PROGRAMME AND ABSTRACTS

Third International Conference on

Computational and Financial Econometrics (CFE 09)

http://www.dcs.bbk.ac.uk/cfe09/

and

Second Workshop of the ERCIM Working Group on

Computing & Statistics (ERCIM 09)

http://www.dcs.bbk.ac.uk/ercim09/

Grand Resort Hotel, Limassol, Cyprus
October 29-31, 2009

University of Cyprus
http://www.ucy.ac.cy

Cyprus University of Technology
http://www.cut.ac.cy

CFE 09 Co-chairs: G. Gallo, E.J. Kontoghiorghes, Y. Omori, H.K. Van Dijk

CFE 09 Programme committee:

A. Amendola (IT), G. Barone-Adesi (CH), L. Bauwens (BE), D. Belsley (US), M. Billio (IT), C. Chen (TW), J. Coakley (UK), J. Davidson (UK), I. Demetriou (GR), C. de Peretti (FR), K. Fokianos (CY), P. Foschi (IT), C. Francq (FR), A.-M. Fuertes (UK), M. Gilli (CH), Z. Hlavka (CZ), G. Kapetanios (UK), L. Khalaf (CA), J.F. Kiviet (NL), C. Kleiber (CH), D. Kuhn (UK), O. Linton (UK), A. Luati (IT), T. Lux (DE), D. Maringer (CH), A. Michaelides (UK), S. Mittnik (DE), M. Ooms (NL), M. Paolella (CH), A. Patton (UK), D.S.G. Pollock (UK), T. Proietti (IT), M. Reale (NZ), M. Riani (IT), E. Ruiz (ES), B. Rustem (UK), O. Scaillet (CH), M. Schroeder (DE), W. Semmler (US), K. Sheppard (UK), M. So (HK), G. Storti (IT), D. Veredas (BE), M. Wagner (AT), P. Winker (DE), M. Wolf (CH), P.L.H. Yu (HK), P. Zadrozny (US), A. Zeileis (AT).

ERCIM 09 Co-chairs: A. Colubi, E.J. Kontoghiorghes, D.S.G. Pollock, S. Van Aelst, P. Winker

ERCIM 09 Programme committee:

M. Alfo (IT), A. Amendola (IT), M. Avalos (FR), F. Battaglia (IT), C. Bekas (CH), D. Boehning (UK), M. Chiarandini (DK), A. Christmann (DE), R. Coppi (IT), C. Croux (BE), T. Denoeux (FR), R. Diaz-Uriarte (ES), D. Dubois (FR), F. Ferraty (Fr), P. Filzmoser (AT), K. Fokianos (CY), R. Fried (DE), L.A. Garcia-Escudero (ES), C. Gatu (RO), K. Giannopoulos (UA), M.A. Gil (ES), J. Godolphin (UK), G. Gonzalez (ES), A. Gordaliza (ES), P. Groenen (NL), M. Hubert (BE), J. Jaenicke (DE), E. Kalotychou (UK), F. Klawonn (DE), A. Kondylis (CH), R. Kruse (DE), U. Ligges (DE), W.G. Manteiga (ES), D. Maringer (CH), I. Moustaki (GR), M. La Rocca (IT), I. Molina (ES), D. Morales (ES), J. Niland (US), S. Paterlini (IT), C. Perna (IT), M. Salibian-Barrera (CA), A. Staszewska (PL), K. Triantafyllopoulos (UK), W. Trutschnig (ES), H. Turner (UK), L. Ugarte (ES), R. Viertl (AT), P. Vieu (FR), I. Vrontos (GR), P. Yanev (CH), J.M. Zakoian (GR), A. Zeileis (AT).

Local organizing committee:

D. Bampou, C. Chappas, C. Gatu, M. Hofmann, E.J. Kontoghiorghes, M. Kapsos, P. Kleniati, M. Lyra, M. Markidou, K. Panagidou, A. Savvides, E. Tsinti.

Dear Friends and Colleagues,

We warmly welcome you in Limassol, Cyprus, for the Third International Conference on *Computational and Financial Econometrics* (CFE 09) and the Second Workshop of the ERCIM Working Group on *Computing & Statistics* (ERCIM 09). As many of you know, this yearly conference has been established as a leading joint international meeting for the interface of computing, empirical finance, econometrics and statistics, and is endorsed by the journal 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 in general. The CFE 09 & ERCIM 09 programme consists of 98 sessions, 3 plenary talks and around 460 presentations. Peer reviewed papers will be considered for publication in special issues of the journal Computational Statistics & Data Analysis.

The Co-chairs have endeavored to provide a balanced and stimulating programme that will appeal to the diverse interests of the participants. The local 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 of many individuals and organizations. The Co-chairs, the Scientific Programme Committee, the Local Organizing Committee, and many volunteers have contributed substantially to the organization of the conference. We acknowledge the support of our sponsors, and particularly the CSDA journal, ERCIM, University of Cyprus and Cyprus University of Technology.

Looking forward, the CFE 10 & ERCIM 10 will be organized by the London School of Economics, UK, and Queen Mary, University of London, UK. It will take place during the second week of December, 2010. The CSDA Annals on Computational and Financial Econometrics will be inaugurated during the conference. You are invited and encouraged to actively participate in these events.

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

Co-chairs: A. Colubi, G. Gallo, E.J. Kontoghiorghes (Chair), Y. Omori, D.S.G. Pollock, S. Van Aelst,
H.K. Van Dijk and P. Winker.

ERCIM Working Group on COMPUTING & 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 applications areas where both computing techniques and numerical 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 computational statistics in diverse disciplines are strongly represented. These areas include economics, medicine and 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 numerical methods in 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 are organizing sessions and workshops during the annual WG meeting.

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 300 members and the following specialized groups:

MCS: Matrix Computations and Statistics.

CFE: Computational Econometrics and Financial Time Series.

SSEF: Statistical Signal Extraction and Filtering.

RDM: Robust Analysis of Complex Data Sets.

OHEM: Optimization Heuristics in Estimation and Modelling.

FSA: Fuzzy Statistical Analysis.

AlgSoft: Statistical Algorithms and Software.

SFD: Statistics for Functional Data.

FGen: Functional Genomics.

SEM: Latent Variable and Structural Equation Models.

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 Erricos John Kontoghiorghes at:
matrix@dcs.bbk.ac.uk .

SCHEDULE

All events, except the closing dinner, take place at the Grand Resort Hotel.

Thursday, 29th October 2009

09:10 - 09:20	Opening (Room 1)
09:20 - 10:10	Plenary Session A (Keynote talk by Neil Shephard)
10:10 - 10:50	Coffee Break
10:50 - 12:30	Parallel Sessions B
12:30 - 14:00	Lunch Break
14:00 - 16:00	Parallel Sessions C
16:00 - 16:30	Coffee Break
16:30 - 18:50	Parallel Sessions D
20:00 - 21:30	Reception

Friday, 30th October 2009

08:30 - 10:30	Parallel Sessions E
10:30 - 10:50	Coffee Break
10:50 - 11:40	Plenary Session F (Keynote talk by Christophe Croux)
11:50 - 13:00	Parallel Sessions G
13:00 - 14:30	Lunch Break
14:30 - 16:10	Parallel Sessions H
16:10 - 16:40	Coffee Break
16:40 - 18:40	Parallel Sessions I
20:15 - 24:00	Conference Dinner

Saturday, 31st October 2009

08:45 - 10:20	Parallel Sessions J
10:20 - 10:40	Coffee Break
10:40 - 13:00	Parallel Sessions K
13:00 - 14:30	Lunch Break
14:30 - 16:30	Parallel Sessions L
16:30 - 17:00	Coffee Break
17:00 - 17:50	Plenary Session M (Keynote talk by Siem Jan Koopman)
17:50 - 18:00	Closing (Room 1)
19:15 - 23:45	Closing Dinner

MEETINGS AND SOCIAL EVENTS

SPECIAL MEETINGS by invitation to group members

- COST Management Committee meeting, Wednesday 28th of October, *Room 6*, 17:00-19:30.
- CSDA Editorial Board Dinner, Wednesday 28th of October, 20:00-23:00.
- CSDA Editorial and CFE 10 & ERCIM 10 Co-chairs Meeting, *Anthea Restaurant, Grand Resort*, Friday 30th of October, 16:15 - 18:00.

SOCIAL EVENTS

- *The coffee breaks* will last one hour each (which adds fifteen minutes before and after to the times that are indicated in the program).
- *Welcome Reception, Thursday 29th of October, 20:00.* The reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the venue Grand Resort. You must have your conference badge in order to attend the reception.
- *Lunches* will be served at the *Anthea Restaurant*, Grand Resort. Lunch tickets can be obtained from the registration desk.
- *Conference Dinner, Friday 30th of October, 20:15*. The Conference Dinner will take place at the venue *Grand Resort*. The conference dinner is optional and registration is required. A small number of tickets can be obtained from the registration desk. You must have your Conference Dinner ticket in order to attend the conference dinner.
- *Closing Dinner, Saturday 31st of October, 20:00.* The Closing Dinner will take place at a traditional restaurant with live music. The dinner is optional and registration is required. Tickets can be obtained from the registration desk. The busses will be leaving from the the Grand Resort and Arsinoe Hotel at 19:15, and from Mediterranean Beach Hotel at 19:30.

GENERAL INFORMATION

Lecture Rooms

The paper presentations will take place at the venue *Grand Resort*. There are ten lecture rooms. Two of them (Room 9 and Room 10) are in the first floor, while the other eight are in the ground floor. There will be signs indicating the location of the lecture rooms. For directions and help, please ask at the registration desk.
The Keynote talks will take place in the lecture Room 1, and will last 50 minutes including questions. 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 switch rooms mid-session to hear particular papers. In the case of a no-show, please use the extra time for a break or a discussion so that the remaining papers stay on schedule.

Presentation instructions

The lecture rooms will be provided 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 by email the talks prior to the conference beginning. Presenters must deliver to the session chair ten minutes before each session the files with the presentation in PDF (Acrobat) or PPT (Powerpoint) format on a USB memory stick.
The PC in the lecture rooms should be used for presentations. *The session chairs should have a laptop for backup*.
Please note that Cyprus has identical plugs /power outlets to the UK, and thus differ from those in the rest of Europe and beyond. We cannot provide adapters, so please do not forget to take your adapters if needed.

Internet

There will be limited access to PCs connected to the Internet and wireless Internet connection near the Rooms 6-8.

Messages

You may leave messages for each other on the bulletin board by the registration desks.

SPONSORS

Elsevier
ERCIM (European Research Consortium for Informatics and Mathematics)
University of Cyprus
Cyprus University of Technology
Queen Mary, University of London, UK
Birkbeck University of London, UK
Research Promotion Foundation, Cyprus
Central Bank of Cyprus
Hellenic Bank
Ernst & Young Cyprus Limited
CYTA
Cyprus Tourist Organization
Cyprus Airways

ENDORSED SOCIETIES & GROUPS

Journal Computational Statistics & Data Analysis, Elsevier
ERCIM Working Group on *Computing & Statistics*
COMISEF, Marie Curie Research and Training Network
COST Action IC0702 - SoftStat
The Society for Computational Economics
International Association for Statistical Computing
International Association for Mathematics and Computers in Simulation

PUBLICATIONS OUTLETS

Journal of Computational Statistics & Data Analysis (CSDA)

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

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@dcs.bbk.ac.uk .

- Selected peer-review papers will be published in the 6th special issue on Computational Econometrics of the Computational Statistics & Data Analysis. Submissions for the 6th special issue should contain both a computational and an econometric or financial-econometric component. The deadline for paper submissions is 15 December 2009. The special Issue Editors are:
D. Belsley, C.W.S. Chen, C. Francq, G. Gallo, E.J. Kontoghiorghes and H.K. Van Dijk.

Selected accepted papers will be considered for inclusion in the ***CSDA Annals of Computational and Financial Econometrics***.

- CSDA has already planned special issues for 2009-2010 on the following topics:
  - Fuzzy Sets in Statistical Analysis.
  - Statistical Signal Extraction and Filtering.
  - Machine Learning and Robust Data Mining.
  - Quantile Regression and Semiparametric Methods.

CSDA Aims and Scope

Computational Statistics & Data Analysis (CSDA), the official journal of the International Association of Statistical Computing (IASC), is an international journal dedicated to the dissemination of methodological research and applications in the areas of computational statistics and data analysis. The three sections are divided into the following subject areas:

Computational Statistics Manuscripts dealing with: 1) the explicit impact of computers on statistical methodology (e.g., Bayesian computing, bioinformatics, computational econometrics, computer graphics, computer intensive inferential methods, data exploration, data mining, expert systems, heuristics, knowledge based systems, machine learning, neural networks, numerical and optimization methods, parallel computing, statistical databases, statistical systems), and 2) the development, evaluation and validation of statistical software and algorithms. Software and algorithms can be submitted with manuscripts and will be stored together with the online article.

Statistical Methodology for Data Analysis Manuscripts dealing with novel and original data analytical strategies and methodologies applied in biostatistics (design and analytic methods for clinical trials, epidemiological studies, statistical genetics, or genetic/environmental interactions), chemometrics, classification, data exploration, density estimation, design of experiments, econometrics, environmetrics, education, image analysis, marketing, model free data exploration, pattern recognition, psychometrics, statistical physics, image processing and robust procedures.

Statistical methodology includes, but is not limited to: bootstrapping, classification techniques, clinical trials, data exploration, density estimation, design of experiments, pattern recognition/image analysis, parametric and nonparametric methods, statistical genetics, Bayesian modeling, outlier detection, robust procedures, cross-validation, functional data, fuzzy statistical analysis, mixture models, model selection and assessment, nonlinear models, partial least squares, latent variable models, structural equation models, supervised learning, signal extraction and filtering, time series modelling, longitudinal analysis, multilevel analysis and quality control.

Special Applications Manuscripts at the interface of statistics and computing (e.g., comparison of statistical methodologies, computer-assisted instruction for statistics, simulation experiments). Advanced statistical analysis with real applications (economics, social sciences, marketing, psychometrics, chemometrics, signal processing, finance, medical statistics, environmentrics, statistical physics).

Contents

Thursday, 29.10.2009 09:20-10:10 Room: 1 Keynote talk 1

Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models

Speaker: **Neil Shephard, Oxford University, UK** Co-authors: Thomas Flury
Chair: Richard T. Baillie

Suppose we wish to carry out likelihood based inference but we solely have an unbiased simulation based estimator of the likelihood. We note that unbiasedness is enough when the estimated likelihood is used inside a Metropolis-Hastings algorithm. This result has recently been introduced and is perhaps surprising given the celebrated results on maximum simulated likelihood estimation. Bayesian inference based on simulated likelihood can be widely applied in microeconomics, macroeconomics and financial econometrics. One way of generating unbiased estimates of the likelihood is by the use of a particle filter. We illustrate these methods on four problems in econometrics, producing rather generic methods. Taken together, these methods imply that if we can simulate from an economic model we can carry out likelihood based inference using its simulations.

Friday, 30.10.2009 10:50-11:40 Room: 1 Keynote talk 2

Robust multivariate scale estimators for volatility estimation of financial time series

Speaker: **Christophe Croux, K.U. Leuven, Belgium** Co-authors: Kris Boudt, Sebastien Laurent
Chair: Peter Rousseeuw

For univariate data robust scale estimators as the interquartile range or the median absolute deviation are well known. In a multivariate setting, robust estimators of scale are perhaps less well-known, but also available. Such estimators needs to be computable in reasonable time, affine equivariant and positive definite. They should not lose too much statistically efficiency with respect to the standard estimator, while being highly resistant to outliers. An example of such an estimator is the minimum Covariance Determinant (MCD) estimator. We show how the MCD can be used to estimate daily volatility of financial time series using high frequency data. A popular measure for the intraday volatility of a multivariate price process is the Quadratic Covariation. If the price process contains jumps, then it is of interest to estimate only its continuous part. We show how the MCD can be used to construct a weighted version of the Realized Quadratric Covariation, discarding the jumps when estimating the continuous part of the daily volatility. We derive its properties and compare it with other proposals. We illustrate the method on 15-minute return series of the EUR/USD and GBP/USD exchange rates.

Saturday, 31.10.2009 17:00-17:50 Room: 1 Keynote talk 3

Dynamic factor analysis by maximum likelihood

Speaker: **Siem Jan Koopman, VU University Amsterdam, The Netherlands**
Chair: Herman K. Van Dijk

A new approach to the modelling of the term structure of interest rates is proposed. We consider the general dynamic factor model and show how to impose smoothness restrictions on the factor loadings. We further present a statistical procedure based on Wald tests that can be used to find a suitable set of such restrictions. We present these developments in the context of term structure models, but they are also applicable in other settings. We perform an empirical study using a data set of unsmoothed Fama-Bliss zero yields for US treasuries of different maturities. The general dynamic factor model with and without smooth loadings is considered in this study together with models that are associated with Nelson-Siegel and arbitrage-free frameworks. These existing models can be regarded as special cases of the dynamic factor model with restrictions on the model parameters. For all model candidates, we consider both stationary and nonstationary autoregressive processes (with different numbers of lags) for the latent factors. Finally, we perform statistical hypothesis tests to verify whether the restrictions imposed by the models are supported by the data. Our main conclusion is that smoothness restrictions can be imposed on the loadings of dynamic factor models for the term structure of US interest rates but that the restrictions implied by a number of popular term structure models are rejected.

Thursday 29.10.2009 10:50-12:30 Parallel Session B

ES03 Room 6 ROBUST ANALYSIS OF COMPLEX DATA SETS 1 Chair: Stefan Van Aelst

E034: **Adaptations of the Stahel-Donoho estimator**
Presenter: **Ellen Vandervieren**, University of Antwerp, Belgium
Co-authors: Stefan Van Aelst, Gert Willems

The Stahel-Donoho estimator is a robust affine equivariant estimator of multivariate location and scatter with high breakdown point in any dimension. It is defined as a weighted mean and covariance, where each observation receives a weight depending on its outlyingness. Observations with large outlyingness receive small weights. In high-dimensional, low quality data it becomes unreasonable to assume that the majority of observations is completely uncontaminated. On the other hand it is wasteful to completely downweight an observation because it is contaminated in just a few components. In this talk we discuss adaptations of the Stahel-Donoho estimator that can better handle high-dimensional data. First, we consider a cellwise adaptation of the Stahel-Donoho estimator that makes a distinction between contaminated and non-contaminated components of an observation. Only contaminated components are then downweighted, which avoids losing the information contained in the other components. Second, with a large amount of componentwise outliers, projecting the data may lead to a cluster of intermediate outliers in one-dimensional projections. This may distort the location and scale estimation when computing the outlyingness and cause an unwanted increase of the outlier weights. Therefore, we propose a Huberized version of the Stahel-Donoho estimator that pulls the outliers back to the bulk of the data before the outlyingness is computed.

E073: **New model selection method by k-fold cross validation**
Presenter: **Shuichi Shinmura**, Seikei University, Japan

The Swiss Bank Note data having 6 independent variables and two classes has been previously considered within the context of discriminant analysis. An important feature of this data set is explained. Specifically, it is linearly separable in two dimensions by independent variables (X4, X6). Moreover, optimal linear discriminant function (IP-OLDF) based on minimum misclassification number (MMN) using integer pogramming finds that 16 models including the two-dimensional model with X4, X6 are linearly separable. On the other hand, stepwise methods based on the Cp statistic or AIC select five or six variables models, while the misclassification number of 16 models is zero. I generated 20,000 resampling data, and applied 10- and 100-fold cross-validation for Fisher's linear discriminant function (LDF) and logistic discrimination. Minimum, maximum and mean of CV error rates of the logistic discrimination are minimal for the same 16 models as the ones selected by IP-OLDF. So, this method is useful for model selection.

E131: **Estimators and tests for copulas based on likelihood depth**
Presenter: **Christine Mueller**, Technical University of Dortmund, Germany
Co-authors: Liesa Denecke

Estimators and tests based on likelihood depth for one-parametric copulas are given. For the Gaussian and Gumbel copula, it is shown that the maximum depth estimator is biased. It can be corrected and the new estimator is robust for the Gumbel copula. Tests are derived via the simplicial likelihood depth for hypotheses about the unknown parameter. Since the simplicial likelihood depth leads to biased estimators as well, the power of tests for some hypotheses is low. However, the power can be improved significantly by another bias correction. Simulation studies for the Gaussian and Gumbel copulas show that the power of the new tests is much better. Moreover, outlier robust tests are achieved for the Gumbel copula.

E110: **Robust bootstrap tests in MANOVA models**
Presenter: **Stefan Van Aelst**, Ghent University, Belgium
Co-authors: Gert Willems

The standard one-way ANOVA problem is considered. That is, we want to test the null hypothesis that multiple groups of multivariate observations share a common center. A standard test statistic is Wilk's Lambda which corresponds to the likelihood ratio statistic under the assumption that all groups are multivariate normal with a common covariance matrix. Clearly, the classical Wilk's Lambda test is very sensitive to outliers. Therefore, we consider robust alternatives that are based on robust estimators of multivariate location and scatter. We propose test statistics that are based on S or MM–estimators of location and scatter. To determine the null distribution of the test statistics we would like to use the fast and robust bootstrap (FRB) procedure. So far the FRB has mainly been used successfully to construct robust confidence intervals and standard errors, but much less for robust tests. We develop an FRB procedure that is consistent for the null distribution of the test statistics. We then investigate the finite-sample accuracy of this FRB method as well as properties of the resulting robust test procedures, such as their robustness and power.

ES32 Room 3 SOFT COMPUTING AND STATISTICS Chair: Gil Gonzalez-Rodriguez

E060: **Fuzzy expert maps: development and applications**

Presenter: **R. Jasinevicius**, Kaunas University of Technology, Lithuania
Co-authors: K. Kavaliauskas, R. Krusinskiene, V. Petrauskas

The concept of fuzzy expert maps (FEM) is considered as one of several possible extensions of very well known decision support tools - fuzzy cognitive maps (FCM). In this paper we describe a process of inclusion into the FCM of more sophisticated expert knowledge presented in two forms: 1) in the form of Mamdani-type reasoning and 2) in the form of Takagi-Sugeno-type reasoning. Newly born product here is called as fuzzy expert map (FEM). Two types of FEM are proposed: the FEM_L is developed on the basis of ordinary FCMs, and the FEM_N is developed on the basis of, so-called, nonlinear FCMs. The structures and practical applications are presented in both cases. Proposed ideas were implemented in different decision support tools. We deliver examples of such tools proposed for: a) marine port security systems; b) decision making in international conflict problems solutions; c) medical diagnostics systems based on investigation of human posturograms (what is done for the first time in medical practice); d) training recognition and classification different types of heart diseases. The cases c) and d) were enriched by inclusion of a new hyperinference procedure as well as by a special training procedure based on it.

E125: **Noiseless IFA with soft labels with applications**

Presenter: **Etienne Come**, Universite Paris 1, France
Co-authors: Latiffa Oukhellou, Thierry Denoeux, Patrice Aknin

In Independent Factor Analysis (IFA), latent components (or sources) are recovered from their observed linear mixtures. Both the mixing process and the source densities (that are assumed to be mixtures of Gaussians) are learnt from observed data. This presentation investigates the possibility of estimating the IFA model in the noiseless setting using prior information on the cluster membership of some examples. Semi-supervised or partially supervised learning frameworks can thus be handled but also more general cases such as soft labels. This extension of IFA, based on the Dempster-Shafer theory of evidence, offers a practical way to deal with specific problems such as label noise. The derivation of the criterion used by our method will be detailed and a clear definition of a GEM type algorithm for its optimization will be presented. Furthermore, results from a real-world application concerning fault diagnosis of a railway device will be provided to demonstrate the potential of this solution in partially supervised dimensionality reduction problems.

E188: **Takagi-Sugeno belief models**

Presenter: **Rui Jorge Almeida**, Erasmus University Rotterdam, Netherlands
Co-authors: Uzay Kaymak

We study how to derive fuzzy rule classification models based on the theoretical framework of belief functions. For this purpose we use the recently proposed Evidential c-means (ECM) to derive first order Takagi-Sugeno models solely from data. The proposed method creates models that use if-then rules and logical connectives to establish relations between the variables defined for the model of the system. Evidential c-means is based on the concept of credal partition, by extending the existing concepts of hard, fuzzy (probabilistic) and possibilistic partition by allocating, for each object, a *mass of belief*, to any subsets of possible clusters. The additional flexibility of ECM allows gaining a deeper insight in the data while being robust with respect to outliers, which makes it an attractive tool for exploratory statistics. The possible interpretability and linguistic description of the knowledge in these models is also discussed. Experiments with synthetic and real data sets show the possibilities of the proposed algorithm in the field of machine learning and exploratory statistics. Furthermore, we study the performance of such models in a classification setting and consider their added value in modeling.

E119: **K-sample homoscedasticity test for fuzzy random variables**

Presenter: **Ana Belen Ramos-Guajardo**, European Centre for Soft Computing, Spain
Co-authors: Gil Gonzalez-Rodriguez, Maria Angeles Gil, Ana Colubi

Different problems in testing a statistical hypothesis concerning fuzzy data have been analyzed in the literature. Most of these previous works are based on the study of the fuzzy mean of a fuzzy random variable. Regarding the one-sample hypothesis test about the variance of a fuzzy random variable, valuable tools to solve the test have been developed. In this work a procedure to test the equality of variances (homoscedasticity) of at least two fuzzy random variables is presented. The variance is defined in terms of a generalized distance quantifying the variability of the fuzzy values of the variable about its (fuzzy) expected value. The proposed method is carried out on the basis of the classical ANOVA procedure applied to dependent variables, which is stated in terms of the distances between the variable values in each group and the sample mean of that group. The test statistic is defined considering the within-group variation and the between-group variation and its convergence is analyzed by using asymptotic techniques. In addition, some simulations are presented to show the behavior of the asymptotic test. Finally, the approach will be exemplified through a case study.

ES38 Room 5 EVOLUTIONARY PROCEDURES Chair: Irene Poli

E200: **Evolutionary experimental designs for synthetic proteins**
Presenter: **Matteo Borrotti**, Universita di Bologna, Italy
Co-authors: Davide De Lucrezia, Giovanni Minervini

To design synthetic proteins we address the problem of searching in high dimensional and complex spaces. The combinatorial nature of proteins, the non-linear interactions among their components and the complexity of the optimization process make the design and the engineering of new synthetic proteins extremely hard. In this paper we propose an evolutionary approach that combines classical statistical design of experiments with meta-heuristic algorithms. More specifically we develop a class of evolutionary experimental designs where the evolution is lead by ant colonies algorithms. Encouraging results from simulations studies will be presented. This methodology could impact the field of protein engineering and design to improve existing proteins or design novel ones for specific applications in different fields such as medicine and fine-chemicals production.

E199: **Evolutionary Bayesian networks for high-dimensional stochastic optimization**
Presenter: **Debora Slanzi**, University Ca' Foscari of Venice, Italy
Co-authors: Irene Poli

High-dimensional and complex stochastic problems are difficult to analyze with classical statistical procedures and optimization techniques. In this paper we address this problem by developing an evolutionary Bayesian network procedure. Bayesian networks use the language of graphs to facilitate the representation and the resolution of complex problems, and the required probability theory to quantify the uncertainty. We present a probabilistic model-based evolutionary approach for the analysis and the design of biochemical complex experiments where the evolution is governed by the information achievable by statistical graphical models on the experimental results. The basic idea consists of inducing a Bayesian network from the best-found experimental set of solutions, estimating new solutions and evolving the model. Simulating the whole combinatorial space will allow us to explore, in a virtual way, all the possible combinations and test the performance of the approach before being adopted for real experiments. Moreover the dependence and conditional independence relations of complex interactions can be elucidated. Comparisons with the classical genetic algorithm approach show the better performance of the evolutionary Bayesian networks in several different experimental scenarios.

E198: **The optimisation of building envelopes with evolutionary procedures**
Presenter: **Giovanni Zemella**, University Ca' Foscari of Venice, Italy
Co-authors: Davide De March

Among all the different sources of energy consumption, buildings are responsible for the largest proportion of CO2 emission. Therefore the improvement of the energy efficiency of buildings can have a very significant impact on the reduction of the overall carbon emission. Facades have direct influence on heat losses, cooling energy requirements and the necessity of artificial lighting. In order to design the envelope of sustainable buildings, it is necessary to apply a holistic approach that takes into account all the aspects related to the energy consumption. Hence, the designer needs to identify the best combination of variables, describing the facade, which leads to the minimisation of carbon emissions. The evolutionary algorithms have proved to be very effective in dealing with optimisation problems. Nevertheless, the potential of these techniques has not been deeply developed in the field of building engineering. This paper develops a comparison between Particle Swarm Optimisation (PSO) and Evolutionary Neural Network (ENN) methods for the evaluation of the optimum facade.

E201: **Spline regression for an evolutionary approach to experimental design**
Presenter: **Claudio Pizzi**, University Ca' Foscari of Venice, Italy
Co-authors: Francesca Parpinel, Margherita Soligo

It is well known that the aims of the design of experiments are both the identification of the factors explaining the level of the response variable and capturing the level reaching the optimal response value. The number of possible experiments increases dramatically with the number of the factors and of their levels. In some cases the high dimension does not allow that all the possible experimental tests can be run because of budget and time constraints. We deal with the problem of experiments complexity in order to reduce the number of tests necessary to obtain information regarding the design variables and the levels of these factors. We propose to use the evolutionary paradigm jointly with a nonparametric approach. More precisely, the proposed procedure is split into two phases; first we select the experiments performed using a genetic algorithm, then we use spline regression to estimate the response surface. To verify if the proposed procedure is able both to identify correctly the factors and to provide their optimal level, we have applied the evolutionary design to a simulated biological problem with high dimensionality.

ES42 Room 7 TIME SERIES MODELLING AND COMPUTATION 1 Chair: Konstantinos Fokianos

E132: **Weak dependence, models and some applications**
Presenter: **Paul Doukhan**, Cergy-Pontoise, France

The basic features of weak dependence are presented. Our main aim is to propose several new models of time series and random fields for which this theory applies beyond mixing. Nonlinear, nonMarkov stationary models will be proposed. A fast evocation of applications will be provided. Extreme values theory is considered, namely we give precise conditions in terms of weak dependence assumptions and detail the way to subsample quantiles of the limit distributions so called (VaR). Another application is given by subsampling variances for weakly dependent sequences. This variance can be expressed in terms of a complicated series and various techniques yield its estimation; we exhibit some cases for which even if the series may be expressed explicitly, a subsampling method seems to be more convenient.

E154: **Phase estimation for fluctuation processes**
Presenter: **Rainer Dahlhaus**, University of Heidelberg, Germany
Co-authors: Jan Neddermeyer

A new class of stationary time series models is introduced. The aim is to model time series with a specific oscillatory pattern but an unobserved phase process in the background. The goals are to estimate the unknown phase process and the oscillatory pattern. An example is the curve of an electrocardiogram recording. The model can be written as a general state-space model treating the phase, amplitude, and baseline as latent Markov processes. For the estimation we suggest a Rao-Blackwellized particle smoother that combines the Kalman smoother and an efficient sequential Monte Carlo smoother. Sequential Monte Carlo smoothers can be applied to nonlinear, non-Gaussian state space models and are based on the idea to approximate the smoothing distribution of the latent states by weighted samples. For the estimation of the oscillatory pattern we develop a nonparametric estimation procedure.

E153: **Extensions of the Lee-Carter model for mortality projections**
Presenter: **Udi Makov**, University of Haifa, Israel
Co-authors: Shaul Bar-Lev, Yaser Awad

Mortality projections were dominated in the 1990's by the Lee-Carter model which assumes that the central death rate for a specific age follows a log-bilinear form, allowing for variations in the level of mortality over time. This model, with its inherent homoscedastic structure, was later extended by a Poisson model governed by a similar log-bilinear force of mortality. Here we will discuss potential extensions to the Lee-Carter model along the following lines: (a) Present the model as a state-space model (b) Adaptation of time index variation (c) Adaptation of Bayesian approach for estimating model parameters, and (d) Adaptation of Bayesian model choice techniques.

E039: **Linear and loglinear Poisson autoregression**
Presenter: **Konstantinos Fokianos**, University of Cyprus, Cyprus

Geometric ergodicity and likelihood based inference for linear and loglinear Poisson autoregressions are considered. In the linear case the conditional mean is linked linearly to its past values as well as the observed values of the Poisson process. This also applies to the conditional variance, implying an interpretation as an integer valued GARCH process. In a loglinear conditional Poisson model, the conditional mean is a loglinear function of its past values and a nonlinear function of past observations. Under geometric ergodicity the maximum likelihood estimators of the parameters are shown to be asymptotically Gaussian in the linear model. In addition we provide a consistent estimator of the asymptotic covariance, which is used in the simulations and the analysis of some transaction data. Our approach to verifying geometric ergodicity proceeds via Markov theory and irreducibility. Finding transparent conditions for proving ergodicity turns out to be a delicate problem in the original model formulation. This problem is circumvented by allowing a perturbation of the model. We show that as the perturbations can be chosen to be arbitrarily small, the differences between the perturbed and non-perturbed versions vanish as far as the asymptotic distribution of the parameter estimates is concerned.

CS01 Room 10 BUSINESS CYCLE: MODELLING AND FORECASTING Chair: Gian Luigi Mazzi

C040: **Multivariate nearest neighbours approach to forecast business cycle**
Presenter: **Patrick Rakotomarolahy**, University of Paris 1 Pantheon - Sorbonne, France
Co-authors: Dominique Guegan

We focus on the use of the nearest neighbor method (univariate and multivariate setting) in order to provide robust forecasts for economic indicators. We present in detail the nearest neighbors method providing new theoretical results that justify to use this approach for forecasting. Using some economic indicators -which are important to obtain fast estimates of GDP- we compare

their nearest neighbors estimates with another method: linear regression modelling and a radial basis function. We show their accuracy in forecasting.

C087: **Alternative specification of business cycle and growth cycle coincident indicators**
Presenter: **Leonardo Carati**, GRETA, Italy
Co-authors: Jacques Anas, Monica Billio, Gian Luigi Mazzi

The ability of promptly and correctly detecting turning points, especially in periods of high uncertainty or recession, is of primary importance for policy makers and analysts. In this paper, we present univariate and multivariate Markov Switching based alternative specifications of two euro area coincident indicators, for business and growth cycle respectively. We simulate in real time their performances by using the Eurostat real time database. Results of the simulations are assessed against the turning points identified in the euro area reference chronology. Particular attention is paid to the number of false signals or missing cycles characterising each indicator. By using the QPS and the Concordance Index, we are able to rank the alternative specifications and to identify the two outperforming specifications for growth and business cycle respectively. Main findings of the paper are that the selected indicators are not characterized by relevant false signals or missing cycles; they appear nevertheless slightly lagging with respect to the reference chronology.

C327: **On the estimation of common factors in the presence of block structures**
Presenter: **Claudia Cicconi**, Universite Libre de Bruxelles and ISTAT, Italy

Factor models in which the cross-correlation among time series is due to both common and block-specific factors are considered. Under certain conditions, these models are asymptotically equivalent to approximate factor models and common factors can be consistently estimated by both principal components and quasi maximum likelihood. However, since local cross-correlation can slow down the law of large numbers, these two estimators may have poor properties in finite samples. We show by Monte Carlo simulation that maximum likelihood estimates obtained by explicitly modelling the block structure of the data can indeed provide a better estimation of the common factors. The potential advantages of modelling the block structure are then illustrated in an empirical application using sectoral business survey data.

C111: **Evaluation of non-linear time series models for real-time business cycle analysis**
Presenter: **Gian Luigi Mazzi**, European Commission, Luxembourg
Co-authors: Monica Billio, Laurent Ferrara, Dominique Guegan

We aim at assessing Markov-Switching and threshold models in their ability to identify turning points of economic cycles. We evaluate the stability over time of the signal emitted by the models by using vintage data bases that are updated on a monthly basis. In this respect, we have built an historical vintage data base going back to 1970 for two monthly macroeconomic variables of major importance for the short-term economic outlook, namely the IPI and the unemployment rate. Applying two different classes of models, we compare their ability to detect the existence of turning points and their accuracy comparing their results with real time experience.

CS06 Room 4 ENVIRONMENT AND FINANCE Chair: Massimiliano Caporin

C033: **Temperature modellings and weather derivatives**
Presenter: **Dominique Guegan**, University of Paris 1 Pantheon - Sorbonne, France
Co-authors: Abdou Ka Diongue

Weather derivatives have been recently introduced on financial markets. The way to construct such derivatives is still in discussion, but a main point of the problem is the modelling of the underlying asset. Here, we focus on temperatures. Even if there exists a huge literature for modelling temperatures, we propose a novel methodology which takes into account both persistence, seasonalities and volatility associated with existence of asymmetry and leptokurtosis. Finally, we discuss how to build contracts for weather derivatives in an incomplete market.

C034: **An economical view of the carbon market**
Presenter: **Marius Frunza**, University of Paris 1 Pantheon - Sorbonne, France
Co-authors: Dominique Guegan

The aim of this work is to bring an econometric approach upon the CO2 market. We will identify the specificities of this market, but also regard carbon as a commodity. Then we will show the econometric particularities of CO2 prices behavior and the results of the calibration. We explain the reasons of the non-Gaussian behavior of this market and we will focus mainly upon jump diffusion and generalized hyperbolic distributions. We test these results for the risk modeling of a structured product specific to the carbon market, the swap between two carbon instruments: the European Union Allowances and the Certified Emission Reductions. We estimate the counter party risk for this kind of transaction and evaluate the impact of different models upon the risk measure and the allocated capital.

C358: **An empirical analysis of a reduced-form permit price model**
Presenter: **Georg Gruell**, University of Ulm/ University of Essen, Germany
Co-authors: Luca Taschini

In the so-called environmental finance research strand, reduced form models have been proposed with the aim to describe the evolution of the equilibrium price of emission permits. It is demonstrated how two competing reduced form models are related to each other. Then, by means of calibration to historical data it is shown that standard continuous time stochastic models such as NIG outperform reduced form models in the current price-evolution framework.

C074: **Modelling and forecasting wind speed with an example on wind derivatives pricing**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Juliusz Pres

The modeling of wind speed is a traditional topic in meteorological research where the focus is on the short term forecast of wind speed intensity and direction. More recently this theme has received some interest in quantitative finance for its relations with the electricity production by wind farms. In fact, electricity producers are interested in long range forecasts and simulation of wind speed time series for two main reasons: to evaluate the profitability of a wind farm to be built in a given location and to offset the risks associated to the variability of wind speed for an already operating wind farm. Here three approaches capable for forecasting and simulating the long run evolution of wind speed intensity (direction is not a concern given that the recent turbines can rotate to follow wind speed direction) are compared: the Auto Regressive Gamma process, the Gamma Auto Regressive process and the ARFIMA-FIGARCH model. We provide both an in-sample and out-of-sample comparison of the models as well as some examples for the pricing of wind speed derivatives using a model based Monte Carlo simulation approach.

CS24 Room 9 GRAPH BASED MODELLING AND CAUSALITY Chair: Marco Reale

C177: **Partial correlation graphs and structured time series models**
Presenter: **Granville Tunnicliffe Wilson**, Lancaster University, UK

Experiences of the development of structured models for vector time series using an empirical approach based on partial correlation graphs are described. The central role of modeling the series innovations will be considered, with discussions of structure in terms of instantaneous causality, simultaneous equation dependence and underlying continuous time processes. The selection and structure of functions of the past as predictors will also be considered, including both lagged series values and innovations, with discussion of model estimation and algorithmic relationships between the model and partial correlation graphs.

C187: **Comparing models with deformation metrics**
Presenter: **Miriam Hodge**, University of Canterbury, New Zealand
Co-authors: Jennifer Brown, Marco Reale

We propose using a deformation mapping metric to compare implied volatility surface models that result from differing model fitting frameworks. Deformation mapping was developed to measure and compare anatomical images resulting from medical imaging technology such as Magnetic Resonance Imaging (MRI). The goal of deformation mapping in medical imagery is to measure the distance between two images where distance can signify disease. The distance between two images is the minimum amount of energy required to deform one image to another measured by the velocity field required to map each point on the first image to the corresponding point on the second image. We will apply the deformation metric to statistical models in place of images and show that it preserves the intuitive notion of relative distance between the models. We will demonstrate the utility of deformation mapping by comparing three implied volatility models. We will model the SP500 index with a Black-Scholes model, a Karhunen-Loeve decomposition model and Garman-Kohlhagen model. The resulting deformation mappings will allow us to quantify the similarities between these models.

C211: **Identifiability of causal effects with non-ignorable missing data using instrumental variables**
Presenter: **Andrea Mercatanti**, Bank of Italy, Italy

The instrumental variables method has been widely adopted over the last two decades in identifying and estimating causal effects in econometrics. Typical examples are those wherein the template of a natural experiment with all-or-none compliance has been implemented in order to eliminate the bias arising from the self-selection of the units for treatment. In these cases, information is generally collected from surveys, wherein missing data is a common problem. However, the strand of literature dealing with missing data is typically restricted to the field of bio-statistics, where non-ignorable conditions for the missing data mechanism have been proposed for situations of missing data in the outcome, but not in the treatment or in the instrumental variable. There are various situations in econometrics, where data may not only be missing in the outcome, and where usual ignorability conditions for the missing data mechanism are considered to be so restrictive that simple integration of the likelihood function over the unobserved data may be misleading. In this study, I propose a set of conditions under which the non-ignorable missing data mechanism can be introduced. Its relative performance is assessed by simulations based on artificial samples.

C224: **Improving an algorithm for break points detection based on regression trees**
Presenter: **Jennifer Brown**, University of Canterbury, New Zealand
Co-authors: William Rea, Marco Reale

Theoretical regression trees are a fast computational methods to identify the presence of multiple shifts in mean in time series. However it was observed that in some simulations regression trees systematically reported breaks where no breaks existed. In particular, in simulations with a univariate time series called the noisy square wave there were a high number of trees with greater than the expected terminal nodes. By visually examining a plot of the regression trees it was clear that in many cases the regression tree algorithm's solution to the problem of incorrectly splitting the root node was to isolate the few incorrectly classified data points into a very small regime. Under these circumstances it seems intuitively obvious that the small regime ought to be combined with one of its two neighbours which are temporally contiguous with it. We provide an improved algorithm to resolve the issue.

CS39 Room 1 ENERGY ECONOMETRICS Chair: Marius Ooms

C104: **A vector autoregressive model for electricity prices subject to long memory and regime switching**
Presenter: **Frank Nielsen**, Aarhus University and CREATES, Denmark
Co-authors: Morten Nielsen, Niels Haldrup

A regime dependent VAR model is suggested that allows long memory (fractional integration) in each of the regime states as well as the possibility of fractional cointegration. The model is relevant in describing the price dynamics of electricity prices where the transmission of power is subject to occasional congestion periods. For a system of bilateral prices, non-congestion means that electricity prices are identical whereas congestion makes prices depart. Hence, the joint price dynamics implies switching between essentially a univariate price process under non-congestion and a bivariate price process under congestion. At the same time it is an empirical regularity that electricity prices tend to show a high degree of fractional integration, and thus that prices may be fractionally cointegrated. An empirical analysis using Nord Pool data shows that even though the prices strongly co-move under non-congestion, the prices are not, in general, fractionally cointegrated in the congestion state.

C197: **Estimation for unobserved component models with multiple stochastic variances using simulated maximum likelihood**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Siem Jan Koopman, Marius Ooms

Unobserved component (UC) models with multiple stochastic volatility (SV) processes have gained considerable interest in econometrics. For example, a generalization of the UC model with stochastic volatility for U.S. inflation in which not only the variance of the temporary disturbances but also the variance of the permanent disturbances varies over time has been suggested. Another example is the development of a multivariate UC model with stochastic volatility in the common permanent components of multiple exchange rates. The methods of estimation have so far relied upon Bayesian Markov Chain Monte Carlo methods and particle filtering. We develop a maximum likelihood method for estimation. As the loglikelihood function for this UC model is not available in closed form, we rely on importance sampling methods for its evaluation. The method builds on previous work which has introduced methods to construct importance sampling devices for observation densities that are not log-concave. While our method is based on an advanced methodology, we show that it remains feasible in practice. In particular, we illustrate our method for an UC model that has been previously considered. The empirical results are presented for a time series of electricity prices.

C173: **Dynamic factors in periodic time-varying regression models**
Presenter: **Virginie Dordonnat**, VU University Amsterdam, Netherlands
Co-authors: Marius Ooms

We consider dynamic multivariate periodic regression modeling for high- frequency data. The dependent univariate time series is transformed to a lower frequency multivariate time series for periodic regression modeling. For hourly series we specify one equation per hour of the day. The regression coefficients differ across equations and vary stochastically over time. As the unrestricted model contains many unknown parameters, we develop a methodology within the state-space framework to model dynamic factors in the coefficients, with common coefficient dynamics across equations. We first present a small-scale simulation, comparing results with a univariate benchmark model. Our dynamic factor component estimates are more precise. We apply our method to French national hourly electricity loads with weather variables and calendar variables as regressors and analyze components and forecasts.

C262: **Short-term electricity load forecasting with Generalized Additive Models**
Presenter: **Amandine Pierrot**, EDF R&D, France
Co-authors: Nicolas Laluque, Yannig Goude

Because of the French electricity market deregulation, EDF (Electricite de France) has to experiment with new models, which are more adaptive to changes in its portfolio size than operational forecasting models. The consideration of Generalized Additive Models for load forecasting is part of this search for new methods. The many different effects in electricity demand make its modeling and forecasting a difficult task. A statistical framework like GAM allows us to integrate both a regressive part with explanatory variables and an autoregressive part with lagged loads. The French electricity load being strongly related to the current hour, we consider twenty-four daily time-series and fit one non-parametric model for each hour. The selected variables are one-day-lagged loads, weather variables like the temperature of the current hour and lagged temperatures, calendar variables and a global trend. Thanks to a cyclic spline fitted on the position of the current day during the year, we can model the summer break (a large decrease in the demand due to the summer holiday) directly in our model, which is not possible in operational EDF forecasting models. Our model is fitted over five years. We compute the RMSE over one post-sample year to assess its accuracy for one-day ahead forecast.

CS64 Room 2 PANEL DATA MODELS Chair: Yasuhiro Omori

C066: **Bayesian and non-Bayesian analysis of quantile regression for longitudinal data**
Presenter: **Genya Kobayashi**, Kobe University, Japan
Co-authors: Hideo Kozumi

Quantile regression models for longitudinal data using the asymmetric Laplace distribution are considered. Simple and efficient Markov chain Monte Carlo (MCMC) methods for sampling the posterior distribution of the parameters are developed. A likelihood-based approach to the quantile regression model is discussed and Monte Carlo EM algorithms based on the proposed MCMC approach are presented. The methods are applied to both simulated and real data sets.

C292: **A Bayesian analysis of unit roots in panel data models with cross-sectional dependence**
Presenter: **Loukia Meligkotsidou**, University of Athens, Greece
Co-authors: Elias Tzavalis, Ioannis Vrontos

A Bayesian approach to unit root testing for panel data models is proposed. This is based on the comparison of stationary autoregressive models with and without individual deterministic trends, with their counterpart models with a unit autoregressive root. This is performed under cross-sectional dependence among the units of the panel. Simulation experiments are conducted with the aim to assess the performance of the suggested inferential procedure, as well as to investigate if the Bayesian model comparison approach can distinguish unit root models from stationary autoregressive models under or without cross-sectional dependence. The approach is applied to real GDP data for a panel of G7.

