q. If there is no cancellation, the bounds become equalities, and this solves the identification problem. Our results include, as particular cases, some already known, and improve bounds given in previous papers. Data simulations and an application on foreign exchange rates are given.

C541: **Proximity-structured multivariate volatility models for systemic risk**
Presenter: **Lorenzo Frattarolo**, University of Venice, Italy
Co-authors: Monica Billio, Loriana Pelizzon
Several measures of systemic risk based on connectedness have been proposed. We analyze how these measures can be helpful also in analyzing how systemic risk spread around in the financial system. In particular, we consider systemic risk measures based on connectivity and we define accordingly proximity matrices helpful for introducing structured multivariate volatility models since they mitigate or solve the curse of dimensionality problem for volatility models. We present the application of a BEKK specification and discuss spillovers effects.

C486: **Hedge fund tail risk and marginal risk contribution in fund of hedge funds**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Kaleab Y. Mamo, Loriana Pelizzon
Hedge fund tail risk for a fund of hedge funds is studied. As funds of hedge funds primarily invest in hedge funds, we develop a risk measure that captures the tail risk of single hedge fund strategies and the tails risk contribution of these hedge fund strategies to the overall portfolio tail risk of a fund of hedge funds. We show that, especially during crisis, all the different hedge fund strategies are contributing to the tail risk of the portfolio of hedge funds, mostly because of the exposure hedge fund strategies face to liquidity and credit risk. Even the dedicated short bias strategy is contributing to tail risk rather than reducing it during crisis.

CS39 Room 3 BAYESIAN FINANCIAL ECONOMETRICS Chair: Yasuhiro Omori

C260: **Volatility and quantile forecasts of returns using realized stochastic volatility models with generalized hyperbolic distribution**
Presenter: **Toshiaki Watanabe**, Hitotsubashi University, Japan
Co-authors: Makoto Takahashi, Yasuhiro Omori
The realized stochastic volatility model, which incorporates the stochastic volatility model with realized volatility, makes it possible to adjust the bias in realized volatility due to the market microstructure noise and non-trading hours. This model is extended such that the asymmetry in volatility and the dependence of the bias on the level of the true volatility are allowed and a wider class distribution, the generalized hyperbolic skew Student's t-distribution, is assumed for financial returns. The extended model is applied to quantile forecasts of financial returns such as value-at-risk and expected shortfall as well as volatility forecasts using the Bayesian estimation scheme via Markov chain Monte Carlo method. The volatility and quantile forecasts are evaluated by several backtesting procedures. Empirical results with Spyder, the S&P 500 exchange-traded fund, show that incorporating realized volatility improves both volatility and quantile forecasts, that heavy tail and skewness of daily returns are important for quantile forecasts but to a lesser extent for volatility forecasts, and that allowing for the dependence of the bias in realized volatility on the level of the true volatility does not substantially improve the model fit nor volatility and quantile forecasts.

C513: **Multivariate realized stochastic volatility model with leverage**
Presenter: **Tsunehiro Ishihara**, Hitotsubashi University, Japan
Co-authors: Yasuhiro Omori
A joint model of multivariate returns and realized measures of covariance is proposed. The model of returns is described by a multivariate stochastic volatility model with leverage. The matrix exponential transformation is used to keep the time varying covariance matrices positive definite. The measurement equation of the multivariate realized measure is formulated as a matrix log-linear form. An efficient Bayesian estimation method using Markov chain Monte Carlo is discussed. The proposed model and estimation method are applied to stock return data.

C797: **Bayesian analysis of structure credit risk models with micro-structure noises and jump diffusion**
Presenter: **Sau Lung Chan**, The Chinese University of Hong Kong, China
Co-authors: Kwok Wah Ho, Hoi Ying Wong
There is empirical evidence that structural credit risk models significantly underestimate both the default probability and credit yield spreads. Two potential sources of the problems in traditional structural models are: 1. The Brownian model driving the firm asset value process fails to capture extreme events because of the normality assumption; and 2. The market micro-structure noise distorts the information contained in equity prices within estimation process. These problems are separately handled in the literature. A Bayesian approach is proposed to estimate jump-diffusion firm value process and micro-structure noise from equity prices based on structural credit risk models. As the firm asset value is not observed but the equity price does, the proposed Bayesian approach is useful in the estimation with hidden variable and Poisson shocks, and produces posterior distributions for financial analysis. The proposed approach enables us to check whether the bias of structural credit risk model is mainly caused by the firm value distribution or the micro-structure noise of the market. By filtering out micro-structure noises, the Bayesian approach facilitates the calculation of default probability and credit value adjustment. A simulation and empirical study will be conducted to ascertain the performance of our model.

C893: **Bayesian quantile regression for financial factor models**
Presenter: **Kenichiro McAlinn**, Keio University Graduate School of Economics, Japan
Co-authors: Teruo Nakatsuma
Factor models have been widely used in financial researchers and practitioners for more than four decades. However, factor models only estimate the factor loading in the mean function. Considering the rising interest in downside risk for financial risk management (Value-at-Risk, for example), there are very few researches that go beyond simply utilizing a normal error distribution as a proxy for its quantile, rather than estimating the factor loading in the quantile function. We investigate the factor loadings of different fundamental factors, including the Fama-French three factors and numerous macro factors, on industrial indexes in respect of their quantiles. In particular, we employ a Bayesian method with asymmetric Laplace distribution to examine how some factors that have negligible factor loadings in the mean can have significant factor loadings in the lower quantiles and vice versa.

CS17 Room 1 STATISTICAL SIGNAL PROCESSING APPLIED TO ASSET MANAGEMENT Chair: Serge Darolles

C395: **A survey of filtering techniques applied to hedge fund replication**
Presenter: **Guillaume Weisang**, Clark University, United States
A survey of filtering techniques applied to hedge fund replication is provided. The investor's problem for hedge fund replication in a dynamic setting is formulated. A rigorous explanation linking the investor's problem to Bayesian filtering is provided through a concise expose of optimal control theory and dynamic programming with a particular emphasis on the duality between optimal control of the tracking error and optimal estimation of the state. A review of the results of applying Kalman filter to hedge fund replication is provided as well as a review of minimax filtering and its comparison to Kalman filtering. Several additional topics and results are covered such as the replication of nonlinearities with particles filters and the treatment of option-like factors and their impact on the performance of the filters. To this aim, the ζ-statistic is introduced and its empirical use in detecting mis-specified replication models is covered. Finally, the problem of selecting the factors is considered with an emphasis on the Bayesian approach. Outstanding research questions and future directions for research are also outlined.

C411: **An agent-based model for microstructure noise and trade impacts.**
Presenter: **Emmanuel Bacry**, Ecole Polytechnique, CNRS, France
Co-authors: Sylvain Delattre, Marc Hoffmann, Jean-François Muzy
A new stochastic model is introduced which accounts for both the variations of asset prices at the tick-by-tick level and the trades arrival of anonymous agents. The construction is based on marked point processes and relies on mutually exciting stochastic intensities as introduced by Hawkes. By coupling suitably the stochastic intensities of upward jumps, downward jumps, trades arrival at best bid and trades arrival at best ask, we can simultaneously reproduce microstructure noise and statistics related to impact of trade arrivals on the price. The impact of a specific labeled agent is studied. The model accounts for the characteristic shape of the impact function of a meta-order: a concave impact during the execution of the meta-order (i.e., the so-called square-root law) followed by a relaxation period leading to either a permanent or a non-permanent impact. Non parametric estimation of the different kernel functions involved in the model are performed on real high-frequency data.

C498: **A regularized version of the Kalman filter for risk management and portfolio hedging**
Presenter: **Emmanuelle Jay**, QAMLab, France
Co-authors: Patrick Duvaut, Serge Darolles
The detection and the estimation performances of a *regularized Kalman filter* called *rg*KF(OD, E-l^q) which was previously derived by the authors is analyzed. This algorithm is based on the setup of a standard Kalman filter where a term is added in the observation equation to take into account the potential and sparse occurrence of exogenous outliers. The recursive equations of the standard Kalman filter need then to be regularized and *rg*KF(OD, E-l^q) proposes a simple but efficient two-step algorithm: an *Outlier Detection* step (OD) is first applied followed by a l^1 or a l^2 *Estimation* step (E-l^q) conditional to the detection of the outlier. *rg*KF offers an alternative solution to the very tricky problem of finding an accurate value of the regularization parameter. The OD-step of *rg*KF determines implicitely its value which has to be very low if an outlier is detected. We apply this methodology on several applications and we show, for example, that this algorithm allows to separate accurately the systemic risks from the idiosyncratic risks, either to analyze the management style of a fund manager or to hedge efficiently a portfolio on a daily basis.

C537: **A new perspective on dependence within financial markets**
Presenter: **David Matteson**, Cornell University, United States
Co-authors: Ruey Tsay
Financial instruments are inherently multivariate in their origination. A rigorous empirical valuation requires simultaneous analysis of many components. Financial analysis is further confounded with nonlinear and inter-temporal dependencies, extreme events, and vast datasets. Independent component analysis is a statistical signal processing method for finding suitable representations of the complex multivariate information within financial markets. We introduce a novel statistical framework for independent component analysis of multivariate data. Our methodology withstands all the fundamental challenges associated with financial analysis: robustness, requires only finite second moments; the curse of dimensionality, identifies non-redundant components that are as statistically independent as possible; predictability, a simultaneous test for all forms of inter-temporal multivariate dependence; sufficiency, statistics for testing the existence of mutually independent components; non-normality, complex distributions are projected onto independent subspaces. Finally, we illustrate the proposed framework on several important applications in asset management.

CS41 Room 2 ESTIMATING THE EFFECTS OF FISCAL POLICY Chair: Evi Pappa

C120: **Fiscal policy in open economy: Estimates for the euro area**
Presenter: **Massimiliano Pisani**, Bank of Italy, Italy
Co-authors: Lorenzo Forni
The aim is to estimate a dynamic general equilibrium model of an open economy to measure the effects of alternative expansionary fiscal shocks on the trade balance in the euro area. We do find that expansionary fiscal shocks -both those on the expenditure and revenue side- induce a joint deterioration of the public sector balance and trade balance. The output multipliers are always below one, while the deterioration of the trade balance and the trade leakages are relatively small. The largest value of trade deficit is equal to -0.1 percent of output and is obtained conditionally to a public consumption or a consumption tax shock. The deficit is generally driven by the increase in imports and decrease in exports, partially compensated by the improvement in the terms of trade. Only in the case of the expansionary labor tax shock, higher imports and the deterioration in the terms of trade are partially compensated by higher exports.The monetary policy stance and the type of fiscal financing do not greatly affect the size of trade balance deterioration, even if they magnify the impact of fiscal shocks on gross imports and international relative prices.

C352: **Mafia and public spending: Evidence on the fiscal multiplier from a quasi-experiment**
Presenter: **Saverio Simonelli**, University of Naples Federico II, Italy
Co-authors: Antonio Acconcia, Giancarlo Corsetti
A law issued to contrast political corruption and mafia infiltration of city councils in Italy has resulted in episodes of unanticipated, large, temporary contractions in public spending. Using these episodes and province-level data, we estimate the multiplier when spending cuts cannot be compensated by monetary expansions, holding the tax burden constant. In our IV-estimates, this multiplier is 1.2. The effects of lagged spending, under the assumption that this is exogenous to current output, bring it to 1.8. These results are consistent with the fiscal transmission in models of small open economies where constraints on monetary policy interact with financial imperfections.

C400: **Spillovers from US monetary and fiscal policy**
Presenter: **Ethan Ilzetzki**, London School of Economics, United Kingdom
Co-authors: Keyu Jin
The spillovers from macroecnomic policy shocks in the US on emerging economies are estimated. We find that shocks to the US Federal Funds rate has the exact opposite effect on emerging market economies than on other OECD economies. A monetary contraction in the US causes an increase in emerging market inustrial production and an appreciation in emerging markets' exchange rates. This is in contrast to the spillovers to other OECD countries, which show responses in the opposite direction. Results are robust to a number of commonly used identification methods. We show that these results are a puzzle in that they are difficult to reconcile with standard open-economy macroeconomic models. We also estimate the spillovers from US fiscal policy.

C582: **On the reaction of the Governmentto nature's attacks**
Presenter: **Evi Pappa**, EUI, UAB and CEPR, Italy
Co-authors: Rodica Calmuc, Yuliya Kulikova, Saverio Simonelli
Natural disasters are regularly observed phenomena both across time and across countries that have significant and often devastating effects. Besides their frequency and their consequences there is no readymade answer by economists to which fiscal instrument is the best to use for alleviating affected economies. We use US data on natural disasters and regional economic data to access the economic size of Natural disasters and to study whether and how fiscal policy can alleviate the affected regions. We use the FEMA assistance programs to distinguish between different types of assistance and using panel regressions techniques, we establish whether individuals assistance is preferable to public programs for decreasing the size of the recession originating from the natural forces.

CS97 Room 4 COINTEGRATION Chair: Robert Kunst

C390: **Variance ratio testing for fractional cointegration in presence of trends and trend breaks**
Presenter: **Andreas Dechert**, University of Wuerzburg, Germany
Modeling fractional cointegration relationships has become a major topic in applied time series analysis as it steps back from the traditional rigid $I(1)/I(0)$ methodology. Hence, the number of proposed tests and approaches has grown over the last decade. The aim is to study a nonparametric variance ratio approach previously for the case of fractional cointegration in presence of linear trend and trend breaks. The consideration of trend breaks is very important in order to avoid spurious fractional integration, so this possibility should be regarded by practitioners. The p-values are proposed to be computed by means of Γ-distributions. Response regressions parameters for their asymptotic moments are given. In Monte Carlo simulations the power of the approach is compared against a Johansen type rank test suggested, which is robust against trend breaks but not fractional (co-)integration. As the approach also obtains an estimator for the cointegration space, it is compared with OLS estimates in simulations. As an empirical example the validity of the market expectation hypothesis is tested for monthly Treasury bill rates ranging from 1958-2011, which might have a trend break around September 1979 due to change of American monetary policy.

C827: **A simple way to test for multiple cointegration with structural breaks**
Presenter: **Andrea Monticini**, Catholic University of Milan, Italy
Testing for cointegrating relations that economic theory would predict, and find that the null hypothesis of noncointegration is not rejected it is a very common situation in empirical analysis. To overcome this problem one method adopted in the literature has been to assume a switch in intercept and/or slope, and so run regressions containing dummy variables for each possible break point within a feasible interval of the sample (say, $[0.15T]$ to $[0.85T]$, where T is sample size) and tabulate the extrema of suitable test statistics under the null hypothesis. Since the correct modelling of data containing conjectured structural breaks is an elaborate and tricky undertaking, we explore a different approach. We compute a Johansen's statistics by choosing an appropriate subsample for the test. The main advantage is twofold. First, there is no need to estimate the model under the alternative, so that the type of break, and the number of breaks, do not need to be specified (it must be assumed that the dates of breaks, as well as their number and their type, are unknown). Second, there is the potential to detect breaks anywhere in the sample. For the case of a single break, at least, there is no limitation on the location of the break date. We provide the critical values and Monte Carlo simulations for power evaluation.

C890: **Cointegration pairs trading strategy on derivatives**
Presenter: **Lai Fan Pun**, The Chinese University of Hong Kong, China
Co-authors: Ngai Hang Chan, Pak Kuen Lee
The notion of cointegration has been widely used in finance and econometrics, in particular in constructing statistical arbitrage strategy in the stock market. The arbitrage trading strategy for derivatives based on cointegration is generalized to account for the volatility factor. Pairs of short dated at-the-money straddles of European options with positive net carry (i.e. theta) are considered, so as to capture the mean reversion property of the linear combinations of implied volatilities. A performance analysis is conducted based on historical data, which is expected to provide better returns than the pair-wise strategy based on the Hang Seng Index constituents. From empirical results, the portfolio based on this strategy makes profit, where Vega plays a major role. Furthermore, modeling and forecasting realized volatility are also considered as a supplement to the trading strategy. Empirical study for the renewed strategy is implemented using a 3-year historical data of currency options.

C1000: **Joint modeling of cointegration and mean reversion in a continuous time approach to statistical arbitrage**
Presenter: **Tomasz Dubiel-Teleszynski**, London School of Economics and European Central Bank, Germany
A model for statistical arbitrage is proposed. The formulation of the model proceeds in a linear continuous time system framework. An Ornstein-Uhlenbeck process is integrated into a triangular first order continuous time system with observable stochastic trend. An exact discrete analog is derived. The full system Gaussian estimates are obtained together with subsystem OLS estimates. Finite sample properties of the estimates are considered. A test based on a conditional error correction model formulated in discrete time is used to test for cointegration and mean reversion jointly. The asymptotic distribution of the underlying test statistic is identified and investigated via Monte Carlo simulations. A continuous time model for statistical arbitrage is then applied to pair trading and the efficient market hypothesis is tested on historical data from the Warsaw Stock Exchange. A trading strategy based on the estimates of volatility, cointegrating and mean reversion parameters of the continuous time model is implemented. Stopping times of the Ornstein-Uhlenbeck process form a simple trading rule. Cointegrated pairs are identified in-sample and

respective trading strategies turn profitable out-of-sample in the majority of cases, hence the examined market is shown to be inefficient over the span of historical data used herein.

CS98 Room 6 FINANCIAL TIME SERIES I Chair: John Galbraith

C464: **Influence of weather variability on the orange juice prices**
Presenter: **Florent Mc Issac**, Paris 1 Sorbonne University, France
Co-authors: Frank Laudicina, Marius-Cristian Frunza

The returns of the frozen concentrated orange juice (FCOJ) future prices and the dependency with the temperature variation in the city where the exchanged is based (New York) are explored. Intuitively, one could consider that when temperatures are higher (resp. lower) than the average seasonal temperature, the consumption of orange juice increase (resp. decrease). We do not search a weather index that will account for the real consumption weighted by each regions temperature and number of potential consumers, but most likely the impact thorough a behavioural effect. In fact it is well known that the liquidity on futures market is mainly driven by speculators that use anticipations in the underlying physical market in order to provide with exceeding risk adjusted return. Psychologically FCOJ is apprehended as a weather market and we prove that this apprehension implies a weather linked patter in FCOJ returns. Working on both the demand effect and behavioural influence of weather, we explore dependence between the return of the orange juice price and the variation in temperature. As an application we present an alpha providing investment strategy based on long/short positions on the front month FCOJ futures depending on the temperature anticipations.

C708: **The application of stochastic processes in exchange rate forecasting: Benchmark test for the EUR/USD and the USD/TRY**
Presenter: **Giray Gozgor**, Dogus University, Turkey

The aim is to investigate the short-time exchange rate predictability in a developed and in an emerging market. For this purpose we consider the Euro/United States Dollar (EUR/USD) and the United States Dollar/Turkish Lira (USD/TRY) exchange rates. We apply the benchmark test and compare the results of daily out-of-sample forecasting by Brownian Motion (BM), Geometric Brownian Motion (GBM), Ornstein-Uhlenbeck Mean-reversion (OUM), Jump Diffusion (JD) stochastic processes, Vector Autoregressive (VAR), Autoregressive Integrated Moving Average (ARIMA) models and Uncovered Interest Rate Parity (UCIP) against the Random Walk (RW). We conclude that none of these models or stochastic processes displays superiority over the RW model in forecasting the USD/TRY exchange rate. However, GBM, BM and OUM processes beat the RW model in forecasting the EUR/USD exchange rate. Furthermore, we show that these findings are robust and not time-specific. When we separately examine the pre-crisis and the post-crisis periods, results remain unchanged.

C1004: **Efficiency of stock markets: A GARCH application**
Presenter: **Sultan Harbi**, Swansea University, United States

Stock return and volatility have been modelled in the Saudi Stock Market using four GARCH models (GARCH, IGARCH, GJR-GARCH and APARCH) with four distributions (Normal, Student-t, Skewed student and GED). The data used is a daily data from 1994 to 2011 and divided in three sub-samples (1994-2001), (2002-2008), and (2009-2011). The results show that the SSM is symmetric over the period (1994-2001) and asymmetric in the rest of the sample. Moreover, the SSM is a weak-form inefficiency.

C845: **Forecasting Colombian spot exchange rate: A comparison of different machine learning and data mining approaches**
Presenter: **Julio Cesar Alonso**, Universidad ICESI, Colombia
Co-authors: Darwin Mercado Polo, Carlos Hernan Fajardo Toro

Forecasting the spot exchange rate has become one of the most challenging issues for practitioners. The difficulties that structural models have to forecast short run spot exchange rates have been previously discussed. This observation has propelled a vast literature that tries to improve the out-of-sample properties of econometric models. Our purpose is to apply different machine learning and data mining techniques as a basis to build a future hybrid expert system, where the results obtained from the application of the different data mining technique are the input for the machine learning techniques. The results and those obtained with other statistical techniques are compared. Specifically, different clustering techniques, as well as several neural networks topologies and methods for time series models are considered. We use traditional forecast metrics to compare the out-of-sample behavior of the new approach versus the traditional ones.

C940: **Partial symbolic transfer entropy**
Presenter: **Angeliki Papana**, University of Macedonia, Greece
Co-authors: Catherine Kyrtsou, Cees Diks, Dimitris Kugiumtzis

Partial Symbolic Transfer Entropy (PSTE) is an extension of Symbolic Transfer Entropy. PSTE accounts only for the direct causal effects of the components of a multivariate system. It is an information theoretic measure, which does not suffer from the shortcomings of a model mis-specification. PSTE is defined on the ranks of the components of the reconstructed vectors of the time series, instead of the reconstructed vectors. It is applicable to time series that are non-stationary in the level (mean), since slow drifts do not have a direct effect on the ranks. PSTE is evaluated on multivariate time series of known coupled systems. The statistical significance of PSTE is assessed by a randomization test making use of surrogate time series. As real application, the causal effects among economic time series are investigated. Three time series of returns are considered, all stationary at level, while one of them is variance non-stationary. For comparison reasons, Conditional Granger Causality Index (CGCI) is also estimated. Both measures are estimated for the time series of the returns, and the filtered time series of returns from VAR and GARCH models. PSTE seems to outperform CGCI giving consistent results before and after filtering the data.

CS03 Room 5 CONTRIBUTIONS TO FINANCIAL MARKETS AND MACRO Chair: Willi Semmler

C630: **Financial markets as critical networks and computational models for allocating assets**
Presenter: **Calin Valsan**, Bishops University, Canada

Using the insights from research in computational genetics, a model based on networked agents, acting on their own personal preferences, and emulating the behavior of other agents in their social proximity, is proposed. Each agent's individual preference p, is defined as the minimum payoff at which the agent is willing to hold the asset A. Individual preferences can be purely subjective, or the result of rational expectations. Each agent is seen as a network node that can only be in two states: 0 when not holding the asset; and 1 when holding the asset. Depending on the number of inputs received from other agents, the network can be sparse or dense. The resulting behavior of the system is nevertheless complex. Complex system display an array of important behavior characteristic, such as emergence, homeostasis, nesting, learning, and adaptation. Such an approach has important consequences for the ability to make market and economic predictions, to measure economic performance, and to optimize the allocation of resources through economic policies and managerial decision making.

C836: **Factors driving market's perception of sovereign credit risk in the Euro area during the financial crisis**
Presenter: **Gonzalo Camba-Mendez**, European Central Bank, Germany
Co-authors: Dobromil Serwa, Thomas Werner

Economic factors, political factors and investor's sentiment appear to have all played a role in explaining markets' perceptions of credit risk

associated with Euro area sovereign debt during the financial crisis. When disentangling the components of credit risk into the probability of defaulting and the recovery rate, empirical results suggest that recovery rates are less driven by economic fundamentals and are more closely associated with political factors. Evidence of the impact of sentiment is found in the form of contagion from developments in the perceptions of credit risk in one country to other countries. Results presented suggest that the nature of contagion was not as unidirectional and simple as usually communicated by the financial media, namely from Greece to the rest. Contagion is found to match fairly closely the patterns of cross-country exposures to debt holdings across Euro area countries. Evidence is also found in support of the view that the collapse of Lehman Brothers, while not directly linked to Euro area country risk, shifted markets perception of sovereign creditworthiness in the Euro area and increased market expectations on potential losses of sovereign defaults well before the start of the discussion on private sector involvement in sovereign debt restructuring.

C1062: **Technological progress and financial stability**

Presenter: **Daryna Grechyna**, University of Auckland, New Zealand

The aim is to examine the sources of the small shocks that drive periodic financial crises. We explore the role of firms heterogeneity in building up the riskiness of economic system. The analysis is based on the conjecture that production possibilities in each of the sectors are limited by the capacity limit, defined by the state of the sector's technology. The firms differ by their probability of successful production and are financed by the loans from the banks. The banks have limited liability and therefore finance risky investments. This allows less efficient firms to enter the market. Economic expansions are characterized by growing number of risky firms, and increasing risk-taking by the banking sector. With riskier firms entering the financial markets and production process the probability of bank runs and consequent financial crisis increases.