C073: **A panel examination of long-range dependence in DAX volatilities**
Presenter: **Matei Demetrescu**, Goethe University Frankfurt, Germany
Co-authors: Adina Tarcolea

The persistence of the daily volatility of the German stock market is analyzed in a panel of 30 DAX stocks. Since structural breaks can be mistaken as long memory, the paper tests the null of short memory while allowing for an unconditional shift in the mean at unknown time under the null. For individual stocks, we compute the lag-augmented LM test for fractional integration for all possible break points and take as test statistic the smallest of the resulting values. The test statistic obtained this way is shown to have a standard normal asymptotic distribution. In small samples, the test is biased towards antipersistence, so a small-sample correction is provided. The panel test statistic is built by combining the significance of individual test statistics with a correction for cross-unit dependence; this way, breaks at different times can be considered without increasing the computational complexity of the overall procedure. Examining daily absolute returns in 2004 and 2005, the volatility of DAX stock returns is found to have both a shift in the mean and long memory.

CS65 Room 8 QUANTITATIVE RISK MANAGEMENT 1 Chair: Marc Paolella

C185: **An exercise in stress testing for retail credit cards**
Presenter: **Tony Bellotti**, University of Edinburgh, UK
Co-authors: Jonathan Crook

Stress tests are becoming increasingly important for evaluating consumer credit risk and they are recognised as a key tool in helping financial institutions make business strategy, risk management and capital planning decisions. We present a simulation-based method for stress testing using discrete survival analysis to build a dynamic model of default at the account level which

includes macroeconomic conditions as risk factors. We discuss methods to generate plausible but extreme economic simulations within this framework and apply Monte Carlo simulation to compute empirical distributions of estimated default rates. We present experimental results for a large data set of UK credit card accounts that show that bank interest rates, production index, retails sales index and the FTSE 100 index are all statistically significant systemic risk factors for default. Stress tests using this model generate right skewed distributions of default rates and yield plausible extreme values of risk as measured by Value at Risk and expected shortfall. Finally we discuss statistical validation of stress tests through back-test of the loss distribution.

C284: **Estimation of distortion risk measures**

Presenter: **Hideatsu Tsukahara**, Seijo University, Japan

The class of distortion risk measures with convex distortions coincides with the set of coherent risk measures that are law invariant and comonotonically additive. The class includes the renowned expected shortfall which has many nice features and is of frequent use in practice. To implement the risk management/regulatory procedure using these risk measures, it is necessary to estimate the values of such risk measures. For a distortion risk measure, its form suggests a natural estimator which is a simple form involving L-statistics. Previous work has shown good asymptotic properties with i.i.d. data. This paper investigates the large sample properties of the estimator based on weakly dependent data. The conditions involve the strong mixing property, and are satisfied e.g., by GARCH sequences and stochastic volatility models. Related issues such as semiparametric estimation with extreme value theory and backtesting are briefly addressed.

C260: **Fund-of-funds construction by statistical multiple testing methods**

Presenter: **Michael Wolf**, University of Zurich, Switzerland
Co-authors: Dan Wunderli

Fund-of-funds (FoF) managers face the task of selecting a (relatively) small number of hedge funds from a large universe of candidate funds. We analyse whether such a selection can be successfully achieved by looking at the track records of the available funds alone, using advanced statistical techniques. In particular, at a given point in time, we determine which funds significantly outperform a given benchmark while, crucially, accounting for the fact that a large number of funds are examined at the same time. This is achieved by employing so-called multiple testing methods. Then, the equal-weighted or the global minimum variance portfolio of the outperforming funds is held for one year, after which the selection process is repeated. When backtesting this strategy on two particular hedge fund universes, we find that the resulting FoF portfolios have attractive return properties compared to the $1/N$ portfolio (that is, simply equal-weighting all the available funds) but also when compared to two investable hedge fund indices.

C012: **Fast estimation of highly parameterized GARCH models**

Presenter: **Jochen Krause**, University of Zurich, Switzerland
Co-authors: Marc Paolella, Markus Haas

A new estimation method for the mixed normal GARCH process is devised which is far faster than maximum likelihood and numerical Bayesian methods, and potentially more accurate. Its procedure is reminiscent of the non-MLE closed form techniques used in the estimation of ARMA models and based on the information contained in both the empirical quantiles and the sample autocorrelation function. Based on both simulated and actual returns data, the new method is found to be competitive to the MLE in terms of forecasting quality and vastly superior in terms of speed. The method is general and can be extended to other classes of GARCH models, some of which are also discussed.

Thursday 29.10.2009 14:00-16:00 Parallel Session C

ES05 Room 5 STATISTICAL SIGNAL EXTRACTION AND FILTERING 1 Chair: D.S.G Pollock

E217: **Alternative methods of seasonal adjustment**
Presenter: **Stephen Pollock**, University of Leicester, UK
Co-authors: Emi Mise

Alternative methods for the seasonal adjustment of economic data are described that operate in the time domain and in the frequency domain. The time-domain method, which employs a classical comb filter, mimics the effects of the model-based procedures of the SEATS–TRAMO and STAMP programs. The frequency-domain method eliminates the sinusoidal elements of which, in the judgement of the user, the seasonal component is composed. It is proposed that, in some circumstances, seasonal adjustment is best achieved by eliminating all elements in excess of the frequency that marks the upper limit of the trend-cycle component of the data. It is argued that the choice of the method seasonal adjustment is liable to affect the determination of the turning points of the business cycle.

E070: **Low-pass filter design using locally weighted polynomial regression and discrete prolate spheroidal sequences**
Presenter: **Tommaso Proietti**, University of Rome "Tor Vergata", Italy
Co-authors: Alessandra Luati

The design of nonparametric low-pass filters that have the property of reproducing a polynomial of a given degree is addressed. Two approaches are considered. The first is locally weighted polynomial regression (LWPR), which leads to linear filters depending on three parameters: the bandwidth, the order of the fitting polynomial, and the kernel. We find a remarkable linear (hyperbolic) relationship between the cutoff period (frequency) and the bandwidth, conditional on the choices of the order and the kernel, upon which we build the design of a low-pass filter. The second hinges on a generalization of the maximum concentration approach, leading to filters related to discrete prolate spheroidal sequences (DPSS). In particular, we propose a new class of lowpass filters that maximize the concentration over a specified frequency range, subject to polynomial reproducing constraints. The design of generalized DPSS filters depends on three parameters: the bandwidth, the polynomial order, and the concentration frequency. We discuss the properties of the corresponding filters in relation to the LWPR filters, and illustrate their use for the design of low-pass filters by investigating how the three parameters are related to the cutoff frequency.

E137: **Hyper-spherical and elliptical stochastic cycles**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Tommaso Proietti

A univariate first order stochastic cycle can be represented as an element of a bivariate VAR(1) process, where the transition matrix is associated with a Givens rotation. From the geometrical viewpoint, the kernel of the cyclical dynamics is described by a clockwise rotation along a circle in the plane. This paper generalizes this representation in two directions. According to the first, the cyclical dynamics originate from the motion of a point along an ellipse. The reduced form is ARMA(2,1), as in the circular case, but the model can account for asymmetries. The second deals with the multivariate case: the cyclical dynamics result from the projection along one of the coordinate axis of a point moving in $\Re^n$ along an hyper-sphere. This is described by a VAR(1) process whose transition matrix is a sequence of n-dimensional Givens rotations. The reduced form of an element of the system is shown to be ARMA(n, $n-1$). The properties of the resulting models are analyzed in the frequency domain, and we show that this generalization can account for a multimodal spectral density. The illustrations show that the proposed generalizations can be fitted successfully to some well-known case studies of econometric time series.

E221: **Generalized linear dynamic factor models: an approach via singular autoregressions**
Presenter: **Manfred Deistler**, Vienna University of Technology, Austria
Co-authors: Brian D.O. Anderson, Alexander Filler, Ch. Zinner, Weitian Chen

We consider generalized linear dynamic factor models. These models have been developed recently and they are used for high dimensional time series in order to overcome the *curse of dimensionality*. We present a structure theory with emphasis on the zeroless case, which is generic in the setting considered. Accordingly the latent variables are modeled as a possibly singular autoregressive process and (generalized) Yule–Walker equations are used for parameter estimation. The Yule–Walker equations do not have a unique solution in the singular case, and the resulting complexities are examined with a view to isolating a stable system.

ES24 Room 6 ADVANCES IN ROBUST DATA ANALYSIS Chair: Luis Angel Garcia-Escudero

E076: **Statistical applications of over-fitting due to trimmings**
Presenter: **Pedro C. Alvarez Esteban**, University of Valladolid, Spain

The overfitting effect is considered. Two random samples drawn from the same probability distribution and partially trimmed to make them as similar as possible, will be distinguishable from any other pair of non-trimmed samples of the same sizes. We provide sound and empirical evidence of this affirmation and design a general bootstrap procedure for comparison of two samples or one sample and a given distribution. This statistical procedure is also useful in other frameworks of model validation. As an added value of the principle, we provide an appealing methodology to analyze, from a nonparametric point of view, if we can assume that k samples arise from essentially identical underlying structures.

E097: **Robust analysis of default intensity**
Presenter: **Tiziano Bellini**, Universita Degli Studi di Parma, Italy
Co-authors: Marco Riani

We tackle the problem of estimating and calibrating default intensity through a robust approach. We exploit a discretization of Ornstein-Uhlenbeck stochastic process, that is known in financial literature as the Vasicek model, and we adopt the Kalman filter to calibrate one factor and multi-factor Vasicek model parameters applied to default intensity. In order to obtain robust estimates of the parameters which are not affected by outliers, we embed the Kalman filter algorithm in a forward search context. In this paper we use both simulated and real data from Deutsche Bank, General Electric and Intesa Bank credit spread time series to estimate their default intensities. We use robust confidence envelopes obtained from Monte Carlo simulations to check for the presence of outliers. The most interesting output of the procedure is a series of robust confidence intervals for the single-name intensity of default estimation. In other words, we build up not only a unique single-name intensity of the default curve, as typical in the literature, but a robust interval within which intensity of default is likely to stay.

E082: **An outlier robust Edwards-Havranek procedure for graphical models**
Presenter: **Sonja Kuhnt**, TU Dortmund University, Germany

Graphical models combine multivariate statistical models with a representation of conditional independence properties by a mathematical graph. The main research questions usually concern the association structure between the considered random variables. The process of selecting an appropriate model can severely be influenced by observations which deviate strongly from the main part of the data. We develop an outlier robust model building strategy by combining the well-know Edwards-Havranek procedure with the identification of outliers. A formal definition of outliers with respect to the conditional Gaussian distribution is used, which captures the perception of outliers as surprising observations. One-step outlier identification procedures are based on the classical maximum likelihood estimator and a new modified maximum likelihood estimator. Examples as well as simulation results are given for the new model building procedure.

E123: **Robust partial least squares regression**
Presenter: **Rosario Romera**, Universidad Carlos III de Madrid, Spain
Co-authors: Daniel Pena

PLS regression is a linear regression technique to relate many regressors to one or several response variables. Since the PLS method is known to be very sensitive to outliers, robustified PLS methods have been introduced to reduce or remove the effect of outlying data points. Most of them are just for building linear robust PLS methods. Nevertheless, when dealing with complex systems, linear PLS is inappropriate for describing the underlying data structure because the system may exhibit significant nonlinear characteristics. In this paper we briefly review the most recent proposed robust PLS methods including linear and nonlinear kernel-based methods. Their performance are illustrated by Monte Carlo and by some real data examples.

E061: **Tolerance zones through robust clustering techniques**
Presenter: **Luis Angel Garcia-Escudero**, Universidad de Valladolid, Spain
Co-authors: Alfonso Gordaliza, Agustin Mayo-Iscar

A technique for obtaining (asymptotic) distribution free tolerance intervals has been previously introduced. Those intervals contain the non-trimmed observations when computing the Least Trimmed Squares (LTS) estimator in the univariate location case. These tolerance intervals were also extended to tolerance ellipsoid based on the Minimum Covariance Determinant (MCD) estimator. In a similar fashion, tolerance zones can be obtained from the non-trimmed observations after applying trimmed k-means. Trimmed k-means were introduced with the aim of robustifying the classical k-means clustering method. These tolerance zones arise from trying to summarize the whole distribution through the use of quantization techniques instead of merely considering summary functionals like mean and scatter matrix estimators. The consideration of high values for parameter k provides very flexible and adaptive tolerance zones that turns out to be useful in complex Quality Control problems.

ES34 Room 3 STATISTICAL APPLICATIONS WITH FUZZY SETS Chair: Frank Klawonn

E037: **A fuzzy set-valued stochastic framework for birth-and-growth process. Statistical aspects**
Presenter: **Enea Bongiorno**, University of Milan, Italy
Co-authors: Giacomo Aletti, Vincenzo Capasso

A particular family of fuzzy monotone set-valued stochastic processes is considered. In order to investigate suitable alpha-level sets of such processes, a set-valued stochastic framework is proposed for the well-posedness of birth-and-growth process. A birth-and-growth model is rigorously defined as a suitable combination, involving Minkowski sum and Aumann integral, of two very general set-valued processes representing nucleation and growth respectively. The simplicity of the proposed geometrical approach allows us to avoid problems arising from an analytical definition of the front growth such as boundary regularities. In this framework, growth is generally anisotropic and, according to a mesoscale point of view, is not local, i.e. for a fixed time instant, growth is the same at each point space. The proposed setting allows us to investigate nucleation and growth processes. A decomposition theorem is established to characterize nucleation and growth. As a consequence, different consistent set-valued estimators are studied for growth processes. Moreover, the nucleation process is studied via the hitting function, and a consistent estimator of the nucleation hitting function is derived.

E063: **Reconstruction of flood chronologies on the basis of historical information**
Presenter: **Ana Colubi**, University of Oviedo, Spain
Co-authors: Elena Fernandez, Soledad Anadon

Historical information is seldom considered in flood analysis in spite of its relevance. Traditionally it has been used as a complement to calibrate models or to improve the estimation of extreme events. This is usually due to its limitations in connection with the requirement of a large volume of data, the reliability of the records and its current representativeness. However, these limitations may be overcome. A methodology aimed at the reconstruction of the flood chronology on the basis of historical data is presented here. The event intensity extended back to 1900 and the return period estimation will be addressed. Since some of the historical data are collected from interviews and other sources with different degrees of precision, the statistical variability and the imprecision will be jointly considered. The new methodology involves intervals (ranges), fuzzy sets, and weights to formalize and average the criteria which determine the importance of the different events. On the other hand, we propose to estimate the return period in a flexible and efficient way by considering bootstrap confidence intervals. The methodology is applied in an illustrative case study in a basin in north-western Spain.

E120: **Collecting and analyzing survey data using fuzzy random variables**
Presenter: **Maria Angeles Gil**, University of Oviedo, Spain
Co-authors: Gil Gonzalez-Rodriguez, Ana Belen Ramos-Guajardo

In many random experiments two kinds of uncertainty can be considered. On the one hand, we have the stochastic variability or randomness, which may be handled through probabilistic tools. On the other hand we may be interested in observing on each experimental outcome a non-precisely observed or intrinsically ill-defined characteristic. The lack of precision may be described through fuzzy sets. This work is centered on those situations in which personal perceptions, valuations or opinions about different topics are involved. Specifically, sociological surveys are considered, because they are based on subjective valuations, which most often are naturally associated with a certain degree of imprecision. In this framework, the imprecision will be described through a fuzzy scale instead of the usual one, which commonly ranges from strongly disagree to strongly agree. We propose to use trapezoidal fuzzy sets, defined on a support ranging from nullity to fullness. The difference between the usual and the suggested scales in developing statistical analysis will be discussed. In addition, some statistical techniques involving fuzzy data are explained in order to apply them to this kind of experimental data. To illustrate the approach, the methodology to collect and analyze survey data in a real-life application is described.

E178: **Fuzzy temporal random sets: a probabilistic tool in image and video processing**
Presenter: **Guillermo Ayala**, Universidad de Valencia, Spain

Total Internal Reflection Fluorescence Microscopy (TIRFM) allows the imaging of fluorescent-tagged proteins along time over the plasma membrane. We are interested in the spatial-temporal dependencies between proteins. A natural procedure for image segmentation is thresholding the original gray-level images resulting in a binary image sequence in which a pixel is covered or not by a given fluorescent-tagged protein. This binary logic is not appropriate because it leaves a free tuning parameter to be set by the user which can influence on the inference procedure. We consider pairs of gray-level image sequences, associated with two pairs of fluorescent-tagged proteins simultaneously imaged by TIRFM as realizations of a bivariate fuzzy temporal random closed set. The need for fuzzy set theory arises from the ambiguity within the pixels inherent to gray-level images. These FRACS's can be described by different functions that are good descriptors of the joint behaviour of both proteins. We have to choose for each image of each sequence a possibly different threshold level. We have to establish the correspondence for both

sequences. Furthermore we apply some mean sets associated to a non temporal fuzzy random set to the segmentation of retinal vessels in fundus ocular images.

E100: **Learning ordinal partial class memberships with kernel-based proportional odds models**
Presenter: **Willem Waegeman**, Ghent University, Belgium
Co-authors: Jan Verwaeren, Bernard De Baets

In areas like machine learning, statistics and fuzzy set theory, classification problems have been studied where data instances not simply belong to one particular class, but exhibit a degree of membership to several classes. For example, if we would think of classifying humans into ethnic groups, then it happens that a person with a European mother and an Asian father obtains a partial membership to two classes. Models capable of representing and learning partial class memberships explicitly, are typically referred to as partial, mixed or fuzzy membership models. Using similar ideas as in existing algorithms, we present a supervised probabilistic approach for learning partial class memberships in a slightly more specific setting, namely the ordinal regression setting, where a linear order relation is defined on the classes. To this end, we combine kernel methods and proportional odds models, in which the binary-coded label vectors in the likelihood are replaced by $[0,1]$–valued class memberships. A gradient descent algorithm is implemented to maximize the likelihood. Empirical results on synthetic and real-world datasets confirm that our approach leads to quasi-identical results, but a tremendous reduction of computational complexity, compared to the naive approach, where partial class memberships are casted to discrete labels.

ES41 Room 10 COMPUTATIONAL STATISTICS 1 Chair: Marc Hofmann

E192: **Tree structure for variable selection**
Presenter: **Klea Panayidou**, University of Oxford, UK
Co-authors: Steffen Lauritzen

We aim to identify associations between binary variables. For this purpose we use graphical models and more specifically trees, which are connected, acyclic, undirected graphs. We describe an existing method that can optimally approximate a discrete joint distribution by a tree. Subsequently we focus on variable subset selection with the purpose of predicting the remaining ones. We propose an approximation to Shannon Entropy, termed tree entropy and use the greedy algorithm to optimize it. The methods are illustrated with experiments on genetic data.

E173: **Different variants of model-based coclustering of continuous data**
Presenter: **Mohamed Nadif**, Universite de Technologie de Compiegne, France
Co-authors: Gerard Govaert

Coclustering becomes a fundamental problem that has numerous applications in different fields and particularly in bioinformatics and text mining. This kind of method aims to obtain homogeneous blocks by clustering simultaneously a set of objects and a set of variables. To tackle the coclustering when the objects are described by quantitative variables an adapted mixture model is proposed. Different variants are defined and, to focus on the clustering context, the maximum likelihood and the classification maximum likelihood approaches are considered. Setting these variants in this last approach, derived dissimilarities are developed, their interpretation is discussed and their impact on the quality of clustering is studied.

E179: **Variable selection in joint mean-variance models using H-likelihood**
Presenter: **Christiana Charalambous**, The University of Manchester, UK
Co-authors: Jianxin Pan, Mark Tranmer

We propose to extend the use of penalised likelihood variable selection on hierarchical generalised linear models (hglms) for jointly modelling the mean and variance structures for multilevel social science data. The hierarchical structure increases the need to deal with variation both within and between levels of hierarchy. Misspecifying the variance structure can lead to biased results, which can be avoided by joint modelling of the mean and variance. The basic idea is that both the mean and variance can be modelled by sets of covariates. But how do we choose which covariates to include in the modelling process? In order to attenuate any possible modelling biases, a large number of covariates may be introduced initially. However, analysing high-dimensional data can often prove to be computationally intensive. Variable selection can help overcome that by eliminating the least promising covariates. For simultaneously modelling mean and dispersion, we use hglm methodology, thus not only do we include random effects in our models, but we also do not restrict them to follow a normal distribution. To do variable selection, we use the scad penalty, a penalised likelihood variable selection method which shrinks the coefficients of redundant variables to 0 and at the same time estimates the remaining important covariates.

E047: **Seasonal misspecification in long memory processes: a simulation study**
Presenter: **Yuliya Lovcha**, University of Alicante, Spain

Negative seasonal fractional integration is a common feature of many macroeconomic time series that appears as a result of seasonal adjustment. However, this characteristic is frequently ignored. In this simulation work, I study the influence of the omission of negative seasonal fractional integration on the estimate of the parameter of fractional integration at zero frequency. Through a spectral-density bootstrap approach, I simulate pure and mixed ARFISMA processes. For each of them, I estimate a set of ARFIMA processes, omitting seasonal fractional integration. On the basis of the testing procedure described in the paper I choose the one that fits data better. The bias in the estimate of the fractional integration parameter at zero frequency under seasonal misspecification is positive for the majority of the specifications considered and for all the specifications which can be clearly identified as processes with negative seasonal fractional integration. High negative bias appears only for processes with strong negative short memory (-0.9). In this extreme case, even under the true specification, it is impossible to distinguish this process from one with positive fractional integration at frequency $\pi = 3,14$. Thus, it cannot be clearly identified as a process with negative seasonal fractional integration. This case is studied deeply in the paper. For illustration purposes, I provide two empirical examples.

E031: **The use of the Scan and Grimson disease clusters tests in order to avoid the epidemic's spread**
Presenter: **Dimitrios Nikolakis**, University of Piraeus, Greece
Co-authors: John Goulionis, Vasilios Benos

Many methods have been proposed for the identification of disease clusters. However only a few are used routinely in published investigations. All of these methods have both advantages and limitations. A validation of two temporal disease clusters tests with the simple epidemic model is presented. We conclude with recommendations for which methods should be applied to which types of situations. We prove that when the infection rate is high and the population size is small the Scan method cannot be used for the detection of a simple epidemic. For such cases the Grimson method offers better results. When the size of population increases the Scan method begin to offer more encouraging results. In general, the Grimson method offers more reliable results than Scan.

CS10 Room 2 BAYESIAN ECONOMETRICS 1 Chair: Wai-Sum Chan

C010: **Comparison of two different methods for Bayesian student-t volatility model**
Presenter: **Boris Choy**, University of Sydney, Australia
Co-authors: Jennifer Chan, Joanna Wong

A stochastic volatility (SV) model with leverage is considered. The log-return and the log-volatility are assumed to follow a bivariate Student-t distribution. This paper proposes an efficient method to implement the SV model with leverage. The performance of the proposed method is demonstrated via simulation studies. This paper also shows that the proposed method is very useful in outlier diagnostics. Existing methods cannot tell whether the outliers are outlying in the log-return, log-volatility or both. Our method provides an answer. For illustrative purpose, stock index returns data and exchange rates data are analyzed.

C183: **Evaluating short-run forecasting properties of the KOF employment indicator for Switzerland in real time**
Presenter: **Boriss Siliverstovs**, ETH ZURICH, Switzerland

This study investigates the usefulness of the business tendency surveys collected at the KOF institute for short-term forecasting of employment in Switzerland aggregated in the KOF Employment Indicator. We use the real time dataset in order to simulate the actual predictive process using only the information that was available at the time when predictions were made. We evaluate the presence of predictive content of the KOF Employment Indicator both for nowcasts that are published two months before the first official release and for one-quarter ahead forecasts published five months before the first official release. We produce now- and forecasts of the employment growth rates for the current and the next quarters with help of the autoregressive distributed lag (ARDL) models cast in the Bayesian Model Averaging framework. In doing so, we avoid an, often arbitrary, choice of a single best model and integrate model selection uncertainty in making out-of-sample predictions. We find that inclusion of the KOF Employment Indicator leads to substantial improvement both in in-sample as well as, more importantly, in out-of-sample prediction accuracy. This conclusion holds both for nowcasts and one-quarter ahead forecasts.

C351: **Expected shortfall and Value at Risk via the asymmetric Laplace distribution**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Qian Chen

A parametric approach to estimating and forecasting Value-at-Risk (VaR) and Expected Shortfall (ES) for a heteroskedastic financial return series is proposed. A GJR-GARCH models the volatility process, capturing the leverage effect. To take account of potential skewness and heavy tails, the model assumes an asymmetric Laplace distribution as the conditional distribution

of the financial return series. Furthermore, dynamics in higher moments are captured by allowing the shape parameter in this distribution to be time-varying. The model parameters are estimated via an adaptive Markov Chain Monte Carlo (MCMC) sampling scheme, employing the Metropolis-Hastings (MH) algorithm with a mixture of Gaussian proposal distributions. A simulation study highlights accurate estimation and improved inference of parameters compared to a single Gaussian proposal MH method. We illustrate the model by applying it to return series from four international stock market indices, as well as the exchange rate of the AU dollar to US dollar, and generating 1 step ahead forecasts of VaR and ES. We apply standard and non-standard tests to these forecasts from this model and find that it out-performs many popular competitors.

C352: **Extension of geometric process model to conditional autoregressive range models**
Presenter: **Jennifer Chan**, University of Sydney, Australia
Co-authors: Philip Yu, Connie Lam, Sai-tsang Choy

A geometric Process (GP) model is proposed as an alternative model for financial time series. The model contains two components: the mean of an underlying renewal process and the ratio which measures the direction and strength of the dynamic trend movement over time. Compared with the popular GARCH and SV models, this model is simple and easy to implement using the Bayesian method. In order to allow the dynamic trend movement as well as the time-varying and carryover nature of the volatility, we extend the GP model to the conditional autoregressive range model (CARR) to analyze the intra-day log price range for four Asian stock markets. We assign a distribution, say Gamma, to the underlying renewal process and some autoregressive functions to both the mean and ratio of the GP. Finally, models are selected according to BIC and forecasting is then performed for the best model.

C145: **Structural break in instability of return prediction models with heteroskedasticity**
Presenter: **Cathy Chen**, Feng Chia University, Taiwan
Co-authors: Kevin Lee

Numerous studies in the literature focus on whether stock price changes could be predicted by using past information, thus prediction regression models are developed to analyze stock returns. This study investigates evidence of instability in return models with heteroskedastic dynamic of ex post predictable components related to structural breaks in the coefficients of state variables. GARCH-type dynamics are added to the return prediction model, in order to efficiently examine the relationship between state variables and market return over time. An adaptive Bayesian Markov chain Monte Carlo scheme is designed for estimation and inference of unknown parameters and structural breaks, simultaneously estimating and accounting for heteroskedasticity. The proposed methods are illustrated using both simulated and international stock market return series. An empirical study of Taiwan and Hong Kong stock markets based on our model with oil and gas prices as a state variables provides strong support for instability of return prediction models with heteroskedasticity having recent structural breaks.

CS11 Room 4 QUANTITATIVE RISK MANAGEMENT 2 Chair: Marc Paolella

C248: **Conditional EVT for VaR estimation: comparison with a new independence test**
Presenter: **Paulo Araujo Santos**, Polytechnic Institute of Santarem and CEAUL, Portugal
Co-authors: Isabel Fraga Alves

We compare the out-of-sample performance of methods for Value-at-Risk (VaR) estimation, using a new exact independence test. This test is appropriate for detecting risk models with a tendency to generate clusters of violations and overcomes several limitations of the available tests for evaluating the performance under heteroscedastic time series. We focus the comparison on a two-stage hybrid method which combines a GARCH filter with a Extreme Value Theory (EVT) approach, known as Conditional EVT. Previous comparative studies show that this method performs better for VaR estimation. However, the independence tests used in previous comparative studies, suffer from limitations. Our contribution is comparing the performance with the new exact independence test. This allows us to show, with additional evidence, the superiority of the Conditional EVT method for VaR estimation under heteroscedastic financial time series.

C014: **Saddlepoint approximation of expected shortfall for transformed means**
Presenter: **Simon Broda**, University of Zurich, Switzerland
Co-authors: Marc Paolella

Expected Shortfall, as a coherent risk measure, has received a substantial amount of attention in the literature recently. For many distributions of practical interest however, it cannot be obtained in explicit form, and numerical techniques must be employed. The present manuscript derives a saddlepoint approximation for the Expected Shortfall associated with certain random variables that permit a stochastic representation in term of some underlying random variables possessing a moment generating function. The new approximation can be evaluated quickly and reliably, and provides excellent accuracy. The doubly noncentral t distribution is considered as an example.

C154: **Value at Risk and Expected Shortfall: A forecast combination approach**
Presenter: **Constantinos Kourouyiannis**, University of Cyprus, Cyprus
Co-authors: Elena Andreou, Andros Kourtellos

The recent financial crisis that started in 2007 provides a challenge for improving or proposing new methods to predict risk in the presence of extreme events and structural changes. We study two measures of market risk, Value at Risk and Expected Shortfall used by policy makers, practitioners and academics. We show that using forecast combinations to account for model uncertainty provides relatively more accurate forecasts of risk. To account for model uncertainty our model space includes models of two broad categories, parametric (e.g. normal GARCH) and semi-parametric (filtered historical simulation and extreme value theory). We compare the out of sample performance of individual models and alternative forecast combination methods using international stock market indices. The results in the empirical applications indicate that individual models of forecasting risk cannot generally outperform forecast combinations.

C313: **Measuring market risk in fixed income markets**
Presenter: **Michele Doronzo**, University of Zurich and UBS, Switzerland

Quantifying the market risk of a large fixed-income portfolio presents specific modelling issues which so far have not been adequately addressed. Previous research has been focusing mainly on the credit risk modelling of those securities. The current credit crisis exacerbated the need for accurate and versatile models in order to capture the market risk in the fixed income and credit markets. This study proposes various solutions and compares their out-of-sample performance. A method which utilizes a variation of multivariate filtered historical simulation is the most promising, displaying admirable in- and out-of-sample forecast results.

C285: **Revisiting marginal risk**
Presenter: **Simon Keel**, AERIS Capital AG, Switzerland
Co-authors: David Ardia

An important aspect of portfolio risk management is the analysis of the portfolio risk with respect to the allocations to the underlying assets. Standard concepts to achieve this are known as marginal and component risks in the financial literature. This paper presents a novel approach for analyzing the marginal risk of a portfolio. This new concept, called constrained marginal risk, accounts for the leverage induced when marginally increasing the position in the portfolio. A closed-form formula is derived in the case of an underlying elliptical model and a detailed illustration is provided for a synthetic portfolio. Finally, the financial relevance of the new concept is demonstrated with a portfolio of stocks.

CS12 Room 1 TIME SERIES ANALYSIS AND ECONOMIC APPLICATIONS Chair: Luc Bauwens

C043: **A class of nonstationary yet nonexplosive GARCH models with application to energy prices**
Presenter: **Jean-Michel Zakoian**, CREST and University Lille 3, France
Co-authors: Nazim Regnard

We introduce a class of GARCH(1,1) models in which the time-varying coefficients are functions of the realizations of an exogenous stochastic process. The model combines changes in regimes and nonstationarity. The changes of regimes occur at known dates and are driven by an observed process.Time series generated by this model are in general nonstationary. Necessary and sufficient conditions are given for the existence of non-explosive solutions, and for the existence of moments of these solutions. The asymptotic properties of the quasi-maximum likelihood estimator are derived under mild assumptions. As application the modelling of the volatility of gaz prices is proposed, in which the change of regimes are governed by the temperature.

C016: **Testing linear causality in mean in presence of other forms of causality**
Presenter: **Hamdi Raissi**, INSA RENNES, France

We consider the test for linear causality in mean in the large set of processes given by Vector AutoregRessive (VAR) models with dependent but uncorrelated errors. We see that this framework allows to take into account the possible presence of causality in mean and/or causality in variance. We derive the Quasi Maximum Likelihood Estimator (QMLE). Using the asymptotic normality of the QMLE we propose various modified tests for testing the causality in mean in presence of dependent errors. We study the finite sample performances of the modified tests by mean of Monte Carlo experiments. An application to the daily returns of the exchange rates of U.S. Dollars to one British Pound and of U.S. Dollars to one New Zealand Dollar is proposed to illustrate the theoretical results.

C129: **Extremal behaviour of aggregated economic processes in a structural growth model**
Presenter: **Fred Jouneau-Sion**, Universite Lille Nord de France, France
Co-authors: Stephane Auray, Aurelien Eyquem

We propose a new structural approach to the modelization of economic growth models. In this model, growth is generated by episodes of increasing return to scale induced by positive shocks on the accumulation of capital. The structural model derived from this framework is a Random Coefficient Autoregressive model. We then use the literature on extreme behaviour of dependent processes to related the possibility of sustained growth to the occurrence large swings in the aggregated economic time series. In particular we relate the extremal parameter to structural economic properties.

C204: **Bootstrap prediction mean squared errors of unobserved states based on the Kalman filter with estimated parameters**
Presenter: **Alejandro F. Rodriguez**, Universidad Carlos III de Madrid, Spain
Co-authors: Esther Ruiz

In the context of linear and Gaussian state space models, assuming that the parameters are known, the Kalman filter generates best linear unbiased predictions of the underlying components together with their corresponding prediction mean squared errors (PMSE). However, in practice, some parameters of the model must be substituted by consistent estimates when running the filter. The resulting PMSEs subestimate the true PMSE because they are not taking into account the uncertainty caused by parameter estimation. In this paper, we propose a new bootstrap procedure to obtain PMSE of the unobserved states based on obtaining replicates of the underlying states conditional on the information available at each moment of time. By conditioning on the available information, we simplify the procedure with respect to an alternative bootstrap proposal. Furthermore, we show that the new procedure has better finite sample properties than the alternatives. We illustrate the result by implementing the proposed procedure in both time-invariant and time-variant models. Finally, we obtain bootstrap prediction intervals of the unobserved trend in a time series of Spanish monthly inflation.

C128: **On marginal likelihood computation in change-point models**
Presenter: **Luc Bauwens**, UCLouvain, Belgium
Co-authors: Jeroen Rombouts

Change-point models are useful for modeling time series subject to structural breaks. For interpretation and forecasting, it is essential to estimate the correct number of change points in this class of models. In Bayesian inference, the number of change points is typically chosen by the marginal likelihood criterion, computed by Chib's method. This method requires to select a value in the parameter space at which the computation is done. Motivated by results from empirical illustrations, a simulation study shows that Chib's method is robust with respect to the choice of the parameter value used in the computations, among posterior mean, mode and quartiles. Furthermore, the performance of the Bayesian information criterion, which is based on maximum likelihood estimates, in selecting the correct model is comparable to that of the marginal likelihood.

CS18 Room 8 STOCHASTIC OPTIMIZATION IN FINANCE Chair: Daniel Kuhn

C102: **Threshold accepting for credit risk assessment**
Presenter: **Marianna Lyra**, University Giessen, Germany
Co-authors: Akwum Onwunta, Peter Winker

According to the latest Basel framework of Banking Supervision, financial institutions should internally assign their borrowers into a number of homogeneous groups. Each group is assigned a probability of default which distinguishes it from other groups. This study aims at determining the optimal number and size of groups that allows for statistical ex post validation of the efficiency of the credit risk assignment system. Our credit risk assignment approach is based on Threshold Accepting, a local search optimization technique, which has recently performed reliably in credit risk clustering especially when considering several realistic constraints. Using a relatively large real-world retail credit portfolio, we propose a new computationally less complex technique to validate ex-post the precision of the grading system.

C115: **Decision rule approximations for continuous linear programming**
Presenter: **Dimitra Bampou**, Imperial College London, UK
Co-authors: Daniel Kuhn

We study approximation schemes for continuous linear programs. We compute hierarchies of upper and lower bounds by approximating the functional form of the primal and dual control policies, respectively. We obtain different approximations by restricting the control policies to piecewise constant and polynomial functions of time. These approximations lead to semi-infinite problems, which are reformulated as tractable conic optimization models. Primal approximations are reformulated by using sums-of-squares polynomials and dual approximations are reformulated in terms of the solutions of moment problems. We apply the theory to example problems in management science.

C134: **Asset-liability management under time-varying investment opportunities**

Presenter: **Robert Ferstl**, University of Regensburg, Germany
Co-authors: Alex Weissensteiner

Multi-stage stochastic linear programming for asset-liability management under time-varying investment opportunities are proposed. A first-order unrestricted vector autoregressive process is used in order to model predictability in the asset returns and the state variables, where – in addition to equity returns and dividend-price ratios – Nelson/Siegel parameters are included to account for the evolution of the yield curve. The objective is to minimize the Conditional Value at Risk of shareholder value, i.e. the difference between the mark-to-market value of (financial) assets and the present value of future liabilities. The predictability effects in bond and equity returns are decomposed and shown that the strategy clearly benefits from dynamic asset re-allocation in a multi-stage setting.

C207: **Decision rule approximations for index tracking**

Presenter: **Angelos Georghiou**, Imperial College London, UK
Co-authors: Daniel Kuhn, Wolfram Wiesemann

We study an index tracking problem with the objective to minimize the Conditional Value-at-Risk (CVaR) of the terminal portfolio value while constraining the tracking error between the portfolio and a given benchmark. The resulting decision problem is formulated as a multi-stage uncertain linear program. We employ linear and piecewise linear continuous decision rules to approximate both the original problem and its dual counterpart. By using robust optimization techniques, we reformulate both approximate problems as a tractable linear programs or second-order cone programs which can be solved in polynomial time. The difference between the optimal primal and dual objective values allows us to estimate the underlying approximation error.

C356: **Multi-stage stochastic pension fund management**

Presenter: **Ronald Hochreiter**, University of Vienna, Austria

The optimal management of pension funds is important for handling the growing challenges in keeping stable nation-wide pension systems. In contrast to standard Asset Liability Management, the goal of managing a pension fund is not solely based on a maximization of profits, while ensuring the coverage of liabilities. In addition, the contradictory interests of both the active members and the retired members have to be considered. Furthermore, the set of regulatory constraints is huge, and constantly evolving. A multi-stage stochastic programming model for managing pension funds will be presented - with a special focus on generating realistic scenarios.

CS29 Room 9 MIDAS Chair: Andreas Savvides

C071: **Pooling versus model selection for nowcasting with many predictors: an application to German GDP**

Presenter: **Vladimir Kuzin**, DIW Berlin, Germany
Co-authors: Christian Schumacher, Massimiliano Marcellino

Pooling versus model selection for now- and forecasting in the presence of model uncertainty with large, unbalanced datasets is discussed. Empirically, unbalanced data is pervasive in economics and typically due to different sampling frequencies and publication delays. Two model classes suited in this context are factor models based on large datasets and mixed-data sampling (MIDAS) regressions with few predictors. The specification of these models requires several choices related to, amongst others, the factor estimation method and the number of factors, lag length and indicator selection. Thus, there are many sources of mis-specification when selecting a particular model, and an alternative could be pooling over a large set of models with different specifications. We evaluate the relative performance of pooling and model selection for now- and forecasting quarterly German GDP, a key macroeconomic indicator for the largest country in the euro area, with a large set of about one hundred monthly indicators. Empirical findings provide strong support for pooling over many specifications rather than selecting a specific model.

C097: **Forecasting inflation and economic activity using high frequency financial data**

Presenter: **Andros Kourtellos**, University of Cyprus, Cyprus
Co-authors: Elena Andreou, Eric Ghysels

New specifications of Mixed Data Sampling (MIDAS) regression models are employed in order to forecast key US quarterly indicators of inflation rate and economic growth using a new dataset of daily financial indicators. The advantage of MIDAS to provide new forecasts as daily data become available while mixing lower frequency data such as factors is exploited. In particular, using MIDAS regression models with leads and lags the daily data can let to absorb all revisions and examine how events like the Lehman bankruptcy affected the forecasts in the subsequent months. Two categories of factors are constructed: the first is based on quarterly macro factors and the second is based on daily financial factors using a larger cross section of financial series. Then, it is examined whether a MIDAS model, which involves daily financial and quarterly macro factors provides forecasting gains. The results suggest that simple univariate MIDAS models can outperform traditional univariate and Factor models. Moreover, it is found that, on average, daily financial predictors improve the forecasts of quarterly inflation and economic activity.

C195: **Changes in predictive ability with mixed frequency data**
Presenter: **Ana Galvao**, Queen Mary University of London, UK

This paper proposes a new regression model -a smooth transition mixed data sampling (STMIDAS) approach- that captures regime changes in the ability of a high frequency variable in predicting a variable only available at lower frequency. The model is applied to the use of financial variables, such as the slope of the yield curve, the short-rate, and stock returns, to forecast aggregate economic activity and inflation both in- and out-of-sample. The direct use of high frequency data and allowing for changes in the financial variables' predictive power strengthen evidence that stock returns lead economic activity.

C235: **Forecasting with mixed frequency factor models and MIDAS structure**
Presenter: **Cecilia Frale**, Ministry of the Economy and Finance Italy, Italy
Co-authors: Libero Monteforte

Short term forecasting suggests that there is room for improvement in forecast ability by considering a richer dynamic structure. To address this issue, multivariate factor models are combined with MIDAS dynamics to perform short term forecasts. A dynamic factor model with a mixed frequency is used, where the past observations of high frequency indicators are included following the MIDAS approach. This structure is able to represent, with richer dynamics, the information content of the economic indicators and produces smoothed factors and forecasts. The short term forecasting performance of the model is evaluated against other models in a pseudo-real time experiment, also allowing for pooled forecast from factor models.

C367: **Stock market volatility and the business cycle: a multi factor approach**
Presenter: **Gianluca Moretti**, Bank of Italy, Italy

Recently, a new specification of the GARCH model, the Spline-GARCH, has been proposed, which models the low frequency volatility with a slow-moving component represented by an exponential spline. Once this low frequency component is estimated they show that it is significantly related to some economic variables. We extend this approach in several ways. Firstly, rather than modelling the long-run variance component with a deterministic function, we directly link the conditional variance of the model to long run factors extracted from financial and economic data using a mixed frequency state space model. We show that, apart from accurately tracking movements in the volatility during financial crisis and the recessions, our approach allows us to measure the contribution of the business cycle factor and the financial factor to the overall low frequency stock market volatility. Then, differently from the standard spline GARCH, the model can be easily used to produce forecasts for both short and long term components in real time.

CS32 Room 7 FORECASTING, HEAVY TAILS AND NON-STANDARD INFERENCE 1 Chair: Lynda Khalaf

C092: **Forecasting with weakly identified linear state-space models**
Presenter: **Sebastien Blais**, Bank of Canada, Canada

Normalizing models in empirical work is sometimes a more difficult task than commonly appreciated. Permutation invariance and local non-identification cause well documented difficulties for maximum likelihood and Bayesian inference in finite mixture distributions. Because these issues arise when some parameters are close to being unidentified, they are best described as weak identification (or empirical underidentification) problems. Although similar difficulties arise in linear state-space models, little is known about how they should be addressed. In this paper, I show that some popular normalizations do not provide global identification and yield parameter point estimators with undesirable finite sample properties. At the computational level, I propose a novel posterior simulator for Gaussian linear state-space models, which I use to illustrate the relationship between forecasting performance and weak identification. In particular, Monte Carlo simulations show that taking into account parameter uncertainty reduces out-of-sample root mean square forecast errors when some parameters are weakly identified.

C326: **A quasi-likelihood approach based on eigenfunctions for a bounded-valued Jacobi process with an application**
Presenter: **Pascale Valery**, HEC Montreal, Canada

We consider a discretely sampled Jacobi process appropriate to specify the dynamics of a process with range [0,1], such as a discount coefficient, a regime probability, or a state price. The discrete time transition of the Jacobi process does not admit a closed form expression and therefore the exact maximum likelihood (ML) is infeasible. We first review a characterization of the transition function based on nonlinear canonical decomposition. They allow for approximations of the log-likelihood function which can be used to define a quasi-maximum likelihood estimator. The finite sample properties of this estimator are compared with the properties of other existing estimators, such as the estimator which is a method of moments based on an approximated score function, or with a generalized method of moments (GMM) estimator. The quasi-maximum likelihood estimator is further compared with computer–intensive simulation–based estimation techniques such as the indirect inference estimator, or the simulated method of moments (SMM). Indeed, these techniques naturally arise when the ML estimator is infeasible. With respect to computational efficiency, the QML estimator is less time consuming than the simulation-based techniques. We also focus on

computational issues for simulating the underlying continuous–time path of the Jacobi process from a truncated Euler discretization scheme. Finally an empirical application on bond default probability data from MSCI Barra inc. is performed to assess the performance of the Jacobi process to capturing the dynamic of a probability process.

C228: **A non-parametric model-based approach to uncertainty and risk analysis of macroeconomic forecasts**
Presenter: **Claudia Miani**, Bank of Italy, Italy

This paper deals with evaluating the risks (and ultimately the whole distribution) that surround a point forecast obtained from a given macroeconometric model. The predictive density of the model-based forecast is combined with subjective estimates of risks and uncertainty, as resulting from expert judgement and additional *off-model* information. We propose a non-parametric, model-based simulation approach to estimate the probability distribution of future values of key macroeconomic variables, such as inflation and output growth, around a given baseline forecast. More specifically, making assumptions on the (asymmetric) distribution of specific risk factors, stochastic simulations (bootstrap) from the macroeconometric model are used to deliver the new forecast distribution, that takes into account all the information on risks and uncertainty. The method is non-parametric in the sense that resampling is based on model's residuals (and a simple mechanism to generate skewness), without distributional assumptions. A graphical representation of the results is obtained through a so-called *fan chart*, that -unlike the one popularized by the Bank of England- can be genuinely interpreted in terms of the quantiles of the distribution. We believe our approach provides a clear communication device on forecast uncertainty, balance of risks and scenario analysis. The method is illustrated using the Bank of Italy Quarterly Model.

C109: **Modeling the contemporaneous duration dependence for high-frequency stock prices using joint duration models**
Presenter: **Ba Chu**, Carleton University, Canada
Co-authors: Marcel Voia

We test if duration dependence is present in stock returns by employing duration models. The presence of duration dependence in stocks returns indicates a failure of the efficient markets hypothesis, which states that holding period returns on a risky asset should be serially random. The hypothesis suggests that positive or negative stock returns should have short durations. In order to do the test we employ both discrete and continuous duration models to check the sensitivity of finding duration dependence when data is treated as either discrete or continuous.

C093: **Oil prices: heavy tails, mean reversion and the convenience yield**
Presenter: **Lynda Khalaf**, Carleton, Canada
Co-authors: Jean-Thomas Bernard, Maral Kichian, Sebastien McMahon

Empirical research on oil price dynamics for modeling and forecasting purposes has brought forth several unsettled issues. Indeed, statistical support is claimed for various models of price paths, yet many of the competing models differ importantly with respect to their fundamental temporal properties. In this paper, we study one such property that is still debated in the literature, namely mean-reversion, with focus on forecast performance. Because of their impact on mean-reversion, we account for non-constancies in the level and in volatility. Three specifications are considered: (i) random-walk models with GARCH and normal or student-t innovations, (ii) Poisson-based jump-diffusion models with GARCH and normal or student-t innovations, and (iii) mean-reverting models that allow for uncertainty in equilibrium price and for time-varying convenience yields. We compare forecasts in real time, for 1, 3 and 5 year horizons. For the jump-based models, we rely on numerical methods to approximate forecast errors. Results based on future price data ranging from 1986 to 2007 strongly suggest that imposing the random walk for oil prices has pronounced costs for out-of-sample forecasting. Evidence in favor of price reversion to a continuously evolving mean underscores the importance of adequately modeling the convenience yield.

Thursday 29.10.2009 16:30-18:50 Parallel Session D

ES10 Room 7 TIME SERIES MODELLING AND COMPUTATION 2 Chair: Roland Fried

E040: **Bandwidth selection for functional time series prediction**
Presenter: **Theofanis Sapatinas**, University of Cyprus, Cyprus
Co-authors: Anestis Antoniadis, Efstathios Paparoditis

We propose a method to select the bandwidth for functional time series prediction. The idea underlying this method is to calculate the empirical risk of prediction using past segments of the observed series and to select as value of the bandwidth for prediction the bandwidth which minimizes this risk. We prove an oracle bound for the proposed bandwidth estimator showing that it mimics, asymptotically, the value of the bandwidth which minimizes the unknown theoretical risk of prediction based on past segments. We illustrate the usefulness of the proposed estimator in finite sample situations by means of a small simulation study and compare the resulting predictions with those obtained by a leave-one-curve-out cross-validation estimator used in the literature.