C749: **Macroeconomic forecasting using financial volatility: A MIDAS appraisal during the Great Recession**

Presenter: **Clement Marsilli**, Banque de France, France

Co-authors: Laurent Ferrara, Juan-Pablo Ortega

The global economic recession endured by the main industrialized countries during the period 2008-09 in the wake of the financial and banking crisis, referred to as the Great Recession, has pointed out the importance of the financial sector. In this respect, many researchers have started to reconsider the linkages between financial and macroeconomic areas. We evaluate the leading role of the daily volatility of some major financial variables like commodity and stock prices or interest rate spread in their ability to anticipate the US Gross Domestic Product growth. For this purpose, we implement a MIDAS-based modeling approach that enables us to forecast quarterly growth rate using exogenous variables sampled at higher frequencies. Empirical results show that mixing daily and monthly variables can be useful at the time of predicting GDP growth during the period going from 2007q1 to 2010q4.

C558: **EU severe debt crisis: Strengthened links between interest rates and stock market returns**

Presenter: **Nuno Ferreira**, ISCTE - IUL, Portugal

Co-authors: Rui Menezes, Sonia Bentes

The crisis in the periphery of the Eurozone which deeply emerged the European financial markets is typically one of a long list of sovereign debt crises during the last three decades. Generally, the debt problems have similar characteristics with those of the global periphery. Nowadays, a relevant question is to know how far an unstable economy can be influenced by an endogenous business cycle. The focus of the paper is to analyse the effect impact of the European Markets contamination by sovereign credit (particularly in Portugal, Spain, France and Ireland). The more recent models learn over time to make the necessary adjustments to new levels of peaks or troughs, which allows turning points more predictable comparatively to linear models. The Smooth Regression Model could be seen as a linear and a nonlinear component model which determines over time whether there are only one or both components - an alternative approach to the analysis of the historical behaviour adjustment between interest rates and stock markets indexes. From the results were found evidence in the crisis regime, i.e., large negative returns. Particularly in the case of Portugal which registered the biggest nonlinear threshold adjustment between interest rates and stock market returns.

Monday 03.12.2012 10:40 - 12:20 Parallel Session N – ERCIM

ESI03 Room Auditorium ADVANCES IN NONPARAMETRIC AND ROBUST MULTIVARIATE METHODS Chair: Anthony C. Atkinson

E256: **Weighted-mean regions and depth statistics**
Presenter: **Karl Mosler**, Universitaet Koeln, Germany
Weighted-mean regions are the level sets of a class of depth functions, the weighted mean (WM) depth functions. Given a probability distribution in Euclidean d-space, they form a family of set-valued parameters that reflect the distribution's location, dispersion and shape. Based on a sample they provide set-valued statistics. The main properties of WM regions and their associated depth statistics are discussed. Special cases include the zonoid regions and the ECH (expected convex hull) regions. The WM regions of a sample are random convex polytopes in $\mathbb{R}^d$. A law of large numbers applies. The polytopes can be computed for any dimension d by exact and approximate algorithms. Recent applications to linear programming under stochastic constraints are given; as well as to set-valued measures of multivariate risk and economic inequality.

E435: **Robustness and projective shape analysis**
Presenter: **John Kent**, University of Leeds, United Kingdom
Ideas from two branches of statistics are combined. The first is Tyler's robust estimator of a scatter matrix for multivariate data, equivalent to the maximum likelihood estimate for the angular central Gaussian distribution. The second is the theory of projective shape analysis, which is concerned with camera images and the information in a set of landmarks that is invariant under changes of the camera view. It turns out that Tyler's estimate is exactly the tool needed to create a standardized version of projective shape, which in turn facilitates the statistical comparison of different projective shapes. The simplest example is the cross ratio for four points on a line.

E1041: **Convex optimization, shape constraints, compound decisions, and empirical Bayes rules**
Presenter: **Ivan Mizera**, University of Alberta, Canada
Estimation of mixture densities for the classical Gaussian compound decision problem and their associated (empirical) Bayes rules is considered from two new perspectives. The first introduces a nonparametric maximum likelihood estimator of the mixture density subject to a monotonicity constraint on the resulting Bayes rule. The second proposes a new approach to computing the Kiefer-Wolfowitz nonparametric maximum likelihood estimator for mixtures. In contrast to prior EM-algorithm methods for the latter problem our new approach is based on a reformulation of the Kiefer-Wolfowitz estimator as a convex optimization problem that can be efficiently solved by modern interior point methods. The resulting reduction in computational effort, several orders of magnitude for typical problems, greatly expands the practical applicability of the Kiefer-Wolfowitz estimator. Our new procedures are compared with several existing empirical Bayes methods in simulations; some further comparisons are made based on prediction of baseball batting averages.

ES03 Room 11 IMPRECISION IN STATISTICAL DATA ANALYSIS Chair: Ana B. Ramos-Guajardo

E312: **Forecasting insurance ratings under uncertainty. A multi criteria approach based on fuzzy and linguistic modelling**
Presenter: **Raquel Florez-Lopez**, University Pablo Olavide of Seville, Spain
Co-authors: Juan M. Ramon-Jeronimo
In last decade, regulations as Basel II and Solvency II have highlighted the significance of external rating systems to evaluate the solvency of financial institutions as banks and insurers. While some information has been provided on numerical variables used as rating determinants, the classification model that governs risk assessments is deeply vague, imprecise, and based on human judgments. So the early prediction of financial ratings becomes a difficult task, which prevents implementing pro-active risk and severely impacts on firm's profitability and market value. While the use of fuzzy-rule based systems (FRBS) has proved to produce successful results when modelling such problems with non-probabilistic uncertainty, a good balance between its interpretability and accuracy is difficult to obtain. Two main areas have been distinguished in modelling with FRBS: fuzzy modelling and linguistic modelling. Fuzzy modelling closely approximates imprecise data while providing poor understood rules; linguistic modelling provides clear, well-interpretable rules even if the accuracy is penalised. In this paper the advantages of both areas are compared and exploited to forecast European insurance ratings along time, departing from numerical and linguistic determinants. Alternative models are developed and improved using tuning processes based on linguistic modifiers, Simulated Annealing, and rule cooperation. Models are compared in terms of reliability, validity, and interpretability using an out-of-sample strategy; 632+ bootstrapping confidence intervals are provided for statistical inference. Results show that hybrid approaches that combine fuzzy modelling with linguistic modifiers provide simple, good-balanced rating forecasts in terms of accuracy and interpretability.

E868: **Fuzzy volatility elicitation through information merging**
Presenter: **Gianna Figa' Talamanca**, University of Perugia, Italy
Co-authors: Andrea Capotorti
A development of a previously introduced methodology for fuzzy volatility membership elicitation is proposed. The elicitation is based on the interpretation of the membership itself as a coherent conditional probability assessment. Previous numerical examples in the Black and Scholes framework, supported the feasibility of the elicitation method. However, there are cases in which two or more memberships naturally stem. This happens when there are multiple most plausible scenarios and a proper order relation among them should be applied to have a unique answer. A similar problem arises when several sources of information about volatility are taken into account. The merging of memberships stemming from different sources is addressed through specific operators. Among these, an interesting approach is to consider the volatility as a fuzzy random quantity whose fuzzy realizations are weighted through a suitable fuzzy mean. Consequences on the computation of the fuzzy Black and Scholes option price are investigated.

E918: **On the relation between probabilistic fuzzy systems and Markov models**
Presenter: **Rui Jorge Almeida**, Erasmus University Rotterdam, Netherlands
Co-authors: Nick Verbeek, Nalan Basturk, Uzay Kaymak
This study investigates the probabilistic fuzzy system (PFS) and its relation to discrete state space Markov chain models. A probabilistic fuzzy system deals simultaneously with linguistic uncertainty as well as probabilistic uncertainty. It consists of a set of rules whose antecedents are fuzzy conditions, the consequents are probability distributions and a stochastic mapping between the antecedent and the consequents. The rules specify a probability distribution over a collection of fuzzy sets that partition the output domain and also express linguistic information. A Markov chain model can be defined through a set of realized states and a set of transition probabilities that depend on the past state realizations. We show that a PFS can be obtained by extending Markov chain models to accommodate fuzzy states. Special attention is given to PFS model's stability and the required conditions for stationarity of the input-output process in relation to the ergodicity of Markov chains.

E939: **Estimating the parameters of a interval-valued multiple regression model**
Presenter: **Marta Garcia-Barzana**, University of Oviedo, Spain
Co-authors: Ana Colubi, Erricos John Kontoghiorghes
A multiple linear regression model for interval data is introduced. The least squares estimation is derived from solving a quadratic optimization problem with inequality constraints. That guarantees the existence of the residuals. However, transforming the quadratic programming problem into a cone projection one (as proposed in previous works), the inequality constraints will become non-negativity constraints, simplifying the resolution of the problem. An algorithm to estimate the regression coefficients based on the QR decomposition is developed. The performance of the approach will be illustrated by means of simulations and a real-life example.

E567: **Simulation study to compare Likert, fuzzy converted Likert and fuzzy rating scales**
Presenter: **Sara de la Rosa de Saa**, Universidad de Oviedo, Spain
Co-authors: Maria Angeles Gil, Gil Gonzalez-Rodriguez, Maria Teresa Lopez, Maria Asuncion Lubiano
Comparisons have been previously made between the fuzzy rating scale (a fuzzy assessment in accordance with a free response format) and either the Likert scales (identified with integer numbers) or fuzzy conversions of the most usual Likert labels. A furhter triple empirical comparison of the scales is now considered. This comparison is based on: i) the representativeness of the population/sample by means of its mean value (quantified in terms of the mean squared error using a standardized metric, which has been introduced to avoid the possible effect of the order of magnitude of the scales), ii) and simulations of trapezoidal fuzzy numbers for which the characterizing 4-tuples have been generated from certain reasonable distribution models inspired and validated by several real-life examples. Once the fuzzy rating scale-based values are simulated, a 'Likertization' criterion (also based on the same standardized metric) as well as some fuzzy conversions are applied. The empirical results have been quite conclusive: in most of the cases (around 100% if simulated samples are large enough) the mean of the valuations is more representative when the fuzzy rating is used than when either the Likert scale or its fuzzy conversion is considered.

ES13 Room 9 RESAMPLING PROCEDURES FOR DEPENDENT DATA Chair: Andres M. Alonso

E254: **A goodness-of-fit test for the functional linear model with scalar response**
Presenter: **Wenceslao Gonzalez-Manteiga**, Santiago de Compostela University, Spain
Co-authors: Eduardo Garcia-Portugues, Manuel Febrero-Bande
A goodness-of-fit test for the null hypothesis of a functional linear model with scalar response is proposed. The test is based on a generalization to the functional framework of a previous one, designed for the goodness-of-fit of regression models with multivariate covariates using random projections. The test statistic is easy to compute using geometrical and matrix arguments, and simple to calibrate in its distribution by a wild bootstrap on the residuals. The finite sample properties of the test are illustrated by a simulation study for several types of basis and under different alternatives. Finally, the test is applied to two datasets for checking the assumption of the functional linear model and a graphical tool is introduced.

E307: **Simple and powerful bootstrap tests for conditional moment restrictions**
Presenter: **Juan Carlos Escanciano**, Indiana University, United States
A class of tests for conditional moment restrictions of stationary time series is proposed. The tests are consistent against all alternatives not collinear to model's scores, allow for non-smooth moments and are implemented with a wild-bootstrap-type procedure. An important contribution is that we extend the scope of the wild-bootstrap and related methods to specification tests of non-smooth and non-separable conditional moment restrictions, including quantile regressions, for which wild-bootstrap methods were not available. A distinguishing feature of the new tests is their simplicity. They allow for any root-n consistent estimator that does not need to be re-estimated in each bootstrap replication, and are simple quadratic forms in the residuals, making the proposed procedure a powerful and simple tool for practitioners. A Monte Carlo experiment shows that the new method presents more accurate size and higher power than subsampling procedures. In an empirical application we study the dynamics in mean and variance of the Hong Kong stock market index and we evaluate models for the Value-at-Risk of the S&P 500. These applications highlight the merits of our approach and show that the new methods have higher power than popular backtesting methods.

E315: **Estimation risk in portfolio optimization**
Presenter: **Alberto Martin-Utrera**, Universidad Carlos III de Madrid, Spain
Co-authors: Victor DeMiguel, Francisco J. Nogales
Portfolio optimization is an area widely studied within the financial and econometrics literature. It is well known that using sample estimators to construct optimal investment strategies provide portfolios that perform very poorly out-of-sample. In fact, it is know that the benefits from an optimal diversified portfolio do not offset the loss from estimation error. We study different approaches to mitigate the effects of estimation error within optimal investment strategies. We provide analytical and empirical results from both parametric and nonparametric approaches.

E333: **The estimation of prediction error in functional settings through bootstrap methods**
Presenter: **Pedro Galeano**, Universidad Carlos III de Madrid, Spain
Problems in which a functional or real variable Y is predicted by a functional variable X through a model $m(\cdot)$ are considered. The interest is in to evaluate the prediction capability of the model $m(\cdot)$, estimated from a training sample T, on independent test data. Assessment of this capability is important as gives us a measure of the quality of a given model. In particular, we focus on estimation of the prediction error, given by $E\left[E\left[L(Y,\widehat{m}(X))\,|T\right]\right]$, where $L(Y,\widehat{m}(X))$ is a loss function for measuring errors between Y and $\widehat{m}(X)$. We describe a number of bootstrap methods for estimating the prediction error in both situations in which the response variable Y is real or functional. Several models $m(\cdot)$, estimation methods and loss functions will be covered.

ES22 Room 7 NONPARAMETRIC AND ROBUST STATISTICS Chair: Andreas Christmann

E198: **Reducing the mean squared error of quantile-based estimators by smoothing**
Presenter: **Dina Vanpaemel**, KU Leuven, Belgium
Co-authors: Mia Hubert, Irene Gijbels
Many univariate robust estimators are based on quantiles. It is known that smoothing the empirical distribution function with an appropriate kernel and bandwidth can reduce the variance and mean squared error (MSE) of some quantile-based estimators in small data sets. This idea can be applied on several robust estimators of location, scale and skewness. The procedure involves selecting the bandwidth robustly and reducing the bias. A simulation study shows that the use of this smoothing method indeed leads to smaller MSEs, also at contaminated data sets. In particular, better performances for the medcouple are obtained, which is a robust measure of skewness that can be used for outlier detection in skewed distributions. The smoothed medcouple can be used to adapt the adjusted boxplot in oder to detect outliers in small data sets of which the underlying distribution is skewed. This will be illustrated on a real data example.

E361: **General M−estimators in nonlinear models with responses missing at random**
Presenter: **Ana Maria Bianco**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Paula Mercedes Spano
Linear models are one of the most popular models in Statistics. However, in many situations the nature of the phenomenon is intrinsically nonlinear and so, the data must be fitted using a nonlinear model. Besides, in some occasions the responses are incomplete and they are missing at random. It is well known that, in this setting, the classical estimator of the regression parameters, which is based on least squares, is very sensitive to outliers. A family of simplified general $M-$estimators is proposed to estimate the regression parameter in a nonlinear model with responses missing at random. Under mild conditions, these estimators are Fisher–consistent, consistent and asymptotically normal. To study local robustness, their influence function is derived. Monte Carlo simulations illustrate the finite sample behavior of the proposed estimators in different settings in both contaminated and uncontaminated samples.

E485: **Robust estimators for additive models with missing responses**
Presenter: **Matias Salibian-Barrera**, The University of British Columbia, Canada
Co-authors: Graciela Boente, Alejandra Martinez
Additive models are widely used to avoid the difficulty of estimating regression functions of several covariates without using a parametric model (this problem is known as the curse of dimensionality). Different estimation procedures for these methods have been proposed in the literature, and some of them have also been extended to the case of data sets with missing responses. It is easy to see that most of these estimators can be unduly affected by a small proportion of atypical observations, and thus we are interested in obtaining robust alternatives. We consider robust estimators for additive models with missing responses based on local kernel M-estimators, and we also study a robust approach using marginal integration. If time permits, we will introduce a robust kernel estimator for additive models via the back-fitting algorithm.

E917: **A new resampling method: nonsingular subsampling for mixed-type data**
Presenter: **Manuel Koller**, ETH Zurich, Switzerland
Simple random subsampling is an integral part of many algorithms such as bootstrap and algorithms for computing S-estimators. These algorithms usually require subsamples to be nonsingular. A simple and popular strategy to avoid singular subsamples is to simply discard them and to generate a new subsample. This makes simple random subsampling slow, especially if some levels of categorical variables have low frequency, and renders the algorithms infeasible for such problems. This talk presents an improved subsampling algorithm that only generates nonsingular subsamples. We call it nonsingular subsampling. Using nonsingular subsampling to compute S-estimates of linear regression for data with continuous variables is as fast as using simple random subsampling but much faster for categorical or mixed-type data. This is achieved by using a modified LU decomposition algorithm that combines the generation of a sample and the solving of the least squares problem.

ES69 Room 13 MIXTURE MODELS WITH SKEW-NORMAL OR SKEW T-COMPONENT DISTRIBUTIONS Chair: Geoff McLachlan

E652: **Model-based clustering and classification via mixtures of shifted asymmetric Laplace (SAL) distributions**
Presenter: **Brian Franczak**, University of Guelph, Canada
Co-authors: Ryan Browne, Paul McNicholas
A family of shifted asymmetric Laplace (SAL) distributions called the ParSAL family is introduced and used for model-based clustering and classification. This family of models arises through an eigen-decomposition of the component covariance matrices and includes a skewness parameter. An EM algorithm is developed for parameter estimation by exploiting the relationship with the general inverse Gaussian distribution and a novel technique for dealing with the issue of infinite likelihood is presented. The ParSAL family is applied to both simulated and real data to illustrate clustering and classification applications. In these analyses, our family of mixture models are compared to the popular Gaussian approaches. Suggestions for future work are discussed.

E253: **Maximum likelihood estimation in mixtures of skew Student-t-normal distributions and applications**
Presenter: **Tsung-I Lin**, National Chung Hsing University, Taiwan
Co-authors: Hsiu J. Ho, Saumyadipta Pyne
The problem of maximum likelihood estimation for a mixture of skew Student-t-normal distributions is considered. That is a novel model-based tool for clustering heterogeneous (multiple groups) data in the presence of skewed and heavy-tailed observations. We present two analytically simple EM-type algorithms for iteratively computing the maximum likelihood estimates. The observed information matrix is derived for obtaining the asymptotic standard errors of parameter estimates. A small simulation study is conducted to demonstrate the superiority of the skew Student-t-normal distribution compared to the skew t distribution. To determine an optimal number of components for the proposed mixture models, a greedy EM algorithm was applied. We illustrate the approach with a real data sets obtained by flow cytometry.

E445: **Linear regression model with measurement error using finite mixtures of skew-t distributions**
Presenter: **Victor Hugo Lachos Davila**, UNICAMP, Brazil
Co-authors: Celso Romulo Barbosa Cabral, Camila Borelli Zeller
In regression models, the classical normal assumption for the distribution of the random observational errors is often violated, masking some important features of the variability present in the data. Some practical actions to solve the problem, like the transformation of variables to achieve normality, are often kept in doubt. We present a proposal to deal with this issue in the context of the simple linear regression model when both the response and the explanatory variable are observed with error. In such models, the experimenter observes a surrogate variable instead of the covariate of interest. We extend the classical normal model by modeling jointly the unobserved covariate and the random errors by a finite mixture of a skewed version of the Student t distribution. This approach allows us to model data with great flexibility, accommodating skewness, heavy tails and multimodality. We develop a simple EM-type algorithm to proceed maximum likelihood inference of the parameters of the proposed model, and compare the efficiency of our method with some competitors through the analysis of some artificial and real data.

E259: **On some recent results for mixtures of skew normal and skew t-distributions**
Presenter: **Geoffrey McLachlan**, University of Queensland, Australia
Some recent results on applications of mixture models with component distritubutopns that are either skew normal or skew t are presented. We consider a number of versions of these component distributions and discuss the relationships between them. Particular emphasis is placed on comparing mixture models with a restricted version of the skew t-distribution and an unrestricted version. Some examples are given using datasets from flow cytometry.

ES43 Room 8 COMPOSITIONAL DATA ANALYSIS: ADVANCES AND APPLICATIONS Chair: J. Antoni Martin-Fernandez

E323: **Logratio versus traditional methods in statistical analysis of compositional data**
Presenter: **Josep Antoni Martin-Fernandez**, University of Girona, Spain

Broadly speaking, compositional data are multivariate vectors that contain only relative information, i.e. vectors whose variables keep the importance of the parts of a whole. Typical examples are: data from time use surveys (amount of time people spend doing various activities) and data from households budged surveys (households' expenditure on goods and services). In 1879 Karl Pearson stated that some kind of spurious correlation may arise between variables that measures relative information. More than a century later, researchers from many different fields still apply traditional methods to analyse compositional data. Typical methods do not verify basic principles of compositional data analysis: scale invariance and subcompositional coherence. Common statistical techniques as principal components, cluster, and discriminant analysis can provide misleading results and hence suggest erroneous conclusions. These difficulties appear because most of underlying elements in the methods are based on the traditional Euclidean geometry (e.g., arithmetical mean). Instead, these methods become appropriate when are applied to log-ratio coordinates. In others words, they are adequate when one assumes that the simplex, sample space of compositional data, is governed by the Aitchison geometry. This geometry has been developed during the last decade from the approaches introduced by John Aitchison in the 80's.

E162: **A unified approach to classical and robust regression for compositional data**
Presenter: **Karel Hron**, Palacky University, Czech Republic
Co-authors: Peter Filzmoser, Matthias Templ, Gerald van den Boogaart, Raimon Tolosana-Delgado

Regression analysis gives answers to one of the basic problems in statistics, namely, how the response variables are related to the explanatory variables. The methodology for regression analysis with covariates and responses that carry absolute information, i.e. an interval scale based on the Euclidean geometry in real space, is well known. However, many data sets from geochemistry, economics and many other sources are often of compositional nature. Compositional data are multivariate observations carrying relative information on the importance or weight of a set of parts in a total (represented usually as proportions or percentages). Hereby, the standard regression model assumptions fail, because compositional data induce their specific scale and geometry, the log-ratio Aitchison geometry. To perform regression analysis adequately, the compositions have to be expressed in some coordinates (with respect to this Aitchison geometry) and the regression analysis must be built on these coordinates. Finally, one must interpret regression results attending to the natural restrictions of these coordinate representations and to the compositional scale of the analysed data. A concise approach to regression analysis will be presented, where either response variables, covariates or both of them are of compositional nature. We also face the problem with outlying observations that might depreciate the obtained results. Robust alternatives to the classical estimators are therefore presented.

E321: **Simplicial differential equations for applications in some economic problems**
Presenter: **Eusebi Jarauta-Bragulat**, UPC, Spain
Co-authors: Juan Jose Egozcue

Shares of a market or a portfolio, proportions of populations, distribution of GDP per capita by countries are compositional data and can be represented by a vector in the simplex. Frequently, these proportions change in time and modelling their evolution is a primary goal. There is a lack of models in which the proportions are treated jointly and satisfying the principles of compositional data: scale invariance and subcompositional coherence. Scale invariance require analyses to be invariant under change of units. Subcompositional coherence demands that ratios between parts do not change when less number of parts is analysed. First order simplicial linear differential equations satisfy the mentioned requirements and provide flexible enough models for low frequency evolutions. These models can be fitted to data using least squares techniques on coordinates of the simplex. The matrix of the differential equation is interpretable, thus providing a powerful analytical tool. The evolution of the Spanish population is used an illustrative example. Population is divided into 4 classes (roughly speaking, children, employees, non-employees, retired) and its evolution in the period 1976-2009 is analysed. The fitted model reveals a remarkable instability even using only years previous to the present crisis.

E318: **From Herfindahl and Atkinson indices to a compositional concentration index**
Presenter: **Juan Jose Egozcue**, Technical University of Catalonia, Spain
Co-authors: Vera Pawlowsky-Glahn

Concentration indices are in use in Biology and Demography, economical and social sciences for more than a half century. Concentration indices, and also information measures, are normally related to diversity indices by an inverse relationship. Well known examples of concentration/diversity indices are those of Gini, Simpson, Herfindahl, Hirshman, and Shannon. They are closely related and can be interpreted as expected values of proportions. Shannon information and the Atkinson concentration index are based on a set of principles which describe their main properties. Concentration is defined on a vector of proportions assigned to a number of categories -companies, countries, population groups, market shares- and can be viewed as a composition in the simplex. The principles of compositional data analysis can be used to define a new concentration index in a form similar to that of the Atkinson index. The simplex has its own Euclidean structure, known as the Aitchison geometry, which allows the representation of compositions by orthonormal coordinates and provides standard concepts as distance and norm. The proposed index is the norm of a composition made of the proportions assigned to categories.

ES09 Room 12 STATISTICAL ALGORITHMS AND SOFTWARE Chair: Cristian Gatu

E430: **Software and computational methods for two-sex branching models in random environment**
Presenter: **Manuel Molina**, University of Extremadura, Spain
Co-authors: Manuel Mota, Alfonso Ramos

Branching process theory has become an active research area of interest and applicability. In particular, it plays a major role in modeling general population dynamics. We are especially interested in the development of stochastic models to describe the demographic dynamics of populations with sexual reproduction. We consider the class of two-sex branching models in random environment with respect to the number of progenitor couples and/or the mating function. We provide some inferential results about the main parameters and we present the software in R that we have developed for this study. Also, in order to address some inferential problem related to this class of branching models, we propose approximate Bayesian computation methods. As illustration, we include an application in population dynamics.