E026: **On the hyperparameter estimation of time varying Poisson model for Bayesian WWW traffic forecasting**
Presenter: **Daiki Koizumi**, Waseda University, Japan
Co-authors: Toshiyasu Matsushima, Shigeichi Hirasawa

The aim is to contribute to traffic forecasting problems using a Bayesian approach for the time varying Poisson model. This model can be regarded as an application of the Simple Power Steady Model (S.P.S.M.). In this model, time variation of the parameter is formulated by a transformation function and its degree is caught by a real valued hyperparameter k ($0 < k < 1$). However, in S.P.S.M. it has not yet been proposed any definite parameter transformation function nor methods for estimation of the hyperparameter k. These two points are considered. Especially for the latter point, it has been empirically observed that World Wide Web (WWW) traffic forecasting performance strongly depends on the accuracy of the estimate of the hyperparameter k. Here at least takes two approaches for estimating the hyperparameter k, namely maximum likelihood and quasi-Bayes. Their effects on the traffic forecasting are discussed. According to the obtained results, the quasi Bayesian estimate gives satisfactory traffic forecasting under sufficiently large number of subintervals for integration with relatively low complexity.

E041: **Representing hidden semi-Markov models as hidden Markov models**
Presenter: **Roland Langrock**, University of Goettingen, Germany
Co-authors: Walter Zucchini

We show how it is possible to formulate any hidden semi-Markov model (HSMM) as a hidden Markov model (HMM) with a special structure that ensures the state dwell–time distributions are – at least approximately – the same as those of the HSMM. The HMM formulation has the same number of parameters as the HSMM and the approximation can be made arbitrarily accurate. In case of dwell–time distributions with finite support, the HMM formulation is equivalent to the HSMM. The main advantage of formulating the model as an HMM stems from the fact the parameters can be estimated by direct numerical maximization of the likelihood function. This makes it feasible to fit a much richer variety of HSMMs than was previously possible. In particular it becomes practical to fit stationary HSMMs, or to incorporate trend, seasonality and covariates in the model. Indeed one can make use of all the tools that have been developed for HMMs but which were not available for HSMMs. The new class of HMMs is applied to model daily rainfall occurrence for a number of sites in Bulgaria. We illustrate how it is possible to incorporate seasonality in different ways.

E135: **Automated likelihood based inference for stochastic volatility models using AD model builder**
Presenter: **Hans J. Skaug**, University of Bergen, Norway
Co-authors: Jun Yu, David Fournier

The Laplace approximation is commonly used to fit stochastic volatility (SV) models. We argue that the use of a numerical technique known as automatic differentiation in combination with the Laplace approximation greatly simplifies the practical fitting of SV models. Automatic differentiation, not to be confused with symbolic differentiation, evaluates first and higher order derivatives of computer programs (such as the likelihood of an SV model) accurately and efficiently. The approach is implemented in the open source software package ADMB (http://admb-project.org/). By exploiting the state-space structure of SV models we can implement the basic SV model in ADMB using less than 40 lines of C++ code. We also consider different variations over the SV model and techniques for improving the accuracy of the Laplace approximation.

E164: **Macroeconomic forecasting with leading indicators: Penalized nonlinear regression using kernels**
Presenter: **Peter Exterkate**, Erasmus University Rotterdam, Netherlands
Co-authors: Patrick J.F. Groenen

Kernels have become widely popular in the area of machine learning. They allow modeling highly nonlinear relations by mapping the observed predictor variables nonlinearly to some high dimensional feature space. Prediction takes place in this high dimensional space. To avoid overfitting, a penalty term is added. In many contexts, this form of kernel ridge regression performs very well. In spite of these attractive features, this method is not commonly applied in a time-series context. In this study, kernel regression is employed to forecast the Conference Board's Composite Coincident Index in various model specifications. This CCI, a weighted average of macroeconomic variables, is generally taken as a measure of the overall performance of the U.S. economy. The forecast performance is compared to that of more conventional, linear methods. Our results show that, especially when forecasting at longer horizons, allowing for nonlinearities improves predictive accuracy.

C036: **Estimation and validation in count time series models**
Presenter: **Robert Jung**, Universtaet Erfurt, Germany
Co-authors: Andrew R. Tremayne

This paper is concerned with models for time series of small counts based upon random operations to preserve integer structure. A leading case would be the first order integer autoregressive model, INAR(1), based upon a binomial thinning operator; this has received extensive attention in the literature. We focus on its second order generalization and a companion model based upon a different random operation. Parameter estimation, principally by maximum likelihood, is considered and simulation evidence of the performance in finite samples of all estimators considered is provided. An aim of this computational exercise is to provide some guidance to applied workers as to the sample sizes needed so that inferential procedures can be expected to behave reliably. In addition, a number of model validation exercises are advanced. One main approach stems from the use of parametric bootstrap analysis designed to assess the fit of the dependence structure in the observed data. A second arises from appropriate use of probability integral transformations, suitably adjusted to acknowledge the integer nature of the data, generally based on predictive distributions from models. The paper is completed by introducing a novel data set taken from financial econometrics and applying all foregoing methods to it.

ES14 Room 5 GENERALIZED MIXED MODELS Chair: Heather Turner

E050: **On the geometry of generalised linear mixed models**
Presenter: **Karim Anaya-Izquierdo**, The Open University, UK
Co-authors: Critchley Frank, Marriott Paul, Vos Paul

The geometry of Generalised Linear Models is well understood and has proved helpful in understanding their higher order asymptotic properties. Furthermore, the geometry of random effect models has been previously considered. New developments in geometry for statistical science have emerged in recent years to meet the challenge of understanding practical statistical problems. These new theoretical developments have been accompanied by a growing realisation of the potential of computational geometry and associated graphics to deliver operational tools. Our project uses computational information geometry to develop diagnostic tools to help understand sensitivity to choice of GLM by building an appropriate perturbation space. Our long-term aim is to engage with statistical practice via appropriate R software encapsulating the underlying geometry, whose details the user may ignore. In this talk, simple examples of extensions of our computational information geometry work to the GLMM case will be discussed.

E080: **Testing for misspecification in generalized linear mixed models: a SAS macro**
Presenter: **Saskia Litiere**, Hasselt University, Belgium
Co-authors: Ariel Alonso, Geert Molenberghs

The development of software tools such as the SAS procedures MIXED, NLMIXED and GLIMMIX, or the R functions LME, NLME and GLMM have greatly facilitated the use of the Generalized Linear Mixed Model (GLMM) for the analysis of non-Gaussian longitudinal data. However, this subject-specific model may not always be robust against departures from its model assumptions, like the choice of the random-effects structure. Hence, diagnostic tools for the detection of such misspecification are of the utmost importance. In this talk, we present a family of diagnostic tools along the ideas of White's Information Matrix Test, based on the information matrix equality and several equivalent representations of the model information matrix. We will evaluate the power of these tools using some theoretical considerations as well as via simulations, while focusing on misspecification of the random-effects structure. Additionally, we will illustrate how these tools can be used easily in a practical situation through the availability of SAS macros in which these diagnostic tools are readily implemented.

E185: **Flexible random effects in ordinal regression models**

Presenter: **Georgios Papageorgiou**, National University of Ireland, Galway, Ireland
Co-authors: John Hinde

Generalized linear mixed models have been used extensively in the analysis of clustered data. The random effects incorporated in the linear predictors of these models account for the correlation present in the data. Many approaches have been proposed for modeling the random effect distribution: the nonparametric approach with its resulting discrete estimate of the random effect distribution; the overly restrictive fully parametric approach; and also the so-called semi-nonparametric and smooth nonparametric approaches that do not make strong assumptions about the shape of the random effect distribution, but which allow some control over the smoothness of the resulting estimate. Here the class of multivariate generalized linear models is extended to include in the linear predictor random effects that are generated by a smooth density. The focus here is on ordinal regression models with a cumulative logit link. Maximum likelihood estimation of the model parameters is carried out using a Monte Carlo EM algorithm which uses a rejection sampling scheme and automatically increases the MC sample size after iterations in which the EM step is swamped by MC error. The methods are illustrated and compared by simulation and application to a data set.

E209: **Random effects in cumulative mortality models**

Presenter: **John Hinde**, National University of Ireland, Galway, Ireland
Co-authors: Silvia de Freitas, Marie-Jose Martinez, Clarice Demetrio

In toxicological experiments for biological pest control, experiments frequently involve the study of cumulative mortality in a groups of insects measured at various time points. Samples of different strains, or isolates, of the compound under study are used, typically with some replication. The example considered here is from a study of a microbial control to insect damage in sugar cane. Cumulative multinomial models provide an obvious approach to the analysis of these data, however, the basic model needs to be extended to account for overdispersion. Here we consider various random effect models, including the Dirichlet-multinomial and models with random effects in the linear predictor. For some simple time trend models, fitting using a generalized estimating equation approach leads to a surprisingly simple modification of the multinomial fit and casts light on the behaviour in more complex models. Other approaches considered include quadrature methods for normal random effects. We also consider modelling isolates as a random effect, with a mixture distribution to cluster isolates with similar action, together with other replicate random effects for overdispersion.

E142: **A model for correlated paired comparison data**

Presenter: **Manuela Cattelan**, University of Padova, Italy
Co-authors: Cristiano Varin

Paired comparison data are binary data that indicate which one of two objects under comparison is preferred or wins the comparison. In general, traditional models developed for the analysis of this type of data are applied assuming independence among all observations. This may be unrealistic as it is sensible to expect that the results of paired comparisons with a common object will be correlated. A new model is proposed that extends the existing models in a straightforward manner introducing correlation among observations with a common element. This is achieved through the addition of object specific random effects. Unfortunately, the computation of maximum likelihood estimates requires the solution of a high dimensional integral. The problem is overcome by means of composite likelihood techniques. Simulation studies show that composite likelihood gives satisfactory results. The model is illustrated through two examples. The first example considers results of volleyball matches and the second regards soccer data. The main difference between the two data sets is that volleyball matches end in a win for one of the competing teams, while in soccer the final result may also be a draw. In this case, an additional threshold parameter is necessary to specify the probability of draws.

E144: **Mixed Bradley-Terry models**

Presenter: **Heather Turner**, University of Warwick, UK
Co-authors: David Firth

In a basic Bradley-Terry model for paired comparisons, a separate 'ability' is estimated for each 'player' and the odds of player i beating player j is given by the ratio of these abilities. The model may be represented as either a logistic regression model or a log-linear model and fitted using standard glm software or more specialised packages, such as BradleyTerry or prefmod in R. Often however, the substantive interest lies not in estimating the abilities of a particular set of players, but modelling player abilities by player covariates. Usually this is implemented by replacing the individual player abilities by linear predictors in the glm model. An important drawback of this approach is that it does not allow for variability between players with the same covariate values. Clearly this can be overcome by incorporating random effects for the players. Whilst mixed Bradley-Terry models could be handled in principle by any glmm software, specialised software is desirable to enable natural representation of the models and provide useful summary functions. Here we present recent developments of the BradleyTerry package, which extend its capabilities to fit mixed Bradley-Terry models and also to incorporate contest-specific effects.

ES15 Room 6 ROBUST ANALYSIS OF COMPLEX DATA SETS 2 Chair: Andreas Christmann

E104: **Robust inference in semiparametric models**

Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Daniela Rodriguez

Semiparametric models contain both a parametric and a nonparametric component. In particular, generalized partially linear models (GPLM) are semiparametric versions of the generalized linear models where the response is to be predicted by covariates. They are useful when we suspect the linear model is insufficient to explain the relationship between the response variable and its associated covariates. To avoid the curse of dimensionality, GPLM allow most predictors to be modeled linearly while one or a small number of predictors enter the model nonparametrically. In particular, these models include partly linear regression models when the link function is the identity function. Using generalized profile likelihood, root-n consistent estimates for the parametric component can be obtained when the usual optimal rate for the smoothing parameter is used. However, as is well known, such estimators fail to deal with outlying observations and so does any test statistic used to make inferences on the regression parameter. Robust alternatives to some of the estimators and tests considered in the literature will be described. A general approach which includes a robust version of the deviance and a robustified quasi–likelihood is considered.

E202: **Multivariate statistical methods based on spatial signs and ranks**

Presenter: **Hannu Oja**, University of Tampere, Finland
Co-authors: Klaus Norhausen

Classical multivariate statistical inference methods (Hotelling's tests, multivariate analysis of variance, multivariate regression, etc.) are based on the use of an L2 criterion function. Here, alternative L1 criterion functions (with Euclidean norm) are used to extend the classical univariate sign and rank methods to the multivariate case. The first objective function, the mean deviation of the residuals, is the basis for the so called least absolute deviation (LAD) methods; it yields different median-type estimates and sign tests in the one-sample, two samples, several samples and finally general linear model settings (spatial sign methods). The second objective function is the mean difference of the residuals which generates Hodges-Lehmann type estimates and rank tests for different location problems (spatial rank methods). We briefly review the theory of the multivariate spatial sign and rank methods, tests and estimates, in the one sample, several samples and, finally, multivariate linear regression cases. A transformation-retransformation technique is used to obtain affine invariant tests and equivariant estimates. Different properties of the tests and estimates are discussed and compared. An R package entitled MNM will be available for implementing spatial sign and rank methods The theory is illustrated with several examples using the MNM package.

E219: **Plots for the detection of influential observations in dependent data**

Presenter: **Anne Ruiz-Gazen**, Universite Toulouse 1, France
Co-authors: Marc Genton

We introduce plots in order to detect influential observations in dependent data. The first plot is called a hair-plot and consists of all trajectories of the value of an estimator when each observation is modified in turn by an additive perturbation. A hair-plot is a version of the empirical influence function with replacement with a particular parameterization of the perturbation. From the hair-plot, we define two measures of influence: the local influence which describes the rate of departure from the original estimate due to a small perturbation of each observation; and the asymptotic influence which indicates the influence on the original estimate of the most extreme contamination for each observation. The second plot is called a disc-plot and consists in plotting discs with radii proportional to the rate of departure from the original estimate due to a small contamination at each observation. Open discs denote an increase of the value whereas closed discs denote a decrease. The cases of estimators defined as quadratic forms or ratios of quadratic forms are investigated in detail. Sample autocovariances, covariograms and variograms belong to the first case. Sample autocorrelations, correlograms, and indices of spatial autocorrelation belong to the second case. We illustrate our approach on various datasets from time series analysis and spatial statistics.

E121: **On consistency and robustness properties of support vector machines for heavy-tailed distributions**

Presenter: **Arnout Van Messem**, Vrije Universiteit Brussel, Belgium
Co-authors: Andreas Christmann

Consistency and robustness results of SVMs were recently derived for non-negative convex losses L of Nemitski type, if a weak moment condition for the joint distribution P is valid. However, this condition excludes heavy-tailed distributions such as the Cauchy distribution. We weaken this condition on P to only a condition on the marginal distribution P_X, such that the applicability of SVMs can be extended to heavy-tailed conditional distributions. To this purpose, we shift the loss downwards by some function independent of the estimator. More precisely, we define the shifted loss $L^\star(x,y,f(x)) := L(x,y,f(x)) - L(x,y,0)$. Obviously, this new loss can be negative. We first discuss the used $L^\star$-trick, give properties of the new loss $L^\star$, and show that the solution to the shifted problem exists and is unique. Furthermore, this solution coincides with the decision function of the

original formulation if the latter exists. We then give a representer theorem and results on risk-consistency and consistency of the solution. Finally we show that the solution of the shifted problem is robust in the sense of influence functions if a bounded kernel and Lipschitz continuous loss is used. This result holds for both Hampel's influence function and the Bouligand influence function.

E007: **On recent results for support vector machines**

Presenter: **Andreas Christmann**, University of Bayreuth, Germany

Modern learning theory is still a fast developing topic in statistics and support vector machines (SVMs) play an important role in this area. The original SVM approach was derived from the generalized portrait algorithm. However, general SVMs are nowadays usually considered in the context of regularized empirical risk minimization over reproducing kernel Hilbert spaces. Besides classical applications for classification and regression purposes, SVMs are also used for automatic fraud detection, web-mining, text-mining, and pattern recognition in images. The talk will first briefly summarize some facts on general SVMs obtained within the last years. Then some recent results on SVMs will be shown such as (i) a general representer theorem and consistency results both for heavy-tailed distributions and (ii) probabilistic bounds on the sparsity of SVMs that use the epsilon-insensitive loss. The sparsity of SVMs is important from an applied point of view, because sparsity decreases the computation time of SVMs both for the training data set and for the test data set. Finally, some results for dependent data will be given.

ES18 Room 9 BIOINFORMATICS Chair: Kostas Triantafyllopoulos

E054: **Sequence database search with compositionally biased amino acid sequences**

Presenter: **Vasilis Promponas**, University of Cyprus, Cyprus

Co-authors: Ioannis Kirmitzoglou

Statistical significance estimation by BLASTP relies heavily on the background frequencies of amino acid sequences, i.e. the amino acid composition of the database searching against. Query sequences containing segments of extreme amino acid compositions (compositionally biased segments - CBSs) clearly deviate from the aforementioned distribution, thus making the statistics employed for significance estimation unreliable. Diverse techniques have been proposed to identify CBSs in query sequences and filter (mask) them prior to database searches to avoid including false positive hits unrelated to the query under investigation. For this task, a multitude of algorithms based - among others - on information theory, dynamic programming, word statistics, regular expressions or suffix trees are combined with domain-based heuristics. In this contribution, we describe the statistical and biological motivation for masking amino acid sequences prior to database search, and the most popular methods proposed to tackle this problem. We present unpublished results, based on ROCn analysis of standard benchmark datasets, illustrating that the CAST algorithm (which incorporates information of the background and target distributions through the use of amino acid substitution matrices) in a novel two-way fashion may yield improved computational performance (quantified by CPU time or output file size) without sacrificing sensitivity.

E222: **A mixed effects model for differential expression analysis in longitudinal designs**

Presenter: **Giovanni Montana**, Imperial College London, UK

Co-authors: Maurice Berk

We propose a statistical approach for the detection of differentially expressed genes. Specifically, we envisage a longitudinal study design where multiple biological replicates are available and each individual is observed at more than one time point. We present a nonparametric mixed effects model based on smoothing splines and introduce a visual-distance metric for detecting time course profiles that differ across experimental conditions. Parameter estimation is carried out through the EM algorithm whereas model search is performed using a simplex optimization algorithm. Finally we present applications of the suggested model using both simulated and real data sets.

E136: **Clustering of time-course gene expression data using a Bayesian infinite mixture model based approach**

Presenter: **Daniela De Canditiis**, C.N.R., Italy

Co-authors: Claudia Angelini, Marianna Pensky

A novel Bayesian method for clustering longitudinal data obtained from time-course microarray experiments is presented. The statistical problem is formulated as follow: given N sampled curves representing the genes' expression profiles measured under noise, we want to identify the number of different groups hidden in the data and to assign the genes which have a similar 'true' expression profile to the same group. This problem can be recast as functional clustering where the additional challenge is to address some of the technical issues specifically due to the nature of the data, for example the time series are very short, the number of genes is high, there is a significant presence of replicated and/or missing data, the grid design are non-uniform and the tails of the noise are higher than Gaussian. In a regression context this problem is translated into the problem of assessing groups of genes whose coefficients are sampled from the same distribution and it is handled by the means of a Bayesian infinite

Dirichlet Process mixture model. An MCMC split and merge scheme is used for sampling from the posterior and the empirical MAP (Maximum Posterior) is selected as the final configuration of the data.

E189: **An alternative marginal likelihood estimator for phylogenetic models**
Presenter: **Serena Arima**, SAPIENZA University of Rome, Italy
Co-authors: Luca Tardella

Bayesian phylogenetic methods are generating noticeable enthusiasm in the field of molecular systematics. Several phylogenetic models are often at stake and different approaches are used to compare them within a Bayesian framework. One of the most widely used approaches is based on the so-called marginal likelihood of each model or, equivalently, on the Bayes factor for two competing models. However, the computation of the high dimensional integral corresponding to the marginal likelihood is a challenging problem. Several computational solutions have been proposed, none of which can be considered overall superior in terms of both computational complexity and precision. We show the successful implementation of the Inflated Density Ratio approach which is an alternative implementation of the Generalized Harmonic Mean (GHM) method. The new estimator shares the original simplicity of the GHM estimator, recycles MCMC simulations but, unlike it, overcomes the infinite variance issue. The proposed method is applied to some standard phylogenetic models and produces satisfactory results when compared with other commonly used approaches.

E205: **Estimation of Gaussian graphical model with partially known graph information**
Presenter: **Donghyeon Yu**, Seoul National University, Korea (ROK)
Co-authors: Kyunga Kim, Johan Lim, Joongho Won

A sparse graph structure among a set of variables is an important issue in many applications including genetics, psychiatrics, and finance. In Gaussian random variables, the non-zero entry in the concentration matrix represents the conditional independence between its corresponding two variables. A number of methods have been proposed to find better estimates of the Gaussian graphical model. We often know a significant amount of pre-information on the network from various sources. For example, gene regulatory networks and human protein interaction networks are available in the public databases that were derived from previous bioinformatics analyses. However, existing methods do not take into account this pre-information. In this paper, we consider the pre-information on the graph to improve the estimate of the Gaussian graph. We first numerically study the gain from the pre-information in estimating the graph structure. We consider three underlying models (an AR, a star, and a circle network). We estimate the models with 10% and 30% of the known connected edges. We compare the false positive (mistakenly connect) and false negative (mistakenly disconnect) probabilities of the estimates with/without the pre-information. We then apply the proposed method to estimating a gene regulatory network related to prostate cancer.

E030: **Stochastic models to educational and psychological measurements**
Presenter: **Panagiotis Matzioros**, Piraeus, Greece
Co-authors: John Goulionis

Distribution of human capital is a recent addition in the literature to the list of a few fundamental determinants of growth. The purpose of this work is to examine the chance of test misgrading using appropriate statistical models. The estimation of inadvertent grading errors can serve as a basis for quality control in educational and psychological measurements, where the test misgrading can be treated as a specific type of incident. A test is graded by a total number of graders and for any grader the threshold depends of the quality control requirement. Limitations of traditional Poisson models have been reviewed to highlight the need of introducing new models using well established negative binomial distributions (NBDS). Results of this investigation can be employed to ensure the number of misgrading events below a threshold. Finally we give bounds for the variance of these negative binomial estimators.

ES20 Room 4 SPARSENESS AND FUNCTIONAL DATA Chair: Wenceslao Gonzalez-Manteiga

E101: **Cluster number choice in functional data clustering**
Presenter: **Pai-Ling Li**, Tamkang University, Taiwan
Co-authors: Jeng-Min Chiou

Various approaches to cluster number selection in multivariate cluster analysis have been investigated. However, these multivariate data approaches to cluster number selection are not designed for functional data clustering. In this study, we focus on methods of cluster number selection in functional data analysis. We consider cluster number selection in conjunction with the subspace projected functional data clustering method proposed previously, the k-centers functional clustering. Here, clusters are defined via the projection of cluster subspaces, and each cluster subspace comprises the mean function and eigenfunctions of the covariance kernel. We propose a forward testing procedure to determine the number of clusters based upon functional hypotheses tests. The tests use bootstrap resampling schemes to construct reference distributions for comparing the means and modes of

variation differentials between cluster subspaces. The searching procedure starts with a small initial cluster number and then is increased in steps of one until no additional clusters can be identified. It aims at selecting the maximum number of distinct clusters while retaining significant difference between clusters. The performance of proposed approach is compared with other proposed cluster number selection criteria, extended from methods of multivariate data, by simulation studies and is illustrated through applications to clustering gene expression profile data.

E169: **Model selection and wavelet approximation in functional principal component regression with functional response**
Presenter: **Ana M. Aguilera**, University of Granada, Spain
Co-authors: Francisco A. Ocana, Mariano J. Valderrama

It usually happens in practice that we have two related functional variables whose observations are two sets of independent curves. An example is the level of lupus and the level of stress observed daily over a period of time. In this context, the functional lineal model provides the relationship between two functional variables and predicts a functional response from a functional predictor. Sample curves are observed at irregularly spaced time points that could be different among the sample individuals. Wavelet approximation and functional principal component analysis are considered to reconstruct the functional form of sample curves and to avoid the problem of multicollinearity affecting the model estimation. The functional model is then reduced to multivariate principal component regression. To select the best model we present a criterion that introduces pairs of principal components by taking into account not only the variability explained by pcs of the functional predictor but also their ability to forecast the pcs of the functional response. Different criteria as CV (leaving-one-out), BIC, Cp and MSE will be also considered to select the number of pairs. The good performance of the proposed methodology is tested on simulated and real data.

E226: **High dimensional regression analysis when regressors are heavily correlated**
Presenter: **Pascal Sarda**, University Paul Sabatier, France
Co-authors: Alois Kneip

We consider linear regression when the number of regressors p is possibly much larger than the sample size n. This case has been previously studied and usually the regressors are supposed to be almost uncorrelated whereas one assumes sparseness of the vector of coefficients. At the opposite, we envisage high correlations between the regressors. We derive in a first step an upper bound for the L2 error of prediction of the estimator of the coefficients obtained by means of principal component regression. Using ideas from Functional Data Analysis, we show that the bound depends on the rate of decreasing of the eigenvalues of the covariance matrix of the regressors. We then consider a general model where the vector of regressors can be decomposed in a sum of a first term of highly correlated variables and a second term of uncorrelated variables. We apply the Dantzig selector estimator after a decorrelation stage by means of principal component analysis. We show that the conditions are fulfilled and thus, the same rates of convergence hold.

E225: **How to select design points for prediction with functional data predictors**
Presenter: **Philippe Vieu**, Universite Paul Sabatier, France
Co-authors: Frederic Ferraty, Peter Hall

This talk is concerned with the regression problem when the predictor is functional. Such a problem appears in many fields of applied statistics (e.g. environmetrics, geophysics and economics). Chemometrical examples will be used to illustrate the proposed methodology. The first idea that will be presented consists in selecting a few design points (among the numerous discretized points where the functional predictor is observed), by means of a variable selection approach based on local linear smoothing. A few theoretical properties will be presented, and the good predictive power of this procedure will be shown in practice. In a second attempt, it will be shown how the previous design points selection method can be combined with a fully nonparametric functional procedure with important improvement in terms of predictive power. As a matter of conclusion, one could say that the high complexity of functional data makes boosting ideas very attractive. For instance the presented ideas suggest that in functional prediction one needs to capture very different effects (e.g. nonparametric effects, linear effects, multivariate effects and functional effects) and there is an high interest for combining various different prediction methods.

E206: **Some results on lasso logistic regression: application to gene expression data**
Presenter: **Manuel Garcia-Magarinos**, University of Santiago de Compostela, Spain
Co-authors: Anestis Antoniadis, Ricardo Cao, Wenceslao Gonzalez-Manteiga

Use of penalized regression methods is increasing in those fields where the number of covariates p under study is large in comparison with the sample size n ($p \gg n$ problem). Gene expression datasets generally comprise about a few dozens of samples for thousands of genes under study. Need for sparse models is imperative to achieve the aim of better understanding the genetics of disease, besides obtaining accurate classification. Here, we developed a lasso logistic regression (LLR) approach to obtain sparse, interpretable models in problems involving gene expression data with binary response. Specific penalizations are tried for each covariate (gene) and their asymptotic properties have been studied. Our approach is based on the generalized soft-threshold (GSoft) estimator and makes use of the Cyclic Coordinate Descent (CCD) algorithm. The choice of algorithm is

essential for overcoming optimization issues related with non-derivability problems of the objective function. Our classification results involve simulated and real data (leukemia and colon cancer datasets). Both are competitive, improving in some cases those presented in previous articles. An approximation of the covariance matrix for the estimated coefficients is also developed, giving rise to accurate estimates of the variance-covariance values.

ES25 Room 3 STATISTICS WITH FUZZY OR INCOMPLETE DATA Chair: Wolfgang Trutschnig

E093: **Induction of conditional Gaussian probabilistic decision graphs from incomplete data**

Presenter: **Jens Dalgaard Nielsen**, University of Castilla-La Mancha, Spain

Co-authors: Antonio Salmeron

Probabilistic decision graphs (PDGs) are probabilistic graphical models able to capture context specific independence sometimes in a finer way than Bayesian networks. It sometimes makes PDGs more efficient for computing probabilities than other graphical models. PDGs have been only defined in a discrete case, assuming a multinomial joint distribution over the variables in the model. We extend PDGs to incorporate continuous variables, by assuming a Conditional Gaussian (CG) joint distribution. The CG model can be factorised as a product of conditionals. The conditional distribution of each discrete variable is multinomial while for each continuous variable it is Gaussian. A limitation of this model is that the conditional distribution of a discrete variable cannot to depend on continuous variables. For instance, in the framework of Bayesian networks, it means that discrete variables cannot have continuous parents. However, the resulting models can be used in important situations, as in classification problems, where the response variable is discrete. Our proposed approach allows to use CG PDGs as classification models with discrete or continuous explanatory variables. We also propose an EM-based algorithm for parametric learning of CG PDGs from incomplete databases.

E092: **Early warning from car warranty data using fuzzy statistics**

Presenter: **Mark Last**, Ben-Gurion University of the Negev, Israel

Co-authors: Yael Mendelson, Sugato Chakrabarty, Karishma Batra

Car manufacturers are interested to detect evolving problems in a car fleet as early as possible so they can take preventive actions and deal with the problems before they become widespread. The vast amount of warranty claims recorded by the car dealers makes the manual process of analyzing this data hardly feasible. A fuzzy-based methodology for automated detection of evolving maintenance problems in massive streams of car warranty data is described. The empirical distributions of time-to-failure and mileage-to-failure are monitored over time using the advanced, fuzzy approach to comparison of frequency distributions, which automates the visual process of comparing the central tendency of two different histograms. Our fuzzy-based Early Warning Tool builds upon an automated interpretation of the differences between consecutive histogram plots using a cognitive model of human perception rather than crisp statistical models. We demonstrate the effectiveness and the efficiency of the proposed tool on warranty data that is very similar to the actual data gathered from a database within General Motors.

E124: **The estimation of missing data in presence of outliers: computational aspects**

Presenter: **Peter Filzmoser**, Vienna University of Technology, Austria

Co-authors: Heinrich Fritz, Karel Hron, Matthias Templ

Incomplete data cause problems for most statistical analyses. The natural way out is to estimate the missing information, and to perform the analysis for the completed data. We will present two examples of statistical analyses with incomplete and outlying data. The first example is concerned with the estimation of principal components in presence of missing values and possible outliers. In the second example we present a method that is able to robustly estimate missing information for compositional data, a type of data that frequently appears in environmental sciences and official statistics.

E045: **Iterative robust model-based Imputation of complex data**

Presenter: **Matthias Templ**, Vienna University of Technology, Austria

Co-authors: Alexander Kowarik, Peter Filzmoser

Most of the existing methods for the imputation of missing values assume that the data originate from a multivariate normal distribution. This assumption becomes invalid as soon as there are outliers in the data. In addition to that, almost all imputation methods are designed to impute data with continuous scaled variables, and only few methods (e.g. EM-based regression imputation by MIX or Iveware and nearest neighbor imputation) are available to impute data including different scaled variables, namely continuous, semi-continuous, ordinal and nominal variables. We show that our proposed iterative method based on robust estimation and stepwise regression outperforms these methods by a simulation study but also within applications to real-world data such as the European Survey of Income and Living Condition (EU-SILC).

E088: **SAFD - An R-package for statistical analysis of fuzzy data**
Presenter: **Wolfgang Trutschnig**, European Centre for Soft Computing, Spain
Co-authors: Asuncion Lubiano

An R-package called SAFD (Statistical Analysis of Fuzzy Data) was built in order to provide a basis for doing statistics with one-dimensional fuzzy sets having non-empty compact intervals as level sets. The package operates on polygonal fuzzy sets, i.e. fuzzy sets with piecewise linear membership function, having an arbitrary number of supporting points. In case a fuzzy set is not of polygonal type the package allows converting it to an approximating polygonal one having a user-defined number of supporting points. Given a sample of polygonal fuzzy sets SAFD provides in particular functions for calculating the Minkowski sum, the Minkowski mean and the variance of the sample, whereby the variance is calculated based on the well-known Bertoluzza distance. Since, especially from the inferential point of view, in many situations it is very important to be able to generate samples of fuzzy random variables the package also allows to simulate fuzzy random variables in a realistic, semi-parametric way. The functionality of the package will be explained and illustrated by various examples.

E089: **Approaches to hypothesis testing and regression estimation for fuzzy random variables using the R-package SAFD**
Presenter: **Asuncion Lubiano**, University of Oviedo, Spain
Co-authors: Wolfgang Trutschnig, Gil Gonzalez-Rodriguez

Functions for hypothesis testing about the mean of a fuzzy random variable (FRV) have been developed for the R-package SAFD (Statistical Analysis of Fuzzy Data). Specifically, a one-sample bootstrap test that allows testing the hypothesis that the mean of an FRV is a given fuzzy set, as well as multi-sample bootstrap tests for testing the equality of means of two or more FRVs has been implemented. In all cases the null hypothesis is expressed in terms of the Bertoluzza et al. metric, whose calculation is included as a separate function in the R-package. Moreover, a function to estimate a simple linear regression model for two FRVs has been added to SAFD. Thereby the fitting is done based on least-squares minimization with respect to the Bertoluzza et al. metric, by considering the constraints that allow to guarantee the existence of the residuals. The underlying theoretical approaches for hypothesis testing and for linear regression will be described, the corresponding implemented functions will be explained and their main properties will be illustrated by using both simulated and real-life data.

CS34 Room 8 COMPUTATIONAL ECONOMETRICS AND APPLICATIONS Chair: Tommaso Proietti

C121: **SUTSE models: non linear temporal disaggregation and the EM algorithm**
Presenter: **Filippo Moauro**, Eurostat, Luxembourg

Maximum likelihood estimation of unknown elements of a large system of seemingly unrelated time series equations, or SUTSE models, can be based on the EM algorithm. A convenient solution to the problem has been developed in the STAMP program. Here it is presented a further extension to models with time series subject to non-linear temporal aggregation constraints induced by the logarithmic transformation of observed data. The solution requires the definition of a non-linear state space form and of its linear Guassian approximation. The estimation is carried out matching the conditional mode of the states corresponding to the two state space forms and iterating on the Kalman filter and smoother estimating equations. Results of an extensive application based on data extracted from the Eurostat dataset are presented, with a particular focus on data expression of nominal values.

C206: **Comparing aggregate and disaggregate forecasts of contemporaneously aggregated vector MA processes**
Presenter: **Andrea Silvestrini**, Bank of Italy, Italy
Co-authors: Giacomo Sbrana

The paper compares the performance of aggregate and disaggregate predictors in forecasting contemporaneously aggregated vector moving average (MA) processes. The aggregate predictor is built by forecasting directly the process which results from contemporaneous aggregation of the original data generating process (DGP). The disaggregate predictor is obtained by aggregating univariate forecasts for the individual components of the original DGP. The necessary and sufficient condition for the equality of mean squared errors associated with the two predictors is derived in the case of a bivariate MA(1) process. Furthermore, it is shown that the equality of forecasting accuracy can be achieved using specific assumptions on the parameters of the vector MA process. Indeed, several sufficient conditions for the equality of mean squared errors are provided. In this sense, the structure and parameters of the DGP determine the relative efficiency of the two forecasting methods. A Monte Carlo experiment which takes into account potential problems of model misspecification and estimation uncertainty confirms the theoretical findings and sheds further light on the influence of the DGP's parameters on the efficiency of the predictors. Last, an empirical application dealing with the problem of forecasting aggregate Canadian inventory investments is presented to illustrate the main findings.

C038: **Alternative estimating and testing empirical strategies for fractional regression models**

Presenter: **Jose Murteira**, Universidade de Evora, Portugal
Co-authors: Esmeralda Ramalho

In many economic settings, the variable of interest is often a fraction or a proportion, being defined only on the unit interval. The bounded nature of such variables and, in some cases, the possibility of nontrivial probability mass accumulating at one or both boundaries raise some interesting estimation and inference issues. In this paper we survey the main alternative regression models and estimation methods that are available for dealing with fractional response variables and propose a full testing methodology to assess the validity of the assumptions required by each estimator. We consider both one and two-part models based on several alternative functional forms (logit, cauchit, probit, loglog, cloglog) and estimation methods (maximum likelihood or quasi-maximum likelihood). Regarding the testing methodology, we focus on tests for conditional mean assumptions, which may be also used for choosing between one-part and two-part models. In addition to the tests that are commonly employed in the literature of binary models (RESET and goodness-of-link tests), we suggest a new class of goodness-of-functional form tests and investigate the application of non-nested tests in this framework. Finally, we examine the finite sample properties of most of the estimators and tests discussed through an extensive Monte Carlo study. An application concerning corporate capital structure choices is also provided.

C273: **Monte Carlo simulation of discrete choice models involving large sums**

Presenter: **Zsolt Sandor**, University of Groningen, Netherlands

We study Monte Carlo simulations in a class of models involving integrals where the integrand function is the expected value of a discrete random variable with the number of possible values increasing exponentially in some number of choice units. Such integrals occur in several different classes of econometric problems related to discrete choice models. In these problems the integrals involved need to be estimated several times, since their estimation is part of larger algorithms. Therefore, it is crucial that the estimates of the integrals are sufficiently precise and computable within reasonable time. Our estimation algorithm is based on the feature of the integrals that they can be viewed as the expected value of a joint discrete-continuous random variable. We estimate the integrals by Monte Carlo simulations with respect to both the continuous and discrete part of the joint discrete-continuous random variable. We compare the simulation estimates from pseudo-random and quasi-random samples, where the latter are randomized orthogonal array-based Latin hypercube samples and randomized good lattice points. We find that the quasi-random simulations provide estimates that are more precise than pseudo-random simulations in all the cases we consider.

C340: **Detection of changes in parameters of linear regression models based on ratio type test statistics**

Presenter: **Barbora Madurkayova**, Charles University in Prague, Czech Republic

Procedures for detection of changes in parameters of linear regression models are considered. In particular test procedures based on ratio type test statistics that are functionals of partial sums of residuals are studied. Ratio type statistics are interesting for the fact that in order to compute such statistics there in no requirement to estimate the variance of the underlying model. Therefore they represent a suitable alternative for classical (non-ratio) statistics, most of all in cases when it is difficult to find a variance estimate with satisfactory properties. Previously, the properties of ratio type statistics for detection of changes in linear regression models were only described for the case of independent random errors. We study the possibility of extending these methods for the case of dependent random errors, including ARMA processes. This is an important question from the viewpoint of applications, since dependent observations are often the case when analyzing econometric and financial data. We assume to have data obtained in ordered time points and study the null hypothesis of no change against the alternative of a change occurring at some unknown time point. The performance of proposed methods is demonstrated on simulated data.

C199: **New EuroMInd: a monthly indicator of gross domestic product for the euro area and its member countries**

Presenter: **Rosa Ruggeri Cannata**, European Commission, Luxembourg
Co-authors: Cecilia Frale, Massimiliano Marcellino, Tommaso Proietti, Gian Luigi Mazzi

The methodology for the joint estimation of monthly indicators of real GDP at market prices for the Euro area and its four largest member states is illustrated. This is based on a same bottom up strategy which hinges on the separate disaggregation of the components of GDP by expenditure type and by sector. The model for each GDP component is formulated for the monthly growth rates of the series, which yields a nonlinear temporal aggregation constraint, and aims at disentangling from each series the dynamics that are common across the Euro area and that are specific to each individual countries.

CS48 Room 10 FORECASTING AND APPLIED ECONOMETRICS Chair: Christian Francq

C184: **Discerning the importance of new information**
Presenter: **Romulo Chumacero**, University of Chile / Central Bank of Chile, Chile

A framework for evaluating the importance of the arrival of new information for forecasting, estimation, and decision making is developed. By fusing known and recently developed statistical tests and concepts, the paper provides guidelines for detecting outliers, influential observations, innovations, and possible breaks in the end of the sample. The methodology is applied to analyze the Chilean CPI inflation.

C279: **Forecasting the behavior of the real exchange rate using long spans of data**
Presenter: **Efthymios Pavlidis**, Lancaster University, UK
Co-authors: Ivan Paya, David Peel

The nonlinear modeling and forecasting of the dollar sterling real exchange rate using long spans of data is considered. Our contribution to the literature is threefold. First, we provide significant evidence of smooth transition dynamics in the series by employing a battery of recently developed in-sample statistical tests. Second, we investigate through Monte Carlo simulation the small sample properties of several evaluation measures for comparing recursive forecasts when one of the competing models is nonlinear. Our results indicate that all tests exhibit low power in detecting the superiority of smooth transition over linear autoregressive models. Finally, notwithstanding the above, the nonlinear real exchange rate model outperforms both the random walk and the linear autoregressive model in forecasting the behaviour of the series during the post-Bretton Woods era.

C301: **Forecasting spot electricity prices through combination of forecasts**
Presenter: **Fany Nan**, University of Padova, Italy
Co-authors: Silvano Bordignon, Derek Bunn, Francesco Lisi

As a result of deregulation of most power markets around the world electricity price modelling and forecasting have obtained increasing importance in recent years. For the modelling of electricity prices one cannot rely on conventional models used to model stock prices, due to the distinct features of electricity markets. In this work, we consider different typologies of models to forecast one-day ahead electricity prices of some representative moments of the day of the UK Power Exchange market. Included are linear ARMAX models, different specifications of multiple regression models, non linear Markov switching regression models and time-varying parameter regression models. No model globally outperforms the others in forecasting accuracy: differences in forecasting accuracy depend on several factors, such as model specification, sample realisation and forecasting period. Since different forecasting models seem to capture different features of spot price dynamics, we propose a forecasting approach based on the combination of forecasts. Combination results illustrate the usefulness of the procedure, showing that for the dataset combined forecasts have the potential to produce forecasts of superior or equal accuracy relative to the individual forecasts. At the same time, this technique allows to avoid the choice of a single model.

C310: **Population drift and forecasting in the foreign exchange market**
Presenter: **Nicos Pavlidis**, Imperial College London, UK
Co-authors: Efthymios Pavlidis, Dimitris Tasoulis, Niall Adams, David Hand

The presence of long swings in foreign exchange markets, and the poor predictive ability of structural models of exchange rate determination, are well documented. Recent studies show that the forecasting performance of economic models is time dependent. Moreover, theoretical models were developed in which the link between fundamentals and asset prices is time-varying. These developments suggest that it is necessary to accommodate the possibility of changes in the underlying population distribution (population drift) when studying exchange rates. In discussing population drift it is useful to distinguish between abrupt and gradual change. Most work on abrupt change detection relies on the assumption that the parameters of the population distribution are known and constant before the change occurs. This assumption is unrealistic in foreign exchange forecasting. In contrast, adaptive filtering methods handle slowly changing environments by gradually removing the impact of past data on current parameter estimates. This is achieved by an appropriate definition of the criterion which is optimized to yield the model parameters. We investigate the existence and significance of population drift in foreign exchange rate series and propose methods that are capable of yielding accurate predictions, as well as providing insight about the structural changes that occur.

C210: **Proposal for market similarity evaluation method using stock board**
Presenter: **Hirokazu Nishioka**, Nagoya University, Japan
Co-authors: Fujio Toriumi, Kenichiro Ishii

Recently, artificial markets are attracting the attention of many researchers as a new approach to financial markets. By analyzing the simulation results in artificial markets, researchers can clarify market mechanisms or find new insights into financial markets. However, simulation results are not always grounded in reality. Adequate explanation is required so that the simulation results

can be trusted. Therefore, evaluating the similarity of artificial and actual markets is crucial to improve simulation reliability. In this study, we propose a new evaluation method to estimate similarity during observable market behavior data. Our proposed method focuses on the stock board instead of price data. The proposed method is applied to an evaluation experiment to confirm the accuracy of the evaluation method. We prepared three types of data sets in different time classes. Each data set was derived from market opening times, market closing times, and continuous sessions. The simulation results show that the identification rate between the opening times and the continuous sessions is 86% , and the identification rate between the closing times and the continuous sessions is 74% . The above results confirm the validity of the proposed method. With the proposed evaluation method, we can estimate whether the artificial markets can be trusted.

C208: **Asymmetry in price transmission mechanism between consumer and producer prices in European agricultural markets**

Presenter: **Eleni Zafeiriou**, Democritus University of Thrace, Greece
Co-authors: Theodoros Koutroumanidis, Spyridon Sofios

The present paper surveys the efficiency of the national agricultural markets in E.U. To be more specific we survey the cereals, the meat and the dairy market in Hungary, in U.K., in Greece and in Italy. The asymmetry in the price transmission mechanism was investigated between the producers and the consumers prices in the agricultural markets under preview. For the achievement of this objective the Johansen cointegration technique was applied, while at the same time the effect of the world price on the consumer and producer prices of each market was surveyed. In particular we applied the Granger causality test in order to survey whether the world price Granger causes the national producer or consumer prices. In all cases as expected the world price Granger causes the consumer prices but not the producer prices, given the pricing policy affected by the CAP regime. Furthermore, the asymmetry in the price transmission mechanism was surveyed with the application of the Error Correction Model Engle–Granger (ECM–EG model) as well as the LSE–Hendry Model. According to the results derived with the application of the methodology used, symmetry in the price transmission mechanism was confirmed as an exception and not the rule in the agricultural markets studied. Inefficiencies in the agricultural markets can be explained with the CAP regime under which they function as well as by specific conditions of each market studied.

CS49 Room 2 FINANCIAL MARKETS 1 Chair: Elena Kalotychou

C067: **Pricing of discretely monitored exotic options under NIG dynamics**

Presenter: **Arvid Naess**, Norwegian University of Science and Technology, Norway
Co-authors: Eivind Aukrust

In recent years it has been demonstrated that the Normal Inverse Gaussian (NIG) process can be used to model stock returns. The characteristics of the NIG market model compares favourably with empirical findings in the financial markets. Here it is briefly described how vanilla options can be priced efficiently, and how the corresponding option prices match with actual option surfaces. More importantly, it is demonstrated how discretely monitored exotic options can be calculated very fast and accurately in a simple manner under the NIG market model using the numerical path integration approach. Several numerical examples are presented to highlight accuracy and efficiency.

C242: **Modeling stock price returns and pricing a European option with Le Cam's statistical experiments**

Presenter: **Yannis Yatracos**, Cyprus University of Technology and National University of Singapore, Singapore

The embedding of the stock price return modeling problem in Le Cam's statistical experiments framework suggests strategies-probabilities, obtained from the traded stock prices in $[t_0, T]$, for the agent selling the stock's call option at time t_0 and for the buyer who may exercise it at time T. When the transaction times are dense in $[t_0, T]$ it is shown, with mild assumptions, that under each of these probabilities $\log(S_T/S_{t_0})$ has infinitely divisible distribution and in particular a normal distribution for calm stock; S_t is the stock's price at time t. The price of the stock's European option is the expected cost of the agent obtained using the distribution of $\log(S_T/S_{t_0})$. It is similar to the Black-Scholes-Merton $(B-S-M)$ price, and coincides with it for calm stock after translation. The strike price determined by the agent does not give arbitrage opportunity to the buyer. Additional results clarify volatility's role in the buyer's behavior. The results justify the extensive use of the B-S-M price and provide the possibility of a lower price by estimating market parameters.

C308: **Some methodological issues related to the estimation of systematic risk with reference to the Athens Stock Exchange**

Presenter: **Alexandros Milionis**, University of the Aegean, Greece

This work is concerned with the effect of the interval over which returns are measured (intervalling effect) on the estimation of systematic risk (beta coefficient) with reference to the Athens Stock Exchange (ASE) and discuss related methodological issues. It is found that the intervalling effect is strong and causes serious biases in the estimation of betas for short differencing intervals.

A two stage procedure leading to the estimation of the so-called asymptotic betas is used and the sensitivity of the results on the functional form used to express the convergence of betas to their asymptotic values, as well as on the method of estimation, is examined. By and large, the established methodology in the most developed capital markets of using monthly data and OLS estimates seems to provide reliable beta estimates for ASE. Further, using a simple algorithm stocks are categorized into seven classes according to the character of the variation of their beta with the differencing interval. Finally it is found that the estimated proxy for price adjustment delays (pads) depends on the value of asymptotic betas. This has implications when
the capitalization value is used in order to obtain the so-called inferred asymptotic betas.

C068: **A new solution to the purchasing power parity puzzles**
Presenter: **Michael Arghyrou**, Cardiff Business School, UK
Co-authors: Andros Gregoriou, Panayiotis Pourpourides

We argue that even in perfectly frictionless markets risk aversion driven by exchange rate uncertainty may cause a wedge between the domestic and foreign price of a totally homogeneous good. We test our hypothesis using a natural experiment based on a unique micro-data set from a market with minimum imperfections. The empirical findings validate our hypothesis, as accounting for exchange rate uncertainty we are able to explain a substantial proportion of deviations from the law of one price. Overall, our analysis suggests the possibility of a new solution to the purchasing power parity puzzles.

C239: **The clustering financial time series in applications for main market Stocks returns**
Presenter: **Anna Czapkiewicz**, AGH University of Science and Technology, Poland
Co-authors: Beata Basiura

The time series clustering is aimed at classifying the series under study into homogenous groups. The identification of similarities or dissimilarities in the financial time series has become an important research area in finance. The problem in cluster analysis of financial time series is the choice of a relevant metric. In this study some measures of distance are discussed. First, the measure based on the correlation ratio is considered on the correlation obtain from the Copula-Garch models. The marginal distribution are Garch(1,1) with innovations from t-Student and Hansen skewed t-distribution and skewed generalized error distribution (SGED) respectively. The several copula functions which have different characteristics in terms of tail dependency are: the Gaussian copula, the Student-t copula and the Archimedes Copula. The comparison with other distance measures, in particular the metric based on the disparity between the corresponding fitted autoregressive expansions and the Euclidean distance between the logarithms of the normalized periodograms, is provided The similarities and dissimilarities between countries financial markets are investigated using the discussed clustering methods. Data used in this study are daily stock markets returns for 42 of the major international stock markets.

C298: **Modelling intra-daily volatility by functional data analysis: an empirical application to the Spanish stock market**
Presenter: **Juan Romo**, Universidad Carlos III de Madrid, Spain
Co-authors: Esther Ruiz, Kenedy Alva

We propose recent functional data analysis techniques to study the intra-daily volatility. In particular, the volatility extraction is based on functional principal components and the volatility prediction on functional AR(1) models. The estimation of the corresponding parameters is carried out using the functional equivalent to OLS. We apply these ideas to the empirical analysis of the IBEX35 returns observed each five minutes. We also analyze the performance of the proposed functional AR(1) model to predict the volatility along a given day given the information in previous days for the intra-daily volatility for the firms in the IBEX35 Madrid stocks index.

Friday 30.10.2009 08:30-10:30 Parallel Session E

ES02 Room 7 OPTIMIZATION HEURISTICS IN ESTIMATION AND MODELLING Chair: Sandra Paterlini

E079: **Optimised U-type designs on flexible regions**
Presenter: **Christopher Sharpe**, University of Giessen, Germany
Co-authors: Dennis Lin, Peter Winker

The concept of a 'flexible region' describes an infinite variety of symmetrical shapes to enclose a particular 'region of interest' within a space. In experimental design, the properties of a function on the 'region of interest' is analysed based on a set of design points. The choice of design points can be made based on some discrepancy criterion. This paper investigates the generation of design points on a 'flexible region'. It uses a recently proposed new measure of discrepancy for this purpose, the Central Composite Discrepancy. The optimisation heuristic Threshold Accepting is used to generate low discrepancy U-type designs. The results for the two dimensional case indicate that using an optimisation heuristic in combination with an appropriate discrepancy measure, it is possible to produce high quality experimental designs on 'flexible regions'.

E091: **Sampling schemes for approximating integrals in the efficient design of stated choice experiments**
Presenter: **Peter Goos**, Universiteit Antwerpen, Belgium
Co-authors: Jie Yu, Martina Vandebroek

The semi-Bayesian approach for constructing efficient stated choice designs requires the evaluation of the design selection criterion value over numerous draws taken from the prior parameter distributions assumed in generating the design. The semi-Bayesian D-criterion value of a design is then calculated as the expected value of the D-error over all the draws taken. The traditional way to take draws from a distribution is to use the Pseudo-Monte Carlo approach. However, other sampling approaches are available as well. Examples are Quasi-Monte Carlo approaches using Halton sequences, Faure sequences, modified Latin hypercube sampling and extensible shifted lattice points, a Gauss-Hermite quadrature approach and a method using spherical-radial transformations. Not much is known in general about which sampling scheme is most efficient for calculating semi-Bayesian D-errors when constructing efficient stated choice designs. In this study, we compare the performance of these approaches under various scenarios and identify the most efficient sampling scheme for each situation.

E078: **Robust uniform design**
Presenter: **Peter Winker**, University of Giessen, Germany
Co-authors: Dennis Lin

Uniform design has became a standard tool in experimental design over the last two decades. Its properties are analyzed for a situation when the actual values of the control variables are subject to some error in factor level values. A closed form for the expected discrepancy is established, under some mild assumptions. A thorough Monte Carlo study is conducted under various potential scenarios. Some general properties have been revealed. Both theoretical and simulation results are consistent. The robustness of uniform design under error in control variables is also investigated. It is shown that uniform designs are highly robust with regard to uniformly distributed errors in design variables.

E163: **A comparison of design and analysis methods for supersaturated experiments**
Presenter: **David Woods**, University of Southampton, UK
Co-authors: Christopher Marley

Supersaturated experiments in which there are more factors than runs have been proposed for screening studies on industrial and scientific processes. A variety of design and model selection methods have been developed but little is available on their comparison and evaluation. In addition, a lack of real examples in the literature makes it difficult to assess the usefulness of these methods in practical experiments. We use simulated experiments to evaluate the use of $E(s^2)$-optimal and Bayesian D-optimal designs, and compare analysis strategies representing regression, shrinkage, and a model-averaging procedure. Suggestions are made for choosing the values of the tuning constants for each approach. The aim of our investigation is to attempt to provide guidance on which design and analysis methods should be employed, and when supersaturated designs may be successfully be used: how many runs should be used, how many factors can be accommodated, what number and magnitude of active effects can be detected? In addition, some comments are made on the performance of the design and analysis methods when the assumption of effect sparsity does not hold.

E196: **Regression trees for the visualization of results in optimization heuristic design**
Presenter: **Marco Chiarandini**, University of Southern Denmark, Denmark

Optimization heuristics for solving intractable numerical problems that arise in estimation and modelling are characterized by

a number of components and parameters whose proper configuration is a difficult task. Linear statistical models applied to the analysis of computational experiments of these algorithms offer the mathematical framework to separate the effects of components and parameters and to gain insight in a principled way. Yet, the necessary statistical models are not trivial. They entail nested designs with random factors leading to mixed effects linear models. We study recursive data partitioning procedures for these models. Multiple test procedures are used to determine whether the covariates have no significant effect on the response and the recursion needs to stop. We give account of both parametric and permutation based tests. The final outcome is a graphical visualization of results by means of decision trees that are of easy interpretation for the algorithm designer and a possible synthetic way for reporting results.

ES08 Room 3 FUZZY SETS IN REGRESSION AND CORRELATION PROBLEMS Chair: M. Angeles Gil

E129: **Sensitivity analysis in estimating linear regression between interval data**
Presenter: **Beatriz Sinova**, University of Oviedo, Spain
Co-authors: Ana Colubi, Maria Angeles Gil

A mean-squares approach to the problem of estimating an interval arithmetic-based regression in linear models between two random intervals has been previously developed. The mean-squared estimator of the regression coefficient depends on the choice of the distance between interval-valued data formalizing the error, and so the corresponding MSE also depends on such a choice. A preliminary simulation study is now considered. The goal is twofold, namely: firstly, to compare the MSEs associated with the estimator in several relevant situations for all possible choices within a versatile and convenient family of L2-type metrics (these situations concerning different variations of either the explicative random interval or the random error in the model); secondly, to determine the optimal distance for each of these situations. Some preliminary conclusions will be finally drawn.

E107: **Making regression imprecise for providing a better representation of precise data**
Presenter: **Mathieu Serrurier**, Universite Paul Sabatier, France
Co-authors: Henri Prade

Machine learning, and more specifically regression, usually focus on the search for a precise model, when precise data are available. Moreover, it is well-known that the model thus found may not exactly describe the target concept, due to the existence of learning bias. In order to overcome the problem of learning too much illusionary precise models, a so-called imprecise regression from non-fuzzy input and output data has been proposed recently by the authors mainly on an empirical basis. The goal of imprecise regression is to find a model that has the better tradeoff between faithfulness w.r.t. data and (meaningful) precision. Imprecise regression uses an optimization algorithm that produces linear or non-linear (kernel-based) fuzzy regression functions. These functions associate to a precise input vector a possibility distribution that is likely to restrict the output value. In this paper, we proposed a modified version of the initial approach, try to relate it to the representation of family of probabilities by means of possibility distributions. This approach is compared with classical and fuzzy regression frameworks. Experiments on an environmental database are performed and the interest of fuzzy predictions.

E108: **A formal and empirical analysis of the fuzzy Gamma rank correlation coefficient**
Presenter: **Maria Dolores Ruiz**, University of Granada, Spain
Co-authors: Eyke Hullermeier

A generalization of the well-known Gamma rank correlation measure has been recently proposed. The generalization is based on the use of fuzzy order relations and is mainly motivated by the objective to make the measure more robust toward noise in the data. The goal of this paper is threefold. First, we analyze some formal properties of the fuzzy rank correlation measure. Second, we complement the experiments on a simple artificial data set by a more extensive empirical evaluation using real-world data. Third, we offer an alternative motivation of the measure, based on the idea of equivalence relations induced by limited precision in the measurement of data.

E118: **Integration of different slopes for mids and spreads in a regression model for random intervals**
Presenter: **Angela Blanco**, University of Oviedo, Spain
Co-authors: Norberto Corral, Ana Colubi

In the basic linear model for random intervals based on the interval arithmetic, the same (up to sign) regression parameter for the linear relationship between mids (mid-points) and spreads (radius) of the intervals is transferred from the interval model. With the aim of improving the versatility, the formalization of a more flexible interval linear model is presented. It allows to consider different slopes for the linear regression models involving both pairs of variables. Some theoretical properties of the model are studied, and the estimation problem of its parameters is addressed. The estimation is carried out in terms of the least squares criterion based on a distance between intervals with good operational properties and an intuitive meaning. Some constraints are considered in the corresponding minimization problem in order to obtain solutions coherent with the linear model. Analytic

expressions for the regression estimators are shown, and operational expressions in terms of classic moments of the variables mid and spr of the intervals are obtained. These expressions provide a connection with the real simple linear regression theory. The empirical behaviour of the estimators is illustrated by means of some real-life examples and some simulation studies.

E180: **Kriging and epistemic uncertainty: discussion and developments**
Presenter: **Didier Dubois**, Universite Paul Sabatier, France
Co-authors: Kevin Loquin

Geostatistics is a branch of statistics dealing with spatial variability. Geostatistics in general, and the kriging methodology in particular, is based on the assumption that the uncertainty underlying a spatial phenomenon is perfectly known. More precisely, it is assumed that, under some well-chosen simplifying hypotheses of stationarity, the probabilistic model, i.e. the random function describing spatial uncertainty and spatial dependencies, can be completely assessed from the dataset by the experts. In that framework, kriging consists in estimating or predicting the spatial phenomenon at non sampled locations from this estimated random function. In the usual kriging approach, the data are precise and the assessment of the random function is mostly made at a glance by the experts (i.e. geostatisticians) from a thorough descriptive analysis of the dataset. However, it seems more realistic to assume that spatial data are tainted with imprecision due to measurement errors and that information is lacking to properly assess a unique random function model. Thus, it would be natural to handle epistemic uncertainty and imprecision appearing in both data specification and random function estimation steps of the kriging methodology. Epistemic uncertainty consists of some meta-knowledge about the lack of information on data precision or on the assessed model. The aim of this presentation is to discuss the cogency of the usual random function approach to model uncertainty in geostatistics, to propose a brief critical survey of the already existing attempts to introduce imprecision or epistemic uncertainty in geostatistics and to present preliminary results on current developments of new tractable methods that may handle uncertainty due to incomplete information about the variogram model and the available data in the kriging methodology.

ES22 Room 5 MODEL SELECTION AND VOLATILITY MODELS IN TIME SERIES Chair: Jean-Michel Zakoian

E013: **Merits and drawbacks of variance targeting in GARCH models**
Presenter: **Christian Francq**, University Lille 3, France
Co-authors: Lajos Horvath, Jean-Michel Zakoian

Variance targeting estimation is a technique used to alleviate the numerical difficulties encountered in the quasi-maximum likelihood (QML) estimation of GARCH models. It relies on a reparameterization of the model and a first-step estimation of the unconditional variance. The remaining parameters are estimated by QML in a second step. This paper establishes the asymptotic distribution of the estimators obtained by this method in univariate GARCH models. Comparisons with the standard QML are provided and the merits of the variance targeting method are discussed. In particular, it is shown that when the model is misspecified, the VTE can be superior to the QMLE for long-term prediction or Value-at-Risk calculation. An empirical application based on stock market indices is proposed.

E077: **Quasi maximum likelihood estimation and linear ARCH processes**
Presenter: **Lionel Truquet**, CREST/University Paris 1, France

Parametric estimation for short memory Linear ARCH processes is considered. These processes have been introduced in the long range dependence setting for their ability to model the leverage property of financial returns. The volatility of LARCH processes writes as a linear combination of past values and the conditional variance may be not bounded away from zero. Regarding the parametric estimation of the coefficients, arbitrary small values of the conditional variance are possible and this makes difficult to apply the usual Quasi Maximum Likelihood Estimation. Alternative methods have been studied by several autors. We investigate a smoothing version of the usual QMLE which is strongly consistent and asymptotically normal under certain assumptions. We study the behaviour of the asymptotic variance of the procedure as well as the asymptotic variance of the weighted conditional least squares method, an alternative method to the smoothed QMLE. Such a behaviour is linked to the existence of the pseudo Fisher information and an asymptotically efficient estimator may not exist for LARCH processes. The usual chi-square asymptotic behavior of an appropriated diagnostic Portmanteau test is also established.

E004: **Model selection and randomization for weakly dependent time series forecasting**
Presenter: **Pierre Alquier**, Universite Paris 7/ CREST, France
Co-authors: Olivier Wintenberger

Observing a stationary time series, we propose a new procedure in two steps for the prediction of the next value of the time series. The idea is to propose a prediction that does not depend too strongly on hypothesis on the true model of the series. Following a machine learning theory paradigm, the first step consists in determining randomized estimators, or experts, in (possibly numerous) different predictive models. It is not necessary that one of them is the true model. In the second step

estimators are obtained by model selection or randomization associated with exponential weights on these experts. We give oracle inequalities for both estimators that proves that our prediction is almost the best prediction among all the experts and provide some applications for linear and some non-linear predictors.

E005: **Consistent ranking of multivariate volatility models**
Presenter: **Francesco Violante**, FUNDP Namur, Belgium
Co-authors: Sebastien Laurent, Jeroen Rombouts

A large number of parameterizations have been proposed to model conditional variance dynamics in a multivariate framework. This paper examines the ranking of multivariate volatility models in terms of their ability to forecast out-of-sample conditional variance matrices. We investigate how sensitive the ranking is to alternative statistical loss functions which evaluate the distance between the true covariance matrix and its forecast. The evaluation of multivariate volatility models requires the use of a proxy for the unobservable volatility matrix which may shift the ranking of the models. Therefore, to preserve this ranking, conditions with respect to the choice of the loss function have to be discussed. To do this, we extend existing (univariate) conditions to the multivariate framework. By invoking norm equivalence we are able to extend the class of loss functions that preserve the true ranking. In a simulation study, we sample data from a continuous time multivariate diffusion process to illustrate the sensitivity of the ranking to different choices of the loss functions and to the quality of the proxy. An application to three foreign exchange rates, where we compare the forecasting performance of 16 multivariate GARCH specifications, is provided.

E116: **Goodness of fit test for interest rate models: an approach based on empirical process**
Presenter: **Abelardo Enrique Monsalve Cobis**, Universidad de Santiago de Compostela, Spain
Co-authors: Wenceslao Gonzalez-Manteiga, Manuel Febrero Bande

In this work a new test for the goodness of fit of one parametric form of the drift and volatility functions of the interest rate models is proposed. The test is based on a marked empirical process of the residuals. More specifically, the marked process empirical is built using one estimators of the integrated regression function in the case of the drift function and the integrated conditional variance function in case of the volatility function. Under one assumed model, the distribution of the processes can be approximated using bootstrap or the limit distribution function corresponding to a certain zero-mean Gaussian processes. The test is applied to simulated classical financial models and is illustrated in one empirical application to the EURIBOR data set.

ES27 Room 2 STATISTICAL SIGNAL EXTRACTION AND FILTERING 2 Chair: Tommaso Proietti

E014: **Robustness issues in Kalman filtering revisited**
Presenter: **Peter Ruckdeschel**, Fraunhofer ITWM, Germany

We present some optimality results for robust Kalman filtering. In a general state space model framework, we extend the ideal model setup allowing for both system-endogenous and -exogenous outliers, inducing somewhat conflicting goals of tracking and attenuation. We solve corresponding minimax MSE-problems for both types of outliers separately, resulting in corresponding saddle-points. Linearity of conditional means in the ideal model would simplify the situation drastically, so quantification of the deviation from linearity is important. We obtain a somewhat surprising characterization of linearity for the conditional expectation in this setting. As operational solution, insisting on recursivity, we obtain the rLS filter and variants of it specialized to endogenous, to exogenous outliers, and a hybrid version that can cope with both types –after a certain delay. Corresponding procedures are available in an R-package robKalman.

E062: **Towards an interval based deconvolution in signal processing**
Presenter: **Olivier Strauss**, LIRMM, France
Co-authors: Agnes Rico

A signal is usually perceived via a sensor device. When assuming the measurement process to be linear, it is possible to reconstruct the original signal from its observations by using a so-called deconvolution technique. Within digital signal processing, this reconstruction is usually base on inverting a linear equation by minimizing a regularized least square criterion. Such an inversion needs a perfect knowledge of the impulse response of the sensor involved in the signal measurement. The lower this knowledge, the more biased the reconstruction. In practice, identifying the impulse response is a difficult task that involves several arbitrary choices. When the knowledge on the impulse response is imprecise, then the deconvolution should be imprecise too. The deconvolution technique, we present in this paper, is based on a new modeling of the measurement process that extends the concept of convolution to account for an imprecise knowledge of the impulse response of the sensor. It consists in extending the least square inversion principle to this new modeling, via the Schultz iterative inversion scheme. The reconstructed signal is a real interval valued vector, whose imprecision reflects the error induced by the imprecise knowledge of the impulse response of the sensor.

E148: **Global, regional and country factors for the world economy: a dynamic factor approach**
Presenter: **Stefano Grassi**, Univeristy of Rome Tor Vergata, Italy
Co-authors: Borus Jungbacker, Siem Jan Koopman

Recently a new method for likelihood inference for large scale factor models has been proposed. In particular, it has been demonstrated that, when the number of series is larger than the number of factors the computational efficiency of Kalman Filter can significantly be improved by a simple computational device. This new device is based on the projection of the data on the reduced dimensional factor space. The methodology has been later extended to deal with missing data, by a suitable state space formulation. We apply the new methodology for estimating the world cycle and area specific cycles, plus country specific effects. The dataset concerns the GDP growth of a large number of countries. The individual time series feature a large percentage of missing data. The speed and reliability of the state space algorithms is crucial in this framework. The paper provides a systematic assessment of the estimation strategy and discusses the empirical evidence in the light of the previous literature.

E172: **A state space approach to productivity and efficiency measurement: the Italian economy, 1950-2003**
Presenter: **Camilla Mastromarco**, University of Salento, Italy
Co-authors: Ulrich Woitek

Based on the modification of the Kalman filter prediction equations, we estimate efficiency and total factor productivity for the postwar Italian economy in a Bayesian framework. The state space approach allows to estimate a stochastic frontier model from univariate time series. This is an advantage in a historical context, when data quality and comparability is an issue. Moreover, it is possible to model inefficiency as autoregressive process. The technical efficiency measure can be interpreted as output gap which allows for cyclical factor utilization. We find evidence for the well known productivity slowdown, as well as an increase in the output gap over the entire observation period.

E214: **Inference for hidden Markov diffusions with applications to rainfall-runoff models**
Presenter: **Franz Konecny**, BOKU - University of Natural Resources and Applied Life Sciences, Vienna, Austria

Stochastic dynamical systems are widely used to model systems across the sciences and engineering. Such models are convenient to formulate and can be analyzed mathematically and numerically. In most situations, however, the model cannot be fully specified beforehand, and some parameters need to be calibrated based on observed data. The rainfall–runoff model considered in this talk is a conceptual stochastic model, formulated in continuous–discrete state space form. We use a maximum likelihood approach based on an EM algorithm. In this approach, the key ingredient is the computation of smoothed additive functionals of hidden states. Sequential Monte Carlo methods (SMC), also called particle methods, are used for smoothing in the state space model.

ES28 Room 1 SMALL AREA ESTIMATION 1 Chair: Isabel Molina

E109: **A survey of mixed model extensions**
Presenter: **Ren Ohinata**, Georg-August-Universitaet Goettingen, Germany
Co-authors: Stefan Sperlich

While mixed effects models are widely available effective tools in small area estimation, complexity of real data structure and necessity of models that are specifically tailored for the objectives of data analysis require a variety of mixed model extensions. We provide a review of extensions of the classical Gaussian linear mixed models. Our focus is mainly set on giving an overview of a variety of major model extensions rather than of ongoing researches of their asymptotic properties. Being motivated by realistic assumptions that area–specific random effects correlation may underlie neighboring small areas and that heteroscedasticity may exist in variances of the within-area errors, we make a survey of the relaxation of the classical covariance structure for the error terms. We also present semiparametric extensions of the functional relationship between the response variable and covariates in the context of the linear as well as the generalized linear mixed models. Kernel regression and splines techniques are reviewed for semiparametric modeling of the predictors of regression. Our survey also includes parametric and nonparametric relaxation of the classical normality assumption for the random effects terms.

E126: **Resistance to outliers of M-quantile and robust random effect small area estimation models**
Presenter: **Monica Pratesi**, University of Pisa, Italy
Co-authors: Nikolaos Tzavidis, Caterina Giusti, Nicola Salvati

The issue of outlier robust small area estimation has been recently considered. In this work the Robust EBLUP (Empirical Best Linear Unbiased Predictor) is contrasted with M-quantile predictors of the small area mean. Starting from the basic M-quantile predictor, we also consider a bias-adjusted M-quantile predictor and a robust version of the M-quantile predictor. In particular, we report some results from simulation experiments. In the model based simulation study robustness is evaluated under four different scenarios, characterized by the presence of outliers deriving from a contamination at individual level and/or at area

level. The second set of simulations, the design based simulations, are performed using real data. In addition, the contrast is enriched by an application to real income data from the EU-SILC survey (European Survey on Income and Living Conditions).

E159: **Application of model with random regression coefficient to small area estimation**
Presenter: **Tomas Hobza**, Czech Technical University, Czech Republic
Co-authors: Domingo Morales

We introduce a nested error regression model having both fixed and random regression coefficients to estimate linear parameters of small areas. Two types of the mentioned model are supposed: the first one assumes there is no correlation between random regression coefficients and random intercept, the second one allows this correlation. Algorithms and formulas to fit the models, to calculate EBLUP and to estimate mean squared errors are described. Behavior of the two studied models is illustrated by a Monte Carlo simulation experiment. A motivating application to a real data set is also given.

E009: **Bayesian overdispersed models with smoothing splines for spatial age-specific data**
Presenter: **Giovani Silva**, Technical University of Lisbon, Portugal
Co-authors: Charmaine Dean

Disease rates by small areas have usually been represented through maps in order to identify spatial and temporal effects and risk factors of the diseases. Because of the need for producing reliable estimates for these rates, recent developments in disease mapping have had tremendous impact in health policy and epidemiology. The aim here is to develop methods for modelling and mapping of disease mortality and incidence rates by age-groups over public health regions. These methods account for (i) over-dispersion (ii) spatial correlation (iii) nonlinear age-specific effects, using hierarchical Bayesian spatial models with age-specific smoothing. In addition, we can use mixtures of Dirichlet process to relax the distributional assumption of the smoothing and age-specific trend coefficients. That process has been the most widely used nonparametric Bayesian model because one can easily obtain posterior estimates using standard Markov chain Monte Carlo methods. An analysis of spatial age-specific lung cancer rates in the province of Ontario motivates and illustrates the methods developed. That cancer is the leading cause of death due to cancer in Canada, representing an estimated 30% of the cancer deaths in males and 25% of the cancer deaths in females.

E042: **Some nearest neighbor methods for detection of disease clustering**
Presenter: **Elvan Ceyhan**, Koc University, Turkey

In recent years, there has been a growing interest on the use of spatial statistics for various purposes in epidemiology. In particular spatial analysis methods are used in modeling and detecting clusters of a certain disease in a population. In this work I discuss some of the methods based on nearest neighbor (NN) relationships for testing the significance of disease clustering. These methods include analysis of nearest neighbor contingency tables based on case versus control settings, Cuzick-Edwards' k-NN methods, their combined versions, and Tango's modification to them, and so on. The null hypothesis considered is the random labeling (RL) of cases and controls to an observed set of locations. I compare the empirical size performance of these tests under RL. Along this line, I also investigate the effect of location clustering (as in clustering of domestic dwellings in certain neighborhoods) and number of such clusters on these tests, and the confounding effects of relative abundance (i.e., the inevitable difference in the number of cases versus controls in the region of interest). Furthermore, I propose various non-random labeling alternatives, under which the power performance of these tests is investigated.

ES37 Room 9 COMPUTATIONAL STATISTICS 2 Chair: Cristian Gatu

E216: **MM-algorithms and MCMC methods in maximum likelihood estimation for Mallows-Bradley-Terry models**
Presenter: **Simplice Dossoue-Gbete**, Universite de Pau et des Pays de l'Adour, France
Co-authors: Amadou Sawadogo

The maximum likelihood estimation of the parameters of the Bradley-Terry parameters when dealing with the Mallows-Bradley-Terry models for ranking data analysis is considered. Rankings with ties are considered along with complete rankings. The maximum likelihood method is avoided until now because of the normalising constant that involves an intractable sum when the number of items to be ranked is large, e.g. 10 or more. Two MM-algorithms are proposed in order to avoid repetitive calculations of the Hessian matrix required by Newton-Raphson type of algorithms. The first algorithm takes advantage on the lower boundness of the Hessian matrix of the loglikelihood function; the second algorithm relies on convexity properties involved by the logarithm of the normalising constant of the probability distributions. Incorporating Markov Chain Monte Carlo scheme in the gradient step of the MM-algorithms proves to be as viable option for the calculation of the maximum likelihood estimates. The performance of the estimates for statistical inference are explored through a simulation study. The method is also valid for models where covariates are associated with the Bradley-Terry parameters through a link function.

E183: **A new method for the fast computation of high dimensional multivariate normal probabilities**
Presenter: **Ioannis Phinikettos**, Imperial College, UK
Co-authors: Axel Gandy

The efficient and fast computation of multivariate normal probabilities is of common interest. There are several methods that deals with these kind of probabilities. In the R environment there exist the packages mvtnorm and mnormt. None of these methods has been explicitly constructed to work with high dimensional normal distributions. High dimensional multivariate normal probabilities are needed for the computation of p-values of certain model checks for the detection of misspecified covariates in generalised linear models. A new method that exploits the singular value decomposition of the covariance matrix is introduced. This decomposition transforms the problem in a way that we have a priority ordering on the integration variables. Given this new form, we will apply various variance reduction techniques for the fast and efficient computation of the multivariate normal probabilities. These include conditional Monte Carlo, importance sampling and splitting. Also antithetic and control variates may be beneficial. We will compare our method with the existing ones.

E127: **Bayesian kernel projections for classification of high dimensional data**
Presenter: **Katarina Domijan**, National University of Ireland Maynooth, Ireland
Co-authors: Simon Wilson

A sparse Bayesian multi-category kernel classification method for high dimensional data is presented. The applications considered are microarray, image processing and near-infrared spectroscopy data, where each sample is defined by hundreds or thousands of concurrently obtained measurements. The proposed classifier is based on the reproducing kernel Hilbert spaces (RKHS) theory which allows for nonlinear generalization of linear classifiers by implicitly mapping the classification problem into a high dimensional feature space where the data is thought to be linearly separable. Due to the reproducing property of the RKHS, the classification is actually carried out in the subspace of the feature space which is of much smaller dimension. The proposed classifier is constructed so that it performs the classification of the projections of the data to the principal axes of the feature space. The advantage of this approach is that the regression coefficients are identifiable and sparse. In addition, large computational savings and improved classification performance were achieved for the high dimensional data sets considered. The Gibbs sampler is implemented to find the posterior distributions of the parameters, thus probability distributions of prediction can be obtained for new data points, which gives a more complete picture of classification.

E095: **A survey of robust methods under common principal components**
Presenter: **Isabel Maria Rodrigues**, Instituto Superior Tecnico-CEMAT, Portugal

In multivariate analysis, we often deal with situations involving several populations, where the assumption of equality of covariance matrices is usually assumed. Yet sometimes, this assumption is not adequate and models for common structure dispersion have been studied. One basic common structure assumes that the k covariance matrices have different eigenvalues but identical eigenvectors. This model became known as the Common Principal Components (CPC) model. One more restrictive structure dispersion, the proportionality (PROP) model, assumes that the scatter matrices are equal up to a proportionality constant. A unified study of the maximum likelihood (ML) estimators under a CPC and the PROP models has been given and likelihood ratio tests for a hierarchy of models for the scatter structure have been studied. Besides, under normality, the asymptotic properties of the quadratic discrimination procedure under hierarchical models for the scatter matrices have been investigated. However, the ML estimators and likelihood ratio tests assuming normality of the data, as well as discrimination rules can be seriously affected by anomalous observations. Here a survey of the existing theory and methods for robust inference under the CPC model is provided. Finally, we will briefly discuss an approach based on aggregate measures to detected outlying observations.

E067: **A new approach to determining estimability and connectivity in m-way designs**
Presenter: **Janet Godolphin**, University of Surrey, UK

The classical problem of ascertaining the connectivity status of an m-way design has received much attention, particularly in the cases where $m = 2$ and $m = 3$. In the general case, a new approach yields the connectivity status for the overall design and for each of the individual factors directly from the kernel space of the design matrix. Furthermore, the set of estimable parametric functions in each factor is derived from a segregated component of this kernel space. The kernel space approach enables a simple derivation of some classical results. Examples are given to illustrate the main results and to compare the information on connectivity status gained by the proposed approach to that gained using a recently proposed method based on Grobner bases.

ES43 Room 10 COMPUTATIONAL ECONOMETRICS AND FINANCIAL TIME SERIES Chair: Alessandra Amendola

E162: **Maximum likelihood and generalized method of moments for vector multiplicative error model**

Presenter: **Hao Ding**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Kai Pui Lam

Vector multiplicative error model (vMEM) handles multiple nonnegative variables, where the innovations are often specified under a copula dependence structure on marginal probability distributions. Model estimation techniques using Maximum Likelihood (ML) and Generalized Method of Moments (GMM) have recently been proposed for vMEM. Unlike the ML method which requires a specified copula (e.g., Gaussian, Student's t) over Gamma-distributed marginal probabilities, the GMM depends only on the first two moment conditions of the innovation terms. This paper compares the ML and GMM on a 2-dimensional vMEM(1,1) through a simulation study for different cases of model mismatch, e.g., mis-specified copula, fat tail, and outliers. As expected, it has been found that the ML is more accurate and efficient when there is no model mismatch. However, the GMM outperforms the ML when the copula structure is different from that specified in ML. A method of using empirical copula in ML is proposed and found to be more robust. For a fat tail phenomenon with Weibull distribution, the ML estimation is robust against the model mismatch, while the GMM seems to suffer from its inherent sensitivity on the moment conditions. By adding outliers to data as shocks for another test of model mismatch, the ML is found to be more robust than the GMM. It has also been found that good initial estimates are important for starting up the GMM and ML; and for our simulation the two-stage ML method is adopted.

E181: **Bayesian estimation of finite time ruin probabilities**

Presenter: **M. Concepcion Ausin**, Universidad Carlos III de Madrid, Spain
Co-authors: Rosa E. Lillo, Michael P. Wiper

We consider Bayesian inference and estimation of finite time ruin probabilities for the Sparre-Andersen risk model. The dense family of Coxian distributions is considered for the approximation of both the inter-claim time and claim size distributions. We illustrate that the Coxian model can be well fitted to real, long-tailed claims data and that this compares well with the generalized Pareto model. The main advantage of using the Coxian model for inter-claim times and claim sizes is that it is possible to compute finite time ruin probabilities making use of recent results from queueing theory. In practice, finite time ruin probabilities are much more useful than infinite time ruin probabilities as insurance companies are usually interested in predictions for short periods of future time and not just in the limit. We show how to obtain predictive distributions of these finite time ruin probabilities, which are more informative than simple point estimations and take account of model and parameter uncertainty. We illustrate the procedure with simulated data and the well-known Danish fire loss data set.

E065: **Using Benford's Law to identify tax-at-risk with the taxpayers**

Presenter: **Vishnu Lagoo**, Federal Government, Canada

This paper proposes to use the underlying principles of Benford's Law of Anomalous Numbers as a tool for the Canadian Government Auditors to identify and rank the tax recovery potential in the taxpayers. In Canada, the taxpayers assess themselves. It is then up to the Government Auditors to identify the self-assessed tax returns for mistake, fraud or misinterpretation of the tax laws etc. and to recover the unpaid tax. The Auditors do this by allocating a score to each tax return; the score being mostly proportional to the tax recovery potential. We test-scored the already audited tax returns. None of them was originally selected for auditing on Benford principles. We selected certain key variables like Investment Carrying Charges or Allowable Expenses, etc whose values were supplied both by the taxpayer and by the Auditors. We used both the values and Benford principles to score the returns. The correlation between the net tax recovery and the test-score was statistically significant, but the numerical value of the correlation was not high. It is most likely because only a few of the significant variables were used scoring. Another scoring system used the difference between the taxpayer supplied and the Auditor allocated values for scoring. Again, the correlation was positive, statistically significant, but numerically small. It turns out that this scoring system is what is currently in use, though not knowingly based on Benford principles.

E074: **Effective boundary correction in kernel density estimation and regression**

Presenter: **Jing Dai**, Georg-August-Universitat Gottingen, Germany
Co-authors: Stefan Sperlich

Boundary effects are a well known problem in nonparametric estimation, no matter if we think of density estimation or regression. Moreover, if the estimation has been performed on transformed covariates, this problem becomes boosted and may distort substantially the final estimates. In spite of their importance and the amount of theoretical studies, boundary correction methods are hardly used neither in density estimation nor regression. One reason is the lack of implementations in statistical software, another could be a disappointingly small improvement they offer. The practical existing methods can be divided into two groups: they either modify the kernels or the bandwidths at the boundaries. Certainly, some promote the combination of both or extend

them by bootstrap. However, practitioners are not willing to apply complex, seemingly non-intuitive methods. Therefore we introduce a quick and easy boundary correction applying local bandwidths. More specific, we fix the kernel but choose bandwidths such that all points become interior points; in the original boundary regions the windows start, respectively end, at the boundary. We give asymptotic insight, a comprehensive simulation study, a comparison with existing methods, and provide a R package. In simulations this method outperforms existing methods in the bias, and competes in the mean squared error.

E090: **A quasi-PORT methodology for VaR based on second-order reduced-bias estimation**
Presenter: **M. Ivette Gomes**, University of Lisbon, Portugal
Co-authors: Fernanda Figueiredo

We consider the estimation, under a semi-parametric framework, of a positive extreme value index, the primary parameter in Statistics of Extremes, and associated estimation of the Value at Risk (VaR) at a level p, the size of the loss occurred with fixed small probability p. The classical extreme value index estimator is then the Hill estimator, based on an intermediate number k of top order statistics. But the Hill estimator and the associated VaR estimator lead to sample paths with a high variance for small k, a high bias for large k, and a very sharp mean squared error pattern, as a function of k. Recently, several authors have been dealing with bias reduction in the field of extremes and simple classes of second-order minimum-variance reduced-bias (MVRB) extreme value index and VaR estimators have appeared in the literature. These VaR estimators do not enjoy the adequate linear property of the quantiles, contrarily to the recently introduced PORT-VaR estimators, with PORT standing for Peaks Over Random Thresholds. We shall here consider quasi-PORT MVRB VaR-estimators, for which the above mentioned linear property is obtained approximately. A large-scale Monte-Carlo simulation study of the proposed estimators' finite sample behaviour and applications in the fields of finance and insurance will be provided.

CS17 Room 8 TIME SERIES FINANCIAL ECONOMETRICS 1 Chair: Ana-Maria Fuertes

C083: **Method of moments estimation of GO-GARCH models**
Presenter: **Peter Boswijk**, University of Amsterdam, Netherlands
Co-authors: Roy van der Weide

We propose a new estimation method for the factor loading matrix in generalized orthogonal GARCH (GO-GARCH) models. The method is based on the eigenvectors of a suitably defined sample autocorrelation matrix of squares and cross-products of the process. The method can therefore be easily applied to high-dimensional systems, where likelihood-based estimation will run into computational problems. We provide conditions for consistency of the estimator, and study its efficiency relative to maximum likelihood estimation using Monte Carlo simulations. The method is applied to European sector returns, and to the correlation between oil and kerosene returns and airline stock returns.

C103: **Estimating VAR-MGARCH models in multiple steps**
Presenter: **Hakan Eratalay**, University of Alicante, Spain
Co-authors: M. Angeles Carnero

This paper analyzes the performance of multiple steps estimators of Vector Autoregressive Multivariate Conditional Correlation GARCH models by means of Monte Carlo experiments. High dimensionality is one of the main problems in the estimation of the parameters of these models and therefore maximizing the likelihood function of the full model leads commonly to numerical problems. We show that if innovations are Gaussian, estimating the parameters in several steps is a reasonable alternative to the full maximization of the likelihood. Our results show that for the sample sizes, usually encountered in financial econometrics, the differences between the volatility estimates obtained with the more efficient estimator and the multiple steps estimators are negligible. However, this does not seem to be the case if the distribution is a Student-t for some Conditional Correlation GARCH models. Then, multiple steps estimators lead to volatility estimates which could be far from the ones implied by the true parameters. The results obtained in this paper are used for estimating volatility interactions in 10 European stock markets.

C255: **Robust Bartlett adjustment for hypotheses testing on cointegrating vectors: a bootstrap approach**
Presenter: **Alessandra Canepa**, Brunel University, UK

Existing correction factors of Bartlett-type for the LR test of linear restrictions on cointegrated vectors rest on the Gaussian assumption for the innovation terms. However, the distribution of most data relating to financial variables are fat tailed and often skewed. So there is a need to examine small sample inference procedures that require weaker assumptions for the innovation term. This paper shows that using a non-parametric bootstrap to approximate a Bartlett-type correction provides a statistic that does not require specifications of the innovation distribution. This bootstrap statistic can be used by applied econometricians to perform a small sample inference procedure that is less computationally demanding than estimating the p-value of the observed statistic.

C269: **Goodness of fit test for discretely observed diffusion processes**
Presenter: **Ilia Negri**, University of Bergamo, Italy
Co-authors: Yoichi Nishiyama

Diffusion processes are widely used to model financial variables such as asset prices, interest or exchange rates. We propose a nonparametric goodness-of-fit test for the drift coefficient of a one-dimensional diffusion process, where the diffusion coefficient is a nuisance parameter which is estimated. We study the properties of the test statistics under two different sampling schemes. The first scheme is based on time discretization, namely the high frequency data sampling. The second one is based on data observed discretely in space, that is, the so-called tick time sampled data. We prove that in both sampling schemes the limit distribution of the test is the supremum of the standard Brownian motion, and thus the test is asymptotically distribution free. Moreover, the tests are consistent for any fixed alternative. We illustrate the performance of the proposed test statistics applying them to some simulated trajectories of solutions of stochastic differential equations, both when the asymptotic assumption can be considered reached and also in the case when asymptotic conditions are not reached.

C350: **Methods for modeling nonlinear time series with long memory: theory and financial applications**
Presenter: **Richard Baillie**, Michigan State University, United States Of America

Many economic and financial time series exhibit both nonlinearity and long memory. This paper considers various alternative models for the specification and estimation of various forms of nonlinearity, including specific smooth transition autoregressive regime specifications (e.g. ESTAR and LSTAR), (ii) generic approximations based on Artificial Neural Nets and (iii) approximations based on Flexible Fourier Forms. In each case the nonlinear model is appended with a strongly dependent, long memory process. We show that time domain QMLE is appropriate for many of these models and that the simultaneous estimation of the long memory, linear and nonlinear parameters have desirable asymptotic properties. The BIC information criteria is shown to provide consistent model selection procedures. A particularly useful aspect of these results is a time domain semi parametric estimate of the long memory parameter in the presence of generic long memory where the parametric form of nonlinearity is not specified. We show the estimator is consistent and present some simulation evidence which is generally favourable to the method. We also consider the use of some initial semi-parametric estimates of the long memory parameter and subsequent estimation of the short run parameters. In general these procedures are not as efficient as the QMLE approach. We illustrate the methodologies with detailed application to real exchange rates and examine the implied half lives of the return to parity, and also to G7 monthly inflation rates.

CS45 Room 4 QUANTITATIVE RISK MANAGEMENT 3 Chair: Marc Paolella

C035: **Localized realized volatility modelling**
Presenter: **Uta Pigorsch**, University of Mannheim, Germany
Co-authors: Wolfgang Haerdle, Chen Ying

With the recent availability of high-frequency financial data the long range dependence of volatility regained researchers' interest and has lead to the consideration of long memory models for realized volatility. The long range diagnosis of volatility, however, is usually stated for long sample periods, while for small sample sizes, such as e.g. one year, the volatility dynamics appears to be better described by short-memory processes. The ensemble of these seemingly contradictory phenomena point towards short memory models of volatility with nonstationarities, such as structural breaks or regime switches, that spuriously generate a long memory pattern. In this paper we adopt this view on the dependence structure of volatility and propose a localized procedure for modeling realized volatility. That is at each point in time we determine a past interval over which volatility is approximated by a local linear process. Using SP500 data we find that our local approach outperforms long memory type models in terms of predictability.

C085: **A multivariate generalization of the Ornstein-Uhlenbeck stochastic volatility model**
Presenter: **Christian Pigorsch**, University of Bonn, Germany
Co-authors: Robert Stelzer

Using positive semidefinite processes of Ornstein-Uhlenbeck type a multivariate Ornstein-Uhlenbeck (OU) type stochastic volatility model is introduced. We derive many important statistical and probabilistic properties, e.g. the complete second order structure and a state-space representation. Noteworthy, many of our results are shown to be valid for the more general class of multivariate stochastic volatility models, which are driven by a stationary and square-integrable covariance matrix process. For the OU type stochastic volatility our results enable estimation and filtering of the volatility which we finally demonstrate with a short empirical illustration of our model.

C049: **Estimating the degrees of freedom of the realized volatility Wishart autoregressive model**
Presenter: **Matteo Bonato**, Swiss Finance Institute and University of Zurich, Switzerland

An in-depth analysis of the estimation of the realized volatility Wishart Autoregressive model is presented. We focus in particular on the estimation of the degrees of freedom. A new estimator is proposed. Monte Carlo simulations show that this novel estimator is more efficient and more robust to the presence of outliers when compared to the standard estimator proposed in literature. To conclude, an empirical application to the SP 500 - NASDAQ 100 futures realized variance-covariance series is presented. It confirms that the estimated degrees of freedom, first, result sensitively lower when extremely high values in the volatility process are present and secondly, they increase with the sampling frequency.

C299: **Strange facts about the marginal distributions of processes based on the Ornstein-Uhlenbeck process**
Presenter: **Ray Brownrigg**, Victoria University of Wellington, New Zealand
Co-authors: Estate Khmaladze

The Ornstein-Uhlenbeck process is particularly useful for modeling stochastic processes in financial applications. Further, functions of such a process can be used to model random volatility of other processes, resulting in more flexible models for financial risk variables. The distribution of such a financial risk variable is of particular interest in Value at Risk analysis. As we know, the far quantiles of the distribution function provide information on the level of capital reserves required to accommodate extreme stress situations. This paper presents an approximation for the distribution function which in some situations works surprisingly well for even the far tails of the distribution. While theoretically unjustified and strange, it may still be very useful in practice.

C368: **Recursive-design wild bootstrap trace test**
Presenter: **Silika Prohl**, University of Zurich, Switzerland

We show in numerous Monte Carlo experiments that the asymptotic trace test has moderate finite sample performance in presence of the heteroscedastic innovations. The recursive-design wild bootstrap solution to the inference problem is shown to have favorable properties, reducing the empirical size problems associated with the standard asymptotic trace test. When applying this methodology, we provide robust evidence on linkages between the credit default swap (CDS) spreads and credit spreads on the same reference company. Finally we show how this methodology can be extended to compute the Value-at-Risk of CDS positions.

CS56 Room 6 COPULA METHODS IN TIME SERIES ANALYSIS 1 Chair: Alessandra Luati

C070: **A multivariate version of Hoeffding's Phi-Square**
Presenter: **Martin Ruppert**, Universtiy of Cologne, Germany
Co-authors: Sandra Gaisser, Friedrich Schmid

We introduce a measure of association that extends the bivariate measure of Hoeffding to the multivariate case. Phi-Square is building on a Cramer-von Mises functional of the copula associated to a random vector, exacting the squared L2-norm of the difference between the copula and the independence copula. The measure is normalized to [0,1] and its properties are established. We conduct a two step-analysis to measure multivariate association of financial time series: First, we fit univariate ARMA-GARCH models and filter data. Second, we estimate multivariate Phi-Square based on the filtered data. Statistical inference for Phi-Square is based on the empirical copula process. We derive a nonparametric estimator for Phi-Square and its asymptotic distribution. In case of the independence copula, convergence of the Phi-Square estimator with non-normal asymptotic distribution is established. In cases other than the independence copula, asymptotic normality of the estimator is proven using the functional delta method. The asymptotic variance is estimated using consistent bootstrap methods. Tests for Phi-Square following from the results on its asymptotic behavior are provided. We illustrate our methodology within an event-study of multivariate financial contagion prior to and after the bankruptcy of Lehman Brothers.