E691: **A programming to calculate a Groebner basis with R**
Presenter: **Toshio Sakata**, Kyushu University, Japan
Co-authors: Rieko Sakurai, Manabu Iwasa, Toshio Sumi, Mitsuhiro Miyazaki

The Groebner basis theory is now widely used in algebraic statistics. For example, it is used for the Markov basis theory on the high-dimensional contingency tables and the experimental designs. Recently, the schools of Takemura and Takayama developed "a holonomic gradient descent method", which can calculate a maximum of the likelihood by solving a related Pfaffian differential equation. This method is a quite powerful and now producing many new results for the problems which have been considered intractable yet. The method uses the Groebner basis on the differential ring with rational polynomial coefficients. In reality, there are several algebraic softwares for the calculation of Groebner basis, however,

considering the importance of the methods for the statistical world, we consider that there is an urgent need for a package of R for Groebner basis calculation. In fact, it is some disadvantage to calculate Groebner bases by using such algebraic softwares and then return to the R to make some inference for a statistical model. It is shown that a programming of R for calculating such algebraic entities are easy more than expected. We deal both with the polynomial ring and differential ring with the coefficient field of rational polynomials.

E973: **Modelling dynamical systems with a DSL**
Presenter: **Erik Mathiesen**, Unaffiliated, United Kingdom
Being able to quickly prototype models describing dynamical systems is a common need in many fields. We present a domain specific language for describing and simulating both stochastic and deterministic systems modelled by differential equations. The language enables us to easily explore, prototype and select models for a given system. Besides straightforward trial and error, the DSL also allows us to explore the model space by the use of various optimisation techniques. The theory behind the approach will be described and examples illustrating its practical usage will be discussed.

E1056: **survPresmooth: An R package for presmoothed estimation in survival analysis**
Presenter: **Maria Amalia Jacome Pumar**, Universidade da Coruna, Spain
Co-authors: Ignacio Lopez de Ullibarri
The survPresmooth package for R implements the nonparametric estimators of the main functions studied in survival analysis (survival, density, hazard and cumulative hazard functions) using, besides the classical estimators, presmoothing techniques. The presmoothed estimators are computed by giving mass to all the data, including the censored observations, so more information on the local behavior of the lifetime distribution is provided. Also, using an auxiliar function, the available information can be extrapolated to better describe the tail behavior. The beneficial effect of presmoothing depends on the choice of the presmoothing bandwidth b_1. When the optimal bandwidth is used, the presmoothed estimators have smaller asymptotic variance and, therefore, a better Mean Squared Error performance. The simulation studies confirm this gain in efficiency under small and moderate sample sizes. Moreover, they also show that the presmoothed estimators are better than the classical ones, not only for the optimal value of the bandwidth but for quite wide ranges of values of b_1. The survPresmooth package provides plug-in and bootstrap bandwidth selectors, also allowing the possibility of using fixed bandwidths.

E653: **Optimization modeling using R**
Presenter: **Ronald Hochreiter**, WU Vienna University of Economics and Business, Austria
Simplifying the task of modeling optimization problems is important. Many commercial products have been created to support the optimization modeling process, but none of these products has been adopted by a significantly large number of users. As soon as real-world decision problems under uncertainty have to be modeled, flexible and quick changes to the underlying model are necessary. Simplifications are crucial to implement such optimization models into business processes successfully. Furthermore, the learning overhead for users should be minimized. We outline an approach on how to simplify optimization modeling using R and external optimization modeling languages as well as by building model generators for specific application problems. Examples from the areas of Finance and Energy will substantiate the applicability of the chosen approach.

ES60 Room 10 STATISTICAL METHODS IN GENOMICS AND PROTEOMICS Chair: Ruben Zamar

E743: **Calculating the false discovery rate in MS/MS based workflows**
Presenter: **Markus Mueller**, Swiss Institute of Bioinformatics, Switzerland
MS/MS peptide fragmentation spectra are measured to identify peptides from a biological sample in large repositories such as sequence databases or spectrum libraries. In order to increase the identification rate we use workflows that combine sequence and spectrum library searches. The spectrum library searches are performed in a open modification mode in order to find modified variants of the spectrum library peptides.The confidence in these identifications is usually expressed by the false discovery rate (FDR), which can be estimated using decoy repositories. We outline several issues that arise when estimating the FDR. First the FDR is sensitive to the tail structure of the decoy score distribution and therefore depends on the way the decoy database was created and the MS/MS search was performed. Secondly, the FDR has to be estimated differently on the spectrum, peptide, and protein level. Finally, when combining results from several search engines or datasets false matches may accumulate and the overall FDR increases and has to be corrected.

E813: **Robust instrumental variables estimators in proteomics**
Presenter: **Gabriela Cohen Freue**, University of British Columbia, Canada
Co-authors: Ruben Zamar, Bruce McManus
Recent advances in genomic and proteomic technologies have stimulated a large number of molecular biomarker discovery studies of various diseases. The number and the quality of the technical resources available for these biomarker studies are well recognized. However, the development of tailored statistical methods has lagged behind, dramatically reducing the pace, quality and precision of biomarker studies. This talk will be focused on the problem of measurement errors in mass spectrometry proteomic quantitation, which may affect the identification of protein biomarkers in a discovery study. As protein levels are regulated in part by gene expression, related genomic data can be integrated to address this problem through the implementation of instrumental variables estimators. Instrumental variables estimators are designed to provide unbiased and consistent regression parameter estimates using additional information provided by the instrumental variables (e.g., genes). We propose a new robust instrumental variables (RIV) estimator that is highly resistant to outliers and has attractive theoretical properties. The proposed methodology exploits, in an intuitive way, the plausible mechanisms from existing biological knowledge that relate genes, proteins, and diseases. We use RIV to identify human plasma proteomic biomarkers of cardiac allograft vasculopathy using related gene expression data as instruments.

E822: **Model-based gene set enrichment analysis**
Presenter: **Julien Gagneur**, LMU Munich University, Germany
Co-authors: Sebastian Bauer, Peter Robinson
In genomics, data-driven analysis often involves a search for biological categories that are enriched for the responder genes identified by the experiments. Single-category enrichment analysis procedures such as Fisher's exact test tend to flag large numbers of redundant categories as significant, which can make interpretations arduous and subject to arbitrary choices. I will present Model-based Gene Set Analysis (MGSA) in which we tackle the problem by turning the question differently. Instead of searching for all significantly enriched categories, we search for a minimal set of groups that can explain the data. Probabilistic inference based on the Metropolis-Hastings algorithm is used to identify the active categories. The model takes category overlap into account and avoids the need for multiple testing corrections met in single-category enrichment analysis. Simulations results as well as applications on yeast and the HeLa cell line demonstrate that MGSA provides high-level, summarized views of core biological processes and can correctly eliminate confounding associations. An open source R implementation, mgsa, is freely available as part of Bioconductor.

E1037: **Web Science 2.0 - in silico research in the Web, from hypothesis to publication**

Presenter: **Mark Wilkinson**, Universidad Politecnica de Madrid, Spain

The term Singularity - traditionally used to describe the merger of human and machine - was recently used to describe the state in which, the moment a new discovery is made, the knowledge embodied in that discovery is instantly disseminated, influencing other's research immediately and transparently. This clearly requires that research become tightly integrated into the Web, however the technologies necessary to achieve this kind of "Web Science" do not yet exist. Novel approaches to in silico investigation recently published by our laboratory demonstrate that we are enticingly close, and we believe that the path towards Web Science is now becoming clear. Web Science will not only simplify in silico scientific exploration for individual bench researchers, but in a broader sense, will facilitate scientific discourse by enabling researchers to easily see their data through another's eyes, explicitly compare disparate hypotheses to precisely identify differences in opinion, automatically evaluate those hypotheses over novel data-sets to investigate their validity, and integrate the resulting knowledge directly into the community knowledge-pool. Finally, it will enhance the rigor with which experiments - in particular high-throughput experiments - are executed, improving on the current state of often sparsely-documented investigations, thus enforcing a proper and rigorous scientific process, including transparency, accuracy, and reproducibility.

Monday 03.12.2012 12:30 - 13:30 Parallel Session P – CFE

CS77 Room 1 UNIT ROOT **Chair: Roderick McCrorie**

C872: **Stochastic integration and cointegration tests under a randomized and a weak bilinear unit root process**
Presenter: **Julio Angel Afonso Rodriguez**, University of La Laguna, Spain
The stochastic cointegration modelling framework represents a generalization of the standard stationary cointegration analysis, allowing some or all of the system variables are conventionally or heteroskedastically integrated. The concept of heteroskedastic integration is very close to that of a stochastic (or randomized) unit root, in the sense that generalizes the case of a fixed unit root (or difference stationary) process, by incorporating errors that display nonstationary variances but with a transitory effect on the level. It is shown that the test procedure for detecting a heteroskedastic integrated process is also consistent against other forms of nonstationarity, such as a randomized unit root process under a local heteroskedastic integrated parameterization and a simple bilinear (non Markovian) unit root process with a weak bilinear parameter. Also, it is explored the properties of the recently proposed estimation method called Integrated-Modified OLS under the stochastic cointegration setup. In the case of a standard static and stationary cointegrating relation this method allows to simultaneously correct for exogeneity bias and to obtain a conditional asymptotic normality result.

C887: **A new approach for parametric unit root testing**
Presenter: **Dimitrios Vougas**, Swansea University, United Kingdom
A new class of two-step methods for full parametric unit root testing is proposed and examined, whereby the parameter of interest is fully identified and estimated. The new method relies on one round/iteration, feasible generalised least squares (GLS) estimation of the simple autoregression, along with an additional autoregressive part of latent parameters. The advantage of the proposed method is that it avoids non-identification, while employing full, efficient estimation of all relevant parameters. It also allows reliable testing for the presence of a hidden (in the innovating autoregressive part) unit root. First step estimators for the latent parameters are available, so an extensive simulation study is undertaken to evaluate their relative effects on second step estimation of the parameter of interest. The resulting new parametric unit root tests have finite sample size distortions, although they are correctly sized asymptotically, and have better power than existing unit root tests.

C910: **Gaussian power envelope for panel unit root tests in the presence of cross-sectional dependence**
Presenter: **Irene Gaia Becheri**, VU University Amsterdam, Netherlands
Co-authors: Feike Drost, Ramon van den Akker
The aim is to derive the (local and asymptotic) power envelope for tests of the unit root hypothesis in Gaussian panels with cross-sectional dependence. Our setting allows for heterogeneous panels and heterogeneous alternatives. We consider the asymptotic scheme in which the number of cross section units and the length of the time series tend to infinity jointly. The power envelope is derived using the limiting experiment approach. In particular, we first consider the submodel in which all parameters (but the autoregression coefficient) are known and show that it is locally asymptotically normal. The power envelope for the submodel is thus easily computed thanks to Le Cam's theory. Then, we construct a test statistic, valid for the model of interest, which attains the power envelope of the submodel. As a consequence the constructed test is asymptotically efficient in the model of interest. Finally, we provide Monte Carlo results to assess the finite sample performances of the test.

CS78 Room 10 PORTFOLIO RISK MEASURES **Chair: Joern Sass**

C531: **Time scaling properties of multivariate dynamic higher order moments using Cornish-Fisher expansion**
Presenter: **Sorin Dumitrescu**, Institute for Economic Forecasting, Romania
Co-authors: Radu Lupu
The higher order moments of skewness and kurtosis are used in Cornish-Fisher expansions for the computation of Value-at-Risk measures. Previous research provided evidence on the higher performance of VaR measures using dynamic higher moments for individual and multivariate series of stock returns. We use a Cornish-Fisher approximation of the bivariate distribution to compute the quintiles of the distribution as a function of its higher moments and a normally distributed variable. Using the roots of this polynomial we find a likelihood function that can be used for the computation of the time varying coefficients of autoregressive higher moments (skewness, kurtosis, coskewness and cokurtosis). After fitting this model on series with different frequencies (from 5 minute returns to weekly returns) on data representing a set of European stock market index returns, we build a model for the behavior of these coefficients when moving from high frequencies to low frequencies. These results will be useful for VaR computations for multivariate portfolios in which higher-order moments will be taken into account.

C846: **Using machine learning techniques to forecast the variance-covariance of a portfolio: Application to VaR's estimation**
Presenter: **Carlos Hernan Fajardo Toro**, Universidad Sergio Arboleda, Colombia
Co-authors: Ernesto Camilo Diaz Estrada, Julio Cesar Alonso
Risk measurement has become a key aspect of risk management. By far Value-at-Risk (VaR) is one of the most popular ways to measure market risk of a portfolio. VaRs conceptual simplicity contrasts to the sophistication that can be behind its estimation. In particular VaR for large portfolio implies forecasting a large variance-covariance matrix. This variance-covariance matrix may be estimated with computational intensive models such as the ones from the family of Multivariate GARCH models. We propose a different approach to estimate and forecast it using neural networks. VaR's estimation involves non stationary data, and the chaotic and imprecise nature of these data rises different uncertainty levels, situation that compels to use the model that gives the best fit to this features. The neural networks have this ability, and depending on their topology and architecture they can fit efficiently to chaotic and stochastic data. The purposes is to applied and make a comparison between different techniques of machine learning for the estimation of VaR for a large portfolio, using different topologies of neural networks and other artificial intelligence techniques. Based on these results, some bases to propose an expert system to obtain these risk measurement are established. The results obtained has been compared with other statistical techniques traditionally used to estimate the VaR.

C886: **Unbiased evaluation of the portfolio model performance by the expected utility efficient frontier**
Presenter: **Petr Jablonsky**, University of Economics, Czech Republic
An unbiased method for the evaluation of the portfolio model performance using the expected utility efficient frontier is introduced. We apply the asymmetric behavioural utility function to capture the behaviour of real market investors. The new theory is applied on stock markets of United States and Germany in 2003 to 2012 to test the efficiency of nine portfolio models. The tested models include the Markowitz model and several alternative models that use different risk measures such as semi-variance or mean absolute deviation models. As the Markowitz model is the leading market practice we focus on investigation whether there were any circumstances in which some of the tested models might provide better performance than the Markowitz model.

CS79 Room 6 FINANCIAL ECONOMETRICS: MODELS AND APPLICATIONS Chair: Ard Den Reijer

C264: **The dPPS distribution in the modelling of financial returns**
Presenter: **Antonio Jose Saez-Castillo**, University of Jaen, Spain
Co-authors: Faustino Prieto, José María Sarabia
Resumen The dPPS distribution is considered to fit non-zero returns of four stocks traded on continuous markets (BBVA, Microsoft, Tele5 and Uralita) observed at three different time lags: daily, weekly and monthly. This distribution has shown a high accuracy modelling economic indicators as the current account balance, characterized by a high skewness. Thus, the fit achieved by the dPPS distribution permits to infer skewness properties of the trades returns, even with non-equal proportions of positive and negative returns. Goodness of fits are compared to those provided by other well-known distributions in this context. The results show that the dPPS distribution provides the best fit when the returns are observed daily and weekly; when monthly returns are considered, the dPPS provides the best fit of Microsoft and Uralita returns and it is also close to the best fit in BBVA and Tele5 returns.

C349: **International yield curve dynamics and interactions**
Presenter: **Kavita Sirichand**, Loughborough University, United Kingdom
Co-authors: Simeon Coleman
The existing yield curve literature typically examines the curve of a country in isolation. They report strong evidence of a link between the yield curve and the macroeconomy, with some suggestions of evidence of instability in the dynamics of the US yield curve. We investigate the dynamics and interactions between the term structure of interest rates across four areas - Canada, the UK, US and the euro area by estimating the co-movement, at high frequency, between the yields over various maturities. Our results establish the time varying properties in the dynamics of the yield curve of each region and, further, demonstrate cross-region co-movements over time.

C937: **Using property price distribution to detect real estate bubbles**
Presenter: **Takaaki Ohnishi**, The Canon Institute for Global Studies and the University of Tokyo, Japan
Co-authors: Takayuki Mizuno, Chihiro Shimizu, Tsutomu Watanabe
It is proposed making use of information on the cross-sectional dispersion of real estate prices to detect real estate bubbles. During bubble periods, prices tend to go up considerably for some properties, but less so for others, so that price inequality across properties increases. In other words, a key characteristic of real estate bubbles is not the rapid price hike itself but a rise in price dispersion. Given this, we examine whether developments in the dispersion in real estate prices can be used to detect bubbles in property markets as they arise, using data from Japan and the U.S. First, we show that the land price distribution in Tokyo had a power-law tail during the bubble period in the late 1980s, while it was very close to a lognormal before and after the bubble period. Second, in the U.S. data we find that the tail of the house price distribution tends to be heavier in those states which experienced a housing bubble. We also provide evidence suggesting that the power-law tail observed during bubble periods arises due to the lack of price arbitrage across regions.

CS80 Room 11 MARKOV SWITCHING MODELS Chair: Christian Francq

C495: **Cyclicality and bounce back effect in financial market**
Presenter: **Songlin Zeng**, University of Cergy Pontoise and ESSEC Business School, France
Various shapes of the recoveries in financial market may exhibit within a Markov Switching model are explored. It relies on the bounce back effects first analyzed and generalized with application on business cycle to allow for a more flexible bounce back types. When the bouce back markov-switching models are applied to stock market returns across G7 counties, the "bounce-back" form of intra-regime dynamics is statistically significant and large. The type of bounce back is tested using LR test based on general BBF Markov switching model, and hence, the permanent impact of bear market is measured.

C500: **Hamilton smooth particle filters**
Presenter: **Frederic Karame**, University Evry Val d Essonne, France
A new particle filter to estimate Markov-switching stochastic volatility models with leverage is proposed. Our first contribution relies on its global structure that is based on the well-known Hamilton filter. This strategy allows for a simple treatment of the unobserved regime. We use a sequential importance sampling particle filter to approximate the unobserved log-volatility and calculate the conditional likelihood necessary for the regime probabilities update and parameters estimation. In order to easily update particles and implement likelihood-based inference, we propose to use the smooth resampling approach. Our third contribution relies on the choice of the proposal distribution for the importance sampling step. After a description of our new approach and some simulation experiments, we present the estimation results on IBOVESPA index for comparison with existing works.

C710: **Modelling ratings impacts on stock return distributions within a multivariate regime switching long memory framework**
Presenter: **Hung Xuan Do**, Monash University, Australia
Co-authors: Robert Brooks, Sirimon Treepongkaruna, Eliza Wu
A framework that allows a multivariate system of long memory processes to be conditional on specific regimes is proposed. The model is motivated by the necessity to investigate the impacts of sovereign credit quality assessments on financial return distributions. While the consistent and robust estimates of moments of the financial return distribution (i.e., realized moments) somewhat exhibit the long memory behavior, the regime switching feature of sovereign ratings has been widely documented. We illustrate our model by investigating the effects of overall European Union (EU) sovereign creditworthiness assessments on the EU stock return distributions via their first four realized moments. We find dissimilar effects of sovereign rating actions across regimes, implying the usefulness of our proposed model in accommodating both long memory and regime switching features. Further, we note that the total effects, including both direct and indirect forces, of the sovereign credit quality assessments on the realized moments can be different to their direct effects. Hence, we derive a function, which can capture these total effects, to investigate which agency has the greatest impact on the EU stock return distributions. We find that the rank orders of agencies are not unique across regimes and even in each realized moment.

CS81 Room 13 MODELS FOR FINANCIAL AND COMMODITY MARKETS Chair: Gonzalo Camba-Mendez

C875: **Hedging effectiveness of Malaysian crude palm oil futures contracts: The extended mean-Gini framework**
Presenter: **Kin-Boon Tang**, The University of Nottingham Malaysia Campus, Malaysia
Co-authors: Ju-Yau Tarn

The stochastic dominance model has long been proven to exhibit greater theoretical advantages than the Minimum-Variance (MV) approach in futures hedging. The aim is to demonstrate the superiority of th Extended Mean-Gini (EMG) framework which is consistent with the second-order of stochastic dominance theory. The merits of the EMG model are credited to its specifications where the distributional assumptions of normality are not required. The study provides a comprehensive analysis of investors' distinct risk adverse behaviour towards optimal futures hedging strategy. The empirical distribution function method and the more efficient kernel estimation method are employed in the estimation of EMG hedge ratios. Furthermore, in order to investigate the suitability of 'buy and hold' strategy in futures hedging, the moving data window procedure is used to examine the stability of the dynamic hedge ratios. The research is conducted on Malaysian Crude Palm Oil (CPO) and CPO Futures (FCPO) markets for the period of 16th March 1995 to 28th June 2011. The empirical results show that the EMG approach is apparently more appropriate than the MV approach where EMG framework incorporates the risk aversion factor. The study also shows the instability of dynamic hedge ratios across time horizons hence not favourable to investors who adopt the 'buy and hold' strategy.

C955: **Forecasting commodity currencies with dynamic Bayesian models**
Presenter: **Daniel Buncic**, University of St Gallen, Switzerland

Recently developed Dynamic Model Averaging/Selection (DMA/DMS) and Bayesian Model Averaging/Selection (BMA/DMS) methods are used to forecast a number of commodity currencies using a set of around 25 standard macroeconomic as well as financial variables that include the VIX and also various commodity price indexes. We find that the selected DMA and DMS specifications perform very well when evaluated out-of-sample not only when compared to common simple benchmarks such as random walk, exponential smoothing and low order autoregressive models, but also when compared to static BMA and BMS models and also various forms of simple time varying parameter autoregressive models. Our study finds further that allowing for fast changes in the time variation of the model size, ie., the number of predictor variables that are included in a given time period, as well as in the magnitude of the coefficients is an important feature of exchange rate data that substantially improves the out-of-sample forecast performance of the DMA and DMS methods, where the speed of change in the model size and coefficients is specified by the magnitude of the forgetting factor. These results are robust to various loss functions that are used to evaluate the forecasts as well as rolling and expanding forecast window sizes.

C990: **Tractable and estimable general model of limit order markets**
Presenter: **Martin Smid**, Institute of Information Theory and Automation, Prague, Czech Republic

Recently, several models of limit order markets have been proposed, all of them assuming constant arrival rates of market and limit orders (the latter possibly depending on a distance to the best quotes) and constant cancellation rate (also possibly depending on the distance to the quotes). Even if those models comply with several stylized facts, some of their properties (conditionally exponential durations between transactions, zero-intelligence behavior of the agents) contradict the empirical evidence. We propose a more general model (the existing models being its special cases) allowing a general behavior of the best quotes (hence the price) while keeping the assumption of the Poisson behavior of the order books. Contrary to the existing models, non-unit order sizes are allowed. An analytical formula for the (conditional) distribution of the order books (given the best quotes' process) is formulated, enabling precise econometrics of the order books based on either L1 or (possibly incomplete) L2 data. The econometrics is demonstrated on the order book data from LSE.

CS69 Room 9 FINANCIAL MARKETS AND COPULAS Chair: Jean-David Fermanian

C623: **Latin American exchange rates dependencies: A regular vine copula approach**
Presenter: **Luis Melo**, Central Bank of Colombia, Colombia
Co-authors: Ruben Loaiza

A regular vine copula methodology is studied to estimate the dependence structure of the exchange rates of six Latin American countries (Argentina, Brazil, Chile, Colombia, Mexico and Peru) over the period June 2005 to April 2012. The results from the estimation of the tail dependence coefficients by regular vine copula methodology indicate that these countries are divided into two groups. The first one is composed by Brazil, Colombia, Chile and Mexico, their exchange rates exhibit the largest dependence coefficients among themselves; whereas the second block is composed by Argentina and Peru, for which the exchange rate's dependence coefficients with other Latin American countries are low. We also found that most of the Latin American exchange rate pairs exhibit an asymmetric behavior characterized by non-significant upper tail dependence and significant lower tail dependence. These results imply that the probability that Latin American exchange rates move together against dollar in periods of large appreciation is higher than the probability that they move together in periods of high depreciation.

C735: **Commodity currencies and commodity prices: Modeling static and time-varying dependence**
Presenter: **Natalia Ponomareva**, Macquarie University, Australia
Co-authors: Katja Ignatieva

A copula approach is employed in order to study the relationship between combinations of the exchange rates and commodity prices for large commodity exporters. We use daily data for the nominal exchange rates of the four commodity currencies (Australian, Canadian, and New Zealand dollars, and Norwegian krone) against the U.S. dollar and the relevant country-specific commodity price indices. We find positive dependence structure for all pairs of the exchange rates indicating that currencies tend to appreciate or depreciate jointly against the U.S. dollar. We also find positive dependence between the values of currencies and commodity indices with a pronounced increase in the time-varying dependence from the beginning of the global financial crisis. For most combinations of the exchange rates and commodity indices we observe no major asymmetries in the tail dependence.

C888: **Example of using copula functions as a tool for researching contagion in financial markets**
Presenter: **Daniel Papla**, Wroclaw University of Economics, Poland

We define contagion as a significant increase in cross-market linkages after a shock to one country (or group of countries). According to this definition, if two markets show a high degree of comovement during periods of stability, even if the markets continue to be highly correlated after a shock to one market, this may not constitute contagion. According to this definition, contagion occurs if cross-market co-movement increases significantly after the shock. Main goal is to analyse changes in dependence between Polish stock market and chosen world stock market. Following research hypothesis will be verified: dependence between Polish stock market and other stock markets is increasing in periods of rapid decrease in value of stock market indexes. Positive verification of this hypothesis means, that there is a contagion in financial markets.