C072: **On the robustness of goodness-of-fit tests for copulas**
Presenter: **Gregor Weiss**, Ruhr-University Bochum, Germany

The modelling of the dependence structure inherent in a random vector by the use of copulas is usually complemented by a test of the goodness-of-fit of the chosen parametric copula. Although two recent simulation studies examine the performance of copula GoF-tests, both of these studies only consider uncontaminated data simulated directly from a prespecified copula. This paper provides a comprehensive simulation study on the robustness of goodness-of-fit tests based on the empirical copula process, Rosenblatt's and Kendall's transform for bivariate elliptical and archimedean copulas. To assess the tests' robustness, I consider perturbations and outliers both in the dependence structure and the observations from the joint distribution. The Monte Carlo simulations show that depending on the underlying true copula, the GoF-test and chosen test statistic, even minor contaminations of the data can lead to a significant decrease in the GoF-test's power. Through the computation of the Value-at-Risk of a portfolio, I illustrate the practical relevance of the results from the simulations.

C294: **Dynamic asymmetric tail dependence: evidence on the German stock market**

Presenter: **Evdoxia Pliota**, University of Essex, UK
Co-authors: Wing Lon Ng

A mixed copula approach that accounts for an asymmetric dependence structure is proposed. The upper and lower tail dependence that cause extreme co-movements and cannot be captured by Gaussian models can now be separately investigated. Studies on financial contagion show that cross market dependence varies over time. Rather than applying multivariate GARCH models, which are restricted to satisfying certain constraints, a mixed copula approach is proposed. The methodology is employed on a rolling window to reveal time varying dependence patterns and especially extreme asymmetric co-movements. We focus on the sensitivity of stock prices' joint behavior with respect to the market's performance and the market's expected future volatility, as their multivariate relationships will heavily affect the joint distribution of returns and hence the investor's portfolio development and its optimization and management. Data from the German stock market have been analyzed, including the stock index DAX, the two volatility indices VDAX New and VDAX Old, and five selected equities. Our results show that stocks from different economic sectors will react differently to the market's trend and its stability as implied by the volatility index - not only with regard to direction (bull/bear) and magnitude (possibly tail dependence), but also in terms of time (speed of contagion).

C320: **Applying dynamic copulas to modeling interdependencies in global financial market during financial crises**

Presenter: **Ryszard Doman**, Adam Mickiewicz University, Poland

In this paper we model the dynamics of linkages between main national stock markets during the periods that include financial crises. The stock markets are represented by the returns on stock indices. Our interests focus on similarities and differences between the patterns of changes in the conditional dependence structure in the case of the crisis that started in 1997 from Asian emerging markets, and the crisis of 2007-2009 that began in the USA. We combine Markov regime switching models with dynamic copulas. Our approach uses parametric copula models with copulas belonging to different families. We allow switching of the copula family according to some homogeneous Markov chain as well as time evolution of the copula parameters that takes into account the dynamics of the corresponding univariate volatilities and location of their levels with reference to some thresholds. The models enable us to distinguish regimes with different extreme dependence structure that can be of asymmetric type. We are thus able to analyze the linkages between the investigated national stock markets focusing on a comparison of the strength and pattern of the conditional tail dependencies between the indices returns during the periods that include the two investigated crises.

C357: **Estimating LGD correlation**

Presenter: **Jiri Witzany**, University of Economics, Czech Republic

A new method to estimate correlation of account level Basle II Loss Given Default (LGD) is proposed. The correlation determines the probability distribution of portfolio level LGD in the context of a copula model which is used to stress the LGD parameter as well as to estimate the LGD discount rate and other parameters. Given historical LGD observations we apply the maximum likelihood method to find the best estimated correlation parameter. The method is applied and analyzed on a real large data set of unsecured retail account level LGDs and the corresponding monthly series of the average LGDs. The correlation estimate comes relatively close to the PD regulatory correlation. It is also tested for stability using the bootstrapping method and used in an efficient formula to estimate ex ante one-year stressed LGD, i.e. one-year LGD quantiles on any reasonable probability level.

Friday 30.10.2009 11:50-13:00 Parallel Session G

ES07 Room 4 SMALL AREA ESTIMATION 2 Chair: Domingo Morales

E094: **Small area estimation using P-spline models**
Presenter: **Ana F. Militino**, Public University of Navarre, Spain
Co-authors: Maria Dolores Ugarte, Tomas Goicoa

During the last few years there has been a quick expansion of model-based small area estimation methods motivated by the increasing demand of precise information for small domains. Much of the research has been conducted on the use of mixed effects models to derive empirical best linear unbiased predictors (EBLUP) of the small area means or totals. Very recently, non-parametric models have been included in the small area estimation toolkit because they can prevent from bias when the functional form of the relationship between the response and the covariates is unknown. In this paper, a small area estimator is built based on a penalized spline model that approximates a non-linear but smooth relationship between a response and a given covariate. The procedure is illustrated with data from the 2006 Spanish Household Budget Survey (SHBS).

E130: **Goodness-of-fit tests for small area estimation models**
Presenter: **Gerda Claeskens**, K.U. Leuven, Belgium

In many surveys the interest lies in providing estimates for small domains within the population of interest. Such 'small area estimation' often proceeds by constructing a linear (fixed effect) model for the mean and a random effect for the small areas. This random effect is assumed independent from the error or noise variable. Mixed models, containing both fixed and random effects, are often estimated on the assumption that the random effects are normally distributed. We construct several formal tests of the hypothesis that the small area random effects and or the errors are normally distributed. The tests are nonparametric in the sense that they are designed to detect a wide range of alternatives to normality. Moreover, in case of rejection they provide an estimator of the alternative distribution.

E182: **Robust variance components in the nested-error model**
Presenter: **Isabel Molina**, Universidad Carlos III de Madrid, Spain
Co-authors: Betsabe Perez, Daniel Pena

Henderson method III for estimating variance components in a nested-error model is based on calculating expectations of certain sums of squares and applying a method of moments. This method does not need a specification of a parametric distribution for the response. Several estimators of variance components are developed by robustification of Henderson method III. These new estimators are expected to be robust against groups of outliers that might mask each other. Simulation results show a large gain in efficiency of these estimators under the presence of mean shift outliers along with small loss under normal distribution without outliers. The resulting estimators are used to improve the definition of basic diagnostic measures for a nested-error model, which will be useful for the detection of general model departures.

ES12 Room 6 ROBUST FUNCTIONAL DATA ANALYSIS Chair: Matias Salibian-Barrera

E161: **Robust functional principal components analysis for skewed distributions and its application to outlier detection**
Presenter: **Liangliang Wang**, University of British Columbia, Canada
Co-authors: Nancy Heckman, Matias Salibian-Barrera

The main motivation is an outlier detection problem for the radiosonde data which were collected by balloons that carried measuring instruments and were launched from the surface into the atmosphere and the stratosphere. The basic data structure is functional data: temperature as a function of pressure. The primary statistical problem is to determine objective ways of detecting unusual observations. To detect atypical observations we use the standardized difference between observed and fitted curves, obtained with functional PCA through conditional expectation (PACE). Unfortunately PACE is not robust and the estimates from it are significantly affected by outliers. We use a straightforward idea to robustify PACE by obtaining robust estimates of the mean and covariance functions. Estimating these functions robustly is challenging because skewness and outliers are present at the same time. Simulation studies have shown that the robust PACE performs similarly to the original PACE when there are no outliers, and that the robust PACE outperforms the original PACE in terms of MSE in the presence of outliers. When applied to the radiosonde data, the robust PACE can identify more reasonable outliers than the original PACE.

E211: **Ordering sparse functional data**
Presenter: **Ying Wei**, Columbia University, United States Of America
Co-authors: Sara Lopez-Pintado

Functional data analysis is a rising area of statistics. The statistical analysis on functional data (curves) can be significantly improved using robust estimators. New ideas of depth for functional data have been studied recently. They provide a way of ordering curves from center–outward, and L–statistics can be defined in a functional context. These methods are designed for trajectories that are observed on a fine grid of equally spaced time points. However, in many applications the trajectories are observed on sparse irregularly spaced time points. Hence, some preliminary smoothing step needs to be applied to obtain a curve. When the number of observations for individual paths is small the smoothing methods do not perform well. We extend the concepts of band depth and modified band depth to sparse functional data. Furthermore we propose a model–based consistent procedure for estimating the depths when the curves are observed on sparse and unevenly spaced points. The methods are based on a two step approach. First, we estimate the conditional distribution of the underlying curve path given the observed measurements. Second, we calculate the depths of the curves through conditional expectations. We apply this procedure to analyze the rates of children growth.

E113: **Robust methods for functional principal components**
Presenter: **Juan Lucas Bali**, Universidad de Buenos Aires and ANPCYT, Argentina

When dealing with multivariate data, like classical PCA, robust PCA searches for directions with maximal dispersion of the data projected on it. Instead of using the variance as a measure of dispersion, a robust scale estimator can be used in the maximization problem in order to get robust estimators. The aim is to adapt the projection pursuit approach to the functional setting and to study the asymptotic behavior of the proposals. Sometimes instead of raw robust functional principal component estimators, smoothed ones can be of interest. We will discuss three approaches to obtain smoothed estimators. Two of them are based on penalizing either the norm or the robust scale function. The third one is related to B-splines and sieve estimation. Through a simulation study, the performance of the different proposals with the classical ones is compared under a Gaussian distribution and different contamination schemes.

ES35 Room 3 UNCERTAINTY MODELLING FOR DATA ANALYSIS AND DATA MINING Chair: Jonathan Lawry

E187: **Summarizing time series with probabilistic fuzzy quantifiers**
Presenter: **Felix Diaz**, University of Santiago de Compostela, Spain
Co-authors: Alberto Bugarin

Although summarization with fuzzy quantifiers is a topic widely discussed in the literature, most of the approaches do not use plausible models of fuzzy quantification. Three main issues in fuzzy quantification can be identified: interpretation, reasoning and summarization. Interpretation is the main basic goal, as the other issues depend on a correct modeling of fuzzy quantifiers, i.e., no plausible behavior can be expected for summarization when using non-plausible models for fuzzy quantifiers. As the modeling problem is previous to the summarization problem, an adequate theory of summarization with fuzzy quantifiers should focus its attention into plausible models. In this work, we show that the probabilistic fuzzy quantification models fulfill a number of properties that make them appropriate for the summarization problem. We focus our investigation in the use of the FA model for data summarization. This model has a very solid theoretical behavior, since it is a (non-standard) Determiner Fuzzification Scheme. Based on the properties of the probabilistic models we propose a summarization algorithm based on linguistic and data mining criteria. Examples of use are provided in the domain of meteorology using daily temperature data sets, obtaining a plausible behavior of the summaries compatible with human expectations.

E186: **Linguistic prototypes for data description and classification**
Presenter: **Ines Gonzalez-Rodriguez**, University of Cantabria, Spain
Co-authors: Jonathan Lawry

We propose to use label semantics theory and clustering techniques to provide prototypical descriptions of a dataset. These prototypes can be used to explore data via linguistic queries and to perform classification using conditional probabilities. The label description of a dataset is a vector of mass assignments on sets of labels, the means across all data instances of their value description in terms of the given labels. For such description, we can obtain a joint mass assignment, later used to evaluate linguistic queries. Computing this joint mass assignment presupposes an independence assumption that may not be appropriate in some cases. Thus we partition the data into a number of disjoint sets or clusters where the elements in each cluster are assumed to be sufficiently similar to allow an independence assumption. We cluster directly on the space of multidimensional label descriptions, so each cluster means constitutes a label prototype, representing an amalgam of similar points. Similarity between label descriptions is measured based on comparing the different features using t-norms. To form the clusters, we adapt the c-means algorithm to label descriptions. In classification problems, prototypes are obtained for each of class and used to classify unseen instances.

E207: **Semantic cells: a random set and prototype theory interpretation of linguistic labels in rule-based systems**
Presenter: **Jonathan Lawry**, University of Bristol, UK
Co-authors: Yonchuan Tang

A semantic cell model for linguistic labels is proposed where semantic cells correspond to random set neighbourhoods of prototypes as described in the prototype interpretation of label semantics. A label L is defined according to a set of prototypes P such that L said to be 'appropriate' to describe element x, provided x is 'sufficiently close' to P. Clearly, 'sufficiently close' is an imprecise requirement and in our approach we model this by a random set (semantic cell) corresponding to those elements within an uncertain threshold of P. Here similarity is quantified by a distance metric and the threshold epsilon is a random variable with associated probability density function delta. This paper introduces an algorithm for automatically learning linguistic rules from data, where labels are represented by semantic cells. The conditional rules generated have the simple form 'If X is about P then Y is about f(X),where f is linear function of the input variables X. The proposed algorithm determines a semantic cell for each input label in the form a single prototype P, a Gaussian density function delta and a set of linear coefficients for the output function f, as estimated by using a hyperplane clustering algorithm. The potential of this approach is then demonstrated through its application to a number of well-known benchmark regression problems and also to a real-time flood forecasting problem.

ES36 Room 7 TIME SERIES MODELLING AND COMPUTATION 3 Chair: Roland Fried

E149: **Improved prediction limits for a general class of Gaussian models**
Presenter: **Paolo Vidoni**, University of Udine, Italy
Co-authors: Federica Giummole

We consider the problem of prediction for a general class of Gaussian models, which includes ARMA time series models, linear Gaussian state space models and Gaussian Markov random fields. Although prediction is usually performed by giving a point predictor, in most applications it can be useful to define prediction intervals or prediction limits. The simplest frequentist approach, based on the estimative prediction limit, may be rather inaccurate since both the conditional and the unconditional coverage probability differ from the nominal level by a term usually of order $O(n^{-1})$. In this paper a general method for correcting estimative prediction limits, in the context of Gaussian processes is provided. The proposed prediction limits are modifications of the estimative ones and present coverage error reduced to order $O(n^{-1})$. The analytic expression for both the conditional and the unconditional improved prediction limits are derived. The modifying terms have a quite simple form and they involve, respectively, the conditional and unconditional bias and mean square error of the plug-in estimators for the conditional expectation and the conditional variance of the future observation. The associated predictive distribution functions and densities are also given.

E193: **Nonparametric time series with sudden changes in structure**
Presenter: **Jurgen Franke**, University of Kaiserslautern, Germany
Co-authors: Jean-Pierre Stockis, Joseph Tadjuidje

We consider nonparametric time series alternating between finitely many different phases. A simple example is the following simple nonlinear autoregressive mixture model of order 1 with 2 different phases: $Y_t = I_t m(X_t) + se_t + (1-I_t)m*(X_t) + s*e_t$. The innovations e_t are i.i.d. with mean 0 and variance 1, I_t is a hidden Markov chain assuming values in $\{0,1\}$, and X_t denotes a random vector composed of some observations of the time series up to time t. We discuss various nonparametric methods based on local smoothing and on feedforward neural networks for estimating the autoregressive functions m, $m*$ and identifying the change points between different phases, i.e. in the example the changes of the state process I_t from 0 to 1 and vice versa. We extend this framework to more than 2 phases and to heteroscedastic time series, and we present some applications to modelling pathological ECG data and to forecasting stock prices and getting trading signals for portfolio management.

E166: **Shifts in individual parameters of a GARCH model**
Presenter: **Pedro Galeano**, Universidad Carlos III de Madrid, Spain
Co-authors: Ruey S. Tsay

Most asset return series, especially those in high-frequency, show high excess kurtosis and persistence in volatility that cannot be adequately described by the generalized conditional heteroscedastic (GARCH) model, even with heavy-tailed innovations. Many researchers have argued that these characteristics are due to shifts in volatility that may be associated with significant economic events such as financial crises. Indeed, several authors have investigated the case of pure structural changes, in which all of the parameters in the GARCH model are assumed to change simultaneously. In this paper, we take an alternative approach by studying the case in which changes occur in individual parameters of a GARCH model. We investigate the impacts of such changes on the underlying return series and its volatility, and propose an iterative procedure to detect them. In all cases, the changes affect permanently the level of the volatility, but in some cases, the changes also alter the dynamic structure of the volatility series. Monte Carlo experiments are used to investigate the performance of the proposed procedure in finite samples, and real examples are used to demonstrate the impacts of detected volatility changes and the efficacy of the proposed procedure. Practical implications of the parameter changes in financial applications are also discussed.

CS07 Room 10 MULTIVARIATE MULTIPLICATIVE MODELS AND RELATED DISTRIBUTIONS Chair: David Veredas

C192: **Robust estimation of CCC and DCC GARCH models**
Presenter: **Sebastien Laurent**, Universiteit Maastricht, Netherlands
Co-authors: Kris Boudt, Jon Danielsson

The use of CCC and DCC models has now become standard in financial econometrics. Their estimation is usually done in two or three steps by Gaussian quasi-maximum likelihood. We show that this method is very sensitive to outliers in the data. We propose to use robust estimators for both models. The Monte Carlo study and empirical application document the good robustness properties of this estimation method.

C189: **Flexible models obtained by perturbation of symmetric densities**
Presenter: **Anna Clara Monti**, University of Sannio, Italy

Flexible models obtained by perturbation of symmetric densities are considered. Appropriate choices of symmetric densities and perturbing functions lead to distributions able to fit the distributions of data which exhibit skewness and/or tick tails, in both the univariate and multivariate contexts. Henceforth the adoption of these models provides an alternative to robust procedures when departures from normality are likely to occur. Furthermore, these models have stochastic representations which give rise to interesting applications in the analysis of real data. An overview of these models is provided, and their properties and some inferential issues are discussed.

C046: **Common long-run volatility. A seminonparametric multivariate MEM**
Presenter: **David Veredas**, Universite Libre de Bruxelles, Belgium
Co-authors: Matteo Barigozzi, Christian Brownlees, Giampiero Gallo

We propose a multivariate multiplicative error model for a large number of assets that disentangles between the common long-run movements and the idiosyncratic short-run dynamics. Estimation rests in a meta-t-copula where the marginal densities, which include the long-run and short-run components, are first estimated with generalized profile likelihood. The parameters of the copula are estimated in a second step through a iterative procedure -avoiding hence the maximization of the copula density - that involves a simple test for ellipticity and a distribution-free estimator of the correlation matrix.

CS14 Room 8 COPULA METHODS IN TIME SERIES ANALYSIS 2 Chair: Alessandra Luati

C167: **Forecasting realized volatility with a copula-based time series model**
Presenter: **Oleg Sokolinskiy**, Tinbergen Institute, Erasmus Universtiy Rotterdam, Netherlands
Co-authors: Dick van Dijk

This paper develops a novel approach to modeling and forecasting realized volatility (RV) based on copula functions. Copula-based time series models can capture relevant characteristics of RV such as nonlinear dynamics and long-memory type behavior in a flexible yet parsimonious way. This makes it a possible contender to conventional approaches for modeling and forecasting realized volatility, such as the HAR-RV model. In an empirical application to daily realized volatility for the SP500 index futures, we find that the copula-based model for RV (C-RV) outperforms the HAR-RV for one-day ahead volatility forecasts in terms of accuracy (based on RMSE and MAE) and in terms of efficiency (based on Mincer-Zarnowitz and encompassing regressions). Among the copula specifications considered, the Gumbel C-RV model achieves the best forecast performance, which highlights the importance of asymmetry and upper tail dependence for modeling volatility dynamics. Although we find substantial variation in the copula parameter estimates over time, conditional copulas improve the accuracy of volatility forecasts only marginally.

C196: **Vine copulas with asymmetric tail dependence and applications to financial return data**
Presenter: **Aristidis Nikoloulopoulos**, University of East Anglia, UK
Co-authors: Harry Joe, Haijun Li

One goal in the theory of dependence modeling and multivariate copulas is to develop parametric families that are appropriate as models for multivariate data with different dependence structures. In particular, when data display dependence among extreme values and it is necessary to have copula models with reflection asymmetry and flexible lower/upper tail dependence, then vine copulas is the best choice. In finance there is some empirical evidence of asymmetric correlations in stock returns with market indices, so that vine copulas with appropriate choices of bivariate linking copulas can provide models to assess such asymmetries. This talk consists in the application of vine copulas with asymmetric tail dependence to multivariate financial market index returns and the investigation of the following practical issues for vine-copula modeling: (a) type of the vine, (b) type of bivariate copula families as building blocks, and (c) given the vine, the matching of variables to labels/indexes.

C221: **The effects of misspecified frequency, severity and dependence function modeling on operational risk measures**
Presenter: **Dean Fantazzini**, Moscow State University, Russia

The quantitative analysis of Operational Risk is a relative recent field of study within the more general quantitative risk management framework. However, little attention has been place to study the effect on the computation of the operational risk measures when the frequency models, the severity models and the dependence structures are misspecified. Besides, few studies have analyzed the behaviour of these models in small samples. In this work we present and discuss the results of a large-scale Monte Carlo study of different misspecified models in an Operational Risk setting. We show that the misspecification in the marginal frequency and severity distributions more than offset the biases in copula parameters when computing risk measures and small samples are considered, which is the usual case for operational risk. Interestingly, we find that Extreme Value Theory (EVT) works fine only with medium to large datasets, while it delivers poor results when dealing with small samples.

CS31 Room 1 VERIFYING ASYMPTOTIC APPROXIMATIONS BY SIMULATION Chair: Jan Kiviet

C062: **Improved variance estimation of coefficients in stable first-order dynamic regression models**
Presenter: **Garry Phillips**, Cardiff Business School, UK
Co-authors: Jan Kiviet

In dynamic regression models the least squares coefficient estimators are biased in finite samples and so are the usual estimators for the disturbance variance and for the variance of the coefficient estimators. By deriving the expectation of the sum of the initial terms in an expansion of the usual expression for the coefficient variance estimator and by comparing this with an approximation to the true variance, we find an approximation to the bias in variance estimation from which a bias corrected estimator for the variance readily follows. This is also achieved for a bias corrected coefficient estimator which enables one to compare analytically the second-order approximation to the mean squared error of the ordinary least squares estimator and its counterpart after bias correcting the the coefficient estimator to first order. Illustrative numerical and simulation results on the magnitude of the bias in coefficient and variance estimation and on the options for bias reduction are presented for three particularly relevant cases of the ARX(1) class of models. These show that substantial efficiency gains and test size improvements can easily be realized.

C017: **Edgeworth expansions of the QMLEs in the EGARCH(1,1) model**
Presenter: **Dimitra Kyriakopoulou**, University of Piraeus, Greece
Co-authors: Antonis Demos

The last years there has been a substantial interest in approximating the exact distributions of econometric estimators in time series models. Although there is an important and growing literature that deals with the asymptotics of the Generalized Autoregressive Conditional Heteroskedastic (GARCH) models, the asymptotic properties of the estimators in the Exponential GARCH(1,1) process of Nelson have not been fully explored. Comparing to the GARCH process, the advantages of the EGARCH model are well-known, with the main one being the fact that the model captures the negative dynamic asymmetries noticed in many financial series, i.e. the so-called leverage effects. In this paper we develop the Edgeworth expansion of the Quasi Maximum Likelihood Estimators (QMLEs) of the EGARCH(1,1) parameters and we derive their asymptotic properties, in terms of bias and mean squared error. The notions of the geometric ergodicity and stationarity are also discussed in shedding light on the asymptotic theory of the EGARCH models. Additionally, necessary and sufficient conditions for the existence of the Edgeworth approximation are investigated. We also examine the effect on the QMLEs of the variance parameters by including an intercept in the mean equation. We check our theoretical results by simulations. In doing this, we employ either analytic or numeric derivatives and the convergence properties of the methods are also discussed.

C030: **Comparing the asymptotic and empirical (un)conditional densities of OLS and IV in a simultaneous equation**
Presenter: **Jan Kiviet**, University of Amsterdam, Netherlands
Co-authors: Jerzy Niemczyk

Monte Carlo simulation is often used to assess distributional properties of estimators and test statistics in finite sample. The data generating processes involved may include exogenous variables that follow particular typical stochastic processes. If all replications use just one single realization of such processes the simulation yields the conditional distribution of the analyzed inference methods and the unconditional distribution when each replication is based on new independent draws. Both simulation designs may yield useful information, although for practitioners it is often the more specific though more efficient conditional distribution that is of primary interest, provided that it is conditioned on the preferred actual exogenous variables. For consistent estimators their conditional and unconditional limiting distributions are usually the same, but when translating these into an asymptotic approximation to the small sample distribution conditioning matters. However, for inconsistent estimators their limiting distribution may change when conditioning or not. These phenomena are analyzed and numerically illustrated when comparing OLS (ordinary least-squares) and IV (instrumental variables) estimators in a single simultaneous equation, thus supplementing, and occasionally correcting, earlier results. The findings suggest that, especially when one is willing to make an assumption on

the maximum degree of simultaneity in the model, inference based on inconsistent OLS may often be more attractive than that produced by consistent IV when the instruments are relatively weak.

CS41 Room 9 OPERATIONAL RISK Chair: Stefan Mittnik

C309: **Estimation of operational risk: dependence and robustness**
Presenter: **Tina Yener**, Ludwig-Maximilians-University, Germany
Co-authors: Stefan Mittnik, Sandra Paterlini

The modeling of operational risk has mainly concentrated on the marginal distributions of frequencies and severities within the Loss Distribution Approach (LDA). However, a realistic quantitative model should be capable to model the characteristics of the loss distribution (high kurtosis, severe right-skewness, and excessive heavy tails) while providing stable estimates, incorporate dependencies and overcome the too simplistic assumption of a perfect positive correlation among operational losses. Furthermore, the scarcity of real-world available data has so far not allowed developing best practice guidelines among practitioners. In this work, having access to a real-world data set, we analyze the effects of competing state-of-art strategies for dependence modeling on aggregated risk capital estimates and focus particularly on robust methods for model estimation in order to improve the stability of the estimates even when the underlying model is not perfectly satisfied by the available dataset.

C343: **A Bayesian approach to extreme value estimation in operational risk**
Presenter: **Stefan Mittnik**, University of Munich, Germany
Co-authors: Bakhodir Ergashev, Evangelos Sekeris

This paper studies a Bayesian approach to fitting the Generalized Pareto distribution to historically observed operational losses of financial institutions. When loss data is insufficient to accurately estimate the tail behavior of losses, the incorporation of expert opinions plays an important role in reducing the statistical uncertainty around the parameter estimates as well as the resulting operational risk capital. We use the Bayesian approach to include expert judgments in the estimation process as prior assumptions about the the model's unknown parameters. The formulation of the priors is implemented through the elicitation of expert opinions. We also discuss some important challenges of the elicitation process.

CS54 Room 2 INFERENCE FROM ROBUST ESTIMATORS Chair: Davide Ferrari

C341: **Tail modelling in linear models by quantile regression**
Presenter: **Jan Dienstbier**, Technical University of Liberec, Czech Republic

We deal with the description of a variable of primary interest in terms of covariates from the extreme value point of view. We are interested in estimating condition indices, extreme conditional quantiles and small exceedance probabilities. These tasks are motivated by real case studies e.g. in geology or finance. It has been proposed to use regression quantiles as a suitable covariate dependent high threshold or even base extreme value analysis on regression quantile process. We broaden older results by presenting strong approximations of high regression quantiles. Hence following the older results on strong approximations of the empirical tail distribution function we obtain a whole class of consistent and asymptotically normal estimators of extreme value index as Hadamard differentiable location and scale invariant functionals of the regression quantile process. As a consequence we also gain estimators of extreme conditional quantiles and small exceedance probabilities. We illustrate the properties of the estimators on the simulations and on Aon Re Belgium fire portfolio data.

C161: **Higher-order robustness**
Presenter: **Davide La Vecchia**, University of Lugano, Switzerland
Co-authors: Elvezio Ronchetti, Fabio Trojani

The higher–order robustness for M–estimators is introduced and defined. The conditions needed to ensure higher stability of the asymptotic bias are provided by refining the Von Mises bias expansion. Admissible M–estimators featuring second–order robustness are thus introduced. Then, a saddle-point argument is applied in order to approximate the finite sample distribution of second–order robust M–estimators. The link between the stability of this approximation and the second–order robustness is explored. Monte Carlo simulation provides evidence that second–order robust M–estimators perform better than the MLE and Huber–type estimators, even in moderate to small sample sizes and/or for large amounts of contamination.

C259: **A fully parametric approach to minimum power-divergence estimation**
Presenter: **Davide Ferrari**, University of Modena and Reggio Emilia, Italy
Co-authors: Davide La Vecchia

We approach parameter estimation based on power-divergence using Havrda-Charvat generalized entropy. Unlike other robust estimators relying on divergence measures, the procedure is fully parametric and avoids complications related to bandwidth se-

lection. Hence, it allows for the treatment of multivariate distributions. The parameter estimator is indexed by a single constant q, balancing the trade-off between robustness and efficiency. If q approaches 1, the procedure is maximum likelihood estimation; if $q = 1/2$, we minimize an empirical version of the Hellinger distance which is fully parametric. We study the mean squared error under contamination by means of a multi-parameter generalization of the change-of-variance function and devise an analytic min-max criterion for selecting q. Optimal q between $1/2$ and 1 give remarkable robustness and yet result in negligible losses of efficiency compared to maximum likelihood. The method is considerably accurate for relatively large multivariate problems in presence of a relevant fraction of bad data.

CS69 Room 5 FINANCIAL MARKETS 2 **Chair: Elena Kalotychou**

C082: **The persistence of bank profits**

Presenter: **Philip Molyneux**, Bangor University, UK

This paper explores the strength of competitive forces operating in banking systems around the world. Using a data set of banks from 65 countries, we test whether banks located in some countries can persistently earn profits significantly different from the long-run equilibrium value. Overall we find that banks in developing countries exhibit less profits persistence than in developed systems. We find that profits persistence is highest in North America and Western Europe and lowest in East Asia and the Pacific region. Profits are more likely to persist for banks located in countries with high barriers to entry and low levels of competition. Our findings have implications for the extent to which policy makers should intervene to promote competitive outcomes in the banking industry.

C369: **Modelling realized volatility, bipower variance and jumps in energy futures**

Presenter: **Sjur Westgaard**, Norwegian University of Science and Technology, Norway

Co-authors: Per Bjarte Solliebakke, Eirik Haugom, Gudbrand Lien

The introduction of market-based principles in electricity markets around the world has increased the risk for power producers and marketers substantially. That is, one of the most prominent consequences of deregulating the electricity sector is the strong increase in the fluctuations of the prices. Since electricity is a 'commodity' that is non-storable in nature, these fluctuations can reach extreme levels even on a daily basis, and are usually far above the levels observed in other financial – and commodity markets. In order to manage the risk associated with production, trading and investment decisions, actors need techniques that are able to capture the properties of the prices correctly. Previous research on volatility modeling of electricity prices has shown evidence of such clustering. This research has usually utilized historical data for daily returns and often the popular GARCH-type framework is applied. However, these typically used approaches do not reflect the intra day price dynamics. Since electricity future prices are determined several times during the day it would be possible to calculate the intra day variability using the concepts of realized volatility (RV) and bipower variation (within the theory of quadratic variation). Findings from studies in other markets (i.e. finance) have indicated promising results when making predictions based on these measures compared with the popular GARCH approach. Models and forecasts of the non-parametric volatility measures (RV) are obtained utilizing several time series models. The results from our unique energy future intra day data are compared to studies in the stock and FX market.

C200: **The economic value of realized covariance for market timing**

Presenter: **Wei Liu**, City University, UK

Co-authors: Ana-Maria Fuertes, Elena Kalotychou

Volatility modelling and forecasting is important for optimal portfolio selection. From this economic perspective, the present study contributes to the literature by investigating the incremental value of intra day price information for forecasting the covariance matrix of asset returns with a view to selecting optimal portfolio weights. Covariance forecasts are generated using several multivariate GARCH (MGARCH) models that incorporate intra day information in two alternative ways. On the one hand, MGARCH models fitted to daily returns are augmented with realized (co)variance measures to produce one-day-ahead conditional (co)variance forecasts. On the other hand, appropriately deseasonalized intra day returns are directly modelled through MGARCH specifications and the intra–day (co)variance forecasts thus generated are aggregated into one–day–ahead predictions. To economically evaluate the alternative approaches a dynamic asset allocation strategy is deployed, which uses as inputs the forecasted (realized) covariance from the various MGARCH specifications. The performance of the various strategies is evaluated using risk-and cost–adjusted measures and contrasted with that from two benchmark covariance models, the (realized) simple historical and EWMA models. Finally, we also explore the controversial issue of rebalancing frequency on a cost–adjusted basis by following a time–varying approach that allows it to change according to economic and market conditions.

Friday 30.10.2009 14:30-16:10 Parallel Session H

ES16 Room 7 TIME SERIES MODELLING AND COMPUTATION 4 Chair: Konstantinos Fokianos

E133: **Frequency domain tests in multivariate time series**
Presenter: **Efstathios Paparoditis**, University of Cyprus, Cyprus

We consider nonparametric frequency domain tests about the spectral density matrix of a multivariate time series and propose a general bootstrap procedure to approximate the distribution of the corresponding test statistics under validity of the null hypothesis. Under a set of easy to verify conditions, we establish asymptotic validity of the proposed bootstrap methodology. We apply a version of this procedure together with a new statistic in order to test the hypothesis that the spectral densities of not necessarily independent time series are equal. The test statistic is based on a L_2-distance between the nonparametrically estimated individual spectral densities and an overall, 'pooled' spectral density, the later being obtained using the whole set of m time series considered. The effects of the dependence between the time series on the power behavior of the test are investigated. Some simulations are presented and real-life data examples are discussed.

E035: **TFT-Bootstrap: resampling time series in the frequency domain to obtain replicates in the time domain**
Presenter: **Claudia Kirch**, University of Karlsruhe, Germany
Co-authors: Dimitris Politis

A new time series bootstrap scheme, the Time Frequency Toggle (TFT)-Bootstrap, is proposed. Its basic idea is to bootstrap the Fourier coefficients of the observed time series, and then back-transforming them to obtain a bootstrap sample in the time domain. Related previous proposals, such as the 'surrogate data' approach resampled only the phase of the Fourier coefficients, and thus had validity limited to Gaussian processes. By contrast, we show that the appropriate resampling of phase and magnitude of Fourier coefficients yields a bootstrap scheme that is consistent for a large class of linear and nonlinear time series processes. As a main result we obtain a functional limit theorem for the TFT-Bootstrap under a variety of popular ways of frequency domain bootstrapping. Possible applications of the TFT-Bootstrap naturally arise in change-point analysis and unit-root testing where statistics are frequently based on functionals of partial sums.

E019: **Inference of multivariate dynamic generalized linear models**
Presenter: **Kostas Triantafyllopoulos**, University of Sheffield, UK

Inference for a wide class of dynamic generalized linear models for multivariate responses are developed. Our motivation stems from time series data following multinomial distributions. The proposed approach of inference, being Bayesian, adopts the conjugate prior distribution and it proposes an approximate estimation of its parameters, based on Bayes linear methodology. We develop predictive and smoothing inference and we discuss in some detail prior specification and diagnostic tests. We thus provide a complete analogue to inference for univariate dynamic generalized linear models that have been previously suggested. We review several estimation procedures for multivariate dynamic generalized linear models and provide a critical appraisal in comparison with those we are proposing. We discuss in detail the multinomial distribution and we provide Kalman-filter type equations for filtering, smoothing and forecasting. Based on this distribution, our empirical results illustrate that the proposed methodology offers several advantages compared with the existing estimation procedures for multinomial time series. We conclude by suggesting several possible directions of future research and open problems, demonstrating the usefulness and potential of the proposed inferential approach.

E170: **On robust change-point detection in time series**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Ursula Gather, Herold Dehling

Automatic measurement of a multitude of variables in short time lags is common nowadays. Such time series data are often disturbed by a high level of noise and different types of measurement artifacts. A basic question is whether the underlying data-generating mechanism is stable over time. For a reliable statistical analysis of this problem we need to distinguish between short-term fluctuations and (patches of a few) outlying observations on the one hand and long-term changes like level shifts on the other hand. Motivated by online monitoring in intensive care, we have developed robust statistical procedures for shift detection in noisy time series. Robust test statistics can be constructed from the difference of medians within different time windows, or using the median of all pairwise differences between the observations in different windows. For online monitoring of time series with slowly varying trends, we can estimate the underlying trend applying repeated median regression to a moving time window. For the retrospective setting, the asymptotical theory underlying our test statistics can be derived. The usefulness of the approaches is investigated via simulations.

ES31 Room 3 FOUNDATIONS FOR FUZZY STATISTICAL ANALYSIS Chair: Didier Dubois

E051: **A fuzzy sets membership function computation under uncertainty**
Presenter: **Maria Yudaeva**, Saint-Petersburg State University, Russia
Co-authors: Nikolay Hovanov, Dmitry Kolesov

A method of randomized membership functions computation on the basis of expert non-numeric, non-complete, non-precise information is presented. All possible membership functions are assumed to be piecewise constant and determined by corresponding discrete parameter vectors. It is supposed also that NNN-information is represented in the form of system of equalities and inequalities for components of parameter vectors. The expert information helps to proceed to a set of all admissible parameter vectors. Uncertain choice of a parameter vector from the set of all admissible parameter vectors is modeled by random choice and in the result random parameter vector is obtained, random membership function can be derived. Mathematical expectation of the random membership function is regarded as a numeric estimation of membership function value and exactness of this estimation may be measured by the corresponding standard deviation (explicit formulas are obtained). Randomized membership functions technique (RMFT) outlined above can be used for different fuzzy variables estimation under numeric information deficiency. Among other things the RMFT gives us a tool for fuzzy financial and economic indicators estimation on the base of expert NNN-data. Such application of the RMFT is illustrated by dollar-ruble exchange rate forecast computation.

E072: **Statistical inference using belief functions: a reappraisal of the General Bayes Theorem**
Presenter: **Thierry Denoeux**, Universite de Technologie de Compiegne, France
Co-authors: Didier Dubois

The General Bayes Theorem (GBT) as a generalization of Bayes theorem to the belief function framework. In probability theory, Bayes rule is used for three purposes: prediction based on observations, fusion of information, and learning (what statisticians call inference). The GBT supposedly addresses the latter task. It provides a method for deriving a belief function on a model parameter based on observations, when likelihoods are expressed as belief functions, without assuming any prior knowledge of the parameter. The GBT has been applied to diagnosis and classification tasks, but its use in statistical inference has remained very limited. After recalling the derivation of the GBT from first principles and some fundamental results, we try to analyze the potentials and the limitations of this approach to statistical inference. Its application to a very simple problem is discussed and compared to other approaches. A possibilistic counterpart of the GBT is also outlined.

E208: **Impact of fuzziness in measurement scale on basic statistical inference**
Presenter: **Shubhabrata Das**, Indian Institute of Management Bangalore, India

We investigate how statistical inference problem of parameters like mean, variance and proportion are influenced when the available data is fuzzy because of measurement scale device. In particular, we wish to explore if the measure of fuzziness of the scale is useful in providing any guideline for suitable alternation on inference problems. Of interest is a suitable parameter of the distribution of X. However X is un-observable and what we observe are y's , fuzzy form of x's. The y's are considered to be in a k–point Likert scale under different models of fuzziness. The role of k is be de-emphasized here, although we experiment with few alternative values of k. We explore a sufficient condition under which the fuzziness of scale has no adverse impact and investigate how reasonable this sufficient condition is under various structures of fuzziness. When the conditions are not valid, we study the degree of difference in the two means (of x and y), leading to validity of the approximation. Then we obtain the relationship between variance of x and y and study how that impacts the inference problem for variance/mean (in-)directly. Finally we focus on the problem of estimating proportions and discuss relevant estimation issues.

E197: **Further exploring statistical fuzzy entropy**
Presenter: **Jan van den Berg**, Delft University of Technology, Netherlands

Using earlier work as a starting point, we extend the exploration of the properties of the notion of Statistical Fuzzy Entropy and its applications. The exploration is started by considering several cases, among others, the case of a binary statistical fuzzy information source that generates two 'strictly' complementary fuzzy events. It turns out that, for this special case, the statistical and fuzzy uncertainties involved, in the mathematical sense, play an equal role. We perform several analyses and also provide an interpretation of the analytical results. Next it is shown how the notion of SFE can be applied to induce statistical fuzzy decision trees from a fuzzy data set consisting of fuzzy data records each one of which has a fuzzy classification. To be able to induce statistical fuzzy decision trees, the notion of Statistical Fuzzy Information Gain (SFIG) is (re)formulated. Finally it is shown how - given an induced tree - a new, unseen data element can be classified where generally the classification outcomes of all branches need to be aggregated in a weighted manner. This extra calculation power needed is the price we have to pay for generalizing the *crude* statistical crisp world to the 'smoother' statistical fuzzy world.

E227: **A smoothing procedure based on fuzzy F-transform**

Presenter: **Luciano Stefanini**, University of Urbino "Carlo Bo", Italy

The fuzzy transform (F-Transform), as a general tool for approximation, its theoretical properties and several applications have been previously considered. Recently a parametrized class of fuzzy numbers to define flexible general basic functions for the fuzzy transform has been proposed and an estimation of the shapes with better approximation properties has been obtained. An extension of the fuzzy partitions has been suggested and shown that the inverse fuzzy transform based on extended partitions has an important smoothing property. The discrete fuzzy transform on extended parametric partitions is presented. Its smoothing properties (in the unidimensional case) are shown and a numerical procedure to obtain least squares estimates of the parameters defining the basic functions is described. Finally, a generalized cross validation approach for selecting the smoothing bandwidth is suggested and examples are considered.

CS02 Room 8 PERFORMANCE EVALUATION Chair: Dominique Guegan

C048: **Heterogeneity in hedge funds performance persistence**

Presenter: **Serge Darolles**, CREST and SGAM AI, France

Co-authors: Christian Gourieroux, Jerome Teiletche

The ratings of funds based on past performance and the rating persistence are crucial informations for investors. This paper proposes a general econometric framework to investigate the dynamics of performance based ratings of funds, for any risk adjusted measure of performance. We reformulate the questions related to rating persistence at different horizons in terms of Markov Chain analysis, both for individual and joint rating trajectories. The methodology is illustrated by the analysis of hedge funds returns extracted from the TASS database for the period 1994-2008.

C045: **Backward/Forward optimal combination of performance measures**

Presenter: **Monica Billio**, University of Venice, Italy

Co-authors: Massimiliano Caporin

The financial economics literature proposes a large number of alternative performance measures for the evaluation of financial assets and managed portfolios. Despite the large number of works dealing with performance measurement analysis, there is not a general agreement on which is, if there exist one, an optimal performance measures. Some authors compare different PM for the purpose of determining if they all provide the same informative content with some interesting conclusions. In this paper we start from these findings and propose an approach for determining the optimal combination of a restricted list of PM providing partially different information. We develop an algorithm for combining PM based on a backward simulated track record and then apply the combination for the identification of the best assets for the forward construction of a portfolio with an equally weighted allocation. We test our method using historical prices of managed accounts.

C059: **A Monte Carlo test for raters agreement with applications to mutual funds**

Presenter: **Daniele Franceschi**, University of Padova, Italy

Co-authors: Francesco Lisi

The paper deals with the problem of comparing raters agreement. The focus of the analysis is on testing the hypothesis that two ratings are equivalent. This is different from the more usual case of testing the independence between raters. To reach our goal first the notion of beta-equivalence is introduced and then a Monte Carlo test for the hypothesis of beta-equivalence is described. As a case-study we analyze 1763 monthly return time series of US mutual funds for which we consider the Morningstar classification with different risk aversion parameters. Results document that the current Morningstar classification, based on a risk-adjusted measure, accounts for risk only marginally.

C041: **Performance of long/short equally weighted portfolios**

Presenter: **Ludovic Cales**, University of Paris 1 Pantheon - Sorbonne, France

Co-authors: Monica Billio, Dominique Guegan

The aim is to provide a measure for the performance of a Long/Short Equally Weighted portfolio that would be both objective and computable. We consider a portfolio manager investing according a Long/Short Equally Weighted strategy. He has the choice to invest in $(n!/(n/2)!)^2$ different portfolios. In order to provide an objective measure of the goodness of his choice we can compare his portfolio performance with the performances of the alternative portfolios. The rank of his portfolio performance would be the measure of its performance. Such a measure is interesting because: (a) it is independent of the market conditions. Either you are in the top 10% or not, whatever is the performance; (b) no referent portfolio manager is needed. There is no need to compare the performance measure with the performance of a referent who might have traded in a slightly different market.

CS13 Room 5 ECONOMETRIC VALIDATION OF AGENT-BASED MODELS Chair: Thomas Lux

C240: **Socio-dynamic discrete choice: an agent-based approach and issues in estimation**
Presenter: **Elenna Dugundji**, Universiteit van Amsterdam, Netherlands
Co-authors: Laszlo Gulyas

A multi-agent based model of binary choice behavior with interdependence of decision-makers' choices is developed. First the model is docked to analytical results where agent heterogeneity is not explicitly treated. Next two abstract classes of networks are considered to introduce agent heterogeneity via an explicit local interaction structure. Then the model is applied in an example of intercity travel demand using empirical data to introduce individual agent heterogeneity beyond that induced by the local interaction structure. Studying the long run behavior of more than 120,000 multi-agent based simulation runs reveals that the initial estimation process can be highly sensitive to small variations in network instantiations. We show that this is an artifact of two issues in estimation, and highlight particular attention that is due at low network density and at high network density. Suggestions for future research efforts are outlined.

C321: **Economic interpretation of GARCH models: an agent-based simulation study**
Presenter: **Balazs Torma**, MTA-SZTAKI, Hungary
Co-authors: Laszlo Gerencser

GARCH models are technical volatility models with parameters of hardly any economic meaning. Applying an agent-based approach, we discovered fundamental economic effects that may determine the parameter values of the GARCH model that best describes observed market prices. In our market microstructure model trend followers, mean reverter chartists and fundamentalists trade in a stock market. The market structure is defined and parameterized by the distribution of wealth among groups of agents of the three different types. Chartists predict future stock prices using simple technical rules and calculate their transaction requests according to the prediction. An important component influencing the belief of the fundamentalists about the correct stock price is the information they receive. Our model of information process, derived from Hawkes's point processes, exhibits the self-exciting nature of market news we can observe on real markets. Our extensive numerical experiments have shown that market structure is a fundamental determinant of the coefficients of the best-fit GARCH model. In particular, an increase in the wealth of trend followers increases the ARCH-coefficient and an increase in the wealth of fundamentalists increases the GARCH-coefficient as the impact of the self-exciting effect of news increases. Based upon this finding, we have developed a real-time stochastic approximation algorithm using the Minimum Description Length principle to detect abrupt changes in the market structure.

C345: **How bounded is subjects' rationality in a simple experiment**
Presenter: **Josep Domenech**, University Jaume I and Universitat Politecnica de Valencia, Spain
Co-authors: Simone Alfarano, Eva Camacho

It has been widely accepted that the human subjects performance in economic experiments cannot be satisfactorily described assuming complete rationality of their behavior. In the last decade many models of human behavior which account for bounded rationality have been proposed in the literature. The adaptation of genetic algorithms (GA) from the realm of optimization literature to the description of human learning is an example of the creative ability of researchers in introducing bound rational agents. A number of papers are now available in the literature which apply different versions of GAs in order to reproduce the behavior of human subjects in different experimental settings. However, up to now the different contributions are almost entirely based on a rough calibration of the underlying crucial parameters. We provide a method to estimate the key parameters of the GA by means of an extensive simulation-based approach, using an extremely simple experimental setting of a common-pool resources problem. The resulting artificial agent is characterized by processing a relatively large set of potential strategies evaluated using past experience rather than considering his strategic role in the game. This is consistent with a vast literature on the use of simple heuristics in decision making under uncertainty.

CS15 Room 9 GROWTH ECONOMETRICS Chair: Martin Wagner

C249: **Estimation of higher-order panel vector autoregressions with finite time dimension**
Presenter: **Michael Binder**, Goethe University Frankfurt, Germany
Co-authors: Jan Mutl, M. Hashem Pesaran

We consider estimation of higher-order panel vector autoregressions (PVARs) under fixed effects when the time dimension of the panel is finite and the cross-sectional dimension is large. We outline both Quasi Maximum Likelihood (QML) and Minimum Distance (MD) estimators for PVARs, and compare their properties to those of various Generalized Method of Moments (GMM) estimators that are widely used in the applied literature. To aid computation of the QML and MD estimators, we (i) derive non-recursive characterizations of initial observation restrictions both under stationarity and in the presence of unit roots and cointegrating relations, and (ii) obtain analytical derivatives for the likelihood and minimum distance criterion functions. We

also adduce Monte Carlo evidence comparing the finite sample performance of the QML estimator to that of the MD and a variety of GMM estimators. The Monte Carlo evidence strongly favors the QML estimator, both relative to the MD and the GMM estimators.

C339: **On the heterogeneity of regional growth patterns across Europe: A spatial mixture model**
Presenter: **Michele Battisti**, University of Palermo and LUISS, Italy
Co-authors: Giuseppe Arbia, Gianfranco Di Vaio

In applied economic analysis is often useful to make separate estimation and inference, on the sub–samples of a population, avoiding a priori assumptions on the grouping criteria. A consolidated approach considers the use of mixture models although this approach has not been followed frequently in the spatial econometrics. We incorporate the standard spatial models into the mixture regression framework. The proposed spatial mixture regression model (SMRM) deals simultaneously with the spatial dependence and the spatial heterogeneity present in the empirical data. Estimation is carried out through the EM algorithm, which in turn is based on a maximum likelihood approach. Computational aspects are treatable, since the model nests one maximization problem into the other. Inference is obtained through bootstrapping. Finally, an application to the regional economic growth process across Europe, that aims to identify several convergence patterns, is considered.

C076: **Shrinking cross-country growth regressions**
Presenter: **Marek Jarocinski**, European Central Bank, Germany

This paper studies how countries' economic, social and geographic characteristics are related to economic growth. It uses a single cross country growth regression with a large number of explanatory variables and a Bayesian shrinkage prior that prevents overfitting. The employed prior structure has growth-theoretic and econometric advantages over Bayesian Model Averaging used in the previous literature. Bayesian Model Averaging is a limiting case of the employed prior. The main findings are that coefficients of measures of trade openness are large and positive, coefficients of measures of tropical location, malaria prevalence, fertility and investment prices are large and negative, and convergence is weak. These findings are robust to using different vintages of income data.

C061: **Catching growth geterminants with the adaptive LASSO**
Presenter: **Martin Wagner**, University of Vienna, Austria
Co-authors: Ulrike Schneider

The adaptive LASSO estimator is employed to determine the variables important for economic growth. The adaptive LASSO estimator is a computationally very simple procedure that can perform at the same time consistent parameter estimation and model selection. The methodology is applied to three data sets, two global data sets from widely cited papers and one regional data set for the European Union. The results for the former two data sets are very similar in many respects to those found in the published papers, yet are obtained at a fraction of computational cost. Furthermore, the results for the regional data highlight the importance of human capital for economic growth.

CS20 Room 2 TIME SERIES ANALYSIS Chair: Richard Gerlach

C107: **A structural model for credit migration**
Presenter: **Hoi Ying Wong**, Chinese University of Hong Kong, Hong Kong
Co-authors: Jing Zhao, Ngai Hang Chan

Credit transition matrices published by rating agencies constitute an important building block of risk management. This paper proposes a firm-specific structural model for credit migration that incorporates firm capital structure and risk perception of rating agencies. The proposed model incorporates the notion of distance-to-default, which is estimated from accounting information and market variables, and a migration signal duration, which is modeled by an excursion time to reflect the credit history of a firm. This model not only allows one to derive a closed-form credit transition probability, but can also be used to explain the default probability overlap and the slow-to-respond feature of rating agencies. Examples are provided to demonstrate the estimation of perception thresholds using historical credit transition matrices and empirical distributions of rating categories. The transition probabilities of a rated company can then be computed in a firm specific manner. As the financial market is concerned about how agency ratings affect market conditions, the estimated perception threshold provides an effective measure of a rating agency's response to credit migrations.

C274: **Testing for a change in the order of integration of G7 and Euro area inflation**
Presenter: **Andrea Halunga**, University of Exeter, UK
Co-authors: Denise Osborn, Marianne Sensier

We test G7 and Euro area inflation for a change in the order of integration either from I(1) to I(0) or I(0) to I(1). We extend this work by conducting Monte Carlo analysis on the impact of structural breaks in the deterministic components of these tests. We

find that structural breaks in the mean of a series do matter when testing for the order of integration and therefore we allow for possible structural breaks in the level of the series, implementing a search procedure for deterministic level shifts and persistence changes. Allowing for a structural break in the level of the series, we find that all G7 and Euro area inflation series are stationary from the early 1980s.

C231: **Doubly constrained factor models: estimation and applications**
Presenter: **Henghsiu Tsai**, Academia Sinica, Taiwan
Co-authors: Ruey Tsai

Factor models have been widely used in recent years to improve the accuracy of forecasting when many explanatory variables are available. However, the models often encounter the difficulties of over–parameterization and factor interpretation. In this paper, we first consider constrained factor analysis to obtain a parsimonious factor model and propose likelihood ratio statistics to test the adequacy of factor constraints. Real and simulated examples are used to demonstrate the proposed analysis. In an application, we show that the constrained factor analysis can provide a deeper understanding of variations in monthly financial asset returns. We then extend the constrained models to the doubly constrained factor models by incorporating external information on both rows (e.g., subjects) and columns (e.g., variables) of a data matrix. Maximum likelihood estimates and likelihood ratio statistics of the proposed models are derived. Finally, we consider the applications of doubly constrained factor models in economics and finance.

C088: **Modelling investment guarantees in Japan: A risk-neutral GARCH approach**
Presenter: **Wai Sum Chan**, The Chinese University of Hong Kong, Hong Kong
Co-authors: Andrew Cheuk-Yin Ng, Johnny Siu-Hang Li

The variable annuity market in Japan is still young, but growing rapidly. The market has grown from less than USD5 billion in 2002 to its present USD158 billion. Most variable annuities in Japan are sold with one or more investment guarantees, such as a Guaranteed Minimum Accumulation Benefit (GMAB), which guarantees that the ultimate annuity principal will not fall below a pre-set level regardless of the underlying investment performance. Of interest to financial institutions selling variable annuities is the cost associated with such a guarantee. Although the orthodox Black-Scholes option pricing formula can be applied readily, the resulting price might be inaccurate because returns on the Japanese stock price index do not seem to behave as assumed in the formula. In this study, we propose a method for valuing investment guarantees on the basis of a GARCH-type model, which better captures the characteristics of the stock price index. A difficulty in option-pricing with GARCH is the identification of a risk-neutral probability measure. We solve this problem by considering the conditional Esscher transform.

CS23 Room 10 REGIME SWITCHING GARCH MODELS Chair: Leopold Soegner

C042: **Asymmetric multivariate Markov-switching GARCH: structural properties and applications**
Presenter: **Markus Haas**, University of Munich, Germany
Co-authors: Ji-Chun Liu

We develop a multivariate Markov-switching GARCH model which allows for regime-specific interdependent volatility dynamics, leverage effects and correlation structures. Conditions for stationarity and the existence of moments are derived, and expressions for the unconditional moments, the correlation structure of the power-transformed process and closed-form multi-step covariance matrix forecasts are provided. The consistency of the maximum likelihood estimator (MLE) is established. The behavior of the MLE is also evaluated by simulation for sample sizes typical for applications to daily stock returns. The model nests several specifications proposed in the literature. The relationship between these models is discussed, and they are compared in an application to European stock market indices. Both in- and out-of-sample, we find that the dominating model is a two-regime specification with asymmetric volatility dynamics and correlations which are significantly higher in the high-volatility regime.

C276: **A dynamic conditional correlation model with factorial hidden Markov representation**
Presenter: **Philippe Charlot**, GREQAM & Aix-Marseille University, France

The goal of this paper is to present a new multivariate GARCH model with time-varying conditional correlations. Our approach is a Markov-Switching version of the Engle's DCC GARCH. However, we suggest a slightly different approach than the classical Markov-Switching. Indeed, our specification is based on the Ghahramani and Jordan's Factorial Hidden Markov model. This model generalize the previous approach by representing the hidden state in a factored form. Each hidden state is factored into multiple hidden state variables evolving in parallel. Each chain has similar dynamics to a basic hidden Markov model. It makes possible to build a Markov-switching set up for correlation dynamic where each parameters are linking to a hidden Markov chain and allows them to switch independently from others. This Factorial Markov-switching DCC is applied to exchange rate data. We also propose an asymmetric extension. This extended model is applied to equities index returns. Results show that FHM-DCC can have an interesting explanatory power by providing more flexibility in the correlation process.

C361: **Estimation of continuous time Markov switching models**

Presenter: **Joern Sass**, Austrian Academy of Sciences, Austria
Co-authors: Sylvia Fruehwirth-Schnatter, M. Hahn

We consider a multidimensional, continuous time model where the observation process is a diffusion with drift and volatility coefficients being modeled as continuous time, finite state Markov processes with a common state process. I.e., a linear, non-autoregressive Markov switching model is considered. The interest is in the econometric estimation of the states for drift and volatility and the rate matrix of the underlying (unobservable) Markov process as well as the choice of the number of different states. For a continuous time model, where observations are available only at discrete times, a method of moments-type approach may be used, but it requires a large number of observations. As an alternative suitable for medium sizes of observations, taking a Bayesian approach we develop a MCMC method. We discuss problems arising if discrete time methods are applied to estimate the continuous time model, like discretization error and embedding-problem. Employing different methods for parameter estimation, we find some advantages of the MCMC approach. For illustration, we analyse daily stock index quotes from Argentina, Brazil, Mexico, and the US from 1998 to 2008. The estimation results allow for a good economic interpretation.

C100: **Risk analysis and mixture modeling**

Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria
Co-authors: Christian Haefke

The broad objective this paper is to extend and to apply mixture models and switching GARCH models to problems in portfolio optimization and in risk management. Following literature, closed form solutions for the probability distribution function and quantiles for the student t-distribution can be derived in closed form. This article applies and even extends this methodology to skew student-t densities. A further goal is to combine classical mixture models with mixtures of ARCH or GARCH type models Our model extends existing literature in several ways: First, we use t-distributed innovations, secondly, the distribution of the GARCH volatility term and the noise term are allowed to switch between different states. In addition we model correlation by means of copulas. Such a setting also provides an important extension from an economic point of view: The believes of practitioners that asset returns show different behavior in good or bad times can be modeled and even tested by such a setting.

CS36 Room 6 ROBUST METHODS IN ECONOMETRICS Chair: Christophe Croux

C139: **Accurate and robust tests for indirect inference**

Presenter: **Veronika Czellar**, HEC Paris, France
Co-authors: Elvezio Ronchetti

Accurate tests for indirect inference are proposed. Under the null hypothesis the new tests are asymptotically χ^2–distributed with a relative error of order $n^{(-1)}$. They exhibit a much better finite sample accuracy than classical tests for indirect inference which have the same asymptotic distribution but with an absolute error of order $n^{(-1/2)}$. Robust versions of the tests are also provided. Finally, we illustrate their accuracy in various models, including nonlinear regression, Poisson regression with overdispersion, and diffusion models.

C152: **Testing the contribution of jumps to total price variance: A review and a new test**

Presenter: **Kris Boudt**, K.U. Leuven-Lessius University College, Belgium
Co-authors: Thomas Ghys, Sebastien Laurent

Recent research has shown that separate estimates of the continuous and jump component of quadratic variation lead to more accurate volatility forecasts and estimates of financial betas. We present a new approach for disentangling both components based on the Realized Outlyingness Weighted Variance (ROWVar). The ROWVar has a higher efficiency than the Realized Bipower Variation (RBPVar) and a higher finite sample robustness to jumps. We propose jump-robust estimators for the variance of the ROWVar and derive their asymptotic variance. The ROWVar and the estimator of its variance are combined to greatly enhance the power of testing the jump contribution in total price variance with respect to tests based on the RBPVar. In addition, we show that accounting for the intra day periodic component in volatility improves further the performance of these tests. We compare our proposal with other recently proposed alternatives for the RBPVar. The paper concludes with an illustration of the proposed jump detection method for high-frequency stock prices and exchange rates. In both markets, tests based on the ROWVar detect much more jumps than the RBPVar and lead to more accurate volatility forecasts.

C163: **A robust approach to the analysis and forecasting of electricity prices and volatility**
Presenter: **Dennis Tuerk**, FUNDP, Belgium
Co-authors: Kris Boudt, Sebastien Laurent

During the last years several approaches for electricity price modeling under jumps have been proposed. They yield either unsatisfying results or are quite complicated to implement. Recently, non-parametric techniques for daily volatility forecasting using high-frequency data have been proposed. The price changes are centered around an estimate of the price drift and functionals of the centered high-frequency price changes are aggregated into estimates of the continuous and/or jump part of daily price variability. Unfortunately the need for accountancy for the presence of jumps when estimating the price drift was ignored. Furthermore, no satisfying solution was proposed for the fact that the realized volatility estimator must be able to cope with multiple consecutive jumps. We propose robust estimation techniques to satisfy the before mentioned requirements. Preliminary forecasting results demonstrate the superiority to other modelling approaches.

C205: **Spot variance estimation and its application to high frequency jump testing**
Presenter: **Pawel Janus**, VU University Amsterdam, Netherlands
Co-authors: Charles Bos, Siem Jan Koopman

We provide an empirical framework for estimating the spot variance path in the simultaneous presence of intra day periodicity, jumps and microstructure noise. The estimation is done with discretely observed high frequency price data in a model-free manner. The underlying spot variance consists of a deterministic and a stochastic component. Extraction of the intra day periodic component is based on a likelihood method that is robust to possible outliers. The stochastic part is estimated with the pre-averaged bipower variation measure, hence estimates are robust to jumps and microstructure effects. This measure allows also for the negative correlation in the return-variance innovations to accommodate for the leverage effect. We find no noticeable impact of such correlation on the spot variance path estimation. As application, a robust two-step estimating method for the spot variance path provides means to extend an existing high frequency jump test statistics. The extended statistics enables to detect discontinuities on top of local variation due to price continuous component and microstructure noise. Timing of jumps as well as distributional characterizations of jump series are revealed. The procedures are explored through Monte Carlo simulations. An empirical illustration is based on the intra day EUR/USD rate.

CS61 Room 4 COMPUTATIONAL ECONOMETRICS: ESTIMATION AND TESTING Chair: Achim Zeileis

C246: **Sparse principal component regression**
Presenter: **Joseph Ryan Lansangan**, University of the Philippines, Philippines
Co-authors: Erniel Barrios

Modeling of complex systems is usually confronted with high dimensional independent variables. Econometric models are usually built using time series data that often exhibit nonstationarity due to the impact of some policies and other economic forces. Both cases are usually confronted with the multicollinearity problem resulting to unstable least squares estimates of the regression coefficients in a linear model. Principal component regression often provides solution, but in cases where the regressors are nonstationary or the dimension exceeds the sample size, principal components may yield simple averaging of the regressors and the resulting model is difficult to interpret. A sparsity constraint is added to the least squares criterion to induce the sparsity needed for the components to reflect the relative importance of each regressors in a sparse principal component regression (SPCR) model. Simulated and real data are used to illustrate and assess the performance of the method. SPCR in many cases leads to better estimation and prediction. SPCR can be used in modeling high dimensional data and as an intervention strategy in regression with nonstationary time series data.

C283: **A test for m order polynomial versus** $(m+1)$ **convex regression**
Presenter: **Ioannis Demetriou**, University of Athens, Greece

Given n measurements of a smooth function that include random errors, a test is presented for the appropriateness of the least squares polynomial regression model. The null hypothesis is that the regression function is a polynomial of order m and the alternative one that it has nonnegative derivative of order $m+1$ ($m+1$ convexity). In practice, the underlying function in the alternative hypothesis is obtained by a strictly convex quadratic programming calculation that is defined by the regression model subject to nonnegative divided differences of order $m+1$. The following special cases provide useful descriptions. When $m=0$, this test reduces to testing a constant against a monotone regression, a problem that has raised very many publications and applications in the last sixty years. When $m=1$, the test reduces to testing a line again a convex regression (2-convex regression, i.e. a piecewise linear fit with non-decreasing slopes). In the general case, $m>1$, when the errors are independent and normal, it is shown that the exact distribution for a likelihood ratio test statistic is a mixture of beta random variables. Some numerical results illustrate the method of calculation and the efficiency of the test.

C243: **Nonsmooth optimization using classification and regression trees**

Presenter: **Blair Robertson**, University of Canterbury, New Zealand
Co-authors: Chris Price, Marco Reale

Nonsmooth local optimization problems occur in many fields, including engineering, mathematics, and economics. In economics, nonsmooth problems can be found in finance, mathematical economics and production theory. Examples include, nonsmooth utility maximization and exact penalty functions. Here we present a random search algorithm to solve such problems. The method is set up to solve nonsmooth or discontinuous minimization problems in an hyper-rectangular optimization region S. Since discontinuous functions are considered, the optimization boundary can be dealt with directly via a discontinuous barrier function and hence non-trivial regions can be considered. The algorithm alternates between partition and sampling phases. At each partition phase S is divided into a set of high and low subregions, with respect to objective function value, using classification and regression trees. A new batch of points is then distributed mainly or totally into the hyper-rectangular subregions classified as low. Using both the existing and new batch of points, a new partition is then formed. The method is shown to converge to an extreme point of an objective function with probability one. Numerical examples are also presented to demonstrate the effectiveness of the method.

C305: **Model-Based regression trees in economics and the social sciences**

Presenter: **Achim Zeileis**, WU Wirtschaftsuniversitat Wien, Austria

Regression models are the workhorse for empirical analyses in economics and the social sciences. For a wide variety of standard analysis problems, there are useful specifications of regression models, validated by theoretical considerations and prior successful empirical studies. However, in non-standard problems or in situations where data on additional variables is available, a useful specification of a regression model involving all variables of interest might not be available. Here, we explore how recursive partitioning techniques can be used in such situations for modeling the relationship between the dependent variable and the available regressors. We show how different models (linear regression via OLS or WLS and the Bradley-Terry model) can be embedded into a common framework of model-based recursive partitioning. The resulting regression trees are grown by recursively applying techniques for testing and dating structural changes in parametric models. They are compared to classical modeling approaches in three empirical applications: (1) The demand for economic journals is investigated. (2) The impact of professors' beauty on their class evaluations is assessed. (3) Differences in rating scales are captured for students' choice of a university in exchange programmes.

CS63 Room 1 STOCHASTIC VOLATILITY MODELS Chair: Yasuhiro Omori

C217: **Online estimation of time-varying volatility and co-volatility for tick-by-tick data: a Bayesian approach**

Presenter: **Jan C. Neddermeyer**, University of Heidelberg, Germany
Co-authors: Rainer Dahlhaus

For high-frequency trading the online estimation of time-varying (co)-volatilities based on tick-by-tick data is an important task. As a result of the non-synchronous trading and the presence of market microstructure noise this is a challenging problem. In the present article, we develop a novel technique for online estimation of time-varying covariance matrices. Our algorithm updates the covariance estimate immediately after the occurrence of a new transaction. We propose a market microstructure model that considers financial transaction data as observations of a latent efficient price process which are corrupted by (generalized) rounding noise. This model reproduces the major stylized facts of high-frequency data such as the price discreteness and the first-order negative autocorrelation of the returns. It takes the form of a nonlinear state-space model with non-synchronous observations. An original sequential Monte Carlo algorithm is designed that allows the approximation of the filtering distributions of the efficient price. For the estimation of the spot covariance matrix we propose a sequential variant of the EM algorithm. The practical usefulness of our technique is verified through Monte Carlo simulations and through an application to stock transaction data. As an important empirical result we obtain that the correlations of high-frequency stock returns vary significantly over the trading day.

C258: **Estimation of dynamics for Income inequality by stochastic volatility model**

Presenter: **Haruhisa Nishino**, Chiba University, Japan
Co-authors: Takashi Oga, Kazuhiko Kakamu

Income inequality is a great issue also in the Japanese economy likewise the world economy, and the Gini coefficient is a famous measure of economic inequalities. While many previous works calculate a Gini coefficient as a descriptive statistic or estimate the Lorenz curve by nonparametric methods, we estimate Gini coefficients using simple parametric models in order to investigate changes of inequalities in the paper. Since microdata are difficult to access in Japan and also in other developing countries, we use grouped data to estimate Gini coefficients. Assuming simple parametric models including Pareto and Lognormal, we represent Gini coefficients as functions of estimated parameters. It is therefore easy to follow dynamics of Gini coefficients. The use of

asymptotic theory of selected order statistics enables us to estimate the parameters of the parametric models based on grouped data. In order to extend the parametric models to dynamic models, we incorporate them with SV (stochastic volatility) models. We estimate the SV models by MCMC methods, and exploit a model comparison to choose a best model. Finally we apply the method for Japanese data (Family Income and Expenditure Survey by Ministry of Internal Affairs and Communications, Statistics Bureau, Japan).

C227: **Bayesian forecasting using an extended Nelson-Siegel model**

Presenter: **Fuyu Yang**, Humboldt-Universitat zu Berlin, Germany
Co-authors: Nikolaus Hautsch

Central banks control the short-term interest rates to implement monetary policy. In order to model the term structure of interest rates, we re-visit an extended Nelson-Siegel model (NS-SV model), in which all cross–sectional yields are summarized with a few common factors. This NS-SV model could not only capture the shifts in the yield curve by modelling these common factors, but also model the volatility dynamics in these factors using stochastic volatilities. However, the estimations for this appealing NS-SV model was rather inefficient. Here the severe estimation challenge is solved via Markov chain Monte Carlo method. We propose an efficient algorithm to make simulation-based inference in the NS-SV model. With applications to a monthly zero–coupon yields data set, we found the NS-SV fits the data well. Forecast capacities of the NS-SV model is evaluated via out–of–sample forecast exercises.

C060: **Stochastic volatility model with asymmetric heavy-tailed error using GH skew Student's t distribution**

Presenter: **Yasuhiro Omori**, University of Tokyo, Japan
Co-authors: Jouchi Nakajima

Bayesian analysis of a stochastic volatility (SV) model with General Hyperbolic (GH) skew Student's t error distribution is described to consider asymmetric heavy-tailed distributions of asset returns. The efficient Markov chain Monte Carlo estimation methods are developed using the normal variance-mean mixture representation of the GH skew Student's t distribution where the mixing distribution is the gamma distribution. The proposed methods are illustrated using simulated data and daily stock returns, and the model comparison is conducted based on the marginal likelihood in the empirical study. The strong evidence of the asymmetry is found in the distribution of the stock returns.

Friday 30.10.2009 16:40-18:40 Parallel Session I

CS04 Room 3 COMPUTATIONAL PANEL ECONOMETRICS Chair: Jan Kiviet

C078: **Cointegration versus spurious regression and heterogeneity in large panels**

Presenter: **Lorenzo Trapani**, Cass Business School, UK

An estimation and testing framework to identify the source(s) of spuriousness in a large nonstationary panel is provided. This can be determined by two non mutually exclusive causes: pooling units neglecting the presence of heterogeneity and genuine presence of I(1) errors in some of the units. Two tests are proposed that complement a test for the null of cointegration: one test for the null of homogeneity (and thus presence of spuriousness due to some of the units being genuinely spurious regressions) and one for the null of genuine cointegration in all units of the panel (and thus spuriousness arising only from neglected heterogeneity). The results are derived using a linear combination of two estimators (one consistent, one inconsistent) for the variance of the estimated pooled parameter. Two consistent estimators for the degree of heterogeneity and for the fraction of spurious regressions are also provided.

C090: **Structural breaks and unit root tests for short panels**

Presenter: **Elias Tzavalis**, Athens University of Economics & Business, Greece

In this paper we suggest panel data unit root tests which allow for a potential structural break in the individual effects and/or the trends of each series of the panel, assuming that the time-dimension of the panel, T, is fixed. The proposed test statistics consider for the case that the break point is known and for the case that it is unknown. For the first case, the paper calculates the power of the test statistic analytically. The paper calculates the limiting distribution of the test statistic for the case that the break point is unknown based on the correlation matrix of the potential break points in the sample which is derived in closed form solution. Monte Carlo analysis conducted by the paper suggests that the test statistics have size which is very close to the nominal five percent level and power which is analogues to that of other panel data unit root tests with fixed-T test statistics, which do not allow for a structural break. Finally, an empirical application of the paper show how to apply the tests in practice to test the income convergence hypothesis.

C180: **A Monte Carlo estimator for simulable models**

Presenter: **Michael Creel**, Universitat Autonoma de Barcelona, Spain

For a model with parameter ϕ, and an estimator $\tilde{\theta}$ of the parameter of an auxiliary model, Monte Carlo methods can be used to generate pairs (ϕ, $\tilde{\theta}$) for a randomly chosen point ϕ in the parameter space. Nonparametric methods may be use to fit the expectation of ϕ conditional on $\tilde{\theta} = a$, using the Monte Carlo data. It is proposed to use this fitted function evaluated at the $\hat{\theta}$ that is obtained using the real sample data, as an estimator of ϕ. Under certain conditions, this is a consistent and asymptotically normally distributed estimator. Monte Carlo results for dynamic panel data and other applications show that this estimator can have very attractive small sample properties. Confidence intervals can be constructed using the quantiles of the ϕ for which $\tilde{\theta}$ is close to $\hat{\theta}$. Such confidence intervals have very accurate coverage.

C063: **Heuristic optimization methods for dynamic panel data model selection**

Presenter: **Ivan Savin**, Justus-Liebig University, Germany

Co-authors: Peter Winker

Innovations, be they radical new products or technology improvements in the sense of cost efficiency, are widely recognized as a key factor of economic growth. To identify the main factors triggering innovative activities is a main concern for economic policy. This is also an important research area in economic theory and empirical examination. As the evidence on the effectiveness of different stimulating instruments is mixed and the number of hypotheses, which may be tested, is large, the process of model selection becomes a crucial factor in the empirical implementation. The problem is complicated by the fact that we need to take an unobserved heterogeneity and possible endogeneity of regressors into account. A new efficient solution on this problem is suggested by applying threshold accepting and genetic algorithm heuristics, which exploit the inherent discrete nature of the problem. The selection criteria are based on the Akaike, Bayesian and Hannan-Quinn information criteria (the former is not consistent) and on the Sargan test of the overidentifying restrictions. The method is applied on the example of Russian economy constructing a log-linear dynamic panel data model. To illustrate the performance of the method, we also report the results of Monte-Carlo simulations.

C251: **Maximum likelihood estimation of random coefficient panel data models**
Presenter: **Mehdi Hosseinkouchack**, Goethe University Frankfurt, Germany
Co-authors: Michael Binder, Florian Hoffmann

While there are compelling reasons to account for cross-sectional heterogeneity of slope coefficients in panel data models, applied work still regularly ignores to capture such heterogeneity. In the context of random coefficient panel data models, this appears in large parts due to the fact that classical estimation procedures are difficult to implement in practice. In this paper, we propose a simulated annealing algorithm to maximize the likelihood function of random coefficient panel data models. Tayloring the simulated annealing algorithm explicitly to random coefficients panel data models, we proceed to compare the performance of our global maximization approach with that of approaches based on standard local optimization techniques. Our results suggest that simulated annealing - in contrast to standard local optimization techniques - renders estimation of Swamy random coefficient panel data models practical. The structure of the Swamy random coefficient panel data model further allows us to compare the simulated annealing based maximum likelihood estimates with a closed-form Generalized Least Squares (GLS) estimator. We find that the simulated annealing based maximum likelihood estimator performs reasonably well and should be a serious alternative to Bayesian estimators in more complex random coefficient panel data models for which the GLS estimator is not feasible.

CS08 Room 8 ECONOMETRIC METHODS IN DERIVATIVES APPLICATIONS Chair: Panayiotis Andreou

C009: **An alternative model to forecast default based on Black-Scholes-Merton model and a liquidity proxy**
Presenter: **Dionysia Dionysiou**, Aston University, UK
Co-authors: Lenos Trigeorgis, Andreas Charitou, Neophytos Lambertides

Building upon the theoretical Black-Scholes-Merton model, we develop an alternative model to forecast default. Without solving the required nonlinear equations, we deviate from prior approaches by estimating volatility in a simpler manner. We consider the probability of intermediate involuntary default before debt-maturity which we capture via a liquidity proxy. Finally, we use a weighted average life of total debt as time-to-option-maturity. Cox proportional hazard models and several approaches that test the model predictive ability suggest that our alternatives indicate sufficient statistic and ability to forecast default.

C020: **Multistage product development with value-enhancing and pre-emptive options**
Presenter: **Nicos Koussis**, Frederick University, Cyprus
Co-authors: Spiros Martzoukos, Lenos Trigeorgis

We provide a real options framework for the analysis of product development that incorporates R&D, product attribute-enhancing actions with uncertain outcome and pre-emptive options. We derive two-stage analytic formulas and propose a multi-period general solution using a numerical lattice approach. Our analysis reveals that exploration actions are more important when the project is out or at-the-money (near zero NPV) and less important for high project values. In a multi-stage setting, exploration actions are important even for in-the-money projects when follow-on actions exist that can enhance the expected value of the project. With path-dependency early actions are more valuable since they enhance the impact or reduce the cost of subsequent actions. Pre-emptive controls affecting rare event (jump) frequency or innovations that introduce positive jumps are more valuable for firms with higher frequency of competitive threats involving low volatility.

C022: **A new method of employing the principle of maximum entropy to retrieve the risk neutral density**
Presenter: **Leonidas Rompolis**, University of Cyprus, Cyprus

This paper suggests a new method of implementing the principle of maximum entropy to retrieve the risk neutral density of future stock, or any other asset, returns from European call and put prices. Instead of options prices used by previous studies, the method maximizes the entropy measure subject to values of the risk neutral moments. These moments can be retrieved from market option prices at each point of time, in a first step. Compared to other existing methods of retrieving the risk neutral density based on the principle of maximum entropy, the benefits of the method that the paper suggests is the use of all the available information provided by the market more efficiently. To evaluate the performance of the suggested method, the paper compares it to other risk neutral density estimation techniques based on a number of simulation and empirical exercises.

C075: **The forward premium puzzle and unobserved factors day by day**
Presenter: **Kerstin Bernoth**, DIW Berlin, Germany
Co-authors: Jurgen von Hagen, Casper de Vries

The forward premium puzzle (FPP) is the negative correlation between the forward premium and the realized exchange rate return, which has been found in numerous empirical studies based on foreign currency forwards. This paper contributes by using futures data instead of forwards to complete the maturity spectrum at the (multi-) day level. We find that the correlation is positive for very short maturities, and has a negative tendency that slowly becomes stronger as the number of days to maturity is increased

to the monthly level. Using futures data allows us to control for the influence of an unobserved factor present in each contract. Once we do this, we find that the coefficients on the forward premium hover around one. Finally, we show that the latent factor is related to conventional proxies of currency risk and is significantly affected by domestic monetary variables.

C054: **Options pricing via statistical learning techniques: The support vector regression approach**

Presenter: **Panayiotis Andreou**, Durham University, UK
Co-authors: Chris Charalambous, Spiros Martzoukos

In this paper we investigate the performance of epsilon-insensitive Support Vector Regression and Least Squares Support Vector Regression for pricing call options on the SP500 index. Support Vector Regression is a novel nonparametric method that has been developed in the context of statistical learning theory and until now it has been practically neglected in financial econometric applications. This new method is compared with parametric options pricing models using standard implied parameters and parameters derived via deterministic volatility functions. The out-of-sample pricing results indicate that the Least Squares Support Vector Regression models perform better than the epsilon-insensitive Support Vector Regression ones. Moreover, this type of nonparametric approach is also overall best when compared with the widely used Black and Scholes model.

CS21 Room 4 ENERGY AND FINANCIAL ECONOMETRICS Chair: Frederique Bec

C319: **Contagion in electricity markets**

Presenter: **Davide Ciferri**, University of Perugia, Italy
Co-authors: Carlo Andrea Bollino, Paolo Polinori

The existence of contagion effects in electricity markets is investigated. The concept of contagion has been developed for high frequency financial markets. The paper presents a canonical, econometric model of contagion and investigates the conditions under which contagion can be distinguished from mere interdependence. The empirical analysis is based on different regional markets in the Italian Power Exchange (IPX). The previous literature has focused much on the typical characteristics of electricity prices such as high volatility and very large, or extreme, price changes but, so far, it has ignored the question whether contagion exists among different electricity markets. The analysis and identification of contagion requires that each individual market equations contains market specific regressors, consequently we have to involve market specific variables in structural equations in order to correctly specify the model. Consistency is obtained by including regional market specific fundamentals and using instrumental variables (IV) estimations. The most important conclusions of this paper are that contagion can be identified separately from interdependence and that effects are asymmetric in the IPX market and we find that price crisis occur when dynamic patterns are spiking suddenly downward and not upward.

C089: **The power of weather**

Presenter: **Chen Zhou**, De Nederlandsche Bank, Netherlands
Co-authors: Francesco Ravazzolo, Christian Huurman

This paper examines the predictive power of weather for electricity prices in day-ahead markets in real time. We find that next-day weather forecasts improve the forecast accuracy of Scandinavian day-ahead electricity prices substantially in terms of point forecasts, suggesting that weather forecasts can price the weather premium. This improvement strengthens the confidence in the forecasting model, which results in high center-mass predictive densities. In density forecast, such a predictive density may not accommodate forecasting uncertainty well. Our density forecast analysis confirms this intuition by showing that incorporating weather forecasts in density forecasting does not deliver better density forecast performances.

C188: **Density forecasts of crude-oil prices using option-implied and ARCH-type models**

Presenter: **Esben Hoeg**, Aalborg University, Denmark
Co-authors: Leonidas Tsiaras

The predictive accuracy of competing crude-oil price forecast densities is investigated for the period 1994-2006. Moving beyond standard ARCH models that rely exclusively on past returns, we examine the benefits of utilizing the forward-looking information that is embedded in the prices of derivative contracts. Risk-neutral densities, obtained from panels of crude-oil option prices, are adjusted to reflect real-world risks using either a parametric or a non-parametric calibration approach. The relative performance of the models is evaluated for the entire support of the density, as well as for regions and intervals that are of special interest for the economic agent. We find that non-parametric adjustments of risk-neutral density forecasts perform significantly better than their parametric counterparts. Goodness-of-fit tests and out-of-sample likelihood comparisons favor forecast densities obtained by option prices and non-parametric calibration methods over those constructed using historical returns and simulated ARCH processes.

C286: **Bayesian analysis of time-varying integration in energy markets**
Presenter: **Arco Van Oord**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide, Herman Van Dijk

We investigate the degree of market integration in global energy commodity prices, including crude oil, natural gas and coal. Our specific focus is on the speed of reversion towards the equilibrium relations; increased market integration should be reflected by shorter half-life times of the disequilibrium. We perform a Bayesian analysis within a flexible cointegration model with time-varying strength of the adjustments towards equilibrium conditions. Using simulation based inference we recursively estimate the reversion to cointegration relations between pairs of energy prices. Thereby we provide insight into the exact distribution of the time-varying half-life times without resorting to asymptotic approximations, and assess the evidence in favor of increased integration.

C027: **Term structure and cyclicality of Value-at-Risk: consequences for the solvency capital requirement**
Presenter: **Frederique Bec**, University of Cergy-Pontoise, France
Co-authors: Christian Gollier

This paper explores empirically the link between French equities returns Value-at-Risk (VaR) and the state of financial markets cycle. The econometric analysis is based on a simple vector autoregression setup. Using quarterly data from 1970Q4 to 2008Q4, it turns out that the k-year VaR of French equities is strongly dependent on the cycle phase: the expected losses as measured by the VaR are smaller in recession times than expansion periods. These results strongly suggest that the European rules regarding the solvency capital requirements for insurance companies should adapt to the state of the financial market's cycle.

CS22 Room 2 FINANCIAL ECONOMETRICS: FORECASTING AND DYNAMICS Chair: Giampiero Gallo

C101: **Prediction of stock returns with nonparametrically generated bond yields**
Presenter: **Michael Scholz**, University of Gottingen, Germany
Co-authors: Stefan Sperlich

It is widely documented that expected returns in financial markets vary over time, and that long-horizon returns contain a significant predictable component. In particular, the dividend-price ratio has proven to be a good predictor of future stock returns. We extend in the present work, the fully nonparametric two-dimensional model based on the dividend-price ratio and the lagged excess stock return using in addition also the bond yield of the same year as regressor. The reasoning is quite simple; it is often discussed that the same year's bond yield is basically the prediction error. The problem which obviously occurs is that also the actual bond yields are unknown. To circumvent this handicap we predict them in a prior step, using again linear kernel regression and cross validation. One may think of the inclusion of predicted bond yields when predicting stock returns nonparametrically as a kind of dimension reduction; importing more structure as a proper way to circumvent the curse of dimensionality. The additional use of a simple dummy variable which controls over certain time periods for structural breaks in the market and the knowledge about structure improve the prediction of the stock returns.

C149: **Forecasting performance of nonlinear models for intra day stock returns**
Presenter: **Juan Carlos Reboredo**, Universidade de Santiago de Compostela, Spain
Co-authors: Jose Maria Matias

We studied the predictability of intra day stock market returns using both linear and non-linear time series models. In order to test their ability to forecast intra day SP500 intra day returns, we compared simple autoregressive and random walk linear models with a range of nonlinear models that included smooth transition, Markov switching, artificial neural network, nonparametric kernel regression and support vector machine models for horizons of 5, 10, 20, 30 and 60 minutes. Forecasting performance was evaluated on the basis of both statistical criteria and economic criteria. As statistical criteria, we used in-sample and out-of-sample one-step-ahead forecasts using goodness of forecast measures, proportion of times the signs of returns were correctly predicted, the Pesaran-Timmermann directional prediction test, and the Diebold-Mariano test of equal predictability. The economic criteria were based on model predictions and transaction cost underpinning a simple trading rule. The empirical results indicate that nonlinear models outperformed linear models on the basis of both criteria. Specifically, although weak-form market efficiency was achieved by around ten minutes after return serial correlation receded and returns behaved as a random walk, return predictability still persisted for up to sixty minutes according to nonlinear models. More flexible nonlinear models such as support vector machines and the multi layer perceptron artificial neural network did not clearly outperform the Markov switching, smooth transition and nonparametric kernel regression models. On economic grounds, trading rules based on Markov switching and support vector machine models are the most profitable for short-term horizon returns, whereas the smooth transition and multi layer perceptron models behave better for longer periods, even though profitability decreases as time elapses.

C084: **Forecasting stock market returns along financial cycles**
Presenter: **Marco Willner**, Goethe University Frankfurt, Germany

Given the recent slump in global stock markets, we revisit the evidence on the forecastability of the US equity premium with one critical question in mind: How do various predictors perform in upswing and downswing periods? We propose a dating scheme which divides the financial cycles into three phases (downturn, recovery and growth). The forecasting models are evaluated by their ability of anticipating the direction of future price movements. We show that models exhibiting state-dependent or time-varying coefficients tend to improve the forecasting performance, especially in downturn and recovery periods. The performance of the real-time average turns out to be poor whereas even simple regression models do not uniformly underperform naive models as it is commonly found in the literature.

C007: **Modelling and estimating the forward price curve in the energy market**
Presenter: **Boda Kang**, University of Technology, Sydney, Australia
Co-authors: Carl Chiarella, Les Clewlow

By enhancing a multi factor forward price framework which is consistent not only with the market observable forward price curve but also the volatilities and correlations of forward prices, we propose a two factor stochastic volatility model for the evolution of the gas forward curve. Both volatility functions contain a common seasonality adjustment term, but one volatility function is positive and declining with increasing maturity to a constant; another volatility function captures the overall tilting of the forward curve where the short maturity contracts move in the opposite direction to the longer maturity contracts. We allow the parameters of the volatility functions to take different values in different states of the world. The dynamics of the states of the world, for example an on-peak or off-peak time for gas are represented by a finite state Markov Chain. We propose and implement both off line, namely the Markov Chain Monte Carlo (MCMC) and online, namely the particle filter approaches to estimate the parameters of the above forward curve model. Applications to simulated and market gas forward data indicate that the proposed algorithms are able to accommodate more general features, such as regime switching of forward volatility functions.

C263: **A moment based approach to the combination of volatility forecasts**
Presenter: **Alessandra Amendola**, University of Salerno, Italy
Co-authors: Giuseppe Storti

The availability of accurate volatility forecasts is a prerequisite for many financial applications such as risk measurement or portfolio optimization. The wide range of methods available makes model uncertainty a critical problem for anyone interested in volatility prediction. Combining volatility forecasts obtained from different models would appear as a natural solution. However, the application of classical forecast combination techniques is not straightforward. We show that the GMM can be used to define an effective procedure for the combination of volatility forecasts. One of the strengths of the approach is its generality in the sense that it is not specific to a given class of models. Also, it can be applied to combine univariate as well as multivariate volatility forecasts. In multivariate applications a relevant problem is related to the high number of moment conditions. In order to overcome this difficulty and deal with large dimensional systems, a two-step estimation strategy is adopted. The finite sample properties of the GMM estimator of the combination weights are assessed by means of a Monte Carlo simulation. Finally, some empirical evidence on real financial data is presented.

CS25 Room 9 FINANCIAL MODELLING AND APPLICATIONS Chair: Hyunchul Lee

C178: **The information contained in the trades associated with the exercise of executive stock options**
Presenter: **Kyriacos Kyriacou**, Brunel University, UK
Co-authors: Bryan Mase, Kul Luintel

This paper investigates whether executives use private information in the trading decisions associated with the exercise of their executive stock options. We find a marked and significant difference in subsequent performance between exercises categorized by the proportion sold at the time of exercise. The difference in post-exercise performance persists after controlling for factors that include option moneyness, the previous stock return and the value of the exercise. Further, the disparity in US and UK executives' trading behaviour at option exercise could be related to the extent to which executive remuneration is linked to shareholder wealth, highlighting the difficulties associated with designing effective remuneration packages for executives.

C261: **Affine term structure model with ARFIMA factors**
Presenter: **Adam Golinski**, Imperial College London, UK
Co-authors: Paolo Zaffaroni

We present a discrete time, essentially affine, two long memory Gaussian factor term structure model. We derive an asset pricing model through the stochastic discount factor with underlying ARFIMA processes, which is a generalization of autoregressive factor models. The model is based on economic theory, so the factors have a clear economic interpretation as real short rate and

expected inflation. We estimate the model in three versions: with two long memory factors, with two short memory factors and with long memory only in expected inflation factor, and find that extension of the model from short memory to long memory factors gives a substantial improvement in the fit of the model and forecast errors. Specifically, it seems crucial to model the expected inflation factor as a long memory process, while we don't find evidence of high persistence in the short rate dynamics. Based on the model estimates we extract the processes of real interest rates, real and nominal risk term premia, and inflation premia for different maturities. All these quantities are time varying and we examine their properties.

C086: **Tests of recent advances in extracting information from option prices**
Presenter: **Jerome Healy**, University of East Anglia, UK

A large literature exists on techniques for extracting probability distributions for future asset prices (returns) from option prices, but no definitive method has been developed. The parametric *mixture of normals*, and non-parametric *smoothed implied volatility* methods, are the most popular approaches amongst practitioners. These are subject to estimation errors due to discretization, truncation, and noise however. Recently, several authors have derived *model free* analytical formulae for computing the moments of the risk neutral density (RND) directly from option prices without first estimating the full density. The accuracy of these analytical formulae is studied here for the first time. The Black-Scholes formula is used to generate option prices, and error curves for the first 4 moments of the RND are computed from these. The resulting error curves are convex, and the errors differ in magnitude across the domains of the input variables. This has implications for empirical applications of these analytical formulae.

C032: **The finance of fraud-computational model of unauthorized trading repression**
Presenter: **Duc Phamhi**, ECE Graduate School of Engineering, France

Today's operational risk most currently used model in banks, the Loss Distribution Approach and its variants, are without time dimension. They cannot represent causality and therefore no control action can be shown to take effect. In the first part of this paper, changes in the LDA simulation mode toward a temporal process, allows description of correlation and causality. Other features like Internal Control Factors are added to let budgetary actions take effect through risk reduction both in severity and frequency. As a consequence the model can then be shown to follow a simplified Hamilton-Jacobi-Bellman equation. In the second part, discussions on how to solve this HJB equation leads to envisage 2 different approaches: linearization and computational exploration, on one hand, transformation of the problem into an adaptive learning problem, on the other hand. For the computational approach, reductions in complexity in the simulation process is obtained via the Single Loss Approximation and use of Extreme Value Theory. In the third part, computational results are displayed and discussed. We observe that there is an optimum budget in risk reduction with regards to each type of control represented in our formalism. Finally, some further directions for investigation are suggested.

C335: **Econometric evidence from the continuous double auction**
Presenter: **Martin Smid**, UTIA AS CR, Czech Republic

A general model of continuous double auctions is considered. Earlier work proposed econometric methods for validation and estimation of the model based only on the history of the best quotes (no order book data). This method is applied to US tick-by-tick stock data. We demonstrate that the model is able to reduce unexplained variability of the prices.

C044: **The time-varying European government bond markets integration and fiscal policy: The role of EMU**
Presenter: **Hyunchul Lee**, University of Essex, UK
Co-authors: Jerry Coakley, Andrea Cipollini

Using panel data methods with time-varying realised correlations and betas, this paper examines the effect of European Monetary Union (EMU) on the dynamic integration of 13 European long term government bond markets over the period January 1992 to December 2006. We find that there has been a clear regime shift in integration among EU government bond markets following the introduction of the Euro. Post EMU, higher gross deficits has lead to increased integration of the EU bond markets suggesting default risk premium comovements between EU bond markets and the German market for pricing EU bonds. However, there is no evidence that gross debt has a significant relation with bond market convergence pre and post EMU implying no credit risk premium for EU bond pricing. A possible explanation for this might be investors' implied expectation of a bail-out by the EU for high debt member states.

CS33 Room 6 FINANCIAL AND ECONOMIC VOLATILITY Chair: Gianluca Moretti

C131: **Potential PCA interpretation problems for volatility smile dynamics**
Presenter: **Dimitri Reiswich**, Frankfurt School of Finance & Management, Germany
Co-authors: Robert Tompkins

The typical factor loadings found in PCA analysis for financial markets are commonly interpreted as a level, skew, twist and curvature effect. Lord and Pelsser question whether these effects are an artefact resulting from a special structure of the covariance or correlation matrix. They show that there are some special matrix classes, which automatically lead to a prescribed change of sign pattern of the eigenvectors. In particular, PCA analysis on a covariance or correlation matrix which belongs to the class of oscillatory matrices will always show n-1 changes of sign in the n-th eigenvector. This is also the case in most PCA results and raises the question whether the observed effects have a valid economic interpretation. We extend this line of research by considering an alternative matrix structure which is consistent with foreign exchange option markets. For this matrix structure, PCA effects which are interpreted as shift, skew and curvature can be generated from unstructured random processes. Furthermore, we find that even if a structured system exists, PCA may not be able to distinguish between these three effects. The contribution of the factors explaining the variance in the original system are incorrect.

C164: **The effects of Interest rate movements on assets' conditional second moments**
Presenter: **Alessandro Palandri**, University of Copenhagen, Denmark

This paper investigates whether the short term interest rate may explain the movements observed in the conditional second moments of asset returns. The theoretical connections between these seemingly unrelated quantities are studied within the C-CAPM framework. Under the assumption that the product of the relative risk aversion coefficient and the marginal utility is monotonic in consumption, original results are derived that attest the existence of a relation between the risk-free rate and the conditional second moments. The empirical findings, involving 165 stock returns quoted at the NYSE, confirm that, at low frequencies, the interest rate is a determinant of the 165 conditional variances and 13530 conditional correlations.