CS92 Room 3 FINANCIAL ECONOMETRICS: INFERENCES Chair: Arvid Raknerud

C636: **Testing combinations in conditional quantiles: The case of Value-at-Risk (VaR)**
Presenter: **Georgios Tsiotas**, University of Crete, Greece
Value-at-Risk (VaR) is a measure of risk evaluation that helps financial institutios to allocate risk. VaR forecast estimation concerns the evaluation of quantiles conditional on current information. Recent development in VaR evaluation incorporates conditional variance in quantile estimation introducing the CAViaR models. However, the amount of alternative CAViaR models raises the issue of finding the optimal quantile predictor. To resolve this uncertainty, we propose a Bayesian encompassing test that evaluates alternative single CAViaR model predictions against the combined CAViaR alternative. This test provides a basis for forecasting combined conditional VaR estimates when the encompassing principle is rejected. We illustrate this test using simulated and financial daily return series. Results show that there is much to gain from using combined conditional VaR estimates when forecasting quantile risk.

C666: **Estimation of the Markov-switching GARCH model with a maximum likelihood approach**
Presenter: **Maciej Augustyniak**, University of Montreal, Canada
The Markov-switching GARCH model offers rich dynamics to model financial data. Estimating this path dependent model is a challenging task because the exact computation of the likelihood is infeasible in practice. This led some authors to propose estimation methods that do not depend on the likelihood. Other authors suggested estimating by maximum likelihood modified versions of the Markov-switching GARCH model that avoid the path dependence problem. To this date, there is no method available to obtain the maximum likelihood estimator without resorting to a modification of the model. A novel approach is developed based on the Monte Carlo expectation-maximization algorithm and on importance sampling to calculate the maximum likelihood estimator of the Markov-switching GARCH model and its asymptotic variance-covariance matrix. Practical implementation of the proposed algorithm is discussed and its effectiveness is demonstrated in simulation and empirical studies.

C643: **Log-logistic function estimation method and its application in finance and economics**
Presenter: **Rafal Siedlecki**, Wroclaw University of Economics, Poland
The application of the modified logistic function (loglogistic function) and its estimation method is presented. This function is most frequently used to describe economic or natural phenomena. In many economic and financial cases it turns out that the logistic function does not work, it concerns mainly unlimited growth phenomenon. Quantities such as GDP, stock market indexes, salaries in enterprises, sales or company value cannot be limited. If their value does not decrease rapidly after the intensive growth phase, then it is followed by a slow increase (its rate should fall to zero) The way to eliminate the logistic function limited growth defect is to modify the function by introducing the ln(t) factor. The modified function is called log-logistic function (logarithmic-logistic). The function was previosly proposed and is expressed by the following formula: $\ln(t)*a/(1+exp(b-c*t)$ where $a>0$, $b>0$and $c>0$

CS94 Room 2 TIME SERIES ECONOMETRICS II Chair: Dennis Kristensen

C867: **An adaptive and robust Portmanteau test for uncorrelatedness under $q-$dependency and heteroskedasticity**
Presenter: **Hsein Kew**, Monash University, Australia
Co-authors: David Harris
Portmanteau tests for uncorrelatedness in the presence of a $q-$dependent process and unconditional heteroskedasticity are considered. Allowing for a $q-$dependent process is important because some financial theories may imply that serial correlation vanishes for, say, $(q+1)-$lag and beyond while allowing the $q-$lag to be serially correlated due to the presence of overlapping observations. It is also well-known that many financial time series exhibit unconditional heteroskedasticity. The standard Box-Pierce Portmanteau tests are shown to suffer from asymptotic size distortions since they are not designed to allow for the presence of a $q-$dependent process and heteroskedasticity. To overcome this problem, a Portmanteau test that employs the Heteroskedasticity and Autocorrelation Consistent (HAC) estimator is proposed. This HAC estimator is shown to produce asymptotically valid inferences in the presence of unconditional heteroskedasticity of an unknown form. However, asymptotically more efficient inferences are possible if a specific form of unconditional heteroskedasticity is known. Utilizing a recently proposed non-parametric variance estimation method, an adaptive test is thus developed. This test is shown to be just as efficient as tests based on a priori knowledge of the form of unconditional heteroskedasticity.

C897: **Strucutral break models under misspecification: Implications for forecasting**
Presenter: **Bonsoo Koo**, Monash University, Australia
Co-authors: Myung Seo
Previous research on inference for structural-break models has made the implicit assumption that the data generating process (DGP) follows a form of structural-break models. However, as the DGP is unknown, uncertainty about the form of the DGP should be taken into account in the model specification. We show that when possible model misspecification is not accounted for, the conventional inference procedures for structural-break models are invalid. In doing so, we establish new distribution theory for structural break models under the relaxed assumption that our structural break model is the best linear approximation of the true but unknown data generating process. Our distribution theory involves cube-root asymptotics and it is used to shed light on forecasting practice. We show that the conventional forecasting methods do not necessarily produce the best forecasts in our setting. We propose a new forecasting strategy which incorporates our new distribution theory. By applying our forecasting methods to numerous macroeconomic data, we show that our forecasting method outperforms standard forecasting methods. In addition, we provide comparative performance of various contemporary forecasting methods including ours, and demonstrate the validity of our method.

C454: **Learning from learners**
Presenter: **Tom Holden**, University of Surrey, United Kingdom
Traditional macroeconomic learning algorithms are misspecified when all agents are learning simultaneously. We produce a number of learning algorithms that do not share this failing, including one capable of learning the stationary MSV solution for all parameters, and one that may converge to any sunspot solution. As a by-product, we are able to show that when an agent knows all other agents are learning by some reasonable rule, all deep structural parameters of standard new-Keynesian models are identified, overturning a previous key result. This holds irrespective of whether the central bank is following the Taylor principle, irrespective of whether the implied path is or is not explosive, and irrespective of whether or not beliefs of other agents converge. If shocks are observed then this result is trivial, so our analysis is carried out in the more plausible case in which agents do not observe shocks.

CS85 Room 8 FINANCIAL MARKETS II Chair: Gaelle Le Fol

C948: **Measuring capital market efficiency: Global and local correlations structure**
Presenter: **Miloslav Vosvrda**, Academy of Sciences of the Czech Republic, Czech Republic
Co-authors: Ladislav Kristoufek
A new measure for the capital market efficiency is introduced. The measure takes into consideration the correlation structure of the returns (long-term and short-term memory) and local herding behavior (fractal dimension). The efficiency measure is taken as a distance from an ideal efficient market situation. The methodology is applied to a portfolio of 41 stock indices. We find that the Japanese NIKKEI is the most efficient market. From geographical point of view, the more efficient markets are dominated by the European stock indices and the less efficient markets cover mainly Latin America, Asia and Oceania. The inefficiency is mainly driven by a local herding, i.e. a low fractal dimension.

C884: **Latency and quote duration: Limits to integration in fragmented markets**
Presenter: **Thomas Katzschner**, University of Jena, Germany
In fragmented markets with overlapping trading hours, the distance between trading platforms has become a barrier for a consolidated market view and order routing, where the latency for collecting order book messages and submitting orders exceeds the duration of quotes. A simulation based on an autoregressive conditional duration (ACD) model demonstrates that an increase of quote changes from 100 to 1000 per minute impedes a consolidated market view with 17.6% of quotes disappearing within 10 milliseconds and 82.3% within 100 milliseconds. Empirical evidence for a limit to integrate quoted prices comes from the German DAX 30 stock Siemens, for which London-based Chi-X Europe constitutes an alternative trading platform to Xetra, the Frankfurt-based platform of Deutsche Boerse. On Chi-X Europe, about 60% of ask and bid prices last for a maximum of 5 milliseconds in September 2010, while the round-trip delay time amounts to 9 milliseconds between London and Frankfurt.

C781: **Wavelet-based correlations: International evidence between stock market and oil returns**
Presenter: **Helena Veiga**, Universidad Carlos III de Madrid, Spain
Co-authors: Belen Martin-Barragan, Sofia Ramos
Financial contagion is of utmost importance for portfolio management as well as policy making since turmoil in one market, usually increases the cross-market linkages due to the propagation of shocks. Despite the interest from distinguishing contagion from interdependence still very few works have being able to do it. Therefore, the aim is: First, to analyze whether oil price shocks affect stocks markets. Second, to test for contagion from oil market to stock markets and then, given oil prices shocks, among stock markets. Often, increases in oil prices lead to negative returns of stock markets of oil importing countries. However, decreases of oil prices of the same magnitude do not seem to affect stock markets. An important question that remains to be analyzed in this area is the duration and the magnitude of these shocks: Are they temporary or have they longer effects? Hence, a third aim is to study the nature of these shocks. In this way, we propose to analyze the sign, the magnitude and the duration of oil price shocks.

CS76 Room Multipurpose CONTRIBUTIONS TO BAYESIAN ECONOMETRICS AND APPLICATIONS Chair: Helga Wagner

C672: **Model averaging in Bayesian periodic autoregressive models to forecast seasonal time series**
Presenter: **Alexander Vosseler**, Institute for employment research IAB, Germany
A Bayesian periodic autoregressive (PAR) model is presented for the prediction of quarterly and monthly time series data. The model admits one structural break in the periodic trend function, where the timing of the break is treated as a further model parameter. As in many instances it is not known whether a structural break has occurred, this lack of knowledge thus constitutes a source of uncertainty in the model building step. In this context the occurrence of a break is treated as an additional parameter. Furthermore the autoregressive lag order is also treated as an unknown quantity, which together with the occurrence of a structural break serves to identify different models in the model space. Instead of resorting to a model selection approach by choosing a single model for prediction, a Bayesian model averaging approach for prediction in PAR models is proposed. As the posterior predictive distribution of multistep forecasts in PAR models is complicated and follows no standard distribution, a Gibbs sampling strategy to generate draws from this distribution under a specific model combination is presented. Finally using the posterior probabilities of each model combination as weights a model averaged posterior predictive distribution is computed. The forecasting performance of this Bayesian approach is then compared with classical periodic and also non-periodic time series models on the basis of various prediction criteria using monthly time series data.

C864: **Forecasting fiscal variables using Bayesian VARs with constant and drifting coefficients**
Presenter: **Angeliki Theophilopoulou**, University of Westminster, United Kingdom
Co-authors: Andrea Carriero, Haroon Mumtaz
Previous research has shown that simple autoregressive models often provide better forecasts of fiscal variables relative to vector autoregressions. This is puzzling because economic theory suggests that fiscal variables should be tightly intertwined, and therefore individual time series should contain useful information about the others. A possible explanation is that the VARs considered by previous studies are small-scale, likely burdened by overparameterization, and do not feature time variation in the coefficients and volatilities. Bayesian VARs (BVARs) instead allow to efficiently summarize the information contained in a large data set, avoid the overparameterization problem, and can allow for time variation in the coefficients and in the volatilities. We explore the performance of BVARs with constant and drifting coefficients in forecasting key fiscal variables such as government revenues, expenditure, and interest payments on the outstanding debt. Using data from Germany, France, UK, and US, we show that BVARs can significantly improve over an autoregressive forecasts. We find that both the adoption of a large system and the introduction of time variation helps in forecasting, with the former playing a relatively more important role than the latter.

C878: **Bayesian estimation of beta type distribution parameters based upon grouped data**
Presenter: **Haruhisa Nishino**, Chiba University, Japan
Co-authors: Kazuhiko Kakamu
The estimation method of generalized beta (GB) distribution parameters based upon small grouped income data is considered from a Bayesian point of view. As the GB distribution includes several kinds of familiar distributions of income such as generalized beta distribution of the second kind (GB2 distribution), Singh-Maddala distribution and so on, it is reasonable to consider the distribution. However, when the number of groups is small, it is sometimes difficult to estimate the parameters of the distribution utilizing the existing estimation methods such as maximum likelihood, minimum chi-square and so on. Thus, we propose a Markov chain Monte Carlo (MCMC) method to estimate the parameters of the distribution. The concept of the selected order statistics is utilized to construct the likelihood function. This method is also applied to the Japanese data. The empirical result captures the change of Japanese economy.

CS68 Room 5 PANEL DATA **Chair: Roberto Leon-Gonzalez**

C380: **Applying projected score methods to panel data models**
Presenter: **Andrew Pua**, Universiteit van Amsterdam and Universite Catholique de Louvain, Netherlands

Panel data allow control of time-constant unobserved heterogeneity of cross-sectional units. As a result, we have an incidental parameter problem in large N, short T panels. Projected score methods are used to construct either fixed$-T$ root$-N$ or root$-NT$ consistent estimators under the assumption that incidental parameters are treated as constants. This is because of the negative results on identification and the possibility of root$-N$ consistent estimation in nonparametric correlated random effects models. The idea of projecting the score equations for the structural parameters onto an appropriate space of functions, capturing the effects of the presence of high-dimensional nuisance parameters, is helpful in estimating these parameters in both linear and nonlinear panel data models. Explicitly calculating these projections can be used to trace and assess the relative impacts of various sources of model misspecification on the resulting estimators. Furthermore, exogeneity conditions play a crucial role in these calculations. As a consequence, projected scores are somewhat harder to calculate but very transparent. Further, the projected scores can give the similar answers as profile score adjustments for varying intercept models found in the literature.

C366: **Direct semiparametric estimation of fixed effects panel data varying coefficient models**
Presenter: **Alexandra Soberon**, Universidad de Cantabria, Spain
Co-authors: Juan M. Rodriguez-Poo

A new technique to estimate varying coefficient models of unknown form in a panel data framework is presented where individual effects are arbitrarily correlated with the explanatory variables in an unknown way. The resulting estimator is robust to misspecification in the functional form of the varying parameters and it is shown to be consistent and asymptotically normal. Furthermore, introducing a transformation, it achieves the optimal rate of convergence for this type of problems and it exhibits the so called oracle efficiency property. Since the estimation procedure depends on the choice of a bandwidth matrix, we also provide a method to compute this matrix empirically. Monte Carlo results indicates good performance of the estimator in finite samples.

C410: **A mixed-cross-section GVAR for countries and banks**
Presenter: **Marco Gross**, European Central Bank, Germany
Co-authors: Christoffer Kok

The aim is to illustrate how a Mixed Cross-Section Global Vector Autoregressive (MCS-GVAR) model can be set up and solved to then be used for the purpose of forecasting and scenario simulation. The two cross-sections that are referred to for exemplifying the idea are countries (sovereigns) and banks. Besides the avail of the model for simulating shock scenarios, the estimated weight matrices for banks vs countries, banks vs banks, countries vs banks and countries vs countries can be a useful means for assessing spill-over channels that link the macro and the financial sphere.

CS25 Room 4 CONTRIBUTIONS ON SYSTEMIC RISK **Chair: Gloria Gonzalez-Rivera**

C729: **The dynamics of spillover effects during the European sovereign debt crisis**
Presenter: **Adrian Alter**, University of Konstanz, Germany
Co-authors: Andreas Beyer

Empirical measures for the strength of spillover effects are developed. Modifying and extending a previous framework, we quantify spillovers between sovereign credit markets and banks in the euro area. Spillovers are estimated recursively from a vector autoregressive model of daily CDS spread changes, with exogenous common factors. We account for interdependencies between sovereign and bank CDS spreads and we derive generalised impulse response functions. Specifically, we assess the systemic effect of an unexpected shock to the creditworthiness of a particular sovereign or country-specific bank index to other sovereign or bank CDSs between October 2009 and July 2012. Channels of transmission from or to sovereigns and banks are aggregated as a Contagion index (CI). This index is disentangled into four components, the average potential spillover: i) amongst sovereigns, ii) amongst banks, iii) from sovereigns to banks, and iv) vice-versa. We highlight the impact of policy-related events along the different components of the contagion index. The systemic contribution of each sovereign or banking group is quantified as the net spillover weight in the total net-spillover measure. Finally, the captured time-varying interdependence between banks and sovereigns emphasises the evolution of their strong nexus.

C721: **Measuring the effect of network externalities on financial distress: A spatial econometrics approach**
Presenter: **Carlos Castro**, Universidad del Rosario, Colombia
Co-authors: Juan Sebastian Ordonez, Carlos Leon

The aim is to propose and estimate a financial distress model that explicitly accounts for the interactions or spill-over effects between financial institutions, through the use of a spatial continuity matrix that is build from financial network data of interbank transactions. Such a setup of the financial distress model allows for the empirical validation of the importance of network externalities in determining financial distress, in addition to institution specific and macroeconomic covariates. The relevance of such a specification is that it incorporates simultaneously micro-prudential factors (Basel 2) as well as macro-prudential and systemic factors (Basel 3) as determinants of financial distress.

C726: **A theoretical and empirical comparison of systemic risk measures**
Presenter: **Sylvain Benoit**, Universiry of Orleans, France
Co-authors: Gilbert Colletaz, Christophe Hurlin, Christophe Perignon

The purpose is to develop a theoretical and empirical comparison of several popular systemic risk measures that can be estimated from public data: *Marginal Expected Shortfall* (MES), *Systemic Risk Measure* (SRISK), and *Delta Conditional Value at Risk* (ΔCoVaR). First, we assume that the time-varying correlation completely captures the dependence between the firm and market returns and we show that (*i*) MES corresponds to the product of the market expected shortfall (market tail risk) and the firm beta (firm systematic risk) and that (*ii*) ΔCoVaR corresponds to the product of the firm VaR (firm tail risk) and the linear projection coefficient of the market return on the firm return. We also derive (*iii*) conditions under which different systemic risk measures lead to consistent rankings of systemically important financial institutions. Second, we relax our fundamental assumption and empirically validate our theoretical findings for a sample of US financial institutions. Collectively, our results indicate that the best systemic risk measure very much depends on the considered objective: MES helps little to *rank* systemically important financial institutions, whereas ΔCoVaR brings limited added value over and above VaR to *forecast* systemic risk. Overall, SRISK offers the best compromise between the too-big-to-fail and too-interconnected-to-fail dimensions.

CS02 Room 7 FORECASTING FINANCIAL MARKETS Chair: Massimiliano Pisani

C697: **Speculative behavior and oil price predictability**
Presenter: **Ekaterini Panopoulou**, University of Piraeus, Greece
Co-authors: Theologos Pantelidis
The aim is to build on the theory of speculative behavior and self-fulfilling expectations to develop two- and three-state Regime-Switching (RS) models and test their forecasting ability for oil prices. Taking advantage of the deviations we periodically observe between the market price of oil and its fundamental value (we refer to this difference as a bubble), our models relate the expected gross return in the oil price to the bubble size and an additional explanatory variable. Specifically, we compare the predictive power (in both statistical and economic evaluation terms) of nine alternative macroeconomic/indicator variables assuming a forecast horizon of one month. Our findings indicate substantial benefits, in terms of forecasting accuracy, when RS models are employed relative to the Random Walk (RW) benchmark, especially in the case of three-state RS models. Moreover, in most cases, RS models enriched with one of the predictors proposed in this study outperform simple RS models that contain no predictors (other than the bubble size).

C725: **Combination forecasts of bond and stock returns: An asset allocation perspective**
Presenter: **Sotiria Plastira**, University of Piraeus, Greece
Co-authors: Ekaterini Panopoulou
The out-of-sample predictability of U.S. bond and stock returns for a variety of horizons ranging from the short-run (1 month) to the long-run (3 years) is assessed. Our list of predictors contains the typically employed financial and macroeconomic variables along with the value (HML), size (SMB) and momentum factors. We construct a variety of portfolios of forecasts based on combination methods and the aforementioned variables, thus investigating the power of the combining methods in conjunction with the power of predictors. Given that statistical significance does not imply economic significance, we assess the economic value of the predictive power of our proposed models by investigating the utility gains accrued to investors who exploit the predictability of bond and stock returns relative to the no-predictability alternative. Our findings suggest that combination forecasts of suitably decomposed value, size and momentum factors display superior forecasting ability. The robustness of our results is assessed for the European and Japanese bond and stock markets.

C825: **Determining currency risk premiums using a bilateral arbitrage free Nelson Siegel term structure model**
Presenter: **Sarah Mouabbi**, Queen Mary, University of London, United Kingdom
Empirical evidences have so far rejected the use of the uncovered interest rate parity in the determination of exchange rates due to the forward premium bias. Its failure has been accredited to the existence of a time varying risk premium. An attempt is made to renew the interest in determining the depreciation of exchange rates through interest rate differentials, and ultimately to extract the term structure of the currency risk premium. The forecasting method proposed consists in developing a bilateral affine arbitrage free class of dynamic Nelson-Siegel term structure models (BAFNS), which in turn will be used in determining the discount rate variations. Assuming linearity and gaussianity, the estimation exercise comprises of a state space analysis through the Kalman filter. The imposition of the dynamic Nelson Siegel (DNS) structure allows for a tractable and robust estimation, offering significant computational benefits, whilst no-arbitrage restrictions enforce the model with theoretically appealing properties.

CS83 Room 12 FINANCIAL MODELLING AND APPLICATIONS Chair: Leopold Soegner

C952: **Valuing CDS options under double exponential jump diffusion**
Presenter: **Nedim Handzic**, University of New South Wales, United Kingdom
Co-authors: Ramaprasad Bhar
It is demonstrated how the Double Exponential Jump Diffusion (DEJD) process can be used to value iTraxx IG CDS options based on historical returns of the underlying CDS index. In the first step we find Maximum Likelihood estimates for the volatility of the normal component of returns, the Poisson jump frequencies and mean sizes of upward and downward exponential jumps. The DEJD provides a better fit than either the normal or single-jump specifications. We take the additional step of using parameter estimates as inputs to a semi-closed form European option pricing formula under the DEJD. Rolling parameter estimates are used to generate a time series of option prices across strikes and compared to the market implied volatilities. Our findings suggest that the DEJD provides a realistic description of the joint role of positive and negative economic surprises in credit markets. In addition, the use of Maximum Likelihood parameters is a practical way to analyse divergence between historical and market-implied distributions of credit returns, and can be used to check pricing.

C1019: **Excess financial development and economic growth**
Presenter: **Lorenzo Ductor Gomez**, Massey University, New Zealand
Co-authors: Daryna Grechyna
The possible negative influence of the financial development on the economic growth is investigated. We define excess finance as the difference between the financial and real output growth of the economy under which the aggregate output decreases. Based on a panel data of 33 OECD economies, we show that in order to have a smooth economic development, a balanced growth of both the real and the financial sectors is required. Whenever the financial development exceeds the development of the productive industries by 4.5% (when measured in terms of growth rates of the two sectors output), there is a threat of reaching the productive capacity bound of the economy, with a consequent "financial" crisis. The existence of excess financial development may be justified by the theory of informational overshooting.

C1066: **The optimal dividend yield range: the case of Europe and Spain during the period 2000-2009**
Presenter: **Teresa Corzo Santamaria**, Comillas Pontifical University, Spain
Co-authors: Maria Dolores Lagoa-Varela
Using monthly data during the period 2000-2009 for 163 European and Spanish companies, we show a clear, non-linear relationship between dividend yields and total stock returns earned by stockholders. This relationship is an inverted, U-shaped relationship in the sense that stocks within a specific range of Dividend Yield (DY) have superior returns and are the only ones with positive average returns throughout our sample period. We also find that stocks within a portfolio with superior and positive returns are the ones that show lower risk when risk is measured by the portfolio's average standard deviation. The relationship between DY and risk exhibit a U shape. Our results contradict the irrelevance theorem and reveal the existence of an optimal dividend yield policy. These relationships persist when we relate current DYs with returns and risks one month, three months, six months, and one year ahead, especially in the case of Spanish companies. Our ten-year sample includes a period with financial market downturns and recoveries, which helps reduce biased results related to bull or bear markets. Additionally, we found that the nonlinear relationship between DY and risk, and DY and returns, cannot be explained by different market capitalizations.

Monday 03.12.2012 15:00 - 16:40 Parallel Session Q – ERCIM

ES33 Room 7 ADVANCES IN DISTANCE-BASED METHODS AND APPLICATIONS Chair: Aurea Grane

E327: **Hierarchical linear discriminant analysis**
Presenter: **Jose Berrendero**, Universidad Autonoma de Madrid, Spain
Co-authors: Javier Carcamo
We consider the binary supervised classification problem and provide a decomposition of any quadratic classification rule in terms of certain products of hyperplanes plus a remainder. To obtain the hyperplanes we only need to diagonalize an appropriate symmetric matrix, so that the procedure is quite standard from a computational point of view. Some applications of such a decomposition are: (a) To detect when a quadratic classification rule can be replaced with a linear one without any loss. This would be the case when the remainder term is negligible, and one of the hyperplanes of the product is irrelevant for the classification problem; (b) To identify linear combinations of the original variables that are relevant to discriminate between the two classes. If the number of relevant combinations is much smaller than the number of original variables, the dimensionality of the problem can be substantially reduced; (c) To define new piecewise linear, easy to interpret, classification rules. Using some kind of regularization, the new rules can be defined even when the number of variables is greater than the number of observations.

E329: **Relative predictor importance in distance-based generalized linear models**
Presenter: **Eva Boj**, Universitat de Barcelona, Spain
Co-authors: Anna Esteve, Josep Fortiana, Teresa Costa
Assessing the relative importance of each predictor and calibrating its influence on the response, a straightforward operation in linear prediction, becomes a non trivial issue for non linear or non parametric methods. In distance-based (DB) prediction this difficulty is compounded, due to the indirect relationship between predictors and response, through a metric. Two partial approaches have been proposed to this problem for DB regression (DB linear model with a continuous response). Firstly, there exists a version of the F-test statistic for selecting explanatory variables, with bootstrapped p-values. Secondly, a set of influence coefficients, that is, local -valid in a neighborhood of a given point in predictor space- measures of the relative importance of each observed variable. This second procedure is extended to DB generalized linear models. The reach and performance of these influence coefficients is demonstrated on case studies drawn from actuarial (insurance) problems, in which quantifying the relative leverage of risk factors is of utmost importance.

E412: **Analysis of dependence evolution in Spanish disabled population through funcional data analysis**
Presenter: **Ana Arribas-Gil**, Universidad Carlos III de Madrid, Spain
Co-authors: Pablo Alonso, Irene Albarran
In a health context dependence is defined as lack of autonomy in performing basic activities of daily living that require the care of another third person or significant help. However, this contingency, if it happens, may surely change through the lifetime. In fact, empirical evidence shows that, once this situation appears, there is almost impossible to return to the previous state and in most cases and increase in the intensity is suffered. The evolution of the intensity in this situation is studied for the Spanish population affected by this contingency. Evolution in dependence can be seen as sparsely observed functional data, where for each individual we get a curve only observed at those points in which changes in his/her dependence situation occur. Here we have used functional data analysis techniques such as curve registration, functional data depth or distance-based clustering to analyze this kind of data. This approach proves to be useful in this context since it takes into account the dynamics of the dependence process and provides more meaningful conclusions than simple pointwise or multivariate analysis. The database analysed comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations, EDAD 2008, (Spanish National Institute of Statistics, 2008). The evaluation of the dependence situation for each person is ruled in Spain by the Royal Decree 504/2007 that passes the scale for assessment of the situation set by Act 39/2006. The paper has been prepared using the scale value reached by each individual included in EDAD 2008. Differences between sex, ages and first appearance time have been considered.

E460: **A sequential distance-based approach for imputing missing data: The forward imputation**
Presenter: **Nadia Solaro**, University of Milan-Bicocca, Italy
Co-authors: Alessandro Barbiero, Giancarlo Manzi, Pier Alda Ferrari
A recurring problem in multivariate data analysis (MVDA), potentially sparing no field of application, is the treatment of incomplete information. The subject is vast and complex, and has originated a literature rich of very different approaches. In an exploratory framework distance-based methods and procedures involving MVDA techniques can treat the problem properly. The nearest-neighbour imputation (NNI) method is distance-based in that detects sets of "donors" for incomplete units on the basis of their mutual nearness measured by a specific metric. MVDA techniques, such as PCA, through an iterative minimization of a loss-function, can recover values for incomplete units taking into account associations between variables. Both approaches have attractive features. In NNI, the metric and the number of donors can be chosen at will. The MVDA-based approach expressly accounts for variable associations. The approach here proposed, called forward imputation, ideally meets these features. It is developed as a distance-based approach that imputes missing values sequentially by alternating a MVDA technique and the NNI method. The MVDA technique could be any. Given the wide range of possibilities, attention here is confined to PCA. Comparisons with alternative imputation methods are then performed in presence of different data patterns.

ES65 Room 11 LONGITUDINAL/SPATIAL DATA ANALYSIS Chair: Xuming He

E224: **Partial linear single index models for clustered data**
Presenter: **Shujie Ma**, University of California-Riverside, United States
Co-authors: Hua Liang, Chih-Ling Tsai
There has been increasing interest in study of partial linear single-index models (PLSiM) because they can allow one to parsimoniously consider the interaction within some covariates through a nonlinear relationship with the response variable. We propose to use PLSiM for modeling complex clustered data and estimate the parameters by the quadratic inference function (QIF) with profile principle. We approximate the nonparametric functions by the polynomial spline, and derive the QIF-based estimators for the linear coefficients. We establish asymptotic normality for the resulting estimators of the linear coefficients, and the optimal convergence rates for the estimators of the nonparametric function. The proposed method improves estimation efficiency and increases statistical power for correlated data through incorporating the uncertainty in the correlation structure. We further propose a penalized procedure for variable selection, which can simultaneously identify non-zero linear and obtain final estimation results. The proposed procedure is shown to have the oracle property. Extensive Monte Carlo studies have been conducted and show the proposed procedure works effectively with moderate sample sizes.

E491: **Joint analysis of longitudinal and time to event data: A case study**
Presenter: **Masha Kocherginsky**, University of Chicago, United States
Co-authors: Zora Baretta
In clinical studies, patients are followed from disease diagnosis until an event of interest, such as disease progression or death, occurs. During this period of time, clinical parameters (biomarkers) related to the disease are collected, and it is often of interest to evaluate the effect of such longitudinally measured biomarkers on the time to event outcome. Statistical literature has focused on the development of methods for continuous longitudinal biomarkers (e.g., CD4 cell counts in HIV studies, or quality of life scores and other biomarkers in cancer studies). Such data are most commonly analyzed by including the longitudinal measurements as a time-varying covariate in the Cox proportional hazards model, or by using joint models for the longitudinal and time to event data. However, less attention has been given to the analysis of binary or categorical repeated measurements and time to event data, and currently no software is available to estimate the joint models. Practical considerations for the analysis of such data will be presented and illustrated using data from an observational study evaluating the effect of treatment induced amenorrhea, a binary covariate evaluated at regular time intervals during treatment, on progression-free and overall survival in premenopausal breast cancer patients.

E589: **Informative estimation and selection of correlation structure for longitudinal data**
Presenter: **JIanhui Zhou**, University of Virginia, United States
Co-authors: Annie Qu
Identifying informative correlation structure is important in improving estimation efficiency for longitudinal data. We approximate the empirical estimator of the correlation matrix by groups of known basis matrices which represent different correlation structures, and transform the correlation structure selection problem to a covariate selection problem. To address both the complexity and informativeness of the correlation matrix, we minimize an objective function which consists of two parts; the difference between the empirical information and a model approximation of the correlation matrix, and a penalty which penalizes models with too many basis matrices. The unique feature of the proposed estimation and selection of correlation structure is that it does not require the specification of the likelihood function, and therefore it is applicable for discrete longitudinal data. We carry out the proposed method through a group-wise penalty strategy which is able to identify more complex structures. The proposed method possesses the oracle property and selects the true correlation structure consistently. In addition, the estimator of the correlation parameters follows a normal distribution asymptotically. Simulation studies and a data example confirm that the proposed method works effectively in estimating and selecting the true structure in finite samples, and it enables improvement in estimation efficiency by selecting the true structures.

E270: **Spatial mapping of ground-based ozone observations with the SPDE approach**
Presenter: **Serge Guillas**, University College London, United Kingdom
Co-authors: Kai-Lan Chang
Modelling in spatial statistics are usually specified through the covariance function. The Matérn covariance function is a popular choice for modelling dependence in spatial data. However, standard Matérn covariance models are often computationally demanding for large data sets. We employ a method for explicit, computationally efficient, continuous Markov representations of the Matérn covariance family, using the fact that a random process on $\mathbb{R}^d$ with a Matérn covariance function is a solution to the stochastic partial differential equation (SPDE) $(\kappa^2-\Delta)^{\alpha/2}X(s)=\phi\mathcal{W}(s)$, where $X(s)$ is the latent field, $\mathcal{W}(\mathbf{s})$ is Gaussian white noise and Δ is the Laplace operator. For more general manifolds Ω, such as the sphere, the spectral and Matérn representations are not generally applicable, but the SPDE formulation provides a natural generalisation. It facilitates nonstationary extensions by allowing the SPDE parameters to vary with space. We apply this technique to the Total Column Ozone (TCO) data from World Ozone and Ultraviolet Radiation Data Centre. Comparing with the ordinary spatial kriging, the SPDE model has better fit for the generalised cross-validation criterion. Spatial kriging shows misfit and over-optimistic predictions of uncertainties.

ES44 Room 13 APPLIED STATISTICS **Chair: Paula Camelia Trandafir**

E176: **Optimal linear discriminant functions based on MNM**
Presenter: **Shuichi Shinmura**, Seikei, Japan
Fisher proposed the linear discriminant function (LDF). After Fisher, several methods have been proposed such as the quadratic discriminant function (QDF), the logistic regression analysis etc. We introduce an Optimal Linear Discriminant Function based on the MNM (Minimum Number of Misclassifications) criterion. This is defined by integer programming. Therefore, it is called as Revised IP-OLDF. Revised IPLP-OLDF is a faster algorithm to look for the estimates of MNM. Four kinds of re-sampling data are generated by certain real data such as Iris data, Swiss bank note data etc. The mean of error rates of the Revised IPLP-OLDF are compared with those of LDF and the logistic regression by 100 fold cross validation for 135 different models. In many cases, the former are less than those of the latter for the training and evaluation data. LDF and QDF cannot recognize the linearly separable data. No method recognizes Swiss bank data is linearly separable, nevertheless the Revised IP-OLDF indicates this is linearly separable by two independent variables. In addition, LDF and QDF cannot discriminate the pass/ fail determination of exams.

E177: **Quadrature methods for optimal design of blocked and split-plot experiments**
Presenter: **Kalliopi Mylona**, University of Southampton, United Kingdom
Co-authors: Peter Goos
Many industrial experiments involve one or more restrictions on the randomization. The most commonly used experimental designs in those cases are blocked designs and split-plot designs, where the experimental runs are performed in groups. In our algorithm for generating optimal blocked and split-plot designs, we implement several lesser-known but computationally efficient quadrature approaches for the numerical approximation of a Bayesian optimal design criterion. Quadrature methods are widespread numerical integration techniques. We study several quadrature rules suitable for the evaluation of a one-dimensional integral, for the numerical approximation of the Bayesian optimal design criterion. Specifically, we applied Gauss-Jacobi quadrature, Gauss-Stieltjes-Wigert quadrature and generalized Gauss-Laguerre quadrature. We compare the different quadrature methods in detail to Monte Carlo sampling.

E977: **CDKL- optimality criteria**
Presenter: **Paula Camelia Trandafir**, Universidad Publica de Navarra, Spain
Pharmacokinetics is the study of the evolution in time of the chemical concentration and distribution of administered substances in the various tissues and fluids of a living organism, and of the mathematical relations needed to develop adequate models that interpret the corresponding data. One uses the theory of optimal design to carry out an exhaustive study of these models, taking into account the requirements posed by each particular situation. These include: estimation of the parameters (D-optimality), estimation of a parametric function such as the area under a curve in models of drug absorbance (C-optimality), and discrimination between models (KL-optimality). The models are all non-linear and with a non-Gaussian distribution. We give an equivalence theorem and an iterative algorithm that converges to the optimal design satisfying all of the above criteria, namely, CDKL-optimality, and conclude with some numerical examples of pharmacokinetic models.

E1029: **A covariance-based test for shared frailty in multivariate lifetime data**
Presenter: **Alan Kimber**, University of Southampton, United Kingdom
Co-authors: Shah-Jalal Sarker
The score statistic for testing for shared finite variance frailty in multivariate lifetime data is decomposed into marginal and covariance-based terms. The null properties of the covariance-based statistic are derived for parametric lifetime models. Its non-null properties are compared with those of some competing tests when the underlying lifetime distribution is Weibull. Some examples are used to illustrate the covariance-based test. A case is made for using the covariance-based statistic as a simple diagnostic procedure for shared frailty in an exploratory analysis of multivariate lifetime data and a link to the bivariate Clayton-Oakes copula model is demonstrated.

ES61 Room 12 IMPRECISE PROBABILITIES AND DATA ANALYSIS Chair: Enrique Miranda

E252: **Prior near-ignorance for inferences in the k-parameter exponential family**
Presenter: **Alessio Benavoli**, IDSIA, Switzerland
Co-authors: Marco Zaffalon
A model of prior ignorance about a multivariate variable based on a set of distributions M is proposed. We discuss four properties that a model of prior ignorance should satisfy: translation invariance, near-ignorance, learning and convergence. Translation invariance is important because it guarantees that the set of posterior distributions, obtained by combining via Bayes' rule M and the likelihood model, is invariant to shifts of the sample mean, which is a desirable property for inferences derived from a model of prior ignorance. Near-ignorance ensures that our prior model behaves as a vacuous model with respect to some statistical inferences (e.g., mean, credible intervals, etc.). Learning and convergence ensure that our prior model can learn from data and, in particular, that the influence of M on the posterior inferences vanishes with increasing numbers of observations. By considering the exponential families, we show that these four properties can all be satisfied by a set of conjugate priors. The obtained set M is a model of prior ignorance with respect to the functions (queries) that are commonly used for statistical inferences and, because of conjugacy, it is easy to elicit and tractable. Applications of the model to some practical statistical problems are presented.

E876: **Different strategies to combine binary imprecise classifiers**
Presenter: **Benjamin Quost**, Université Technologique de Compiegne, France
Co-authors: Sebastien Destercke
A simple strategy for combining binary classifiers with imprecise probabilities as outputs is proposed. It consists in computing a set of probability distributions that are compatible with the classifier outputs, by solving an optimization problem with constraints that are defined by these outputs. Consistency problems arise when the classifier outputs are jointly incoherent: then, the set of probability distributions satisfying all the constraints is empty. We propose to restore the consistency by discounting the classifier outputs. More precisely, we compute the smallest discounting rates by solving linear optimization problems, thus making the set of associated constraints non-empty. Two discounting strategies may be used. All constraints may be relaxed to the same level, using a single discounting rate; alternatively, each classifier may be associated with a discounting rate: thus, some constraints may be relaxed more than others. This makes it possible to preserve more information, by concentrating the discounting process on the most conflicting constraints. Experiments show that the local discounting strategy gives very good results, compared both to the global one and to single classifier approaches. Beyond these results, our approach can be applied to any method using binary classifiers. The case of iterative choice ranking will be shortly discussed.

E280: **Imprecise statistical preferences**
Presenter: **Ignacio Montes**, University of Oviedo, Spain
Co-authors: Enrique Miranda, Susana Montes
In decision making under uncertainty, it is not uncommon to encounter situations with vague or conflicting information about the probability distributions of the rewards associated to the different alternatives. In such situations, the elicitation of a unique probability model for each of the alternatives may be difficult, and its use questionable. One of the solutions that have been proposed for situations like this is to model our uncertainty about these probabilities by means of imprecise probability models. Here, two different optimality criteria have been considered: stochastic dominance and statistical preference. The former is based on the comparison of the associated distribution functions, while the latter uses a probabilistic relation and can be seen as a robust alternative to expected utility. These two methods are applied to elicit the preferences between sets of random variables instead of single ones. Some characterizations are provided, and a connection between imprecise statistical preference and the aggregation functions used in voting theory is presented. Finally, a particular case is studied: that of the comparison of random sets. It is proven that under some conditions these methods can be characterized by means of the Choquet integral.

E276: **Stochastic dominance with non-additive measures**
Presenter: **Enrique Miranda**, University of Oviedo, Spain
Co-authors: Alejandro Garcia-Huergo
In decision problems, it is not uncommon to encounter situations of imprecise knowledge about the probabilities of the states of nature associated to the different alternatives. One possible model for this situation is to consider sets of probability measures, which in turn can be equivalently represented by different types of non-additive measures. The aim is to study a number of extensions of the criterion of stochastic dominance to the case where, instead of single probability measures, the credal sets associated to two non-additive measures are to be compared. The cases of coherent lower probabilities, 2-monotone capacities, belief functions and possibility measures are considered. It is shown that the different extensions of stochastic dominance are able to encompass different attitudes towards risk. Moreover, the results can be used to compare random sets by means of stochastic dominance, providing thus a model for dealing with imprecision on the utilities. The results are illustrated by means of Ellsberg's paradox and a number of variations.

ES52 Room 9 MODELS FOR TIME SERIES AND LIFE FORECASTING Chair: Cira Perna

E488: **Inference on the effect of a time-dependent therapy in tumor growth**
Presenter: **Giuseppina Albano**, University of Salerno, Italy
Co-authors: Virginia Giorno, Patricia Roman-Roman, Francisco de Asis Torres-Ruiz
A time non-homogeneous Gompertz-type diffusion process is able to describe the dynamics of a tumor population in which the intrinsic rates are modified by means of a time-dependent therapy. In this context, assuming that the effect of the therapy modifies both the birth and death rates of the tumor cells, we consider a diffusion process $\{X(t), t > t_0 \geq 0\}$ defined in $\mathbb{R}^+$ described by the stochastic differential equation (SDE): $dX(t) = \big[h(t)X(t) - g(t)X(t)\ln X(t)\big]dt + \sigma X(t)\,dW(t)$. Here $h(t)$ and $g(t)$ are positive functions of the time representing the growth and the death rates modified by an exogenous term. Moreover, σ is a positive constant representing the width of random fluctuations and $W(t)$ is a standard Brownian motion. We present a study about this process including: i) The distribution of the process, from the point of view of the SDE's and the Kolmogorov equations, as well as a its main statistical indexes. Also conditions for the asymptotic behaviour are discussed; ii) A procedure in order to estimate the functions $h(t)$ and $g(t)$ when longitudinal data are available. This problem arises in a natural way in experimental studies

since generally it is not immediate to deduce the functional form of these functions, that represent changes in the growth rates starting from a treatment protocol. To fit $h(t)$ and $g(t)$ can be useful in order to understand how a treatment can affect the growth and forecast tumor dynamics. Some simulation results are developed in order to illustrate the study of the process and to show the validity of the proposed procedure.

E520: **Modelling streamflow data by a multi-step estimation approach**
Presenter: **Marcella Corduas**, Univiversity of Naples Federico II, Italy

Conventional fitting methods, relying on the minimization of the one step ahead error sum of squares, may cause the inadequacy of widely used model types (such as ARIMA models) because the main data feature may be missed. Although, other methods, which take multi-step ahead forecast error into account, have been considered, the main focus of such approaches is still prediction instead of fitting. In this respect, a feature matching approach to estimation has been recently proposed. Specifically, this introduces an alternative criterion for time series model estimation based on a weighted sum of least squares loss functions referred to L-step ahead forecast errors (for varying horizons). The feature matching approach is investigated with reference to streamflow series modelling in order to improve data fitting and the description of the observed dynamics. Interest is specifically addressed to the implementation of an approach which can lead to the construction of simulated series. Some methodological problems will be investigated and a special attention will be paid to the choice of the weights needed for the definition of the estimation criterion to be optimized.

E619: **Insurance architecture within a risk-profit sharing structure**
Presenter: **Marilena Sibillo**, University of Salerno, Italy
Co-authors: Emilia Di Lorenzo, Albina Orlando

From the actuarial valuation viewpoint, the participating business, particularly if connected to a pension annuity system, gives rise to a complex contractual structure. The participating survival-indexed annuities within a riskiness context are introduced, taking into account both financial and demographic risk components, with the aim of a risk-adjusted performance measurement based on the portfolio surplus. The surplus is the main quantity to be defined, valued and then distributed in a plan spreading over a wide time interval. The annuitants participate in both the investment profits and in the mortality experience, in this way transferring the demographic risk, and not only the financial risk, to capital markets so considering survival as an asset as well. The Actuarial Return on Equity (AROE) and the Actuarial Return on Gain (ARG) are proposed as suitable and meaningful profitability indexes. The numerical evidences are centred on the outputs of simulation procedures, employed to the financial and demographic processes involved in index valuations. The simulated AROE and ARG values have been got throughout the contract duration: several illustrations are shown and the conditional quantile analysis is developed for descripting the worse scenario situation for addressing management strategies.

E621: **Forecasting healthy life expectancy**
Presenter: **Maria Russolillo**, University of Salerno, Italy
Co-authors: Valeria D'Amato

Developed countries' pension systems are currently being severely challenged by two complementary trends: ageing population and growing longevity. This phenomenon has twofold impact on public finances and company business. In particular, the demographic change involves the firms' workforce: the companies have as employers older workers and are marketers of goods and services to older consumers. The combination of greater longevity and falling birth rates poses many problems for individuals approaching retirement, for societies and governments dealing with the rising pension and healthcare costs. As the longevity increases, strongly age-related conditions such as chronic diseases and disability will consequently become more prevalent. In this context, it is crucial to get a true picture of the longevity trend and especially of the population health. The aim is to investigate computational methods for forecasting Healthy Life Expectancy and Disability-free Life Expectancy weighting the population distribution in terms of region, age group and sex. Healthy Life Expectancy, is defined as expected years of life either in good or fairly good health or free from limiting long-standing illness. The empirical results are presented using a range of graphical analyses.

ES57 Room 8 NONPARAMETRIC INFERENCE FOR LEVY PROCESSES Chair: Frank van der Meulen

E425: **Confidence sets in nonparametric calibration of exponential Lévy models**
Presenter: **Jakob Soehl**, Humboldt-Universitaet zu Berlin, Germany

The calibration of exponential Lévy models is considered, that is the log price of an asset is described by a Lévy process. We assume that the Lévy process has a jump component of finite intensity and that the jump measure is absolutely continuous. The observations of the calibration method are prices of European put and call options with fixed maturity and different strike prices. First, the volatility, the drift and the intensity are estimated and then nonparametrically the jump density. The estimators are based on a cut-off scheme in the spectral domain. We show that the estimators of the volatility, the drift and the intensity are asymptotically normally distributed. We also derive asymptotic normality for the pointwise estimation of the jump density at finitely many points and study the joint distribution of these estimators. Finally, the results on the asymptotic distribution of the estimators allow us to construct confidence intervals and confidence sets. We assess the finite sample performance of the confidence intervals by simulations and apply them to empirical data.

E458: **Nonparametric inference for discretely sampled Lévy processes**
Presenter: **Shota Gugushvili**, Vrije Universiteit, Netherlands

Given a sample from a discretely observed Lévy process $X = (X_t)_{t\geq 0}$ of finite jump activity, we consider the problem of nonparametric estimation of the Lévy density ρ corresponding to the process X. We propose an estimator of ρ that is based on a suitable inversion of the Lévy-Khintchine formula and a plug-in device. Our main results deal with upper risk bounds for estimation of ρ over suitable classes of Lévy triplets. The corresponding lower bounds are also discussed.

E717: **On the correlation between jumps in price and volatility**
Presenter: **Gernot Mueller**, Technische Universitaet Muenchen, Germany
Co-authors: Jean Jacod, Claudia Klueppelberg

Some modern continuous-time models which are designed to describe time series from financial markets postulate correlation or even fixed functional relationships between price and volatility jumps. But are jumps in price and volatility correlated at all? To answer this question an asymptotic test is developed. The test only investigates the bigger jumps of the observable price process and uses local volatility estimates calculated from observation windows before and after the jumps under consideration. The performance of the test is checked in a simulation study. Finally, the test is applied to several high-frequency data sets.

ES73 Room Auditorium MIXTURE MODELS Chair: Marco Riani

E371: **Avoiding spurious local maximizers in mixture modeling**
Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain
Co-authors: Alfonso Gordaliza, Carlos Matran, Agustin Mayo-Iscar
The maximum-likelihood estimation in mixture fitting problems is often an ill-posed problem that is treatable, in practice, through the EM algorithm. The existence of singularities and non-interesting local maximizers, often called spurious solutions, make the choice of a suitable solution a hard problem for non-expert practitioners. A generalization to the multivariate case of Hathaway's constrained mixture fitting approach is considered. We give mathematical evidence of the appropriateness of the approach, but, which is more relevant in practice, we also give an algorithm that makes it feasible the monitoring of the solutions in terms of the constant involved in the restrictions. This leads to a natural way to discard spurious solutions, thus, being a valuable tool for the practitioner in this framework. A trimming approach can be also considered in order to avoid this spurious local maximizers.

E394: **Robust clustering of regression lines in highly dense regions**
Presenter: **Andrea Cerioli**, University of Parma, Italy
Co-authors: Domenico Perrotta, Francesca Torti
The aim is to address clustering issues in presence of densely populated data points exhibiting linear structures with high degree of overlapping, which often arise in the analysis of international trade data. To avoid the disturbing effects of highly dense areas, we retain (and then cluster) a sample of data with a process preserving the general structure of the observations. This approach can be beneficial for clustering data with traditional as well as robust techniques, such as TCLUST or the Forward Search. In particular, when the groups exhibit linear structures, data can be assigned to the components of a linear mixture model estimated robustly on the retained subset. Datasets of declarations collected by the Customs offices from international traders and simulated data are used to show the properties of the suggested approach.

E393: **Random coefficient based dropout models: A finite mixture approach**
Presenter: **Alessandra Spagnoli**, Sapienza Universita di Roma, Italy
Co-authors: Marco Alfo
In longitudinal studies, subjects may be lost to follow up (a phenomenon which is often referred to as attrition) or miss some of the planned visits thus generating incomplete responses. When the probability for nonresponse, once conditioned on observed covariates and responses, still depends on the unobserved responses, the dropout mechanism is known to be informative. A common objective in these studies is to build a general, reliable, association structure to account for dependence between the longitudinal and the dropout processes. Starting from the existing literature, we introduce a random coefficient based dropout model where the association between outcomes is modeled through discrete latent effects; these latent effects are outcome-specific and account for heterogeneity in the univariate profiles. Dependence between profiles is introduced by using a bidimensional representation for the corresponding distribution. In this way, we define a flexible latent class structure, with possibly different numbers of locations in each margin, and a full association structure connecting each location in a margin to each location in the other one. By using this representation we show how, unlike standard (unidimensional) finite mixture models, an informative dropout model may properly nest a non informative dropout counterpart. Sensitivity measures will also be discussed.