C198: **Indirect inference in non-Gaussian stochastic volatility models for exchange rate data**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Oivind Skare

The financial crisis of 2008 has led to increased focus on the topic of market risk and volatility. Standard Gaussian models for derivative pricing, such as the celebrated Black-Scholes-Merton model, does not seem appropriate and the need for implementing more realistic models of market risk seems more important than ever. This paper aims to contribute to this research by developing new methods for inference (estimation and prediction) in stochastic volatility models for exchange rate data based on non-Gaussian Ornstein-Uhlenbeck (OU) processes. Our approach uses indirect inference methods: First, an auxiliary model based on an approximate Gaussian state space representation of the OU-based model is estimated. Then simulations are made from the underlying OU-model for given parameter values. For each simulation, the auxiliary Gaussian model is re-estimated by maximizing the quasi-likelihood function corresponding to the simulated data. The parameter value (in the underlying OU model) which gives the best match between the quasi-likelihood estimate corresponding to the actual data and the quasi-likelihood estimates corresponding to the simulated data, is chosen as the estimate of the true parameter vector.

C215: **Volatility spillovers among the Gulf Arab emerging markets**
Presenter: **Ramzi Nekhili**, University of Wollongong in Dubai, United Arab Emirates
Co-authors: Naeem Muhammad

The volatility spillovers among Gulf Arab emerging markets is examined. Multivariate VAR-GARCH model of daily returns, with BEKK specification based on the conditional variances and conditional correlations, is estimated for all six GCC equity markets of Saudi Arabia, Kuwait, UAE, Qatar, Oman, and Bahrain. The results show high own-volatility spillovers and a high degree of own-volatility persistence in all GCC markets. Moreover, there are significant cross-volatility spillovers and cross-volatility persistence among all GCC markets, with stronger evidence of spillovers from all GCC markets to the Saudi market. Such evidence could be explained by the sensitivity of the banking sector, which is seen as the dominant sector in most GCC economies, and the existence of uncertainties surrounding various Gulf bank exposures to certain Saudi business groups as well as the downward movement of oil prices.

C317: **An alternative approach to test for financial market contagion**
Presenter: **Marco Gallegati**, Universita Politecnica delle Marche, Italy

The existence of contagious effects in financial markets has been one of the most debated topics in the recent literature of international financial economics. Nonetheless, the presence of contagion and/or interdependence after a shock to an individual country is still controversial. We propose a wavelet-based approach to test whether contagion and/or interdependence occurred during

the most recent international financial crisis: the US subprime crisis of 2007. In particular, we apply wavelet correlation analysis to bivariate systems composed by the return in the country where the crisis originated (the ground zero country) and the return in another country for the crisis and non-crisis periods. The new approach based on the distinction between the transitory and permanent nature of the shift in cross-market linkages after a shock can distinguish between contagion and interdependence, each associated with different specific frequencies (high and low frequencies, respectively), and thus provides a unified framework for testing contagion and interdependence directly and separately. The results indicate that all stock markets display significant levels of contagion and that Japan appears to be the only country where both contagion and interdependence are observed.

C155: **Stock return volatility and radical innovation: the case of pharma**

Presenter: **Mariana Mazzucato**, Open University, UK

Co-authors: Massimiliano Tancioni

Recent finance literature highlights the role of technological change in increasing firm specific and aggregate stock price volatility. Yet innovation data is not used in these analyses, leaving the direct relationship between innovation and volatility untested. Our aim is to investigate more closely the relationship between stock price volatility and innovation using firm level patent citation data. The analysis builds on previous empirical work where it is found that stock price volatility is highest during periods in the industry life-cycle when innovation is the most 'competence-destroying'. Here we ask whether firms which invest more in innovation (more R&D and more patents) and/or which have 'more important' innovations (patents with more citations) experience more volatility. We focus the analysis on firms in the pharmaceutical and biotechnology industries between 1974 and 1999. Results suggest that there is a positive and significant relationship between idiosyncratic risk, R&D intensity and the various patent related measures. Preliminary support is also found for the 'rational bubble' hypothesis linking both the level and volatility of stock prices to innovation.

CS37 Room 10 COMPUTATIONAL ECONOMETRICS Chair: Paolo Foschi

C230: **Bootstrap confidence bands for forecast paths**

Presenter: **Anna Staszewska-Bystrova**, University of Lodz, Poland

VAR models are widely used for forecasting. An obvious element of a forecasting procedure is an evaluation of the uncertainty associated with the forecast. Most of the existing methods are concerned with assessing the uncertainty associated with the forecast for a single period. However, the interest is often in forecasting the path the variables will follow over a sequence of periods. The problem of associating a confidence band with a forecast path has been relatively neglected and the objective of this paper is to provide a method that performs better than methods based on large sample normality. The method proposed and evaluated here is based on the bootstrap. The bands are constructed from forecast paths obtained in bootstrap replications with an optimization procedure used to find the envelope of the most concentrated paths. The method is shown to have good coverage properties in a Monte Carlo study. Indeed in small samples the properties are somewhat better than those of a large sample methods.

C344: **Fast iterated double bootstrap bias correction**

Presenter: **Rachida Ouysse**, University of New South Wales, Australia

A computationally efficient procedure for bias adjustment in the iterated bootstrap is proposed. The new technique replaces the need for successive levels of bootstrap resampling by proposing an approximation for the double bootstrap calibrating coefficient using only one draw from the second level probability distribution. Extensive Monte Carlo evidence suggest that the proposed approximation performs better than the ordinary bootstrap bias correction. In identified models, this fast bootstrap bias correction leads to estimators with lower variance than those based on the double bootstrap. The proposed fast iterated bootstrap performs better than the double bootstrap in all scenarios and especially when the model has the weakest instrument relevance and the highest degree of endogeneity. However, when the estimators have no finite moments and the instruments are weak, the bootstrap does not work well and iterating it makes things worse.

C213: **Efficient bootstrap with weakly dependent processes**

Presenter: **Federico Crudu**, University of York, UK

This paper develops the efficient bootstrap methodology for overidentified moment conditions models with weakly dependent observation. The resulting bootstrap procedure is shown to be asymptotically valid and can be used to approximate the distributions of t-statistics, J statistic for overidentifying restrictions, and Wald, Lagrange multiplier and distance statistics for nonlinear hypotheses. The paper also shows the asymptotic validity of the efficient bootstrap based on a computationally less demanding approximate k-step estimator. The finite sample performance of the proposed bootstrap is assessed using simulations in a simplified model of consumption based asset pricing model.

C277: **The power of bootstrap tests of cointegration rank with financial time series**

Presenter: **Niklas Ahlgren**, Hanken School of Economics, Finland

Bootstrap likelihood ratio tests of cointegration rank are commonly used because they tend to have rejection probabilities that are closer to the nominal level than the rejection probabilities of the corresponding asymptotic tests. The effect of bootstrapping the test on its power is largely unknown. Estimating the power of the bootstrap test by simulation requires a large number of Monte Carlo replications. We show that a new computationally inexpensive procedure can be applied to the estimation of the power function of the bootstrap test. The bootstrap test is found to have a power function close to that of the level-adjusted asymptotic test. Consequently, the bootstrap test estimates the level-adjusted power of the asymptotic test highly accurately. The bootstrap test may have low power to reject the null hypothesis of cointegration rank zero, or underestimate the cointegration rank. In particular, against alternatives close to a unit root, the power of the bootstrap test may be close to its size. The problem with the power of the bootstrap test is illustrated by the expectations hypothesis of the term structure of interest rates applied to the first ten years of data on Euribor interest rates.

C233: **Further developments on unit root tests**

Presenter: **Margherita Gerolimetto**, Ca' Foscari University of Venice, Italy
Co-authors: Claudio Pizzi, Isabella Procidano

For frequentist econometricians, the result of a hypothesis testing procedure is a binary decision, whereas Bayesians can compute the posterior probability of the null, given the data. One critique of 'classical' unit root tests is that they are inferior to Bayesian methods because of the low power of the tests, particularly with trend-stationary alternatives. Frequentists, on the other hand, will argue that disagreement over priors leads to different posterior and results. Moreover, flat-prior analysis of unit root can produce biased estimators. Within this discussion, a set of Bayesian unit root tests has been proposed that overcome many of the traditional unit root tests drawbacks, but do not depend on the researcher's prior opinion. Here we intend to shed some more light on the issue after a few years of dormancy in the literature. We also consider Bayesian cointegration analysis.

CS50 Room 5 ECONOMIC AND FINANCIAL TIME SERIES ANALYSIS Chair: Frederic Jouneau-Sion

C151: **Efficient likelihood evaluation of state-space representations**

Presenter: **Guilherme Moura**, University of Kiel, Germany
Co-authors: David DeJong, Hariharan Dharmarajan, Roman Liesenfeld, Jean-Francois Richard

Likelihood evaluation in state-space models requires the calculation of integrals over unobservable variables. When models are linear and Gaussian, required integrals can be calculated analytically. Departures entail integrals that must be approximated numerically. Here we introduce an efficient procedure for calculating such integrals: the EIS filter. The procedure takes as a building block the particle filter, which employs discrete fixed-support approximations to unknown densities that appear in the predictive and updating stages of the filtering process. Therefore, particle filter's likelihood approximations can feature spurious discontinuities. Moreover, the supports upon which approximations are based are not adapted, giving rise to numerical inefficiencies. Numerous extensions of the particle filter have been proposed to address these problems. Typically, efficiency gains are sought through adaption of period-t densities using information available through period t. However, once period-t supports are established they remain fixed over a discrete collection of points as the filter advances forward through the sample, thus failing to address the problem of spurious likelihood discontinuity. Here we propose an extension that constructs adapted and continuous period-t approximations using efficient importance sampling. Example applications involve analysis of DSGE models and are used to illustrate the relative performance of the particle and EIS filters.

C165: **Dynamic factors of economic data**

Presenter: **Marianna Bolla**, Budapest University of Technology and Economics, Hungary

In our model the components of a multivariate time series are described by a relatively small number of uncorrelated factors. The usual factor model of multivariate analysis cannot be applied immediately as the factor process also varies in time. Hence, there is a dynamic part, added to the classical factor model, the autoregressive process of the factors. The so-called dynamic factors can be identified with some latent driving forces of the whole process. Based on prior work, an efficient algorithm for estimating the model parameters is introduced. The model is applied to economic time series, where the variables are econometric indicators registered yearly within a 30 years period. Some of the most significant factors indicate the main tendencies of the Hungarian economy, possibly showing some hidden causes of the present crisis.

C172: **A complete procedure for estimating hidden Markov models with application in locating structural breaks**
Presenter: **Christos Ntantamis**, CREATES, Aarhus University, Denmark

Testing for structural breaks and identifying their location is essential for econometric modeling. In this paper, a Hidden Markov Model approach is used to perform these tasks. The estimation is conducted using a variant of the Iterative Conditional Expectation-Generalized Mixture algorithm, and its convergence properties are examined. This algorithm allows for different functional forms across regimes. The locations of the breaks are obtained by assigning states to data points according to the Maximum Posterior Mode algorithm, and the Integrated Classification Likelihood -Bayesian Information Criterion identifies the number of regimes. The performance of the overall procedure, denoted IMI, is validated by two sets of simulations; one in which only the parameters differ across regimes, and one with differences in the functional forms. The IMI method performs well in both sets. Moreover, its performance is found to be superior, when it is compared to more standard methods. A second round of simulations will also be examined, with sample sizes corresponding to macroeconomic data applications. The purpose is to ascertain the success of the IMI procedure in such settings, before its application to the examination of monetary policies of the OECD countries.

C234: **Structural breaks and the rank of the spectral density matrix**
Presenter: **Victor Bystrov**, University of Lodz, Poland
Co-authors: Francesco Battaglia, Antonietta di Salvatore

Estimation of the rank of the spectral density matrix is used in the identification of the number of common trends and common cycles in multivariate time series. We employ the factor structure and the perturbation theory to show the bias in the rank of the probability limit of the estimated spectral density matrix, which is caused by a deterministic break in mean of stationary series. The bias induced by the structural break, needs to be taken into account when testing for the number of common cycles.

C364: **A market risk model for asymmetric distributed risk factors**
Presenter: **Kostas Giannopoulos**, The British University in Dubai, United Arab Emirates
Co-authors: Ramzi Nekhili

The behaviour of short term interest rates is of major importance. Understanding the dynamics of short-term interest rates is of fundamental importance for many financial applications. Among others, pricing interest rate derivative and designing hedging strategies. A number of researchers have investigated the link between macroeconomic variables and the short end of the yield curve. In finance, pricing fixed income securities and interest rate derivatives, designing hedging strategies, all depend on the dynamic behaviour of the term structure of interest rate. What makes peculiar a series of short term interest rates are some distributional properties not common among other financial time series. Changes in short term interest rates tend be more volatile and there is evidence for a mean reverting process. But the most atypical property is a lower bound found when the rates are at such low levels like the current US dollar and Japanese Yen. In these circumstances the conditional density function is asymmetric truncated to the left. Understanding the dynamic behaviour of these series and in particular the tails of their density function are of a primary importance in risk management. In this paper we will discuss methods of estimating the conditional density and techniques for evaluating risks. We will also discuss the modelling aspect in these conditions.

CS66 Room 7 TIME SERIES FINANCIAL ECONOMETRICS 2 Chair: Ana-Maria Fuertes

C146: **Minimum-variance autoregressive prediction of nonstationary random walk**
Presenter: **Kai Lam**, Chinese University of Hong Kong, China

Signals perturbed by random noise are often difficult to predict precisely even with prior statistical knowledge on the randomness. There is a lower limit on the prediction error variance bounded by the intrinsic variance of the original random noise process. However, signals with deterministic characteristics such as an autogressive (AR) model specification, can be better predicted if information of the specification is available a priori or can be estimated effectively. Minimum-variance predictors which attain the lower limit are readily constructed. Financial time series belong to an important class of random walk signals which are nonstationary (i.e., with non-constant or drifting mean) and possess strong random and deterministic characteristics. In this paper, we demonstrate how to use nonstationary time series from a deterministic autoregressive model of a random process with either a continuous normal or t distribution, skew-normal or skew-t, discrete uniform or Bernoulli distribution, for constructing effective minimum-variance predictors to attain the lower limit. Similar results on the attainable error variance are obtained for different distributions, thus indicating a precise probabilistic knowledge of the noise process is not required. The histogram of the prediction error is also found identical with that of the original random noise process.

C219: **FTSE-100 volatility index (V-FTSE) and volatility risk premium**

Presenter: **Yue Peng**, Essex University, UK
Co-authors: Sheri Markose

Since the introduction in 1993 of the market traded option implied volatility index, VXO, for the 30 day SP-100 returns, such volatility indexes have become a major instrument for assessing stock market volatility and also in the pricing of volatility swap derivatives for hedging market risk. The volatility risk premium, if positive, is the payoff to a holder of a volatility swap and it is the difference between the realized historical volatility and the volatility swap rate which is the expected realized volatility under the risk neutral measure. Recently, Euronext launched the V-FTSE volatility index for the FTSE 100 based on the so called model free method. As only a finite number of strike prices are available on the FTSE-100 options, the model free method leads to an over estimation bias in the V-FTSE. We propose other methods for the construction of the volatility index. These include an advance on the model free method based on interpolations in the implied volatilities at different money-ness levels of the options and also the implied volatility index derived from the Generalized Extreme Value(GEV) risk neutral density option pricing model. The latter flexibly includes the fat tailed behaviour of stock returns. The Black-Scholes implied volatility is also used as a bench mark. We find that the volatility indexes from the interpolation method and the GEV option pricing model perform better than the Euronext V-FTSE for forecasting realized volatility.

C323: **Statistical inference for the multifractal random walk model**

Presenter: **Cristina Sattarhoff**, University of Hamburg, Germany

This paper improves the Generalized Method of Moments (GMM) estimation procedure for the Multifractal Random Walk model by Bacry and Muzy, by means of an optimal iterated GMM estimator. The moment conditions are given by the lags of the autocovariance function of the model. The weights of the moment conditions are based on the asymptotic covariance matrix of the sample moments, which we estimate using an heteroscedasticity and autocorrelation covariance matrix estimator. The employed GMM estimator performs well in finite samples. We report within the scope of a Monte Carlo study normally distributed estimates for the intermittency coefficient and the variance with small empirical bias and empirical MSE. Yet the estimates for the decorrelation scale are only satisfactory in the case of comparatively large sample sizes. We apply the Multifractal Random Walk to the daily DAX values over the past 50 years. The estimated intermittency coefficient suggests that the DAX values are multifractal, whereas the estimated decorrelation scale implies that the DAX logarithmic increments are dependent for 23 years.

C218: **Credit rating migration in the presence of business cycles**

Presenter: **Fei Fei**, City University London, UK
Co-authors: Ana-Maria Fuertes, Elena Kalotychou

Credit migration matrices play a major role in risk management applications. The conventional discrete cohort and hazard rate estimation methods for rating migrations rely on the Markov and time-homogeneity assumptions which have been questioned recently. Extant research documents time heterogeneity in rating migrations and transition intensities that vary over the business cycle (BC). By relaxing the time-homogeneity assumption, this paper proposes a novel approach that builds on the continuous-time Markov chain model to incorporate the effects of the business cycle on credit rating transition probabilities. Taking advantage of a comprehensive US corporate bond ratings sample over 26 years from Bloomberg, we compute BC-adjusted transition matrix estimates and compare them with simple benchmarks. The latter comprise cohort and hazard-rate estimates that neglect business cycles or account for them in a naive manner by splitting the sample into expansions and recessions prior to estimation. Our findings indicate that ignoring business cycle effects will lead to the underestimation of default risk. The economic regime plays a critical role in estimating rating migration matrices, although its importance decreases as the time horizon increases. Bootstrap simulations reveal that the main efficiency gains of the proposed BC-adjusted hazard-rate approach pertain to economic downturns.

C113: **The role of country, regional and global market risks in the dynamics of Latin American yield spreads**

Presenter: **Alena Audzeyeva**, University of Leeds, UK
Co-authors: Klaus R. Schenk-Hoppe

We analyze the joint impact of country, regional and global market risks on weekly changes in yield spreads of Mexico, Colombia and Brazil. In contrast to previous studies, we consider a homogenous set of liquid Eurobonds which are representative of current emerging bond markets. All risk-factor groups are significant, though relative importance is country-specific. Mexico's spread is mainly driven by global risk while country risk contributes 40% to the explained variance for Colombia and Brazil; another 40% stem from regional risk for Colombia and global risk for Brazil. The sensitivity of spread changes to risk factors varies with bond maturity.

Saturday 31.10.2009 08:45-10:20 Parallel Session J

ES04 Room 9 LATENT VARIABLE AND STRUCTURAL EQUATION MODELS Chair: Gil Gonzalez-Rodriguez

E053: **Critical values for testing an endogenous dummy variable in a bivariate probit model**
Presenter: **Johannes Jaenicke**, University of Erfurt, Germany

This paper analyzes the size distortions of the Wald test in a bivariate probit model with an endogeneous dummy variable. Using Monte Carlo simulation techniques, we find severe size distortions of the Wald test in small samples, declining with increasing sample size. The size of the test statistic is significantly affected by the correlation between the two decision equations. The behavior of the simulated critical values is summarized with response surface functions.

E106: **Weak identification in probit models**
Presenter: **Joachim Wilde**, University of Osnabrueck, Germany

Probit models are by now widely used in applied econometrics. As in linear models, one or more explaining variables might be endogenous. This problem can be solved by using instrumental variables. The resulting estimates can be used to calculate test statistics for the parameters of the model. However, in linear models it is well known that weak instruments may cause considerable size distortions. Wald-type tests like the usual t-tests and F-tests are especially vulnerable to this problem. In probit models a single parameter hypothesis is usually tested by the so-called z-test, i.e. the ratio of a consistent estimate and its asymptotical standard error. This is a Wald-type test. Therefore, big size distortions can be expected. Nevertheless, the topic seems to be largely a white spot in the literature. Therefore, after some introducing remarks, a simulation study is presented that analyzes the problem for different hypotheses and different magnitudes of endogeneity. For ease of exposition a binary probit model is chosen. It is shown that there are considerable size distortions if the magnitude of endogeneity is large. Therefore, the usual test results might be rather misleading. It is concluded with an outline of possible solutions of the problem.

E158: **Structural analysis of linear mixed models with measurement error**
Presenter: **Ruggero Bellio**, University of Udine, Italy
Co-authors: Michela Battauz

The estimation of linear mixed models with covariate measurement error is particularly challenging when the mismeasured covariate enters the random part of the model. An importance instance is given by random slope models, with random effects associated to the mismeasured covariate. Another important case is semiparametric modelling of the effect of the covariate measured with error, that can be formulated as a linear mixed model by following the mixed model approach based on penalised splines. In this work, we propose a structural approach for measurement error modelling, where a normal distribution is assumed for the true unobserved covariate. Suitable algorithms are proposed for obtaining the maximum likelihood estimates of model parameters. In particular, for random slopes we provide a simple algorithm, requiring only one-dimensional integration. For semiparametric regression, the solution proposed is based on one-dimensional Monte Carlo integration of the measurement error and the Laplace approximation for integrating out the random effects.

E215: **Dynamic structural equation model for spatial lattice data**
Presenter: **Pasqaule Valentini**, G. d'Annunzio University, Chieti-Pescara, Italy
Co-authors: Mauro Coli, Lara Fontanella, Luigi Ippoliti

The modeling of data resulting from dynamic processes evolving in both space and time is critical in many scientific fields such as economics, sociological studies and environmental sciences. Factor analysis has previously been used to model multivariate spatial data with temporal behavior modeled by autoregressive components. So far the main developments in the literature focused on simple factor analysis model without causal or structural relationships between latent variables. We propose a multivariate model which is able to capture the main dynamical interactions among a set of aggregate and regional quarterly bank variables described in the Reports of Income and Condition (Call Reports) for all U.S. commercial banks. Specifically, we consider a dynamic simultaneous equation model with latent variables which is a generalization of the static structural equation model with latent variables. The novelty of our proposal is twofold. First, at any given time some multivariate measurements from all observed locations are grouped together. Second, the spatial dependence is modeled by parameterizing the spatial factor loadings. The model is estimated in a unified computational framework by using Markov chain Monte Carlo (MCMC) methods.

ES13 Room 10 MODEL VISUALIZATION AND INTERPRETATION Chair: Heather Turner

E044: **Model visualisation and exploration as graph traversal**
Presenter: **Catherine Hurley**, National University of Ireland Maynooth, Ireland
Co-authors: R.W. Oldford

Visualisation is a key step in data and model exploration. Generally, statistical visualisation is about comparisons. We formalize this using mathematical graphs where nodes are visualisation objects and edges represent comparisons of interest. This representation assists in constructing visualisation layouts, which are then simply graph traversals. Typically, layouts of interest correspond to hamiltonians or eulerians of the mathematical graph. When some comparisons are more important than others, we use graphs with weighted edges and weight-driven traversals to construct visualisations which focus on these important comparisons. In this presentation, we explore applications to model visualization and exploration, including a new display for pairwise comparison of treatment groups, and model comparisons in stepwise regression.

E175: **Profiling the parameters of models with linear predictors**
Presenter: **Ioannis Kosmidis**, University of Warwick, UK

Profiles of the likelihood can be used for the construction of confidence intervals for parameters, as well as to assess features of the likelihood surface such as local maxima, asymptotes, etc., which can affect the performance of asymptotic procedures. In this respect, the profile methods of the R language (stats and MASS packages) can be used for profiling the likelihood function for several classes of fitted objects, such as glm and polr. However, often the likelihood is replaced by an alternative objective for either the improvement of the properties of the estimator, or for computational efficiency when the likelihood has a complicated form. Alternatively, estimation might be performed by using a set of estimating equations which do not necessarily correspond to a unique objective to be optimized (for example, quasi-likelihood estimation). In all of the above cases, the construction of confidence intervals can be done using the profiles of appropriate objective functions in the same way as the likelihood profiles. We present the profileModel R package, which generalizes the capabilities of the current profile methods to arbitrary, user-specified objectives and, also, covers a variety of current and potentially future implementations of fitting procedures that relate to models with linear predictors. We give examples of how the package can be used to calculate, evaluate and plot the profiles of the objectives, as well as to construct profile-based confidence intervals.

E157: **Extended quasi-variances**
Presenter: **David Firth**, University of Warwick, UK

Model summaries that use quasi variances provide an economical route to approximate inference and graphical presentation in situations where the estimable parameter combinations are contrasts. We show how the quasi-variance notion can be usefully extended in various ways, in particular to cover situations where arbitrarily scaled or rotated contrasts are of interest. The definition, computation and graphical display of such generalized quasi-variance summaries will be discussed, with real-data examples.

ES29 Room 3 FUZZY STATISTICAL ANALYSIS 1 Chair: Giulianella Coletti

E046: **Credit scoring analysis by a partial probabilistic rough set model**
Presenter: **Andrea Capotorti**, University of Perugia, Italy
Co-authors: Eva Barbanera

Credit scoring analysis is an important issue, even more nowadays that a huge number of defaults has been one of the main cause of the financial crisis. Amongst the many different tools used to model credit risk, recently Rough Set has shown its effectiveness. Furthermore, original rough set theory has been widely generalized and contaminated by other uncertain reasoning approaches, especially probability and fuzzy set theories. In this paper we try to conjugate Fuzzy Rough Set with Coherent Partial Conditional Probability Assessments. In fact this last model has been shown to be a powerful tool to unify different uncertainty reasoning approaches. In particular, we propose to encompass experts partial probabilistic evaluations inside a gradual decision rule structure, with coherence of the conclusion as guideline. In line with Bayesian Rough Set models, we introduce credibility degrees of multiple premises through conditional probability assessments. Discernibility with this method remains anyhow too fine to reach reasonable graded classification rules. Hence we propose to coarsen the partition of the universe U by equivalence classes based on the arity of positively, negatively and neutrally related criteria. We will use data related to a sample of firms in order to build and test our model.

E083: **Fuzzy similarity in statistical analysis: choosing a measure on the basis of a qualitative point of view**
Presenter: **Giulianella Coletti**, UPMC-CNRS, France
Co-authors: Bernadette Bouchon-Meunier, Marie-Jeanne Lesot, Maria Rifqi

Similarity is a key concept for many problems in statistical analysis and also in fuzzy statistical analysis. Nevertheless the choice of the best measure of similarity in a particular framework is an open problem. We propose to study fuzzy similarity

measures from the point of view of the ordering relation they induce on pairs of objects. Using a classic method in measurement theory, introduced by Tversky, we establish necessary and sufficient conditions for the existence of a specific numerical measure, or a class of measures, to represent a given ordering relation, depending on the axioms this relation satisfies. The interest is particularly focused on different conditions of independence. More precisely, we obtain two kinds of equivalence classes of similarity measures: the first one is given by measures representing the same ordering relation, the second, rougher, definition of equivalence is given by the measures representing orders that are not exactly identical but that possess the same properties and satisfy the same axioms. This definition permits to point out the actual rules we accept when we choose one particular measure of similarity and to make explicit underlying requirements on the induced order.

E081: **A generalized Bayesian inference in building a female avatar starting from crisp and fuzzy information**
Presenter: **Osvaldo Gervasi**, University of Perugia, Italy
Co-authors: Sergio Tasso, Gianni Donati

We present a web application that allows to model a female avatar in terms of body aspect and measures, initially inferred from few fuzzy information provided by the user, and optimized applying some regional statistical information. A set of sliders allows the user to model the body aspect, in order to obtain the most suitable representation of the female avatar, according to the user's likelihood. The application is made representing the virtual world of the female avatar using X3D, The user is allowed to fit the avatar using a set of sliders that through Ajax3D is able to modify in real time the body representation. Presently only one character of the female face is available for which a set of customizations are available (skin colour, eyes colour, hair length). The selected fuzzy approach is based on the interpretation of the membership functions in terms of coherent conditional probability: we choose min and max as T-norm and T-conorm and procedures like Bayes for choosing the most probable elements of database and for updating membership functions, on the basis of the crisp and fuzzy information. The used model allows to manage also partial assessments both for probability and for membership function.

E068: **Generalized Bayesian inference in a fuzzy context**
Presenter: **Barbara Vantaggi**, University La Sapienza, Italy
Co-authors: Giulianella Coletti

We deal with a generalized Bayesian inference procedure able to manage partial fuzzy information, partial probability and possibility assessments. We start from the interpretation of a fuzzy subset of the range $C(X)$ of a variable X, as a coherent conditional probability assessment $P(E(g)|X = x)$, where g is a property of X and $E(g)$ the (Boolean) event You claim that X is g. Some time can be interesting to know the probability of the conditional event $X = x|E(g)$, for instance to find among the elements of a database the most probable under the hypothesis $E(g)$. If a probability distribution on the events $X = x$ is available, then we can easily compute the probability of $E(g)$ and so, by Bayes formula, $P(X = x|E(g))$. Nevertheless in real problems our information about the probability of the elements of $C(X)$ can be partial, that is we can only have at our disposal only an assessment on arbitrary family of its subsets. Moreover we can have not all the values of the membership $P(E(g)|X = x)$, but only some values $P(E(g)|B)$, with B belonging to an arbitrary family of subsets of $C(X)$. In this cases we need to check coherence of the assessments and then to compute its coherent enlargement of P to the events $X = x|E(g)$.

ES39 Room 6 ROBUST ANALYSIS OF COMPLEX DATA SETS 3 Chair: Stefan Van Aelst

E203: **Diffusion driven empirical Bayes estimation of high-dimensional normal means vectors**
Presenter: **Pierpaolo Brutti**, LUISS Guido Carli University, Italy

In this work we consider the problem of estimating independent and possibly high-dimensional normal means vectors in a sparse empirical Bayes framework that glues together a recent manifold-modeling technique called diffusion maps, with a more classical concept of sparsity based on the assumption that most of the unknown coordinates of model parameter are actually 0. More specifically, for the vector valued parameter of interest, we adopt a mixture prior composed by an atom at zero and a completely unspecified density for the non-zero component. The novelty of the proposed method is the way we use the data to implicitly drive the prior specification through a (weighted/iterative) quantized diffusion density estimator of the marginal distribution. In this way, depending on the actual structure of the data at hand, we are able to reach a balance between the two complementary approaches to sparsity mentioned above.

E174: **Clustering of categories in multiple regression with categorical predictors**
Presenter: **Jan Gertheiss**, Ludwig-Maximilians-University Munich, Germany
Co-authors: Gerhard Tutz

The challenge in regression with categorical predictors is the high number of parameters involved. Common shrinking methods like Lasso or Elastic Net, which allow for selection of predictors, are typically designed for metric predictors. If independent variables are categorical, selection strategies should be based on modified penalties. For categorical predictor variables with

many categories a useful strategy is to search for clusters of categories with similar effects. The objective is to reduce the set of categories to a smaller number of categories which form clusters. The effect of categories within one cluster is supposed to be the same, but responses will differ across clusters. In this talk two L1-penalty based methods for factor selection and clustering of categories are presented and investigated. The first approach is designed for nominal scale levels, the second one for ordinal predictors. All methods are illustrated and compared in simulation studies, and applied to real world data from the Munich rent standard.

E128: **Robust regression with optimisation heuristics**
Presenter: **Enrico Schumann**, University of Geneva, Switzerland
Co-authors: Manfred Gilli

Least Squares (LS) has appealing theoretical and numerical properties for estimating the standard linear regression. However the LS estimates are often unstable in the presence of outliers. One approach to deal with such extreme observations is the application of robust or resistant estimators, like Least Quantile of Squares (LQS) estimators. Unfortunately, for many such alternative estimators, the optimisation is much more difficult than in the LS case, as the objective function is not convex and often has many local optima. We apply different heuristic methods like Differential Evolution, Particle Swarm and Threshold Accepting to obtain parameter estimates. These methods are very flexible and thus allow to estimate models under different optimisation criteria, like LQS or LTS. The main purpose of the study is to compare the convergence behaviour for the different optimisation techniques for increasing amounts of computational resources (iterations). We also stress the importance of investigating jointly the quality of the solution with respect to the optimisation (e.g., the quantile of squared residuals associated with a solution), and with respect to the actual aim (e.g., identifying outliers).

E177: **Robust smoothing with asymmetrically distributed errors, with applications to functional data analysis**
Presenter: **Matias Salibian-Barrera**, University of British Columbia, Canada
Co-authors: Liangliang Wang, Nancy Heckman

Consider a stochastic model for functional data where each function is a realization of a (typically Gaussian) stochastic process. Such a model can also incorporate additive measurement errors at each of the observed points which is desirable in many real-life applications. The recently proposed PACE method to perform inference for (potentially sparse) functional data is based on a smooth estimator of the covariance function of the assumed underlying stochastic process. This estimate smooths the observed point-wise covariances, whose expected values are the desired covariances, and then interpolates them smoothly for other intermediate times. Not surprisingly, this approach is highly sensitive to the presence of a relatively small proportion of atypical observations (either entire curves, or individual points along a curve). Motivated by an application to detect outliers in atmospheric data using PACE, in this talk we discuss a modification of robust smoothers (focusing on local polynomial methods) for the challenging case of a model with asymmetrically distributed errors (as it is required to compute the smooth estimate of the covariance function mentioned above).

CS28 Room 7 CONDITIONAL MODELS OF RETURN AND RISK Chair: Giovanni Barone-Adesi

C099: **The time-varying prediction of successful mergers**
Presenter: **Giuseppe Corvasce**, Swiss Finance Institute Lugano, Switzerland
Co-authors: Giovanni Barone-Adesi

The dynamics of the physical probability for firms that undertake a stock swap merger is developed through a simple model. Using a sample of 1090 deals from 1992 to 2008, we show how price movements in target stock prices are informative of the success or failure of a stock swap merger and how movements in bidder stock prices are informative at the beginning of the deal period. Without any assumption on the convergence of the target stock price to the bid offer, our results share the findings of the previous literature. According to our results bidder and target movements represent the thermometer of a deal status.

C105: **A stochastic model for hedging electricity portfolio for an hydro-energy producer**
Presenter: **Rosella Giacometti**, University of Bergamo, Italy
Co-authors: Maria Teresa Vespucci, Marida Bertocchi, Giovanni Barone-Adesi

A stochastic portfolio model for a hydropower producer operating in a competitive electricity market is discussed. The portfolio includes its own production, a set of power contracts for delivery or purchase including contracts of financial nature as forwards/futures to be able to hedge against risks. The goal of using such a model is to maximise the profit of the producer and reduce the economic risks connected to the fact that energy price may be highly volatile due to various different, unpredictable reasons (i.e. very cold winter) and to the possibility of a period of scarcity of raining or snowmelting.

C123: **A time-additive regime switching volatility model**

Presenter: **Antonietta Mira**, Universita' della Svizzera Italiana, Switzerland
Co-authors: Reza Solgi, Giovanni Barone-Adesi

The purpose is to build a time-additive volatility model able to capture the main stylized facts observed in financial time series. We also want a model easy to interpret and parsimonious. The main problem with available diffusion models is that they produce volatility smiles that are too smooth to fit prices in the index option market. Because of this empirical limitation several authors have introduced diffusion-jump models. Some GARCH models are flexible enough to fit the data, however GARCH models suffer of two drawbacks that limit their use for risk management applications. First, they are discrete models, that cannot generally maintain the same form over different time scales. Second, GARCH parameter estimation is often complicated by the collinearity issues. We propose to investigate the possibility of mitigating these problems by simplifying the description of the stochastic process driving market volatility. We assume that volatility changes between discrete levels only. Changes are driven by the past history of returns. The function linking the past history of returns to the auxiliary time series that will determine the state of the market takes, as input, the square of negative returns. The model parameters are estimated using maximum likelihood. We compare the predictive performance of this model with that of the corresponding Bayesian model estimated via MCMC. In particular we use powerful adaptive MCMC methods that have been recently developed.

C373: **Quantitative ambiguity models on the space of measures and utility optimization**

Presenter: **Charalambos Charalambous**, University of Cyprus, Cyprus

An abstract setting ambiguity models via relative entropy and total variation distance on the space of measures is formulated. These models codify the notion of robustness associated with uncertainty of stochastic systems. Ambiguity models based on relative entropy give rise to a utility expressed in terms of the free energy of statistical mechanics. On the other hand, ambiguity models based on total variation distance are shown to be related to linear combination of infinity norm and L1 norm, while a candidate measure which corresponds to the infinity norm is constructed from the class of tilted probability measures. The ambiguity models are then applied to minimax optimization of stochastic systems governed by controlled Ito stochastic differential equations.

CS42 Room 4 MODELLING FINANCIAL TIME SERIES Chair: Paolo Foschi

C216: **Regime-Switching modelling of globalization analysis in international stock markets**

Presenter: **Nuno Ferreira**, ISCTE-IUL, Portugal
Co-authors: Rui Menezes, Diana Mendes

Several experimental research showed that stock markets display periods of marked turbulence and exhibit extreme values more often than one would expect if the series were normally distributed (fat tail property). In this context, in order to better understand this phenomenon, it was developed, between others, the Markov Switching Model. Nowadays, this kind of models has attached much attention in financial and economic modelling, since, ample empirical evidence has been gathered for both nonlinearity and structural changes in the dynamic properties of many observed time series. We employ a smooth transition autoregressive (STAR) model in order to investigate cyclical behaviour of stock returns in five international stock markets. In the last two decades much attention has been related to modelling the conditional variance. However a point which needs to be stressed it is that an adequate modelling of the nonlinear dependence in the conditional mean is necessary in order to avoid misspecification of the conditional variance model. The results clearly show that the stock markets are characterized by the presence of nonlinear patterns. These findings have important implications for empirical finance, investment decisions, pricing of financial derivatives, portfolio optimization, and cross-market transmission of volatility.

C226: **Estimation of a stock market return function with significant ARCH effects**

Presenter: **Diacos Panayiotis**, Intercollege, Limassol Campus, Cyprus

A dynamic specification which we employ for the study of the variations in stock market returns in Cyprus is presented. So as to account for the high volatility in the variance of the innovations, in the regression framework we incorporate an ARCH process. Then, using monthly time series observations, we estimate the parameters of the model. The results support a dependence between current and past real returns in the mean equation and a significant relation between current and past squared residuals in the variance equation. Finally, various conclusions are also derived by comparing the historical return series and the estimated variances of the sample.

C169: **Estimation of a time-varying GQARCH-M Model**

Presenter: **Sofia Anyfantaki**, Athens University of Economics and Business, Greece
Co-authors: Antonis Demos

Time varying GARCH-M models are commonly used in econometrics and financial economics. Yet, no exact likelihood analysis of these models has been provided so far. The analysis of a time-varying GARCH-M model becomes substantially complicated since the log-likelihood of the observed variables can no longer be written in closed form. The main modern way of carrying out likelihood inference in such situations is via a Markov chain Monte Carlo (MCMC) algorithm. Unfortunately a regrettable consequence of the path-dependence in volatility is that standard MCMC algorithms will evolve in $O(T^2)$ computational load, where T is sample size. In this context, we suggest using the MCMC algorithm which allows the calculation of a classical estimator via the simulated EM algorithm, or a simulated Bayesian solution in only $O(T)$ computational operations . The crucial idea is to transform the GARCH model into a first order Markov model. We outline our model and the estimation problem that arises from the fact that we have two unobserved processes. The application of this method –both classical and Bayesian estimation– for weekly returns from three major stock markets is finally illustrated.

C130: **Analyzing and exploiting asymmetries in the news impact curve**

Presenter: **Sven Christian Steude**, Universtiy of Zurich, Switzerland
Co-authors: Jochen Krause, Marc Paolella, Markus Haas

The recently proposed class of mixed normal conditional Heteroskedastic (MixN-GARCH) models couples a mixed normal distributional structure with linked GARCH-type dynamics. It has been shown to offer a plausible decomposition of the contributions to volatility, as well as admirable out-of-sample forecasting performance for financial asset returns. The MixN–GARCH model assumes constant mixing weights. Different specifications with time–varying mixing weights are considered. In particular, by relating current weights to past returns via sigmoid response functions, an empirically reasonable representation of the news impact curve with an asymmetric impact of unexpected return shocks on future volatility is obtained, and large gains in terms of in–sample fit and out–of–sample VaR forecasting performance can be realized.

CS44 Room 8 ASSET PRICES AND MACROECONOMICS Chair: Willi Semmler

C008: **Credit economic capital and high performance predictive analytics**

Presenter: **John Angus Morrison**, Asymptotix, Belgium

The Economic Capital, i.e. the amount of capital which a Financial Institution needs in order to survive in a worst case scenario is considered. Calculation of economic capital is no longer an academic exercise. The Credit Crunch (CC) has seen Central Governments pumping fresh capital into the banks which were clearly undercapitalized. One of the primary causes of the Credit Crunch (CC) was the failure to comprehensively compute risk capital in issued structured instruments. These products cannot be abandoned entirely since that would send the wider economy back to a prehistoric wilderness. Computation of Economic Capital is complicated and has remained in the academic domain, thus the Community aspect of Open Source is eminently applicable to engender economic capital computation into banking, Banking needs Open Source now. R or the R-Project is an open source programming language for statistical computing supported in the commercial domain by REvolution Computing. REvolution Computing in commercializing the economic capital modeling process has brought a further innovative development to the the modeling of Economic Capital. REvolution has expertise in delivering High Performance Computing (HPC) solutions It is this High Performance Open Source approach which can commercialise econometrics in the socially necessary way required today.

C303: **Structural change detection and the predictability of returns**

Presenter: **Vijay Vaidyanathan**, EDHEC, United States Of America
Co-authors: Daniel Mantilla-Garcia

In spite of the intense stream of research on stock return predictability, the evidence remains controversial and acutely sensitive to the choice of predictive variables, dataset, and time periods. While some studies show that return predictability results lack reliability and robustness, others researchers believe that returns are predictable. In order to reconcile these differences, recent research has shown that return predictability is much more pronounced in certain sub-periods than it is over the entire post-war period. Using the Bai-Perron technique, recent research finds that adjusting for structural breaks in the predictor variable significantly improves the returns predictability evidence. Unfortunately, the Bai-Perron technique requires the entire time series in order to detect breaks, which limits its usefulness for real-time out-of-sample returns forecasting. Further, the procedure does not lend itself to estimating the probability of a change point at each point in time. Therefore, we examine the applicability of two alternative approaches to change detection in returns predictability. First, we test a Bayesian approach change point detection algorithm. Next, we investigate the performance of real time change detection algorithms. We show that these two approaches could provide different insights to the returns predictability puzzle. The analysis uses publicly available packages in R.

C302: **Modelling dynamics of aggregate consumption for Lithuanian economy**
Presenter: **Audrone Jakaitiene**, Institute of Mathematics and Informatics, Lithuania
Co-authors: Antanas Zilinskas, Julius Zilinskas

Household consumption accounts for about 65% of spending in Lithuania. Therefore it is important for macroeconomists to be able to explain the determinants of consumer spending via a well-specified consumption function. In Lithuania the analysis of household consumption is relatively scarce. Typically consumption is an integral part of a larger structural macroeconomic model and there is one recent publication devoted only for the modelling of consumption. Specifically it uses consumption as the error-correction type of model. The results are quite close to the general ideas of Friedman's permanent income hypothesis. Here, we model household consumption from the perspective of the modern representative agent-based approaches. Household chooses a stochastic consumption plan to maximize the expected value of their time-additive nonlinear utility function subject to asset budget constraint. This multi-period problem can be solved by using the Bellman equation. The first order condition is the Euler equation which is typically estimated using the general method of moments. We employ numerical methods to compute equilibrium. Empirical analysis is conducted using quarterly Lithuanian data covering period from year 1995 to 2008.

C191: **An econometric model of international asset prices and macroeconomic dynamics**
Presenter: **Jan Bruha**, Czech National Bank, Czech Republic

This paper formulates and estimates a two-country model of international asset prices and macroeconomic latent-variable dynamics. The representative agent in each country is endowed with a pricing kernel depending on the latent variables. The agent uses the kernel to price the assets so that the asset yields in both countries satisfy the non-arbitrage condition. The exchange-rate dynamics are then based on the return parity of any asset in the two countries. For estimation purposes, the model is formulated as a state-space model and is estimated on monthly Czech / German data. Two alternative formulations of the driving model for latent variables are compared.

CS46 Room 5 ECONOMETRIC APPLICATIONS Chair: Reinhard Neck

C306: **Pairwise likelihood for missing data treatment in VAR models**
Presenter: **Luca Grassetti**, University of Udine, Italy
Co-authors: Giovanni Fonseca

In the last two decades the missing data issue in multivariate time series models has played a central role in the econometric analysis of micro and macro economic data. Solutions in the state space model context have been developed but alternative methods have also been proposed in the classical likelihood framework. The present work aims at applying the composite likelihood approach to the vector autoregressive (VAR) model estimation in presence of partial and sparse missingness patterns. Usually, the application of composite likelihood methods aims to avoid computational issues for full likelihood. In the VAR context, the application of composite likelihood can be useful for estimating more complex cases, as, for instance, heteroskedastic and regime-switching models. Moreover, pairwise likelihood can present some further valuable advantages. In fact, the common approaches can suffer a lack of estimator efficiency when missing data are present. For example, the skipping algorithm considers as fully missing the partially observed vectors in time. The pairwise approach partially recovers lost information because the single likelihood components are substituted by considering a cross-sectional pairwise version. The proposed method has been validated considering a simulation study and an application to historic agricultural market data has been developed too.

C271: **Where is an Oil shock?**
Presenter: **Michael Owyang**, Federal Reserve Bank of St. Louis, United States Of America
Co-authors: Kristie Engemann, Hoawrd Wall

Previous studies have concluded that the effect of oil shocks on the U.S. economy are asymmetric, that is, increases in oil prices adversely affect economic activity but decreases in oil prices have no effect. We test the veracity of these results using state-level data and find that, while oil shocks may indeed be asymmetric, the direction and magnitude of the asymmetry varies across states. In particular, states which have a large energy producing sector experience a rise in economic activity when oil prices rise. On the other hand, states which have a large manufacturing sector experience a boom when energy prices fall.

C257: **Monthly labour force survey time series, seasonal adjustment and reconciliation**
Presenter: **Riccardo Gatto**, Istat - Italian National Institute of Statistics, Italy
Co-authors: Tommaso Di Fonzo, Marco Marini

The ISTAT (Italian National Institute of Statistics) is starting the diffusion of monthly data from the Labor Force Survey. There will be auxiliary results while the quarterly figures will still be the main output of the survey. The methodology has been studied in order to respect the perfect consistency between monthly and quarterly data. Seasonally adjusted data will be released at the same time and, wishing to preserve the consistency property of the raw data, a double constraint problem arises: contemporaneous

constraints must be imposed for the coherence among the disaggregated series and their more reliable aggregations; temporal constraints must be imposed for the coherence between the monthly and the more reliable quarterly series. Some recent two-step reconciliation procedures are applied here, according to which the single series are temporally benchmarked at the first step, and then balanced in the second step. The results will be evaluated in terms of both adjustments to the levels and movement preservation performances (i.e., impact on the growth rates of the non-reconciled series).

C267: **Stochastic control of econometric models for Slovenia**
Presenter: **Reinhard Neck**, Klagenfurt University, Austria
Co-authors: Dmitri Blueschke, Viktoria Blueschke-Nikolaeva

We present a new version of the algorithm OPTCON for the optimal control of nonlinear econometric models. It can be applied to obtain approximate numerical solutions of control problems with a quadratic objective function for nonlinear econometric models with additive and multiplicative stochastics. The new version was programmed in C# and allows for deterministic and stochastic control, the latter with open-loop and passive learning information patterns. We demonstrate the applicability of the algorithm by some policy problems using two versions (a linear and a nonlinear one) of a quarterly macroeconometric model for Slovenia. This shows the convergence and the practical usefulness of the algorithm for some problems of stabilization policy under small-sized econometric models.