E676: **Identifying nonstandard group shapes in mixture model clustering through the mean shift algorithm**
Presenter: **Jose E Chacon**, Universidad de Extremadura, Spain
When mixture model analysis is used for clustering it is well known that in many cases clusters are not in one-to-one correspondence with mixture components. For instance, several normal components may be needed to model a single cluster with a skewed shape. Various solutions have been proposed to adapt mixture modeling methodology to the goal of cluster analysis: some of them involve changing the maximum likelihood (ML) objective criterion to a clustering-oriented alternative, like the integrated completed likelihood; but the purpose of most of them is to provide mechanisms to determine when two or more of the mixture components of the model fitted by ML should be merged, on the basis of some proximity criterion. The approach to be presented has also the model obtained by ML as its starting point, but instead of merging clusters, each data point is shifted along the steepest path towards a mode of the estimated density, and all the data points eventually converging to the same mode are assigned to the same cluster, according to the modal concept of cluster. To illustrate the procedure, a comparison of this proposal with the existing techniques on simulated data and real data examples is shown.

ES21 Room 10 STATISTICAL GENETICS/BIOINFORMATICS Chair: Ayse Ulgen

E126: **Biological interpretation of data in the context of GWAI studies**
Presenter: **Kristel Van Steen**, University of Liege, Belgium
From an evolutionary biology perspective, epistasis is a natural phenomenon: phenotypes can only be buffered against the effects of mutations, in the presence of an underlying genetic architecture of networks of genes that are redundant and robust. This is contrasted against the relatively poor detection or replication rate of epistasis discoveries, despite some of these studies being extremely thorough and involving adequate sample sizes. The sharp contrast clearly indicates the network complexity with which multifactorial traits are regulated. It also indicates our potential inability to appropriately apply the available methods. One of the problems is that their abundance and diversity, their differential aims and set-up, complicate comparative studies. Another is the lack of flexibility of software implementations; choosing one approach usually implies the same model assumption throughout the genome. We show that capturing complex phenomena such as epistasis using genome-wide data requires 1) the integration of biological information, 2) appropriate method selection criteria and 3) an ensemble view that exploits complementary benefits of several (equally powerful and Type I error controlling) methods, while reinforcing signals from "performance alike" epistasis detection methods. We illustrate our argumentation via simulated data and a variety of real-life genome-wide data applications.

E530: **Cancer genomes: Development of novel methods for the identification of somatic mutations**
Presenter: **Xose Puente**, Universidad de Oviedo, Spain
Cancer is one of the leading causes of death worldwide. However, the molecular mechanisms involved in the development of a tumor are still poorly understood. The recent development of novel sequencing technologies has opened the possibility to determine the genome sequence of an individual for many laboratories. Thus, by comparing a tumor genome with the genome of non-tumor cells from the same patient it is possible to identify the genetic alterations present in a tumor's genome. However, the analysis of cancer genomes still represents a challenge, due to the complexity of tumor populations and the limitations of sequencing technologies. The development of cancer genome analysis tools has lead to the first comprehensive analysis of the mutational landscape of chronic lymphocytic leukemia, the most frequent leukemia in Western countries. This study has allowed for the identification of novel oncogenes that contribute to the clinical evolution of the disease, and revealed that cancer genome analysis may eventually result in the identification of new therapies for this frequent type of human leukemia as well as other cancers.

E635: **Survival analysis: Finding relevant epistatic SNP pairs using model-based multifactor dimensionality reduction**
Presenter: **Francois Van Lishout**, University of Liege, Belgium
Co-authors: Celine Vens, Victor Urrea, M. Luz Calle, Louis Wehenkel, Kristel Van Steen
Analyzing the combined effects of genes (and/or environmental factors) on the development of complex diseases is quite challenging, both from the statistical and computational perspective, even using a relatively small number of genetic and non-genetic exposures. Several data-mining methods have been proposed for interaction analysis, among them, the Multifactor Dimensionality Reduction Method (MDR). Model-Based Multifactor Dimensionality Reduction (MB-MDR), a relatively new dimensionality reduction technique, is able to unify the best of both nonparametric and parametric worlds, and has proven its utility in a variety of theoretical and practical settings. Until now, MB-MDR software has only accommodated traits that are measured on a binary or interval scale. Time-to-event data could therefore not be analyzed with the MB-MDR methodology. MB-MDR-3.0.3 overcomes this shortcoming of earlier versions. We show the added value of MB-MDR for censored traits by comparing the implemented strategies with more classical methods such as those based on a parametric regression paradigm. The simulation results are supplemented with an application to real-life data.

E536: **Genetic versus environmental trends on asthma related respiratory disorders**
Presenter: **Ayse Ulgen**, University of Southern Denmark, Denmark
Co-authors: Matt McGue
A previous genome scan conducted for asthma-related phenotypes in 295 EGEA families ascertained through asthmatic probands indicated potential linkage of 21q21 region to two asthma-related phenotypes: %predicted FEV1 (forced expiratory volume in 1 second) and SPTQ (number of positive skin test responses to 11 allergens). Principal components analysis of genome-wide lod scores also indicated that these phenotypes may share genetic determinants. To further investigate whether the 21q21 region may contain a quantitative trait locus (QTL) with pleiotropic effect on %FEV1 and SPTQ, we conducted a bivariate linkage analysis using two approaches (1) a bivariate variance components (VC) analysis; and (2) A combined principal components (CPC) analysis; using the ACT program (http://www.epigenetic.org/Linkage/act.html). A fine map of 13 microsatellites (average spacing of 3 cM) covering the whole chromosome 21 was built. A probit transformation was applied to %FEV1 and SPTQ to make the distribution of each of these traits close to normal. Further, family-based association analysis (FBAT) were conducted with PC2 obtained from the original traits as well as %FEV1 and SPTQ, with the microsatellite markers. Additional to the microsatellite data, association analysis with candidate genes in close proximity were also conducted. Univariate linkage analyses showed suggestive evidence of linkage to %FEV1 at the two microsatellites at 25.5 cM and at 43 cM and tentative evidence of linkage to SPTQ at 24.4 cM and (LOD=1.1, =0.01) at 41.1 cM. Bivariate VC linkage analysis of these phenotypes led to an increase in linkage signals with a maximum observed value of ($p = 7.10^{-}5\%$) at 25.5cM and ($p = 4.10^{-}5$) at 43 cM. The peaks obtained by the CPC and by constraining the correlation parameter in the VC method led to similar results as the full bivariate VC approach, significance coming mainly from PC2 at two positions ($p = 7.10^{-}5$ at 25.5 cM and $p = 2.10^{-}5$ at 43 cM). These results indicate that at least two genetic factors close to the 43 cM linkage peak may be involved in SPTQ and %FEV1 variability but further genotyping is needed to better understand these findings.

Monday 03.12.2012 15:00 - 16:40 Parallel Session Q – CFE

CS91 Room 4 COMPUTATIONAL METHODS IN ECONOMETRICS Chair: John M. Maheu

C680: **Sequential Monte Carlo for DSGE models**
Presenter: **Ed Herbst**, Federal Reserve Board, United States
Co-authors: Frank Schorfheide
A Sequential Monte Carlo (SMC) algorithm is adapted for estimating Bayesian Dynamic Stochastic General Equilibrium (DSGE) Models, wherein a particle approximation to the posterior is built iteratively through a tempered likelihood. The algorithm is well suited for the emerging parallel framework of scientific computing, unlike standard Markov Chain Monte Carlo (MCMC) techniques. Using three examples consisting of an artificial state-space model, a standard New Keynesian DSGE model, and a large news shock model we show that the SMC algorithm is better suited for multi-modal and irregular posterior distributions than the widely-used Random-Walk-Metropolis-Hastings algorithm.

C907: **On the use of the free distribution methodology for the analysis of business efficiency**
Presenter: **Felipe Rosa-Gonzalez**, Universidad de La Laguna, Spain
Co-authors: Enrique Gonzalez-Davila, Antonio Arbelo-Alvarez
In the present economic conditions companies should adjust their production systems to obtain the best possible performance. There is no doubt that the elimination of any inefficiency in the management of its resources, costs and profits will improve the operation of the company. Any tool to determine the degree of inefficiency with which a company operates and, furthermore, the power to know about the leader and how inefficient it is, will allow a given firm to raise strategies to correct this situation and, as a consequence, to improve in an economic framework. One method to estimate this inefficiency using data panel is the Free Distribution Method. The aim is to study whether the use of this methodology is always appropriate. Using real data and simulation, a population is generated where the inefficiency was perfectly established. From it variations in the application of the methodology were made and different values of inefficiency were obtained. Finally, different coefficients were calculated in order to determine discrepancies in the measurements with respect to population inefficiency.

C427: **Parallel particle learning for Bayesian state space modeling**
Presenter: **Hiroaki Katsura**, Keio University, Japan
Co-authors: Kenichiro McAlinn, Teruo Nakatsuma
A new parallel algorithm for particle filtering and learning which is suitable for massively parallel computations on GPUs is developed. One of the distinctive features of our new algorithm is that, unlike stratified or systematic resampling, it resamples particles from their exact empirical distribution in massively parallel execution. We refer to this novel resampling procedure as parallel resampling. Furthermore, unlike the previously known parallel filtering algorithms for GPUs, ours does not rely on a specific GPU functionality. Thus it is applicable for any grid computing system in principle. We implement our new algorithm in CUDA and demonstrate that it outperforms conventional non-parallel particle filtering algorithms and it is possible to run particle filtering with millions of particles in seconds. Speed and scalability of our new algorithm would make particle filtering a viable alternative to MCMC and pave the way for real-time updating of the posterior distribution which is necessary in high-frequency trading and other financial applications. As a numerical illustration, we apply our new algorithm to estimation of stochastic volatility models and prediction of volatility indices.

C689: **Bayesian network models of credit risk modeling**
Presenter: **Chee Kian Leong**, University of Nottingham Ningbo, China
Financial risk management is often associated with modeling volatility, portfolio management, pricing of options and bond pricing. Less well-known but no less important is the aspect of credit scoring in consumer lending. Institutions which make monetary loans to customers are exposed to credit risks. A loss occurs when a credit applicant defaults on the loan and the actual loss is the amount at risk and the fraction of this recovered. A default risk can be costly to the lending institutions, as exemplified by the recent sub-prime loan market crisis which resulted in the collapse of many venerable names in the industry. A Bayesian network (BN) model for credit risk scoring is presented. The BN model represents graphically the probabilistic relationships between credit default and its predictors. To ensure parsimony, a Markov blanket correction is applied to the network structure to identify relevant subsets of predictors. Based on the results of both 10-fold cross-validation and a validation sample, the BN model respectively outperforms and matches the logistic regression model in terms of predictive accuracy, sensitivity and precision. The differences in their receiver characteristic curves (ROC) are marginal. Although it is outperformed by a neural network model, the BN model can be built faster and is thus amenable for real-time implementation.

CS96 Room 6 FINANCIAL ECONOMETRICS Chair: Niklas Ahlgren

C674: **Balanced tracking portfolios**
Presenter: **Donatien Tafin Djoko**, University of Neuchatel, Switzerland
Co-authors: Yves Tille
Index-based investment products are becoming increasingly popular among passive managers. So far, empirical studies have focused on complex heuristic-related optimization techniques. We adopt a different perspective and apply a survey sampling framework in the context of stock market tracking. We describe a novel and automatic method that enables us to construct a small portfolio to track the total market capitalization (TMC). The constructed portfolio is randomly selected using a new method of balanced sampling. Empirical studies are performed on the constituents of S&P500. Our findings suggest that balanced sampling portfolios efficiently track the S&P500 TMC.

C711: **Structure in the Italian overnight loan market**
Presenter: **Matthias Raddant**, Institute for the World Economy, Germany
The Italian interbank loan market from 1999 until 2010 is analyzed. The analysis of net trade flows shows a high imbalance caused by a few large net borrowers in the market. The trading volume shows a significant drop starting in 2007, which accelerates with the Lehman default in late 2008. The network, based on trading relationships, is very dense. Hence, we try to identify strong links by looking for preferential lending relationships expressed by discounts in the loan rate. Furthermore, we estimate the dynamics of credit spreads for each bank and find that economically significant spreads for the overnight market only developed in 2010. The analysis of bilateral loan relationships reveals that in the pre-crisis era large net borrowers used to borrow at a slight discount. In the post-Lehman era borrowers with large net exposures paid more than the average market rate, which shows that the risk evaluation of market participants has changed considerably.

C863: **A closed-form solution for arithmetic Asian options under a mean reverting jump diffusion model**
Presenter: **Shing Fung Chung**, The Chinese University of Hong Kong, China
Co-authors: Hoi Ying Wong
There is empirical evidence that commodity prices are mean reverting and exhibit jumps in prices. As some commodity option payoff involves the arithmetic average of historical commodity prices, a closed-form solution to arithmetic Asian options under a mean reverting jump diffusion

process is derived. The closed-form solution is implemented with the fast Fourier based on the joint characteristic function of terminal asset price and the realized averaged value. The accuracy and computational time for the model are examined through an empirical study on commodity products.

C688: **The flat tax in post-communist Europe**
Presenter: **Wayne Tarrant**, Wingate University, United States
Since 1994 sixteen post-communist European nations have adopted some type of flat tax. In some of these countries the personal and corporate tax rates are identical, while others have differentiation between the rates. Among other variables we particularly consider GDP growth, tax revenues, Gini coefficients, and sizes of underground economies before and after the adoption of the flat tax. We then make comparisons across the different flavors of the flat tax that different countries have adopted. This is a preliminary report with a series of papers forthcoming.

C1026: **Probability-possibility transformation and application to high frequency financial returns modeling**
Presenter: **Alfred Mbairadjim Moussa**, Montpellier I, France
Co-authors: Jules Sadefo Kamdem, Michel Terraza
The connection between possibility theory and probability theory has been largely discussed over these last decades. A possibility measure allows encoding a family of probability measures by generalizing the notion of best interval substitute to a probability distribution with prescribed confidence. This transformation is interesting for statistical reasoning when the uncertainty has two important sources: the variability of observations and the incompleteness of information. Therefore, converting a probability distribution into a possibility distribution may be useful in the presence of weak sources of information and imprecision in the observed data. This situation often holds with financial high frequency data because of microstructure noise and the imperfections of the trading process. The aim is handling both with these two sources of uncertainty of financial high frequency returns by a probability/possibility transformation. Then, we introduce some moments of the resulting possibility distribution in order to give a new characterization of returns distribution. Mathematical properties of these moments are discussed and comparative analysis with probabilistic moments is made. Finally, an application illustrated with empirical examples, is carried out for performance evaluation.

CS75 Room Multipurpose FINANCIAL TIME SERIES II Chair: Bent Jesper Christensen

C192: **Financial forecasting accuracy: Exploring the M3-competition**
Presenter: **Ana Jesus Lopez**, University of Oviedo, Spain
Co-authors: Rigoberto Perez
Forecasting availability has widely increased, suggesting the need of analyzing the adequacy of different alternative methods. One of the main empirical researches in this field is the M-Competition, whose last edition (M3) is referred to year 2000 and includes 3003 time series (308 of them related to finance) and 24 forecasting techniques. Since the statistical theory of information provides a suitable framework for the evaluation of econometric models, including their goodness of fit and their forecasting ability, we propose the use of two information-based accuracy measures: the U Theil Index and the Quadratic Information Measure. The obtained results are compared with those provided by the five accuracy measures included in the M-competition: the symmetric mean absolute percentage error, the average ranking, the percentage better, the median symmetric absolute percentage error, and the median relative absolute error). A sensitivity analysis is presented.

C840: **Stock market linkages: New evidence on the dynamics of information transmission**
Presenter: **Robert Maderitsch**, University of Hohenheim, Germany
Co-authors: Robert Jung
Our contribution relates to the ongoing debate on stock market interdependence, spillovers and contagion. We investigate the structural stability in the dynamics of information transmission between markets is Asia, Europe and the United States. Using high frequency time series for the Hang Seng Index, the Euro Stoxx 50 and the S&P 500, we analyse both return and volatility transmission. Within a structural VAR context, we consider the trading chronology in the three markets and employ regressions for subsamples, rolling regressions, structural break tests, Markov-Switching and Threshold Regressions. Concerning return transmission, significant spillover effects as well as changes in the process of information transmission are hard to find over the whole sample. Though, for volatility transmission, significant effects are found for all markets over the whole time. Abrupt changes in volatility transmission originating in the United States indicate evidence for contagion during the financial crisis. The overall evidence for our given sampling period suggests that international volatility transmission is state- and threshold-dependent, whereas return transmission is rather stable and insignificant.

C756: **A nonparametrically improved sharpe ratio guided by prior knowledge**
Presenter: **Michael Scholz**, University of Graz, Austria
Co-authors: Jens Perch Nielsen
A two step approach is proposed to improve one of the most used measures of performance evaluation for investment strategies, the Sharpe ratio. Expected returns and their variability in financial markets vary over time and contain a significant predictable component. Thus, we first predict excess stock returns in a heteroscedastic regression model applying a local-linear kernel smoother improved by prior knowledge, and second we repeat the approach for the conditional variance function. To measure the quality of the prediction model, we apply the validated R^2, where the prediction of our cross-validated model is compared to the simple cross-validated mean. A ratio of the estimates of both steps gives our new estimator for which we also provide statistical properties. Based on this Sharpe ratio estimator, we propose an advanced trading strategy that is superior to different often applied simple investment plans. In an applied part, we show the efficiency of our estimator and strategy using monthly data from the S&P500 in a period from 1953 to 2010.

C971: **Tail dependence breaks in financial time series**
Presenter: **Julia Schaumburg**, Leibniz Universität Hannover, Germany
A test for structural changes in the tail dependence function, or tail copula, of bivariate random variables is introduced. In finance, conducting multivariate extreme value analysis for portfolios that contain different types of assets is widely accepted to assess the risk of large comovements. Most studies consider, firstly, only the tail dependence coefficient which corresponds to the point $(1,1)$ on the copula, and secondly, they assume that the tail dependence of (sometimes prefiltered) returns is constant over time. Crises, policy shifts, or other far-reaching events may, however, change the overall joint extreme behavior of two portfolio components, which is captured by the tail copula. Such structural dependence breaks directly affect potential losses, measured e.g. by Value at Risk or Expected Shortfall, even when the marginal distributions remain the same. We examine the properties of the test in a simulation study comparing different estimators for the variance of the test statistic. Furthermore, we show in a detailed empirical application that the bancruptcy of Lehman Brothers on September 15, 2008 brought along a significant increase in the tail dependence between U.S. equity and CDS markets.

C926: **Unfolding GARCH models**
Presenter: **Richard Luger**, Georgia State University, United States
Co-authors: Xiaochun Liu

Any GARCH model for financial returns specified with innovations following a skewed distribution implicitly makes an assumption about the non-linear dependence between the signs and absolute values of returns. We propose a new approach that decomposes returns into their signs and absolute values, and specifies the joint distribution by combining a multiplicative error model for absolute values, a dynamic binary choice model for signs, and a copula for their interaction. The multiplicative error component may be any GARCH model with a folded (or absolute) innovation distribution. The kurtosis parameter associated with the folded distribution is allowed to vary over time, and the copula parameter, which governs return skewness, is also rendered dynamic. We adopt a Bayesian methodology to make inference and we illustrate the approach using simulated data. The new model is then applied to the daily returns on a major stock market index and compared to a GARCH specification with skewed Student-t innovations, whose degrees-of-freedom and asymmetry parameters are also conditionally time varying.

CS31 Room 5 CONTRIBUTIONS TO TIME-VARYING PARAMETER MODELS Chair: Rodney Strachan

C760: **Nonparametric calibration of time-varying coefficient realized volatility models**
Presenter: **Xiangjin Bruce Chen**, Monash University, Australia
Co-authors: Jiti Gao, Degui Li, Param Silvapulle

The aim is to investigate the nonparametric calibration of time-varying coefficients of heterogeneous autoregressive (TVC-HAR) model for realized volatility of the S&P 500 stock index. We employ both the local linear and local constant methods for estimating the developed model, and construct the confidence intervals for time-varying coefficient functions by using the bootstrap method. Then, based on the generalized maximum likelihood ratio and wild bootstrap method, we provide statistical evidence for rejecting the simple linear heterogeneous autoregressive (HAR) model against the nonparametric TVC-HAR alternative. Furthermore, we consider calibrating the coefficients of the TVC-HAR model by using the approximation of the polynomial functions of time. The optimal calibrated specification is chosen via an automated model selection algorithm, which is developed based on a L2-type function criterion. Lastly, the calibrated TVC-HAR model is evaluated as a forecasting model, against the simple HAR. Using the conditional predictive ability test, we conclude that the calibrated TVC-HAR model outperforms the simple HAR model as a point volatility forecasting model.

C925: **Modelling and forecasting implied volatility surface dynamics**
Presenter: **Kai Ming Lee**, Ortec Finance Research Center, Netherlands

Traded options on a given stock with different expirations and degrees of moneyness give rise to a surface of Black-Scholes implied volatilities. Most research on implied volatility surfaces has been focused on modeling underlying asset prices, typically using diffusion processes, consistent with static surface shapes. However, the historical dynamics of these surfaces exhibit considerable variation, which may be better captured using time series oriented approaches. Recent advances in high dimensional maximum likelihood dynamic factor estimation and factor loading reduction methods make dynamic factor models an attractive candidate for modeling implied volatility surfaces. Alternatively, direct polynomial functions of moneyness and expiration with time varying coefficients can model dynamic smooth surfaces in a simple way with relatively few parameters. In-sample model specification and pseudo out-of-sample forecasts for both model types are easily analysed and compared using standard state space techniques.

C943: **Bayesian model averaging and forecasting using self-perturbing Kalman filters**
Presenter: **Nima Nonejad**, AArhus and Creates, Denmark
Co-authors: Stefano Grassi, Paolo Santucci de Magistris

A novel approach to forecasting time series models which are subject to changes in the parameters is proposed. The approach is a flexible way of modeling changes in the model parameters without requiring the use of Markov chain Monte Carlo (MCMC) methods. Fast tracking and low parametric error properties are verified via simulations. Furthermore, model uncertainty is accounted for using a dynamic extension of Bayesian model averaging (DBMA) which leads to major gains in computation time while at the same time handling a huge number of models. The ranking of the models can change over time and DBMA provides an optimal way to combine this information allowing for the entire forecasting model to change over time while at the same time allowing for the parameters of each model to change over time as well.

C916: **Time varying SVARs, parameter histories, and the changing impact of oil prices on the US economy**
Presenter: **Francesca Rondina**, Institute for Economic Analysis - CSIC and Barcelona GSE, Spain

The changes in the relationships between oil prices and US domestic variables in the postwar period is studied using a time-varying VAR model of the economy. A new framework is proposed, which departs from most of the previous literature with respect to the assumptions on the evolution of the covariance matrix of the reduced form vector of innovations. The VAR model resulting from these assumptions features multivariate stochastic volatility that follows a Wishart autoregressive process. Relative to the standard procedure, this framework is able to capture a larger variety of time-varying features of the data, and provides some additional insights on the intertemporal dependence between the parameters of the VAR. In the environment under analysis, it is found that in specific periods, in particular in the early/mid 1980s, changes in the volatility of the model variables are due to temporary shifts in their contemporaneous relationships rather than to changes in the variance of the exogenous shocks. In addition, the covariances between oil inflation and US core inflation and oil inflation and the US growth rate exhibit a strong intertemporal dependence, which might signal some degree of endogeneity of oil prices to US domestic variables.

C609: **Forecasting GPD growth in the Eurozone**
Presenter: **Jasper de Winter**, De Nederlandsche Bank, Netherlands
Co-authors: Xiaowen Jin, Jos Jansen

A systematic comparison of the short-term forecasting performance of linear statistical models in a pseudo-real time setting is conducted using a large-scale dataset of monthly data and quarterly GDP. Our exercise includes various versions of factor models and mixed-frequency models. We also evaluate subjective GDP forecasts by professional forecasters. The empirical work concerns the Eurozone and its five largest countries in the years 1996-2011. We find that pure factor models and factor-augmented versions of mixed-frequency models perform best, in particular in times of crisis. The dynamic factor model consistently outperforms all other approaches, including forecasts by professionals. From a heuristic point of view, our results suggest that summarizing relevant information in the form of factors is a more promising forecasting strategy than averaging alternative forecasts.

CS42 Room 1 CONTRIBUTIONS TO BUSINESS CYCLE ANALYSIS Chair: Toshiaki Watanabe

C686: **Composite leading indicators as nonlinear functions: Quantitative and qualitative forecasts**
Presenter: **Miroslav Klucik**, VSB Technical University of Ostrava, Czech Republic

Possibilities of forecasting of a small and open economy (case of Slovakia) using composite leading indicators as nonlinear functions are being investigated. In contrast to the traditional composite leading indicators used only for qualitative forecasts (linear aggregates with fixed aggregation weights), these indicators are used simultaneously for qualitative and also for quantitative forecasts of the business cycle. The qualitative forecast represents the prediction of business cycle phases. Forecast of the quarterly gross domestic product for one and two quarters ahead represents the quantitative forecast. The near optimal composition of the leading indicator is found heuristically using genetic programming symbolic regression. Time series and their combinations are optimized for best fit and optimal lead against the reference time series. The complexity of the composite indicators is restricted by maximum tree depth of genetic programming individuals. The quantitative forecasts of the leading indicators compete successfully with linear and nonlinear AR models. The proposed methodology for leading indicators construction provides better model fitting and also better qualitative and quantitative forecasts compared to the traditional models. Main drawback is the lack of interpretation of the final leading indicators composition.

C687: **The impact of foreign cyclical development on small open economy (SVAR analysis)**
Presenter: **Jana Juriova**, VSB Technical University of Ostrava, Czech Republic

Sources of business cycle fluctuations in the case of small open economies can be found mainly in foreign environment. The impact of fluctuations of foreign factors on small open economies of Slovak and Czech Republic are analysed and compared using a long-run restrictions SVAR approach. Two different channels of business transmission are considered – export and financial market. Both economies are represented by basic macroeconomic indicators - gross domestic product, inflation and interest rate. The fluctuations in foreign economy are represented by its foreign counterparts for the most important trading partners and world is characterized by the exchange rate and oil price. The most significant trading partners of both economies are selected according to their maximum contribution to the openness of domestic economy. To employ the economic theory, the structural vector autoregression is used as this approach is better suited for small open economies rather than the more traditional identification method of the VAR. The analysis of reactions of domestic macroeconomic indicators to foreign shocks is performed by means of impulse response functions and the intensity of shocks is compared and discussed between both countries.

C799: **Monetary policy rules and business cycle in China-Bayesian DSGE model simulation**
Presenter: **Lixin Sun**, Shandong University, China
Co-authors: Somnath Sen

Using a benchmark Bayesian Dynamic Stochastic General Equilibrium (Bayesian DSGE) model with Taylor's rule and a modified Smets-Wouters model with a money growth rule, we have simulated China's monetary policy transmission process and the roles of monetary variables and non monetary variables in China's business cycle by incorporating many so-called New Keynesian Macroeconomic (NKM) approaches such as nominal stickiness and market imperfections in the model. The estimated values of the parameters in the model by Bayesian approach based on China's quarterly time series data feature the unique characters of China's economy compared with that in the US and the Euro area. The simulation results in terms of the Taylor's rule and money growth rule (MacCullum Rule) highlight the monetary transmission mechanisms of China's monetary policy and the diverse contributions of monetary shocks and non-monetary shocks to China's business cycle.

C828: **The cross-sectional origins of business cycle asymmetries**
Presenter: **Emiliano Santoro**, Catholic University of Milan and University of Copenhagen, Italy
Co-authors: Roberta Distante, Ivan Petrella

Business cycle dynamics are explored from a cross-sectional perspective. Using firm-level growth data, we measure the reactiveness of firms growing at different rates to both macroeconomic and financial shocks. Our quantile panel-data analysis shows that firms in the left side of the distribution, that is firms that are growing more slowly or declining, are typically more affected by aggregate shocks than those in the right side of the distribution. We reconcile this fact with the emergence of asymmetric business cycle fluctuations. Specifically, shifts in the probability mass on either side of the mode are closely connected with asymmetric degrees of deepness and steepness in contractions and expansions. The sectoral analysis reveals marked heterogeneity in the non-linear properties of sectoral business cycles.

C647: **Monitoring U.S. states business cycles synchronization: When and how a recession is coming**
Presenter: **Danilo Leiva-Leon**, University of Alicante, Spain

A new framework is developed for monitoring changes in the degree of aggregate synchronization between many stochastic processes that are subject to regime changes. An empirical application of this methodology to the interconnectedness of business cycle phases in U.S. states suggest that national recessions can be anticipated by an index that accounts for the prominence of states under an interdependent environment, confirming its reliability by means of real time exercises. Moreover the degree of interdependence between states, estimated trough Bayesian methods, can be monitored month-to-month by relying on an empirical network, in order to assess how much state A would be affected by a shock that will hit or is hitting state B or the nation as a whole.

CS51 Room 2 CONTRIBUTIONS TO JUMPS AND VOLATILITIES Chair: Sebastien Laurent

C706: **Inference on self-exciting jumps in prices and volatility using high frequency measures**
Presenter: **Worapree Ole Maneesoonthorn**, Monash University, Australia
Co-authors: Catherine S. Forbes, Gael M. Martin

The dynamic behaviour of price and volatility jumps are investigated. A standard jump diffusion model for price and volatility is augmented by a bivariate Hawkes process for the respective jump intensities. The latter process allows the individual intensities of price and volatility jumps to be correlated over time, and the intensity of volatility jumps to be directly affected by contemporaneous jumps in the price. A state space representation of the model is constructed using financial returns and nonparametric volatility measures as the observable quantities. Bayesian inference, based on a multi-move Markov chain Monte Carlo algorithm, is used to obtain posterior distributions for the relevant model parameters and latent variables, and to analyze various hypotheses about the dynamics in, and the relationship between, the jump intensities, via Bayes factors. Empirical analysis using data constructed from the S&P500 market index is conducted.

C956: **Multiple time scale volatility patterns before abrupt switching in financial markets**
Presenter: **Teruko Takada**, Osaka City University, Japan
Co-authors: Akimitsu Inoue

Though understanding macroscopic abrupt switchings, so called financial bubbles, is an urgent issue for society, its statistical analysis is hampered by insufficient information, due to very limited number of occurrence. Financial markets frequently experience abrupt switchings of trend direction from upward to downward, on time scales ranging from several years to a few minutes. In addition, the volatility patterns before critical transitions have recently gained increased attention as a significant early warning signal. The objective is to exploit information of statistically abundant

smaller scale bubbles which is useful for understanding financial bubbles. We extensively investigate the multiple time scale volatility patterns, based on very long daily and TAQ price data; New York Stock Exchange TAQ data from 1993 to 2010, daily Dow 30 index from 1896 to 2012, and several other daily indices of major stock markets. Using robust statistical methods appropriate for data distributed as fat-tailed, we find several scale-free and scale-dependent patterns, which are consistent with the conflicting reports on volatility patterns before critical transitions.

C675: **Estimating correlated jumps and stochastic volatilities**
Presenter: **Jiri Witzany**, University of Economics in Prague, Czech Republic
A bivariate stochastic volatility jump-diffusion model with correlated jumps and volatilities is formulated. An MCMC Metropolis-Hastings sampling algorithm is proposed to estimate the model's parameters and latent state variables (jumps and stochastic volatilities) given observed returns. The methodology is successfully tested on several artificially generated bivariate time series and then on the two most important Czech domestic financial market time series of the FX (CZK/EUR) and stock (PX index) returns. Four bivariate models with and without jumps and/or stochastic volatility are compared using the deviance information criterion (DIC) confirming importance of incorporation of jumps and stochastic volatility into the model.

C709: **Hidden decisions behind hidden volume: Analysis of incoming iceberg orders in electronic trading**
Presenter: **Vahidin Jeleskovic**, University of Kassel, Germany
If traders want to hide their interest in an electronic market using iceberg orders, they have to make four decisions simultaneously: i) should an iceberg order be used, ii) if yes, how big the displayed, iii) and the hidden part, and iv) how aggressive the order should be. This is a simultaneous stochastic process resulting in new incoming orders that we empirically investigate. To account for the simultaneousness, we analyze the process of these trader's decisions via an unique simultaneous equations system that includes: a logit model (for the first traders' decision) and censored regressions (for the other ones) with explanatory variables. Our empirical analysis is based on the data from Copenhagen Stock Exchange that support the decomposition of total limit order size into its displayed and hidden parts. We find evidence of a clear trade-off between iceberg order aggressiveness and its quantity of the hidden part, on the one side, and the displayed size, on the other side. This is, iceberg orders are more likely to be placed near to the opposite market side while simultaneously the ratio of the hidden size to the displayed one tends to become larger. Moreover, the bigger the hidden size the higher the probability for an iceberg order.

CS84 Room 3 ENERGY MARKETS Chair: Marc Gronwald

C727: **Modelling dependence of extreme price observations in connected electricity markets using tail copulas**
Presenter: **Adebayo Aderounmu**, University of Queensland, Australia
Co-authors: Rodney Wolff, Helen Thompson
Spot electricity prices are very volatile, particularly due to the fact that electricity cannot be economically stored and requires immediate delivery. However, the inability to store electricity means that fluctuations in demand and supply are often transmitted directly into spot prices of electricity, which leads to occasional extreme price observations, so called price spikes. These price spikes constitute a major source of price risk to market participants. More importantly, for those operating in several regional markets simultaneously, the probability of simultaneous extreme price observations, usually called tail dependence, is of great importance in implementing adequate hedging strategies. For this purpose, the problem of modelling the joint occurrence of extreme price observations in the Australian Electricity Market is considered. We suggest a new method to capture the dependence of extreme price observations across several regional markets. It uses the concept of tail copulas as models for different scenarios of joint extreme outcome. For risk management purposes, our findings point out the substantial implications which the joint extreme price observations may have for hedging decisions of market participants, and therefore, also for the pricing of electricity derivatives like futures and option contracts.

C895: **Measuring risk in electricity forward returns**
Presenter: **Christian Lau**, Martin Luther University, Germany
Co-authors: Joerg Laitenberger
The risk of an investment in electricity forwards is analyzed. We compare different ARMA-GARCH models to fit 18 one year forward time series from NASDAQ OMX Commodities Europe. A descriptive analysis shows that heavy tails are an important feature while skewness is negligible. An ARMA(1,1)-GARCH(1,1) model with t distributed innovations gives the best fit in terms of the Bayesian Information Criterion. We use this model in order to evaluate the Value at Risk of a buy-and-hold strategy in electricity forwards.

C762: **Optimal combined forecasts for electricity prices considering renewable energies**
Presenter: **Eduardo Caro**, Technical University of Madrid, Spain
Co-authors: Carolina Garcia-Martos, Maria Jesus Sanchez-Naranjo
A forecasting model for electricity prices is developed, based on the optimal combination of multiple univariate/multivariate time series methods and considering renewable energy generation as possible explanatory variables. Specifically, wind power and hydro generation are considered. The optimal combination is achieved by solving an efficient linear mathematical programming problem which minimizes the mean absolute percentage error of the forecast.

C683: **The relationship between the carbon market and financial markets: A frequency domain analysis**
Presenter: **Marc Gronwald**, Ifo Institute, Germany
Co-authors: Stefan Trueck
The relationship between European carbon prices on the one hand and commodity prices as well as financial market variables on the other is studied by investigating whether or not a Granger causal relationship is present between the carbon price in the EU ETS and prices of various other commodities and financial market variables. As it is a well-known finding that the extent and the direction of causality can differ across frequencies, a frequency-domain causality test is applied. This procedure yields insights as to whether a short- or long-run causal relationship is present between carbon prices and variables such as electricity prices, energy prices and stock market indices. As establishing emission trading schemes is the preferred policy instrument to tackle the problem of climate change, it is of particular importance to have a sufficient understanding of the determinants as well as the behaviour of the prices observed in this market. A particular feature of markets for emission allowances is that these markets are artificial. It is thus interesting to study whether or not the relationship between this kind of markets and genuine markets is different to the relationship between genuine markets.

Monday 03.12.2012 17:10 - 18:10 Parallel Session S – ERCIM

ES83 Room 13 HEALTH DATA AND MEDICAL STATISTICS **Chair: Peter Congdon**

E768: **Statistical estimation problems in causal inference: Equating propensity scores**
Presenter: **Priyantha Wijayatunga**, Umea University, Sweden
In the statistical literature, particularly in medical domains, causal inference problems for observational data became popular with the emergence of such models as probabilistic graphical models and potential outcome based methods. However researchers using these two techniques are opposing to each other. The equivalence of the two methods is shown along with their need of addressing the same estimation problems. In the potential outcome framework attention is paid to the estimation of the so-called propensity scores and an efficient stratification of data sample using them for estimation of treatment effects. This is usually done by equating some of them together, so that each newly formed stratum has almost the same covariate distribution for both treated and non-treated groups where the covariates are the observed confounders of the treatment and outcome variables. Since in medical domains relatively large databases of observational data are common, using probabilistic arguments it is shown how one could equate propensity scores to preserve the requirement. These score ranges depend on marginal probabilities of covariates too. This raises questions on the estimation of propensity scores using discriminative methods such as logistic regression. It is argued that generative estimation methods of probabilities are needed to preserve required balance within strata.

E928: **The effect of aggregation on disease mapping**
Presenter: **Caroline Jeffery**, Liverpool School of Tropical Medicine, United Kingdom
Co-authors: Al Ozonoff, Marcello Pagano
In public health, studying the relationship between an individual's location and the acquisition of disease can serve to prevent further spread of disease or to guide (re)allocation of resources to improve access to care. In health data, spatial information on cases is either available in point form (e.g. longitude and latitude) or aggregated by an administrative region. Statistical methods developed for spatial data can accommodate either form of data, but not always both. In the case of disease mapping, point data or centroids of aggregated regions can serve as spatial location and produce a smoothed map of estimated risk. However the quality of the mapping is affected by how coarse is the resolution of the spatial information. Previous literature has shown that for cluster-detection methods, power tends to decrease as the spatial information on cases becomes coarser. We study the effect of aggregation on a disease risk mapping method, when the method is used to locate an increase in occurrence of cases in one subregion of the study area. Our simulations in the unit disk show that the accuracy of the mapping diminishes as the resolution of spatial information gets coarser.

E983: **Exploring the correlation structure of inter-nucleotide DNA distances**
Presenter: **Sonia Gouveia**, IEETA/UA, Portugal
Co-authors: Vera Afreixo, Manuel Scotto, Paulo Ferreira
DNA is a long sequence of repeated symbols called nucleotides (A, C, G and T), from which four series of inter-nucleotide distances (IND) are obtained from the consecutive distances between equal nucleotides. Previously, the distributions of IND values were shown to exhibit significant differences with the counterpart geometric reference distributions, which would be obtained if nucleotides were placed randomly and independently. In this work, the goal is to explore the possibility that these differences are due to the IND autocorrelation structure. A simulation study was designed from real data gij33286443jrefjNM-032427.1 gene of Homo sapiens (ftp://ftp.ncbi.nih.gov/genomes/) processed in blocks of symbols. Each block was transformed into a binary sequence (0 or 1, according to the nucleotide) and its autocorrelation function was estimated. Afterwards, the binary sequence was scrambled to obtain a random Bernoulli variable with the same success probability and no autocorrelation. This random variable was then colored (with proper filtering/quantization), mimicking the original autocorrelation structure up to an order p. The optimal order was chosen as that leading to non-significant differences between original and colored IND distributions, assessed by chi-square testing adjusted for sample size. This study focuses on the interpretation of the optimal order changes along the gene sequence.

ES90 Room 5 DESIGNS AND SURVEYS **Chair: Kalliopi Mylona**

E765: **Survey estimates by calibration on dual frames**
Presenter: **Maria del Mar Rueda**, Universidad de Granada, Spain
Co-authors: Antonio Arcos, Maria Giovanna Ranalli, Annalisa Teodoro
Survey statisticians make use of the available auxiliary information to improve estimates. One important example is given by calibration estimation, that seeks for new weights that are as close as possible to the basic design weights and that, at the same time, match benchmark constraints on available auxiliary information. Recently, multiple frame surveys have gained much attention and became largely used by statistical agencies and private organizations to decrease sampling costs or to reduce frame undercoverage errors that could occur with the use of only a single sampling frame. Much attention has been devoted to the introduction of different ways of combining estimates coming from the different frames. We will extend the calibration paradigm to the estimation of the total of a variable of interest in dual frame surveys as a general tool to include auxiliary information, also available at different levels. In fact, calibration allows us to handle different types of auxiliary information and can be shown to encompass as a special case the raking ratio method and the pseudo empirical maximum likelihood approach.

E650: **Optimal experimental designs for conjoint analysis: Estimation of utility functions**
Presenter: **Jose M. Vidal-Sanz**, Universidad Carlos III de Madrid, Spain
Co-authors: Mercedes Esteban-Bravo, Agata Leszkiewicz
In conjoint analysis consumers utility functions over multiattributed stimuli are estimated using experimental data. The quality of these estimations heavily depends on the alternatives presented in the experiment. An efficient selection of the experiment design matrix allows more information to be elicited about consumer preferences from a small number of questions, thus reducing experimental cost and respondent's fatigue. Kiefer's methodology considers approximate optimal design selecting the same combination of stimuli more than once. In the context of conjoint analysis, replications do not make sense for individual respondents. We present a general approach to compute optimal designs for conjoint experiments in a variety of scenarios and methodologies: continuous, discrete and mixed attributes types, customer panels with random effects, and quantile regression models. We do not compute good designs, but the best ones according to the size (determinant or trace) of the information matrix of the associated estimators without repeating profiles as in Kiefer's methodology. We use efficient optimization algorithms to achieve our goal.

E985: **D-optimal two-level factorial designs for logistic regression models with bounded design regions**
Presenter: **Roberto Dorta-Guerra**, Universidad de La Laguna, Spain
Co-authors: Enrique Gonzalez-Davila, Josep Ginebra
Under first order normal linear models, the amount of information gathered through two-level factorial experiments, as measured through the determinant of their information matrix, does neither depend on where the experiment is centered, nor on how it is oriented relative to the contour lines of the surface, and balanced allocations are always more informative than unbalanced ones with the same number of runs. Thus, when planning

for two-level factorial experiments for continuous responses and any number of factors, the only thing that matters is the range of variation of the factors involved. Instead, the issues involved when planning for two-level factorial experiments for binary responses are a lot more complicated. D-optimal factorial designs for the logistic regression model with bounded design regions are considered. We provide the designs that maximize the determinant of the information matrix, within the class of two-level factorial experiments centered at a given point, for first order binomial models with either one or two factors. That allows one to explore how the performance of these experiments on binary responses depends on their location, orientation, range, and relative allocation of the total number of runs to each one of the support points.

ES71 Room 12 SPATIAL DATA ANALYSIS **Chair: Jean-Michel Marin**

E463: **On the effect of neighboring populations on the local population growth over time**
Presenter: **Mengjie Han**, Dalarna University, Sweden
Co-authors: Johan Hakansson, Lars Ronnegard

This study extends over a period when society has changed from a pre-industrial agricultural society to an industrial society with mechanisation and wage labour – and, during the latter part of the period, from an industrial to a post-industrial service-producing society. Parallel with this social transformation, major population changes have taken place. The geographical distribution and redistribution of the population has consequently been a constantly recurring research theme within both geography and other disciplines. The research has touched upon both concentration and dispersion, and has structured these phenomena both on local and regional levels. The distribution of population between labor market regions has only undergone marginal changes over roughly the same period of time. That the redistribution of the population mainly has been a local redistribution from the nearby countryside to the cities is also suggested. This implies that the growth of the cities in Sweden is dependent on the regional population potential. This lends support to hypothesis that the local population growth is dependent on how many people there is in the vicinity to area with population growth. Therefore, our aim is to analyze how and to what extent neighboring populations affect the local population growth. The lowest possible level that is feasible to use for population studies of this kind with this time perspective is parish level. The parish division change over time. An unchanged geographical division over time has been constructed. The unchanged parish division consists of 1800 parishes. The non-separable space-time covariance model is also developed to analyze both the changes and the interactions between space and time.

E695: **Performance evaluation for Bayes rule of Gaussian spatial data classification**
Presenter: **Kestutis Ducinskas**, Klaipeda University, Lithuania
Co-authors: Lina Dreiziene

The problem of classification of Gaussian spatial data into two populations under a deterministic training sample design is considered. The populations are specified by different parametric means and common parametric covariance models. Given training sample, ML estimators of the unknown means and covariance parameters are plugged in the Bayes Classification Rule (BCR). The performance of the proposed classification rule is evaluated by the expected error rate. However, the expressions for the expected error rate are intractable even for the simplest forms of classification rules. Therefore, asymptotic approximations of the expected error rate are especially important. The asymptotic results sometimes yield useful approximations to finite-sample properties. The asymptotic properties of ML estimators under increasing domain asymptotic framework are essentially exploited. The approximation of the expected error rate based on Taylor series expansion is obtained. Furthermore, the obtained approximation is proposed as the spatial sampling design criterion. The case of stationary Gaussian random field with exponential covariance function is used for numerical illustration. Comparison with the plug-in Bayes error rate estimator is also implemented by using software system R. The dependence of the performance of the plug-in BCR on the statistical parameters is investigated. Numerical comparison of several spatial sampling designs by the proposed criterion is carried out.

E769: **Divergence-based tests of homogeneity for spatial data**
Presenter: **Tomas Hobza**, Czech Technical University in Prague, Czech Republic
Co-authors: Domingo Morales, Leandro Pardo

The problem of testing homogeneity in contingency tables when the data are spatially correlated is considered. We derive statistics defined as divergences between unrestricted and restricted estimated joint cell probabilities and we show that they are asymptotically distributed as linear combinations of chi-square random variables under the null hypothesis of homogeneity. Monte Carlo simulation experiments are carried out to investigate the behavior of the new divergence test statistics and to make comparisons with the statistics that do not take into account the spatial correlation. We show that some of the introduced divergence test statistics have a significantly better behavior than the classical chi-square one for the problem under consideration when we compare them on the basis of the simulated sizes and powers.

ES84 Room 3 QUANTILE REGRESSION **Chair: Ivan Mizera**

E677: **On standard deviations in variational approximation estimates for quantile regression**
Presenter: **Elisabeth Waldmann**, Georg-August-University Goettingen, Germany
Co-authors: Thomas Kneib

Quantile regression (QR) has become an important tool when analyzing the impact of covariates on quantiles of a conditional distribution in order to obtain a broader information than just the conditional mean. This has been made possible for complicated data sets by using structured additive models estimated in the Bayesian framework with the asymmetric Laplace distribution as an error distribution. For the estimation variational approximations (VA) are an attractive alternative to Markov chain Monte Carlo (MCMC) techniques since they outperform the latter in terms of speed enormously. The main idea of VAs is to minimize the Kullback-Leibler-distance from a product of distributions of subsets of the parameters to the real posterior . This method was implemented for structured additive QR models and led to an obvious underestimation of standard deviation and therefore overly confident intervals, which is a problem inherent to the method. There are several suggestions to this problem, like grid based VAs, MCMC within VAs or the estimation of the covariance via approximations of a pseudo Fisher information matrix. These will be analyzed on the background of the QR framework.

E755: **On kernel smoothing for extremal quantile regression**
Presenter: **Abdelaati Daouia**, Catholic University of Louvain, Belgium
Co-authors: Laurent Gardes, Stephane Girard

Nonparametric regression quantiles obtained by inverting a kernel estimator of the conditional distribution of the response are long established in statistics. Attention has been, however, restricted to ordinary quantiles staying away from the tails of the conditional distribution. The purpose is to extend their asymptotic theory far enough into the tails. We focus on extremal quantile regression estimators of a response variable given a vector of covariates in the general setting, whether the conditional extreme-value index is positive, negative, or zero. Specifically, we elucidate their limit distributions when they are located in the range of the data or near and even beyond the sample boundary, under technical conditions that link the speed of convergence of their (intermediate or extreme) order with the oscillations of the quantile function and a von-Mises property of the conditional distribution. A simulation experiment and an illustration on real data are proposed. The real data are the American electric data where the estimation of conditional extremes is found to be of genuine interest.

E954: **Component selection in additive quantile regression models**
Presenter: **Eun Ryung Lee**, University of Mannheim, Germany
Co-authors: Hohsuk Noh
Nonparametric additive models are powerful techniques for multivariate data analysis. Although many procedures have been developed for estimating additive components both in mean regression and quantile regression, the problem of selecting relevant components has not been addressed much especially in quantile regression. We present a doubly-penalized estimation procedure for component selection in additive quantile regression models that combines basis function approximation with a variant of the smoothly clipped absolute deviation penalty and a ridge-type penalty. We show that the proposed estimator identifies relevant and irrelevant components consistently and achieves the nonparametric optimal rate of convergence for the relevant components. We also provide some numerical evidence of the estimator, and illustrate its usefulness through a real data example to identify important body measurements to predict percentage of body fat of an individual.

ES55 Room Multipurpose CONTRIBUTIONS TO CIRCULAR AND SPHERICAL DATA ANALYSIS Chair: Arthur Pewsey

E447: **Le Cam optimal inference for rotationally symmetric distributions on hyperspheres.**
Presenter: **Thomas Verdebout**, University Lille Nord de France, France
Co-authors: Christophe Ley, Yvik Swan, Baba Thiam
In general, the main objective of spherical statistics is to forecast the location of a natural phenomenon which happens on the surface of the earth. A traditional assumption on spherical data is the so-called rotational symmetry. Using the Le Cam asymptotic theory, we provide optimal inference on the center of symmetry for such a rotationally symmetric distribution. Due to the fact that the parameter of interest belongs to the unit sphere (and therefore to a non-linear manifold), it is necessary to correctly adapt the Le Cam theory in this context.

E899: **Bandwidth selection for directional densities**
Presenter: **Eduardo Garcia-Portugues**, Santiago de Compostela, Spain
Co-authors: Rosa M. Crujeiras, Wenceslao Gonzalez-Manteiga
Two new bandwidth selection methods for directional density estimation are presented. A directional density has support in the unit q-sphere $\Omega_q = \{\mathbf{x} \in \mathbb{R}^{q+1} : |\mathbf{x}| = 1\}$, being circular and spherical densities particular cases. Assuming that the underlying density follows a directional von Mises model, a suitable estimation of the unknown curvature term in the AMISE expression yields a simple plug-in rule, which works satisfactorily for simple directional models. A better proposal is derived from the exact MISE expression, available for mixtures of directional von Mises densities. Although the true density may not follow such a model, this type of mixtures has proved flexible enough to capture remarkable data features. The performance of the two new rules is assessed in a simulation study, comparing their behavior with other existing methods. Finally, the bandwidth selectors are applied to real data.

E974: **A multivariate hidden Markov model for the analysis of space-time circular data**
Presenter: **Francesco Lagona**, University Roma Tre, Italy
Co-authors: Antonello Maruotti, Marco Picone
Motivated by issues of marine data analysis under complex orographic conditions, a multivariate hidden Markov model is proposed for the analysis of a temporal sequence of spatial lattice circular data. Each spatial series is modelled by a circular random field whose parameters depend on the evolution of a latent Markov chain. The circular random field is specified in terms of a consistent set of conditional von Mises densities, which are Markov with respect to a spatial neighborhood structure. Because the likelihood of the model depends on a intractable normalizing constant, estimation is based on a computationally efficient EM algorithm that iteratively updates the parameters of a pseudo-likelihood function. The model is illustrated on a space-time series of wintertime wave directions in the Adriatic sea.

ES91 Room 1 EXPLORATORY MULTIVARIATE ANALYSIS Chair: Jose R. Berrendero

E881: **Data depth: An algorithm for finding boundary points by interpoint distance**
Presenter: **Hiu Fung Lam**, The University of Hong Kong, China
Co-authors: Pui Lam William Leung
Data depth is one of the most popular nonparametric descriptive statistics in multivariate analysis. A new and relatively fast data depth algorithm is proposed for center-outward clustering including non-convex datasets (unlike traditional methods) using an interpoint distance matrix, which can be used to construct non-parametric tolerance regions, multivariate custom control charts,etc. A simulation study on non-convex data clouds such as U-shape data is presented. Finally, a statistic is suggested to measure the convexity of the data cloud over a standard convex(normal) data cloud.

E980: **PCA visualization impoved by regularization**
Presenter: **Julie Josse**, Agrocampus, France
Co-authors: Marie Verbank, Francois Husson
Principal Component analysis (PCA) is a well-established method, which is commonly used to explore and visualize data. Consider a simple model where the data are generated as a fixed structure corrupted by noise. Under this model, PCA does not provide the best recovery (in term of mean squared error) of the underlying signal. Following the same principle as in ridge regression, we propose a regularized version of PCA that boils down to shrink the singular values. Each singular value is multiplied by a term which can be seen as the ratio of the variance of the signal over the total variance (signal plus noise) of the associated dimension. The regularized terms are chosen a priori and can be justified from different point of views such a Bayesian treatment of the model. As a consequence of the regularized PCA, we propose corrected graphical representation of the individuals and the variables and define new percentages of variance (explained by each dimension). Under the model, the graphical representations are closer to the ones that would be obtained from the true structure. The gap between the PCA and regularized PCA may be very important and leads to different interpretation when data are very noisy.

E833: **Extensions for DD-plot**
Presenter: **Dae-Heung Jang**, Pukyong National University, Korea, South
The DD-plot is a useful graphical exploratory data analysis tool for graphical comparisons of two multivariate distribution or samples based on data depth. We can suggest several extensions for DD-plot, namely, DDD-plot, DD-plot matrix, three-dimensional DD-plot, and dynamic DD-plot.

ES93 Room 6 STATISTICAL MODELLING Chair: Geoff McLachlan

E748: **Modeling interval censored multi-states models using multi-step transition probabilities**
Presenter: **Holger Reulen**, Georg-August-University Goettingen, Germany
Co-authors: Thomas Kneib
Multi-state models play an increasingly important role in the analysis of time to event data. Multi-state models extend duration analysis to the analysis of what happens beyond some first event, by allowing individuals to progress through different states in continuous time. For the estimation of multi-state models, the transition probabilities and their underlying covariate mechanism, it is ideal to observe complete event-histories. However in many situations, e.g. courses of diseases, the status of subjects is observed only at a finite number of visits. This leads to interval-censored observations. If we do not assume further properties to a process with recurrent events we have to deal with the number of transitions being unbounded above. So, making no assumptions on the number of transitions will lead to likelihoods with an infinite number of contributions. We are able to bind the number of possible transitions between two time-points by making the assumption that the true process makes transitions into another state only on an equidistant grid of time points. By that we constrain the number of possible paths, are able to calculate multi-step transition probabilities between the observed time-points and states and use these to construct the likelihood for interval censored multi-state models.

E992: **A bivariate Sarmanov regression model for count data with generalised Poisson marginals**
Presenter: **Vera Hofer**, University of Graz, Austria
Co-authors: Johannes Leitner
A bivariate regression model for count data is presented. It allows for positive as well as negative correlation of the response variables. The covariance structure is based on the Sarmanov distribution and consists of a product of generalised Poisson marginals and a factor that depends on particular functions of the response variables. The closed form of the probability function is derived by means of the moment generating function. The model is applied to a large real dataset on health care demand. Its performance is compared to alternative models. We find that our model is significantly better than or at least equivalent to the benchmark models. It gives insights into influences on the variance of the response variables.

E1040: **Some statistical inference results for lognormal diffusion random fields**
Presenter: **Concepcion Roldan**, University of Granada, Spain
Co-authors: Ramon Gutierrez-Sanchez, Jose Miguel Angulo, Antonio Roldan
Two-parameter lognormal diffusions constitute an important case of random fields which are not intrinsically stationary, and then well-known related techniques cannot be applied in this context. Models in this class are suitable to represent diffusion-type positive valued characteristics, like pollutant indicators in environmental studies. Methods for statistical inference on the parameters weighting the relative effect of exogenous factors are discussed. The usefulness of these methods is illustrated by simulation experiments carried out with R scripts. In particular, the results show how data simulated away from the diffusion origin (that is, where the diffusion effect becomes apparently weaker) are still significant to make inferences about the model parameters.

ES94 Room 4 COMPUTATIONAL STATISTICS Chair: Christine Mueller

E738: **A heuristic algorithm for determining the rank of non-negative matrix factorization**
Presenter: **Mu Zhu**, University of Waterloo, Canada
Co-authors: Yali Wan
Both the singular value decomposition (SVD) and the non-negative matrix factorization (NMF) aim to approximate a data matrix X with the product of a few rank-K matrices, where K « rank(X). We first explain why many heuristics used to determine the proper value of K for the SVD are not applicable to the NMF. We then present a different heuristic for making such a decision. Our algorithm is motivated by the so-called "separable factorial articulation (SFA) family". Using simulated, noisy data known to be outside the SFA-family, we demonstrate that our algorithm is also useful in less ideal situations.

E1024: **Estimators for two ordered normal means with ordered variances under Pitman's closeness and numerical analysis**
Presenter: **Genso (Yuan-Tsung) Watanabe (Chang)**, Mejiro University, Japan
Co-authors: Nobuo Shinozaki
The problem of estimating the ordered means of two normal distributions with unknown ordered variances is considered. We discuss the estimation of two ordered means, individually, under Pitman closeness criterion. We show that in estimating the mean with larger variance, the proposed estimator is closer to true mean than the usual one which ignore the order restriction on variances. However, in estimating the mean with smaller variance, the usual estimator is not improved upon. We also discuss simultaneous estimation problem of two ordered means when unknown variances are ordered. Finally, we use the numerical analysis to illustrate the behavior of proposed estimators.

E757: **Computational possibilities and properties of a robustified version of the mixed least squares - total least squares estimator**
Presenter: **Jiri Franc**, Czech Technical University in Prague, Czech Republic
Classical robust regression estimators, such as Least Trimmed Squares (LTS), are not consistent when both independent and some dependent variables are considered to be measured with a random error. One way how to cope with this problem is to use the robustified version of Mixed Least Squares - Total Least Squares (LS-TLS). Mixed Least Trimmed Squares - Total Least Trimmed Squares (LTS-TLTS) based on trimming and mixed Least Weighted Squares - Total Least Weighted Squares (LWS-TLWS) based on the idea of downweighting the influential points, are proposed. The existence and uniqueness of the solution, breakdown point, consistency and another properties of these estimators are discussed. Different approaches of calculation, such as Branch-and-Bound algorithm, elemental concentration algorithm and simulated annealing, are described and their performances are shown on sets of benchmark instances.

ES80 Room 7 COMPOSITIONAL DATA ANALYSIS AND RELATED METHODS Chair: J. Antoni Martin-Fernandez

E919: **Model-based clustering via Gaussian mixture models for compositional data: protein consumption in Europe**
Presenter: **Marc Comas-Cufi**, Universitat de Girona, Spain
Co-authors: Gloria Mateu-Figueras, Santi Thio-Henestrosa, Josep Antoni Martin-Fernandez
Compositional data appear naturally in many applied fields. Any statistical analysis involving this type of data should fulfill two main principles: scale invariance and subcompositional coherence. Log-ratio methodology provides a successful set of techniques to deal with this kind of data. Three main log-ratio transformations (additive, centered and isometric) are applied here for the Gaussian mixture of distributions. The corresponding three models are analysed for clustering purpose and their different particularities are discussed. A case study is presented using a potrein consumption data set for countries in Europe. The results are compared with those provided by a typical normal mixture. In this case study, the results suggest that log-ratio methodology produces more interpretable groups, according to the geography situation of each country.

E673: ***k*-nearest neighbours for compositional data**
Presenter: **Michail Tsagris**, University of Nottingham, United Kingdom
Co-authors: Simon Preston, Andrew Wood
Compositional data are a special type of multivariate data in which the elements of each observation vector are non-negative and sum to one. Data of this type arise in many areas, such as economics, archaeology and biology amongst others. Analysis of such data may be carried out in various ways, e.g. by transforming the data using logs of ratios formed by the components or by neglecting the compositional constraint and treating the observations as standard multivariate data. Recently, we developed a more general approach which includes both approaches indicated above as particular cases. It involves a Box-Cox type transformation, known as the α-transformation, with a free parameter, α. The goal is to propose a new family of metrics, which is closely related to the α-transformation, for classification of compositional data when using the *k*-nearest neighbours algorithm.

E854: **Bayesian modelling of forensic glass data**
Presenter: **Tereza Neocleous**, University of Glasgow, United Kingdom
Co-authors: Gary Napier, Agostino Nobile
Compositional data frequently occur in chemistry. One example is data points corresponding to the percentage with respect to weight of each chemical element present in a glass sample. Such compositions are subject to a sum constraint which should be taken into account when the data are analysed. In addition, it is not unusual in such elemental compositions to observe zero concentrations for some of the components. We examine approaches to dealing with zeros in glass elemental compositions and develop a Bayesian hierarchical model for glass data with the purpose of classifying glass for forensic purposes and for assessing the evidential value of glass as transfer evidence.

ES85 Room 10 MODELS AND COMPUTATIONAL TOOLS IN DATA ANALYSIS Chair: Simone Vantini

E770: **Decomposition of metropolitan mobile-network data through independent component analysis**
Presenter: **Paolo Zanini**, Politecnico di Milano, Italy
Co-authors: Piercesare Secchi, Simone Vantini
The analysis of population behaviors and dynamics in a metropolitan area is a central topic for urban planning and for the optimal design of service networks. We analyze data relative to mobile phone traffic in the metropolitan area of Milano (Italy). In each pixel of a dense rectangular grid (of step 250m) covering the area, the Erlang is measured every 15 minutes for two weeks of the 2009. The Erlang is a dimensionless unit that, for each pixel and for each quarter of an hour, represents the average number of mobile phones simultaneously calling through the network. This variable is assumed to be a proxy of the number of active people in that area at that time. The Erlang spatial surface varies along time and can be modeled as a linear combination of independent, latent time-invariant surfaces, each surface being associated to a specific temporal weight profile. The principal aim of the analysis is to separate and identify residential and mobility basis elements. A suitable extension of Independent Component Analysis, a statistical method widely used in Blind Source Separation problems, is used. Related residential covariates are obtained from Geographic Information System and introduced in the model.

E877: **Validation of a linguistic summarization approach for time series meteorological data**
Presenter: **Alejandro Ramos**, University of Santiago de Compostela, Spain
Co-authors: Alberto Bugarin, Felix Diaz-Hermida, Senen Barro
Linguistic summaries of data are brief and precise general textual descriptions of (usually numeric) datasets. Computational methods which generate linguistic summaries have been developed over recent years, and their usefulness has been proved in different application domains. However, means to validate them objectively as well as subjectively by experts are still in early development and need to be explored and discussed in depth. It is therefore a challenging open problem where new proposals of measures for testing/validating the linguistic summaries obtained or new methodologies for assessing its quality can be contributed. A heuristic approach is described for the automatic generation of operative weather forecasts for Galicia (NW Spain), that are built as spatio-temporal linguistic summaries that describe the results of the numerical meteorological prediction models executed at the Galician Weather Service. Summaries involve linguistic values of the cloud coverage meteorological variable, fuzzy quantifiers ('a few', 'many', ...), spatio-temporal references (e.g. 'the sky will be cloudy in the south coast'). This realm is used as a case of study for proposing new validation measures and quality assessment criteria and procedures that are applied to the obtained summaries in order to confront them with the ones generated manually by the human experts.

E1011: **Practical approach for determining the number of observations needed analyzing univariate responses of sigmoidal equations**
Presenter: **M.A. Prieto**, Instituto de Investigaciones Marinas, Spain
Co-authors: J.A. Vazquez, I. Prieto, Y. Anders, M.A. Murado
The three parameters sigmoidal equations (3P-SE) have been increasingly applied in different fields (such as medicine, biology, pharmacology, microbiology, etc.) with diverse purposes, but with the identical goal of using the obtained parameters (P) to analyze univariate responses. A usual dilemma, when researchers design an experimental, is the number of observations (NO) needed to obtain a fitting solution in which all P are significant. The confidence interval (CI) of each P is not regulated by an unpredictable behavior and it is statistically dependent on the 3P-SE, NO, number of replicates (NR), data variance (VAR) and the type of experimental error (EE). By simulating data and fitting them (nonlinear least-squares) to seven common 3P-SE as a function of this multivariable system (ranges: 4-100 NO; 1-10 NR; 0-25% VAR) a thousand times for each condition, and averaging the resulting parametric estimations, the likelihood degree (such as 95%) of finding all P of a given 3P-SE significant is found, predicted and compared. Additionally, the minimum requirements (NO and NR) to find a desire level of significance of a certain P (such as the half-life) as a function of the 3P-SE, type of EE and VAR is also predicted and compared.

ES30 Room 11 CONTRIBUTIONS TO OPTIMIZATION HEURISTICS IN ESTIMATION AND MODELLING Chair: Manfred Gilli

E793: **Forcing sparsity in principal component analysis**
Presenter: **Emilio Carrizosa**, University of Seville, Spain
Co-authors: Vanesa Guerrero
Different methods have being proposed in the literature to make principal components sparser. In this talk we present an optimization model, which allows one to minimize the sum of squares of point-projections distance whilst controlling the number of nonzero loadings. The optimization model is solved using heuristics which successfully exploit the mixed-integer-nonlinear structure of the problem. Application to test data sets is reported.

E790: **A heuristic for interpolating missing data with autocorrelation and moments matching**
Presenter: **Alba Victoria Olivares Nadal**, Universidad de Sevilla, Spain
Co-authors: Emilio Carrizosa, Pepa Ramirez-Cobo
Most time series forecasting methods assume the series has no missing values. When missing values exist, interpolation methods, while filling in the blanks, may substantially modify the statistical pattern of the data, since critical features such as moments and autocorrelations are not necessarily preserved. We propose to interpolate missing data in time series by solving a smooth nonconvex optimization problem which aims

to preserve moments and autocorrelations. Since the problem may be multimodal, Variable Neighborhood Search is used to trade off quality of the interpolation (in terms of preservation of the statistical pattern) and computing times. Our approach is compared with standard interpolation methods and illustrated on both simulated and real data.

E904: **Tuning support vector machines using variable neighborhood search**
Presenter: **Belen Martin-Barragan**, Universidad Carlos III de Madrid, Spain
Co-authors: Emilio Carrizosa, Dolores Romero Morales
When using Support Vector Machines (SVM), the choice of some free parameters is a critical, though not sufficiently explored, issue. When the number of parameters to tune is large, the common approach of searching in a grid is too time consuming. For this reason, for models involving a higher number of parameters, different metaheuristics have been recently proposed as an alternative. We customize a continuous Variable Neighborhood Search to tune the parameters of the SVM. The structure of the optimization problem at hand is successfully exploited. This is done by expressing the kernel model at hand as a collection of nested kernel models which can be solved sequentially. Our approach enjoys the following advantages. First, it is more efficient than the usual approach of searching in a grid of the parameter space, since it does not need to examine the whole grid. Second, it is a continuous approach as opposed to the usual discrete one, and it does not require the choice of a step size to define the grid, while it allows to zoom in in certain regions of the parameter space. Third, as algorithmic requirements we only need a routine which, for given parameters, finds the SVM classifier. Hence, as soon as an SVM library or any routine for solving linearly constrained convex quadratic optimization problems is available, our approach is applicable. The effectiveness of the proposed method is illustrated in the framework of Multiple Kernel Learning (MKL). The proposed algorithm, being a general framework for tuning parameters and requiring minimal software requirements, is able to provide results that are comparable with state-of-the-art methods that are specifically designed for MKL.

ES96 Room 8 FACTOR ANALYSIS AND CATEGORICAL DATA Chair: Kohei Adachi

E670: **A matrix decomposition approach to nonmetric factor analysis of multidimensionally quantified categorical data**
Presenter: **Naomichi Makino**, Osaka University, Japan
Co-authors: Kohei Adachi
A nonmetric factor analysis procedure for variables of mixed measurement levels has been previously proposed. In that method, parameters are estimated by the alternate iteration of obtaining the model parameters of factor analysis and performing the unidimensional quantification of categorical variables. However, cases are reasonably thought to exist where categories in a variable cannot be scaled unidimensionally. Considering such cases, we propose an extended nonmetric factor analysis procedure in which multidimensional quantification can be given to categorical variables. A feature of the proposed method is that it is formulated as a matrix decomposition problem in which all parameter matrices of factor analysis are directly fitted to a quantified data matrix in a least squares sense. After illustrating the method, we show that it can also be viewed as an extension of multiple correspondence analysis.

E754: **A new factor rotation method to simplicity with joint estimation of optimal target matrix having perfect cluster structure**
Presenter: **Naoto Yamashita**, Tokyo Institute of Technology, Japan
A new method of factor rotation where factor loading matrix is obliquely/orthogonally rotated toward an optimal target for simplicity is proposed. The rotation matrix and the target matrix are jointly estimated by minimizing a single criterion, subject to the constraint that the target matrix has perfect cluster structure where each variable (row) is associated with only one factor (column). Specifically, in each step of proposed procedure, the target matrix is optimally updated in least squares sense so that it has a perfect cluster structure, and the loading matrix is rotated to this target matrix by an appropriate Procrustes rotation method. The proposed method is expected to perform well when the matrix to be rotated can be expressed as a result of an arbitrary rotation of a matrix having perfect cluster structure. A simulation study and a real data example are presented to empirically investigate the performance of proposed method.

E786: **Generalized quasi symmetry models for ordinal contingency tables**
Presenter: **Maria Kateri**, RWTH Aachen University, Germany
Co-authors: Anna Gottard, Claudia Tarantola
The quasi symmetry (QS) model applies on square $I \times I$ contingency tables with commensurable classification variables and is know as the model under which the $(I-1) \times (I-1)$ table of the local odds ratios is symmetric. New models are introduced by considering the symmetry property for alternative types of odds ratios, appropriate for ordinal contingency tables. Thus, the global QS model is the model of symmetric global odds ratios, while the cumulative and continuation QS models are defined analogously. Furthermore, the conditional QS model is introduced for all types of odds ratios. Classical properties of the standard QS model are discussed for the new introduced models. Extensions of these models to multi-way contingency tables are considered as well. Finally, models are illustrated via characteristic examples.

ES82 Room 9 METHODS FOR APPLIED STATISTICS Chair: Agustin Mayo

E663: **Discrete adjusted FDR control**
Presenter: **Chang-Yun Lin**, National Chung Hsing University, Taiwan
It is important in multiple tests to control test error rates, such as FWER (family wise error rate) or FDR (false discovery rate). Most of FWER and FDR control methods in literature correct error rates based on the Uniform distribution of $p-$values under null hypothesis. However, when the test statistic does not follow a continuous distribution, for instance, Fisher's exact test, it was previously pointed out thar the distribution of its $p-$values is discrete not continuous. It was demonstrated that dichotomizing continuous dataset and implementing Fisher's exact test with discrete adjusted FWER control provided much higher power than using $t-$test for the continuous dataset with FDR control when the number of true alternatives is small. Unfortunately, this discrete adjusted FWER control method decreases test power rapidly when the number of true alternatives increases. The purpose is to develop a discrete adjusted FDR control method, which combines the strengths of Westifall's discrete adjustment and Benjamini's FDR control to increase test power for multiple tests.

E692: **Longitudinal mixture models for partially ranked preferences: Examining changes in postmaterialism over time**
Presenter: **Brian Francis**, Lancaster University, United Kingdom
Co-authors: Regina Dittrich, Reinhold Hatzinger
The use of mixture models to incorporate a random effects structure in the analysis of partially ranked data is discussed. Such data occur in surveys, where respondents may be invited to choose the preferred and second most preferred out of a larger set of items. Two questions on postmaterialism in sweeps of a British survey motivate the study. Respondents were asked to choose the most preferred and next preferred out of a set of four items representing priorities for government. Typical analysis of these data does not use the full ranking information. However, by treating partially ranked data as a set of paired comparisons, all of the data is used. We discuss an extension which allows for repeated responses over time. The model takes account of attrition over the sweeps of the survey through a FIML approach which assumes an underlying MAR process. A nonparametric formulation of the random effects structure (time/individual) is fitted using the EM algorithm. Each discrete mass point

is multivalued, with a parameter for each item. The resultant model is equivalent to a latent class regression model, where the latent class profiles are provided by the mass point components.

E785: **Comparison of methods for clustering of variables defined on the hypersphere**
Presenter: **Adelaide Figueiredo**, University of Porto, Portugal
Methods of clustering of variables based on the identification of a mixture of Watson distributions defined on the hypersphere are considered. For the identification of this mixture, the following iterative methods: the Dynamic Clusters Method, the EM (Estimation-Maximisation) Algorithm and the Principal Cluster Component Analysis, are discussed and are compared using simulated and real data. The performance of the methods is compared for the same initial solution, by evaluating the final solutions obtained in these methods through the calculation of a between-groups variability measure and a within-groups variability measure.

ES78 Room 2 MODEL SELECTION Chair: Ivan Kojadinovic

E734: **Unifying approach to the estimation of the Akaike Information in generalized mixed models**
Presenter: **Benjamin Saefken**, Georg-August University Goettingen, Germany
Co-authors: Thomas Kneib
Model choice in mixed models is becoming increasingly important since also additive models can be represented as mixed models. The Akaike information criteria (AIC) has been widely used for model choice and variable selection and work has been done on how to extend it to mixed models. In these models the AIC can be defined for the marginal and the conditional distribution of the response. Using the marginal distribution as it is common practice, since it is easy to compute, leads to a biased criteria. This bias does not disappear asymptotically and cannot be corrected due to its dependence on the unknown but true variance parameters of the random effects. The AIC based on the conditional distribution for Gaussian responses is analytically accessible by the use of the Steinian method. Up on till now there is no conditional AIC for non-Gaussian responses. A generalization of the Steinian method to exponential families gives a unifying perspective on the AIC for generalized linear mixed models. An analytically accessible AIC for some other exponential family distributions can be obtained and it is shown why this method can not be extended to all exponential family distributions.

E1001: **Repeated weigthed hold-out for variable selection in linear regression**
Presenter: **Agostino Di Ciaccio**, University of Roma La Sapienza, Italy
Variable selection criteria in a linear regression model are analyzed, considering consistency and efficiency properties. A Cross-Validation method, which we called Repeated Weigthed Hold-Out, is introduced. This method estimates the Prediction Error, as the well known Repeated Hold-Out, with the difference that, at each iteration, the units extracted in the training-set reduce their probability of a new extraction. This weighting makes the procedure similar to a repeated K-fold Cross-Validation, in which the units appear with the same frequency in the training sets, obtaining a procedure more flexible than the K-fold CV. It is known that a criterion based exclusively on the estimate of PE is not consistent but efficient, however, using a property of CV, i.e. small training-sets introduce large penalty of the criterion while large training-set introduce small penalty, it is possible to obtain an estimator of the PE inherently penalized. This method dominates several criteria for variable selection, choosing the appropriate training-set size.

Authors Index