CS51 Room 1 FORECASTING, HEAVY TAILS AND NON-STANDARD INFERENCE 2 Chair: Lynda Khalaf

C094: **Weak identification and confidence sets for covariances between errors and endogenous regressors**
Presenter: **Firmin Doko Tchatoka**, University of Montreal and University of Sherbrooke, Canada
Co-authors: Jean-Marie Dufour

In this paper, we focus on structural models and propose a finite-and large-sample projection-based techniques for building confidence sets for the endogeneity parameter between errors and regressors allowing for the presence of weak identification. First, we show that the procedure is robust to weak instruments and provide analytic forms of the confidence sets for endogeneity parameter. Second, we provide necessary and sufficient conditions under which such confidence sets are bounded in finite- and large-sample. Finally, after formulating a general asymptotic framework which allows to take into account a possibility of heteroskedasticity and/or autocorrelation of model residuals, we show that the procedure remains valid: this means that it is asymptotically robust to heteroskedasticity and/or autocorrelation. Moreover, the procedure proposed can be used as pre-test (partial exogeneity test) to improve the estimation of structural parameters.

C098: **Level crossing random walk test robust to the presence of structural breaks**
Presenter: **Alex Maynard**, University of Guelph, Canada
Co-authors: Vitali Alexeev

We propose a modified version of the nonparametric level crossing random walk test, in which the crossing level is determined locally. This modification results in a test that is robust to unknown multiple structural breaks in the level and slope of the trend function under both the null and alternative hypothesis. No knowledge regarding the number or timing of the breaks is required. A data driven method is suggested to select the extent of the localization based on a trade-off between finite sample power and size distortion in a proximate model.

C179: **On robust M-estimation of the tail index**
Presenter: **Dieter Schell**, University of Konstanz, Germany
Co-authors: Jan Beran

A new robust M-estimator of the right tail index is introduced, based on a Pareto type MLE. Consistency and asymptotic normality are derived. The estimator is robust with respect to deviations from the Pareto distribution at lower quantiles, while keeping a square-root-n rate of convergence. Simulations illustrate that the new estimator outperforms classical methods for small and moderate sample sizes.

CS59 Room 2 VOLATILITY MODELS AND APPLICATIONS Chair: Andreas Savvides

C295: **Analysing hedge fund investments: evidence from a multivariate predictive Student-t full factor GARCH model**
Presenter: **Ioannis Vrontos**, Athens University of Economics and Business, Greece

Extending previous work on hedge fund return predictability, this paper introduces the idea of modelling the conditional distribution of hedge fund returns using a Student-t full-factor multivariate GARCH model. This class of models takes into account the stylized facts of hedge fund return series, namely heteroskedasticity and fat tails. For the proposed class of multivariate predictive models, we derive analytic expressions for the score, the Hessian matrix and the Information matrix, which can be used

within classical and Bayesian inferential procedures to estimate the model parameters, as well as to compare different predictive specifications. We propose a Bayesian approach for model comparison because it provides posterior probabilities for different predictive models that can be used for model averaging. Our research design and analysis is motivated by the empirical evidence that accounting for model uncertainty and time-varying covariances improves our ability to construct optimal hedge fund portfolios. We explore potential impacts of our approach by analyzing hedge fund strategies and show that it can be economically important.

C110: **Decomposing realized variance: a point process of relevant price changes with long memory in volatility**
Presenter: **Ping Chen Tsai**, Lancaster University Management School, UK

A simple, empirical-based approach to decompose Realized Variance (RV) is proposed, with supportive theoretical argument and empirical evidence. Under the proposed framework, RV is interpreted as a product of the intensity and variance of relevant price changes. Holding the variance aspect constant, statistical inference on the event intensity is conducted, with the spot intensity described by Hawkes process. Empirical analysis on Spyder returns in 2008 confirms strong performance of the model. An attempt to track down the source of long memory is merited by the simple decomposition of RV, and serves as starting point for future research.

C162: **Consistent and asymptotic normal parameter estimates for stochastic volatility models with leverage effect**
Presenter: **Gianna Figa-Talamanca**, University of Perugia, Italy

The problem in fitting stochastic volatility models to market data is, essentially, that volatility cannot be directly observed or deduced from price observations. Thus, it is necessary to introduce a procedure for filtering the volatility process from market prices. However, many of the approaches to the estimation of stochastic volatility models in continuous time, as the Efficient Method of Moments, the Indirect Inference and the Simulated Maximum Likelihood methods, do not need to recover the entire time series for the variance/volatility. However, intensive simulations is a key ingredient in all these methodologies and the major drawback is the computational effort. We introduce a simple procedure to estimate model parameters in a moment-like fashion without any computational effort, only relying on observations of the log-price. The algorithm is based on limit theorems available from our previous research and, by making use of the so-called delta-method, consistency and asymptotic normality of the estimators is proven. Numerical examples are also given on stock indexes.

C077: **Stock market and foreign exchange volatility**
Presenter: **Maria Matsi**, Cyprus University of Technology, Cyprus
Co-authors: Elena Andreou, Andreas Savvides

This paper investigates two-way volatility spillovers between the stock market and the foreign exchange market of a number of emerging economies during the period 1985-2008. In addition to the domestic stock and foreign exchange markets, the model incorporates volatility spillovers from mature stock markets. A tri-variate VAR-GARCH(1,1) model with the BEKK representation is estimated separately for twelve emerging economies: Argentina, Brazil, Chile, Colombia, Mexico, Venezuela, India, Korea, Malaysia, Pakistan, Philippines and Thailand. Evidence suggests that a two-way spillover does exist for some of these countries but for others the spillover is one-way. The inclusion of the mature market indicates that there exist volatility spillovers to most emerging stock markets and to the foreign exchange markets of nearly all emerging economies. The time period is divided into two sub samples: pre and post liberalization. Evidence from the post liberalization period demonstrates an increase in volatility spillovers for some economies. By applying the concept of shift contagion we find evidence of shifts in the transmission of volatility from the stock market to the foreign exchange market and vise versa before and after the liberalization in almost all economies.

Saturday 31.10.2009 10:40-13:00 Parallel Session K

ES01 Room 3 FUZZY STATISTICAL ANALYSIS 2 **Chair: Thierry Denoeux**

E052: **Testing hypotheses as a fuzzy set estimation problem**
Presenter: **Glen Meeden**, University of Minnesota, United States Of America
Co-authors: Siamak Noorbaloochi

The concept of p-value or level of significance is widely used in practice to measure the strength of evidence against a null hypothesis. Although originally introduced by R. A. Fisher the usual formal justification comes from the Neyman-Pearson theory of testing which assumes a sharp break between the null and alternative hypotheses. This makes little sense in most scientific work. Here we argue that in many scientific applications one should reformulated the problem as estimation of a fuzzy membership function of the set of good or useful or interesting parameter points. Rather than specifying a null and alternative hypotheses one should choose a fuzzy membership function to represent what is of interest in the problem at hand. For standard testing situations sensible families of possible membership functions are introduced. We show that the usual p-value can be interpreted as estimating a particular membership function. An example involving the mating behavior of chimpanzees is considered. We select a fuzzy membership function which allows one to measure how consequent the data is with a proposed theory. In this example standard statistical tests are difficult to employ.

E114: **A relational approach to stochastic dominance**
Presenter: **Bernard De Baets**, Ghent University, Belgium
Co-authors: Hans De Meyer

The notion of stochastic dominance is becoming increasingly popular in economics, finance, social statistics, decision making under uncertainty, multi-criteria decision making, machine learning, and so on. Stochastic dominance imposes a partial order relation on a collection of random variables, simply by comparing their marginal cumulative distribution functions. It ignores the dependence structure and is intolerant for small violations of this functional comparability. We have established a pairwise comparison method for random variables, based on bivariate distribution functions, obtained by coupling the marginal distribution functions by means of copula. This copula is not expressing the real dependence, but has to be seen as a parameter of the comparison method. This comparison results in a reciprocal relation on the given collection of random variables. The transitivity of this reciprocal relation can be expressed in the cycle-transitivity framework, and depends on the copula used. This transitivity allows to identify appropriate cutting levels, which allow to turn the reciprocal relation into a strict order relation. This approach can be seen as an alternative to the notion of stochastic dominance.

E112: **On the use of Hilbert space tools to handle fuzzy random variables**
Presenter: **Gil Gonzalez-Rodriguez**, European Centre for Soft Computing, Spain
Co-authors: M. Angeles Gil, Ana Colubi

Fuzzy random variables (FRVs) in Pury and Ralescu sense are natural models for handling imprecise data in Statistics. Formally they are measurable mappings which associate each element of a probability space with a fuzzy set in a given space. There are different approaches for developing inferential tools for fuzzy random variables. One of the most promising and useful technique consists in embedding the considered family of fuzzy sets onto a closed and convex cone of a Hilbert Space by considering an appropriate class of distances. This embedding allows to identify FRVs with Hilbert-valued random elements, which offers many possibilities. On the one hand the statistical tools developed for Hilbert-valued random elements can be used for FRVs and, conversely, the new advances for FRVs can also be used to analyze other classes of functional data. The aim of this work is to introduce the mechanism linking fuzzy random variables and Hilbert-valued random elements considering a very versatile distance based on a mid-spread decomposition of fuzzy numbers: the recently proposed Theta-distance. The usefulness of this procedure in the development of tools for handling fuzzy data from an inferential point of view will be illustrated by means of several examples.

E069: **CECM : Constrained-Evidential C-Means**
Presenter: **Violaine Antoine**, Universite de Technologie de Compiegne, France
Co-authors: Benjamin Quost, Marie-Helene Masson, Thierry Denoeux

The aim of cluster analysis is to group objects according to their similarity. Some methods use hard partitioning, some use fuzzy partitioning and, recently, a new concept of partition based on belief function theory, called credal partition, has been proposed. It enables to generate meaningful representations of the data and to improve robustness with respect to outliers. All these methods are unsupervised ones, as the similarity between the objects is determined using only a numeric description of the objects. However, in some applications, some kind of background knowledge about the objects or about the clusters is available. To integrate this auxiliary information, constraint-based methods (or semi-supervised) have been proposed. A popular type of

constraints specifies whether two objects are in the same (Must-Link) or in different clusters (Cannot-Link). Moreover, actively selecting object pairs enables to get improved clustering performances using only a small number of constraints. We propose here a new algorithm, called CECM, which combines belief functions and the constrained clustering frameworks. We show how to translate the available information into constraints and how to integrate them in the search of a credal partition. The paper ends with some experimental results.

E210: **M-Estimators and advanced fuzzy clustering**
Presenter: **Frank Klawonn**, University of Applied Sciences Brunschweig/Wolfenbuettel, Germany
Co-authors: Roland Winkler, Rudolf Kruse

Fuzzy clustering approaches use weights to assign data to clusters. In standard fuzzy clustering, a parameter called fuzzifier is introduced which leads to certain disadvantages that can be overcome by a more general approach replacing the power transform of the fuzzifier by more general functions. We extend the investigations on standard fuzzy clustering in terms of M-estimators known from robust statistics and demonstrate that they have better robustness properties than the usual fuzzifier.

E049: **An evidential neural network classifier incorporating contextual discounting**
Presenter: **Vasileios Georgiou**, Universite de Technologie de Compiegne, France
Co-authors: Benjamin Quost, Thierry Denoeux

A novel neural network classifier is presented that consists of a typical Multi–Layer Perceptron (MLP) and Radial Basis Function (RBF) neurons. A cluster of reference patterns (prototypes) extracted from each class of the available data set is considered as an item of evidence regarding the class membership of an input vector. This evidence is quantified by RBF neurons in order to measure the reliability of the MLP's classification output. The output of the RBF neurons is a vector of discounting factors so that each component corresponds to the reliability of the MLP conditionally on each element of the frame of discernment. According to Smet's Transferable Belief Model, the MLP's output is represented by basic belief assignments on the elements of the frame of discernment and is combined using Dempster's rule of combination. Then, the contextual discounting operator is applied to the BBAs in order to decrease their influence according to the level that each source is considered reliable given the input vector. Finally, the betting probabilities are calculated in order to obtain the final classification of an input vector. The aforementioned classifier is applied on the problem of fault detection in railway track circuits.

ES09 Room 10 MIXTURE MODELS Chair: Dankmar Bohning

E008: **Population size estimation under the Poisson-Gamma model**
Presenter: **Irene Rocchetti**, University La Sapienza, Italy
Co-authors: Dankmar Bohning

Estimation of the size of an elusive target population is of prominent interest in many areas in the life and social sciences. Our aim is to provide an efficient method to estimate the unknown population size given the frequency distribution of counts of repeated identifications of units of the population of interest. This counting variable is necessarily zero-truncated, since not identified units are not in the sample. We consider several application examples: in all of them the homogenous Poisson model is not appropriate since it does not account for heterogeneity. The Poisson-Gamma model provides a flexible alternative. It was previously used in the development of the Chao-Bunge estimator. Here we consider ratios of neighboring Poisson-Gamma probabilities: they follow a linear relationship to the count of repeated identifications and occur as posterior means from a non-parametric empirical Bayes approach a la Robbins. We propose a weighted logarithmic regression model to estimate the zero frequencies counts, assuming a Gamma-Poisson distribution for the counts. A detailed explanation about the chosen weights and a goodness of fit index are presented. We compared the results from the proposed estimator with those obtained through the Chao-Bunge one. The implications and limitations of such methods are discussed.

E022: **Some new estimators under a Poisson mixture capture probability in capture-recapture experiments**
Presenter: **Krisana Lanumteang**, University of Reading, UK
Co-authors: Dankmar Bohning

Methods are presented to derive some new estimators for estimating the size of a target population under capture-recapture experiments. These proposed estimators are developed by extending the idea of Chao's estimator using monotonicity of ratios of neighboring frequency counts under a Poisson mixture sampling framework. The new estimators are developed as weighted regression estimators considering the log-ratio of neighboring frequencies as dependent variable. A simulation technique was used to study the performance of the proposed estimators under both homogeneous and heterogeneous Poisson capture probability. Confidence interval estimation was done by means of the bootstrap method, and was found to perform reasonably well. An application of estimating the number of drug users in Bangkok Thailand in the year 2002 was also examined in order to illustrate the use of these methods.

E021: **Capture-recapture estimation of population size by means of empirical Bayesian smoothing**
Presenter: **Dankmar Boehning**, University of Reading, UK

In this note we suggest a smoothed generalization of Zelterman's estimator of the size of an elusive population. This estimator is often used to adjust for undercount in registration lists. The Zelterman estimator is defined as a version of the Horvitz-Thompson estimator where a constant Poisson probability is used. The Zelterman estimator suffers under the fact that all units have the same identification probability attached and it seems desirable to allow an estimator which allows for different probabilities of inclusion conditional upon the number of repeated identifications. The crucial question is how a unit specific Poisson parameter can be estimated. We suggest to use an empirical Bayes estimator on the basis of the Bayes theorem for Poisson counts. Using the idea of nonparametric empirical Bayes one can achieve an estimator of the Poisson parameter without any knowledge of the pior distribution. Other approaches for estimating the prior distribution are possible including the Poisson-Gamma or estimating the prior nonparametrically leading to the NPMLE of a discrete mixing distribution. A simulation study compares these different approaches for choosing a prior distribution and presents some results.

E190: **Capture-recapture with heterogeneous detection probabilities**
Presenter: **Luca Tardella**, Sapienza Universita di Roma, Italy
Co-authors: Alessio Farcomeni

We consider recent advances on capture-recapture model Mh, in which capture probabilities are allowed to be heterogeneous. The non-identifiability of conditional likelihood parameterization is overcome with the use of complete likelihood based on the moments of the unobserved distribution of heterogeneous probabilities so that the unknown population size can be identified and consistently estimated. We show the implementation of MLE estimates based on the identified moment parameters and compare their computational performance with respect to the finite mixture parameterization.

E138: **Clustering dependencies via mixture of copulas**
Presenter: **Dimitris Karlis**, Athens University of Economics and Business, Greece
Co-authors: Veni Arakelian

The impact of mixture models for clustering purposes has considerably increased the last years. Nowadays there are several clustering procedures based on mixtures for certain types of data. On the other hand copulas are becoming very popular models in order to model dependencies. One of the appealing properties of copulas is the fact that they can separate the marginal properties of the data from the dependence properties. Thus, they allow for modelling dependencies in neat way. The purpose of our paper is to put together the two ideas so as to construct mixtures of copulas aiming at using them for clustering with respect to the dependence properties of the data. We also describe estimation using an EM algorithm based on the standard approach for mixture models. A real data application in finance is used to illustrate the potential of our method. The idea is that the dependencies between assets change over time and hence using a mixture of copulas we can capture such changes. Moreover in order to examine the factors that influence such changes we make use of explanatory variables in the mixing probabilities.

E151: **Statistical and biological significance in gene discovery, with an application to multiple sclerosis in Italian twins**
Presenter: **Marco Alfo'**, Sapienza Universita di Roma, Italy
Co-authors: Alessio Farcomeni, Luca Tardella

A robust model for detection of differentially expressed genes which directly incorporates biological significance, i.e., effect dimension is proposed. Using the so-called 2–fold rule we transform the expressions into a nominal observed random variable (three categories: within threshold, above upper threshold or below lower threshold) which we assume generated by a nominal latent variable (three categories: not differentially, over and under expressed), with gene-specific parameters. We show how to obtain estimates of the parameters by maximizing the likelihood with a constrained EM algorithm. Different strategies for gene discovery are discussed and compared. We illustrate the method on an original study on multiple sclerosis.

ES11 Room 9 PARAMETRIC AND NONPARAMETRIC MODEL VALIDITY Chair: Simos Meintanis

E018: **Estimating dynamic panel data models with autocorrelation by restricted regressions**
Presenter: **Savas Papadopoulos**, Democritus University of Thrace, Greece

A novel estimation method for a linear dynamic panel data model with random effects and correlated errors is introduced. The method consists of three stages executed for each of the k variables included in the model. In the first stage, we compute the residuals after regressing each variable on its dynamic term and on its lag of order three. In the second stage, we estimate the coefficient of the kth variable on the mth variable by regressing the residuals of the first stage on the residuals of the third stage from the previous iteration, $k-1$. The regressors also include all the variables with indices between $m+1$ and $k-1$ by restricting their coefficients equal to their estimated values from previous steps. The third stage estimates the coefficient of the dynamic term plus the coefficient of the autocorrelated errors. The regressors include the regressors from the first stage plus all the variables

with indices less than $k-1$, by restricting their coefficients equal to their estimated values from the second stage. Simulations show that our estimates have small bias and small RMSE. The existing methods transformed MLE and GMM indicate large bias under autocorrelation and cannot be applied for large T.

E057: **Tests of normality based on Shepp property, and their efficiencies**

Presenter: **Iakov Nikitin**, Saint - Petersburg University, Russia
Co-authors: Xenia Volkova

The normal law occupies the central place in Probability and Statistics, and testing of normality belongs to most important problems of goodness-of-fit theory. In the past there was found a non-linear transformation of data which preserves normality (Shepp, SIAM Rev. 6, 459-460), and more recently it was proved that in fact it is a characterization of the normal law in a broad class of distributions. This characterization enables us to construct new invariant tests of normality based on U-empirical df's. Test statistics are either U-statistics or the supreme of families of U-statistics depending on a real parameter. We describe limiting distributions of new test statistics and evaluate their large deviation asymptotics under the null hypothesis. This permits to calculate the local Bahadur efficiency of our statistics for natural parametric alternatives (like shift, skew and contamination alternatives) which is high enough. Pitman efficiency has the same values. New statistics look promising for normality testing.

E117: **Inconsistent goodness-of-fit tests with improved power for important alternatives**

Presenter: **Olivier Thas**, Ghent University, Belgium
Co-authors: Bert De Boeck, Jean-Pierre Ottoy

It is generally excepted that goodness-of-fit tests should be omnibus consistent in the sense that they are consistent against any fixed alternative. On the other hand often in applied statistics it is assumed, for example, that observations never come from a normal distribution, and even if they do, imperfect measurements or rounding errors may obscure this property. Therefore it is not surprising that omnibus tests will eventually reject the null hypothesis, particularly with large data sets, but this rejection is therefore not necessarily informative. We develop a class of goodness-of-fit tests that are not consistent against alternatives close to the hypothesised distribution. An advantagous consequence is that these tests have generally larger powers against alternatives that are sufficiently distinct from the hypothesised distribution. We show this both theoretically and empirically through simulation studies. The tests closely resemble smooth tests. We also demonstrate how this class of tests can be tuned so as to obtain a better compromise between type I error rate and power or false discovery rates (FDR).

E143: **Exact goodness-of-fit tests for censored data**

Presenter: **Aurea Grane**, Universidad Carlos III de Madrid, Spain

Previously a goodness-of-fit statistic for complete samples to test the null hypothesis of a completely specified distribution function has been proposed. When there is no censoring, it was found out that the test based on the proposed statistic can advantageously replace those of Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling for a wide range of alternatives. Here this statistic is modified so that it can be used to test the goodness-of-fit of a censored sample. We deduce the exact distributions of three modifications of the statistic and obtain the exact critical values for different sample sizes and different significance levels. We also give some conditions under which the convergence to the normal distribution can be asserted. We study the power of the exact tests based on these statistics for five parametric families of alternative distributions with support contained in the $[0,1]$ interval, and we conclude that the tests based on our proposals have a good performance in detecting symmetrical alternatives, whereas the tests based on the Kolmogorov-Smirnov, Cramer-von Mises and Anderson-Darling statistics are biased for some of these alternatives.

E155: **Bootstrapping in sequential change-point procedures**

Presenter: **Marie Huskova**, Charles University in Prague, Czech Republic
Co-authors: Claudia Kirch

The talk discusses various approximations for critical values for test procedures for detection of a change in sequential setup. One usually tries to utilize limit behavior under no change model. Sometimes such approximations provides reasonable approximation, however in many situations the convergence to the limit distribution is rather slow, eventually, the explicit form of the limit distribution is unknown. Therefore various bootstraps were developed in order to provide good approximations for critical values. Here we develop bootstrap suitable for sequential tests for detection changes location and regression models. Theoretical results show the asymptotic validity of the proposed bootstrap procedures. A simulation study compares the bootstrap and asymptotic tests shows that studentized bootstrap test behaves generally better than asymptotic tests.

E147: **Specification tests for the error distribution in GARCH models**
Presenter: **Simos Meintanis**, University of Athens, Greece

Goodness–of–fit tests are proposed for the innovation distribution in autoregressive conditionally heteroscedastic models. The tests utilize an integrated distance involving the empirical characteristic function (or the empirical Laplace transform) computed from properly standardized observations. A bootstrap version of the tests is utilized in order to study the small sample behavior of the procedures in comparison with more classical approaches. As an example, the tests are applied on some financial data sets.

ES17 Room 8 ANSET (ITALIAN SIS GROUP ON TIME SERIES ANALYSIS) Chair: Cira Perna

E012: **Genetic algorithms for fitting nonlinear nonstationary threshold time series models**
Presenter: **Francesco Battaglia**, Sapienza University, Italy
Co-authors: Mattheos K. Protopapas

Many time series exhibit both nonlinearity and nonstationarity, and linear models are not flexible enough to ensure a satisfactory fit. Features like nonlinear dependence, structural change, and slow modifications in time of the dynamic structure, cannot be accounted for by the widely used autoregressive integrated moving average models. Though both nonlinearity and nonstationarity have been often taken into account separately, few attempts have been proposed for modeling them simultaneously. We consider threshold models, and present a general model allowing for different regimes both in time and in levels, where regime transitions may happen according to self-exciting, or smooth transition, or piecewise linear threshold modeling. Fitting such a model involves the choice of a large number of structural parameters (e.g. orders and thresholds) for which, on one hand, no analytical optimization method is available, and, on the other hand, substantive reasons for motivating the choice are rarely found. Similar problems in statistical inference and time series analysis were successfully addressed using meta-heuristic methods. We propose a procedure based on genetic algorithms, evaluating models by means of a generalized identification criterion. The performance of the proposed procedure is illustrated with a simulation study and applications to some real data.

E027: **Clusters of multivariate time series**
Presenter: **Roberto Baragona**, Sapienza University of Rome, Italy
Co-authors: Sanghamitra Bandyopadhyay, Ujjwal Maulik

In a large set of application fields an observation unit is a collection of time series which has the structure of a vector (or multivariate) time series. Seismic waves data, electroencephalograms and macroeconomic data are examples of distant applications that deal with vector time series. Clustering vector time series is a convenient way to reduce a large data set to a smaller set of vector time series each of which may be assumed as representative of a larger number of items. In addition, special applications in finance concerned with dynamic conditional correlation multivariate GARCH models require assets that share the same dynamics to be grouped in the same cluster. We propose to form clusters of vector time series according to statistical time series features that allow meaningful dissimilarities to be outlined. Two or more cluster indexes of internal validity may be used as criteria that may be simultaneously maximized. Pareto optimality concepts and evolutionary computing are considered for implementing multi objective optimization algorithms. Applications to real data and sets of artificial vector time series will be presented to illustrate our procedure and for comparison purpose.

E165: **Parameter estimation for continuous stochastic volatility models**
Presenter: **Giuseppina Albano**, Universita di Salerno, Italy
Co-authors: Francesco Giordano, Cira Perna

Continuous-time diffusion processes are often used in literature to model dynamics of financial markets. In such kinds of models a relevant role is played by the variance of the process. So assumptions on the functional form of such variance have to be made in order to analyse the distribution of the resulting process and to make inference on the model. In this paper the variance is also modelled by means of a diffusion process. This comes out as continuous time approximation of a GARCH(1,1) process. The existence and uniqueness of the solution of the resulting stochastic differential equation is discussed. Then inference on the parameters and properties of the involved estimators are discussed under different choices of the frequency data. Simulations on the model are also performed.

E194: **Smooth and flexible skew-symmetric distributions using B-splines and penalties**
Presenter: **Patrizio Frederic**, University of Modena, Italy

Recently, there has been a growing interest in Skewed-Symmetric (SS) distributions. A pdf of the class SS is represented by the product of a symmetric pdf around zero and a skewing function (i.e. a function bounded in $[0,1]$, odd with respect to the $y = 1/2$ plane). It is straightforward to show that any product between a symmetric in zero pdf and a skewing function is a valid pdf. SS distributions allow us to generalize many skew distribution models such as the skew-normal distribution, the skew-t distribution, and the so called Generalized Skew-Elliptical distributions. We propose a new flexible approach for modelling skewing functions

via B-splines. B-splines are chosen for having a large number of numerical properties such as computational robustness, and efficiency. Furthermore, we introduce a penalized likelihood estimation method to avoid over-fitting. B-splines SS with penalties provide a flexible and smooth semi-parametric setting that allow estimates to capture many features of the target function such as symmetry and multimodality. After outlining some theoretical results, we propose an effective computational strategy. Finally, we present some empirical results on simulated and real-world data which point out the advantages of the proposed method.

E105: **Investigating the profitability of technical trading rules with the regression trunk approach**
Presenter: **Claudio Conversano**, University of Cagliari, Italy

The profitability of technical trading rules derived from a statistical modeling approach is investigated. The stock price is modeled as a linear function of a set of technical indicators, including past prices and volume transformations. The specific purpose is to detect important interactions among technical indicators. These interactions are expressed as threshold interactions instead of traditional cross products among predictors. In order to detect these threshold interactions the Regression Trunk Approach – RTA – is jointly used with subsampling. The strength of RTA lies in its ability to automatically detect a regression model with multiple main effects and a parsimonious amount of higher order interaction effects. It is based on a new algorithm to estimate a regression trunk model that results more efficient than the main effects model. The relationships between the proposed model and traditional (autoregressive) approaches are also discussed. Financial time series are analyzed and the profitability of the derived trading rules are evaluated. This is achieved by mimicking the behavior of a rationale investor whose decision are based on RTA's forecasting of future prices.

E224: **Information reduction techniques for turning point prediction**
Presenter: **Paolo Foschi**, University of Bologna, Italy
Co-authors: Simone Giannerini, Alessandra Luati

A method for the identification of turning points in economic time series is proposed. The method is based upon the idea of discretizing a continuous state space time series. The key arguments are the dimension d of the reconstructed space and, for each variable $X_1, ..., X_d$, the cardinality of the alphabet, that is, the number of categories in which each variable is quantized. In order to select such cardinalities, an optimality criterion based on the prediction performance is introduced. The validity of the proposal is assessed upon simulated and real time series by using the NBER's dating procedure of the business cycle as the gold standard.

ES33 Room 6 ROBUST METHODS Chair: Peter Filzmoser

E032: **Simulation in robust statistics using the R package simFrame**
Presenter: **Andreas Alfons**, Vienna University of Technology, Austria
Co-authors: Matthias Templ

Due to the complexity of robust methods, obtaining analytical results about their properties is often virtually impossible. Therefore simulation studies are widely used by statisticians as data-based, computer-intensive alternatives for gaining insight into the quality of the developed methods. The R package simFrame is an object-oriented framework for statistical simulation with special emphasis on applications in robust statistics. A key feature is that researchers can make use of a wide range of simulation designs with a minimal amount of programming. Control classes allow a certain proportion of the data to be contaminated. Thereby different contamination models are represented by different control classes and the existing framework may easily be extended with user-defined classes. Furthermore, an appropriate plot method is selected automatically depending on the structure of the simulation results. Hence simFrame is widely applicable in the field of robust statistics.

E122: **Error rates for multivariate outlier detection**
Presenter: **Andrea Cerioli**, University of Parma, Italy
Co-authors: Alessio Farcomeni

It is well known that robust estimation and outlier detection are two essentially equivalent tasks: given robust estimators of location and scatter, outliers are revealed by their large distances from this robust fit. Formal identification rules thus require the choice of reliable cut-off points for the robust distances. The goal of this work is to describe how to obtain the required cut-offs when location and scatter are estimated through the high-breakdown Reweighted Minimum Covariance Determinant (RMCD) estimator. In particular, it will be shown that: (a) an accurate approximation to the distribution of the squared reweighted distances yields good control of the size of the test of no outliers even in small samples; (b) this control does not imply great sacrifice of power, especially when alternatives to the family-wise error rate are considered for multiple outlier detection. The resulting procedures differ in their attitude towards swamping. For instance, in outlier tests based on the FDR criterion the acceptable degree of swamping is allowed to depend on the number of outliers found.

E028: **Robust principal components for compositional data**

Presenter: **Karel Hron**, Palacky University, Faculty of Science, Czech Republic
Co-authors: Peter Filzmoser, Clemens Reimann

Many data sets from environmental sciences are formed by observations where the relevant information is only contained in the ratios between the components. Typically, the data values are expressed in percentages and sum up to 100. This condition does not only lead to algebraic problems, the main objection is a conceptual one. Namely the described data, often called compositional data (or compositions, for short), induce another sample space, the simplex, with the geometry different from the standard Euclidean one. As a way out, a family of logratio transformations was proposed, that moves the compositions from the simplex to the usual Euclidean geometry where standard statistical methods can be applied. The centred logratio (clr) transformation enables to perform principal component analysis (PCA) as well as to use the resulting loadings and scores for a meaningful compositional biplot with an intuitive interpretation. However, the clr transformed compositions are singular and for a robustification of the PCA an auxiliary step is needed. The robust loadings and scores as computed via the isometric logratio (ilr) transformation are regular but not easy to interpret. Therefore they have to be back-transformed to the clr space. Thus, the construction of the robust compositional biplot is possible. The theory will be supported by a practical application.

E099: **Robustifying total least squares**

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

If the orthogonal condition (in the linear regression model) is broken, the (Ordinary) Least Squares (OLS) estimator is generally biased and inconsistent. The classical theory offers then the method of Instrumental Variables. A robust version of instrumental variables was studied by Visek, namely Instrumental Weighted Variables. An alternative approach overcoming the inconsistency is known as Total Least Squares (TLS). Due to the quadratic form of the loss function of TLS, the estimation is vulnerable to influential points. For the Error-in-Variable model a robustified version of TLS based on the idea M-estimators was proposed by Jefferys with studentization of the residuals by a robust version of the scale of the disturbances of all variables. To reach scale and regression equivariance for M-estimators, the residuals have to be studentized by a scale estimator which is scale invariant and regression equivariant. Our proposal for robustifying TLS is based on the idea of downweighting the influential points, but unlike to classical Weighted Least Squares the weights are prescribed to order statistics of squared orthogonal residuals rather than directly to squared residuals. The paper will discuss the inspiration for the proposal, properties of the new estimator and the algorithm for its evaluation. Numerical illustration will also be included.

E033: **Robust resampling methods for time series**

Presenter: **Lorenzo Camponovo**, Lugano, Switzerland
Co-authors: Olivier Scaillet, Fabio Trojani

We study the robustness of block resampling procedures for time series by characterizing their quantile breakdown point in a M-estimation setting. We generally find very low quantile breakdown point indicating a serious robustness problem, which is more important than in the pure iid setting. To solve this problem, we introduce robust resampling schemes for time series, which can be applied to a broad class of resampling procedures. We also propose robust data-driven methods for selecting the block size and the degree of robustness in applications. Monte Carlo simulation and sensitivity analysis confirm the fragility of classical resampling procedures and the accuracy of our robust resampling approach under different types of contamination outliers.

E171: **M-procedures for detection of changes**

Presenter: **Jaromir Antoch**, Charles University, Czech Republic

We will consider the linear regression model describing the situation when the first m observations follow the linear model with the parameter beta and the remaining $n-m$ observations follow the linear model with the parameter $\beta+\delta$. Such problems occur in various situation, e.g., in econometric time series. The parameter m is usually called the change point. Its estimators and their properties will be of prime interest of this lecture. We will concentrate on robust M-estimators of m. Sensitivity of the results and comparison with the classical and Bayesian methods will be demonstrated on both simulated and real data sets. Comparison with a prospective (on-line) approach will be considered as well.

CS05 Room 2 VAR METHODS IN ECONOMICS AND FINANCE Chair: Francesco Ravazzolo

C229: **Model selection and rank estimation in vector error correction models**

Presenter: **Peter Winker**, University of Giessen, Germany
Co-authors: Dietmar Maringer

Model specification might affect the estimated cointegration rank, in particular for small sample sizes as typically used in macroeconomic analysis. Both the specification of deterministic components and of the short run dynamics is not trivial in this context. A data driven selection of the short run dynamics is pursued. The resulting complex optimization problem is tackled using an

optimization heuristic. In contrast to previous standard and heuristic approaches, the model selection and estimation step are treated simultaneously. The proposed method is compared to traditional approaches on a comprehensive set of test cases. The construction of these test cases exhibiting a more complex short run dynamic as in most simulation studies is also based on a heuristic optimization approach. The results provide strong support for the dependence of rank estimation on model selection. In particular for small sample sizes, the proposed heuristic method provides better results than standard sequential procedures.

C025: **Time varying VARs, monetary policy and asset prices**
Presenter: **Samad Sarferaz**, Norges Bank, Norway
Co-authors: Francesco Furlanetto

In this paper we study the relationship between interest rates and asset prices in a time varying vector autoregression (VAR), identified using a combination of short-run and long-run restrictions. This set-up allows us to study whether the estimated response of the monetary policy authority to asset prices (stock prices and real estate prices) has changed over time. In a VAR identified through heteroskedasticity that we estimated in a previous paper, we showed by splitting the sample in two that the response to stock prices has been lower in recent years whereas the response to real estate prices has been higher than in the past. In the current paper we want to investigate this result more deeply in a more sophisticated framework where we allow for time variation in the structural parameters. As far as we know, this has not been done yet in the literature.

C346: **Novel segmentation methods for financial data streams**
Presenter: **Dima Alberg**, Ben Gurion University of the Negev, Israel
Co-authors: Mark Last

Two novel algorithms SW-VAR and SWAB-VAR for an efficient segmentation of financial data streams are presented. The proposed algorithms are based on two state-of-the-art segmentation methods: Sliding Window and SWAB. The proposed algorithms are aimed at improving the efficiency and usability of the original Sliding Window and SWAB algorithms. First, the proposed segmentation algorithms decrease the number of input parameters; second they use the Value at Risk as an additional input parameter, and finally they allow to predict structural changes in financial data streams faster and more accurately than Sliding Window and SWAB. The accuracy and computation time of the proposed algorithms will be compared to the state-of-the-art segmentation methods using empirical experiments with synthetic, non-normally distributed data and Israeli financial indexes TA25 and TA100. We will also show that the proposed algorithms can significantly contribute to the GARCH prediction models accuracy and efficiency because they can reduce the dimensionality of the GARCH model volatility auto regression parameter. Finally, the user interface of the online application implementation will be demonstrated. This interface will allow the user to simulate online financial data stream segmentation depending on the user specified Value at Risk measure.

C250: **Approximate regime shifts in predictability with vector autoregressive models**
Presenter: **Stuart Hyde**, Manchester Business School, UK
Co-authors: Massimo Guidolin

In the empirical portfolio choice literature it is often invoked that through the choice of predictors that could closely track business cycle conditions and market sentiment, simple Vector Autoregressive (VAR) models could produce optimal strategic portfolio allocations that hedge against the bull and bear dynamics typical of financial markets. However, a distinct literature exists that shows that non-linear econometric frameworks, such as Markov switching, are also natural tools to handle portfolio diversification needs arising from the existence of good and bad market states. In this paper we examine when and how simple VARs can produce empirical portfolio rules similar to those obtained under Markov switching by examining the effects of expanding both the order of VARs and the number of predictor variables included in a typical stock-bond strategic asset allocation problem on US data.

C024: **Strategic asset allocation under structurally unstable predictability**
Presenter: **Francesco Ravazzolo**, Norges Bank, Norway

Optimal portfolio decisions force investors to make a number of important decisions concerning the treatment of different sources of uncertainty, the relevant predictor variables, the values of the regression parameters, and their stability. This paper investigates a typical, strategic asset allocation problem when all these three sources of uncertainty are taken into account by an expected utility maximizer. US stock, bond and money market returns are modelled and forecasted jointly by a number of financial and macroeconomic factors in a flexible VAR framework that allows for parameter instability, parameter uncertainty and model uncertainty simultaneously. We investigate if the approach provides statistical gains in forecast accuracy. We also examine the effects of the different sources of uncertainty on asset allocation and portfolio weights. Finally, we test the economic value of the model in active investment strategies. Our results show that parameter uncertainty and model uncertainty play an important role for a short-term horizon power utility investor, and structural instability becomes more relevant only for longer horizons. However, predictability decreases with horizons and in our sample investing all the wealth on the stock market gives higher economic values for investments longer than 1-year.

CS09 Room 4 NONPARAMETRIC VOLATILITY ESTIMATION Chair: Simona Sanfelici

C212: **Specification tests in SPD estimation**
Presenter: **Zdenek Hlavka**, Charles University in Prague, Czech Republic

State price densities (SPD) are an important element in applied quantitative finance. The SPD may be estimated from the observed option prices via a nonparametric estimator of the second derivative of the European call, or put pricing function and it has already been demonstrated that the covariance structure of the observed option prices must be taken into account when such SPD estimators are constructed from the observed intra day option prices. We consider a constrained nonparametric SPD estimator and investigate the validity of the assumed covariance structure using some recently proposed tests of independence in heteroskedastic nonparametric regression. These specification tests are typically based on the bivariate joint empirical distribution function of the strike prices and the standardized residuals. We discuss the usefulness and applicability of these specification tests in the framework of the SPD estimation and the power of these tests against various alternatives that might occur in the SPD estimation is investigated in a simulation study.

C170: **Constrained general regression in Sobolev spaces with application to option pricing**
Presenter: **Michal Pesta**, Charles University in Prague, Czech Republic

State price density (SPD) contains important information concerning market expectations. The SPD is a probability density function that can be expressed as the second derivative of both the Call and Put European option prices with respect to the strike price. In existing literature, a constrained nonparametric estimator of the SPD is found by ordinary least squares in a suitable Sobolev space, i.e., a space of sufficiently smooth functions that allows to transform the problem of searching for the best fitting function in an infinite dimensional space into a finite dimensional optimizing problem. We improve the behaviour of the previously proposed estimator by implementing a covariance structure taking into account the time of the trade and by considering simultaneously both the observed Put and Call European option prices.

C056: **The effect of infrequent trading on detecting jumps in realized variance**
Presenter: **Frowin Schulz**, University of Cologne, Germany
Co-authors: Karl Mosler

The analysis of how accurate an elaborated jump detection methodology for realized variance applies to financial time series characterized by less frequent trading is considered. In this context, it is of primary interest to understand the impact of infrequent trading on two test statistics, applicable to detect significant jumps in realized variance. In a simulation study, evidence is found that infrequent trading induces a sizable distortion of the test statistics towards over-rejection. A new empirical investigation using high frequency information of the most heavily traded electricity forward contract of the Nord Pool Energy Exchange corroborates the evidence of the simulation. In line with the theory, a zero-return-adjusted estimation is introduced to reduce the bias in the test statistics, both illustrated in the simulation study and empirical case.

C126: **Nonsynchronous covariation and high-frequency data**
Presenter: **Takaki Hayashi**, Keio University, Japan
Co-authors: Nakahiro Yoshida

An asymptotic distribution theory of the nonsynchronous covariation process for continuous semimartingales is presented. Two continuous semimartingales are sampled at stopping times in a nonsynchronous manner. Those sampling times possibly depend on the history of the stochastic processes and themselves. The nonsynchronous covariation process converges to the usual quadratic covariation of the semimartingales as the maximum size of the sampling intervals tends to zero. When the limiting variation process of the normalized approximation error is random, the convergence to mixed normality, or convergence to a conditional Gaussian martingale is obtained. A class of consistent estimators for the asymptotic variation process is proposed based on kernels, which will be useful for statistical applications to high-frequency data analysis in finance. A Poisson sampling scheme with random change point is presented as an example.

C365: **Pricing options with realized volatility**
Presenter: **Fulvio Corsi**, University of Lugano, Switzerland
Co-authors: Nicola Fusari, Davide La Vecchia

A stochastic volatility option pricing model that exploits the informative content of historical high frequency data is developed. We propose a simple (affine) but effective long-memory process using the Two Scales Realized Volatility as a proxy for the unobservable returns volatility: the Heterogeneous Auto-Regressive Gamma (HARG) model. This discrete-time process, combined with an exponential affine stochastic discount factor, leads to a completely tractable risk-neutral dynamics. The explicit change of probability measure obtained within this framework allows the estimation of the risk-neutral parameters directly under the physical measure, leaving only one free parameters to be calibrated. An empirical analysis on SP500 option index shows that

the proposed model outperforms competing GARCH models, especially for short term options where our model is able to better capture the smile of the implied volatility curve.

C266: **Quarticity estimation via Fourier method**
Presenter: **Simona Sanfelici**, University of Parma, Italy
Co-authors: Maria Elvira Mancino

The availability of high frequency financial data has motivated an intense study of volatility measurement. The issue of quarticity estimation for diffusion processes is still an open problem and arguably more difficult than integrated volatility estimation because the noise is magnified. Nevertheless, in order to obtain feasible estimators of covariance and of the presence of jumps, it is necessary to have good estimators of the so called integrated quarticity, which is the integral over a given period of the fourth power of the diffusion coefficient of the price process. What is required are efficient estimators of quarticity which are robust to noise. Available estimators are based on multipower variation, which allow testing for the presence of jumps. Averaging over subsamples and pre-averaging of data allow a reduction in the estimation error in the presence of noise. These estimators can be modified by means of a threshold methodology to reduce the bias in finite samples. Fourier analysis can be used to recover the integrated quarticity from the Fourier coefficients of the volatility and, ultimately, from observed returns. This paper investigates the extent to which this approach can provide good estimates of the quarticity robust to high frequency noise.

CS19 Room 7 STOCHASTIC AND ROBUST PORTFOLIO OPTIMIZATION Chair: Ronald Hochreiter

C168: **Dealing with uncertainty in an international portfolio context**
Presenter: **Raquel Joao Fonseca**, Imperial College London, UK
Co-authors: Berc Rustem

The benefits of international diversification for risk reduction have been extensively discussed. We present an international portfolio optimization model where we take into account the two different sources of return of an international asset: the local returns denominated in the local currency, and the returns on the foreign exchange rates. The separate consideration of the returns on currencies introduces non-linearities in the model, both in the objective function (return maximization) and in the triangulation requirement of the foreign exchange rates. The uncertainty associated with the returns is incorporated directly in the model by the use of robust optimization techniques. We show that, by using appropriate assumptions regarding the uncertainty sets and the correlation between local and currency returns, the model can be solved efficiently with standard optimization techniques. While robust optimization provides a guarantee of a minimum return inside the uncertainty set considered, additional guarantees are presented in the form of currency options for cases where the realised foreign exchange rates fall outside the uncertainty sets.

C112: **Worst-case portfolio optimization with skewness and kurtosis: a proposed solution strategy**
Presenter: **Polyxeni-Margarita Kleniati**, Imperial College London, UK
Co-authors: Berc Rustem

In this paper, we address the worst-case, or robust, mean – variance – skewness – kurtosis portfolio optimization problem for discrete rival estimates of asset statistics. Although higher order moments, such as skewness and/or kurtosis, have been often considered in the deterministic portfolio selection problem as a remedy to non-normal return distributions, its robust counterpart remains still undiscovered despite uncertainty underlying the asset statistics estimates. As a solution strategy for the resulting (nonconvex) polynomial optimization problem (POP) we propose a decomposition-based scheme made for POPs. To end with, we present and discuss preliminary numerical results.

C201: **Sparse and stable Markowitz portfolios**
Presenter: **Domenico Giannone**, Free University of Brussels, Belgium
Co-authors: Joshua Brodie, Ingrid Daubechies, Christine De Mol, Ignace Loris

We consider the problem of portfolio selection within the classical Markowitz mean-variance framework, reformulated as a constrained least-squares regression problem. We propose to add to the objective function a penalty proportional to the sum of the absolute values of the portfolio weights. This penalty regularizes (stabilizes) the optimization problem, encourages sparse portfolios (i.e. portfolios with only few active positions), and allows to account for transaction costs. Our approach recovers as special cases the no-short-positions portfolios, but does allow for short positions in limited number. We implement this methodology on two benchmark data sets. Using only a modest amount of training data, we construct portfolios whose out-of-sample performance, as measured by Sharpe ratio, is consistently and significantly better than that of the naive evenly-weighted portfolio.

C236: **Omega optimization as a linear program**
Presenter: **Michalis Kapsos**, Imperial College London, UK
Co-authors: Nicos Christofides, Berc Rustem, Steve Zymler

Omega is a recent performance measure. A new approach to compute Omega as a linear program is derived. While the Omega ratio is considered to be non–convex, an exact formulation in terms of a convex optimization problem and its transformation as a linear program (LP) is shown. Thus, an exact solution can be attained using simplex-based and interior point methods. The convex reformulation for Omega maximization is a direct analog to mean–variance framework, and the fractional linear program to Sharpe maximization. Thus, the Omega ratio becomes easier to understand and implement by investment companies, hedge funds, and pension funds, since it captures the downside and upside potential of the constructed portfolio utilizing all higher order moments.

C333: **Worst-case Value-at-Risk of non-linear portfolios**
Presenter: **Steve Zymler**, Imperial College London, UK
Co-authors: Daniel Kuhn, Berc Rustem

Evaluation of the Value-at-Risk (VaR) of some portfolio requires precise knowledge of the probability distribution of the asset returns, which is rarely available in practice. Moreover, it is well known that VaR is a non-convex function and is therefore difficult to optimize in practice. In order to overcome both these issues the notion of Worst-Case VaR which is a conservative and convex approximation of VaR has been introduced. However, Worst-Case VaR bases itself only on the first– and second–order moments of the asset returns, and therefore cannot accurately represent the non–linear relationships between option returns and the underlying asset returns. We propose a more flexible Worst–Case VaR model which approximates these non-linear relationships by a (possibly non-convex) quadratic function of the underlying asset returns. We show that this model can be cast as a convex Semidefinite Program.

C166: **A realized conditional correlation model for large-scale portfolio optimization**
Presenter: **Jonathan Cornelissen**, K.U. Leuven, Belgium
Co-authors: Kris Boudt, Christophe Croux

In recent years new covariance matrix estimators based on high-frequency data have been developed, enhancing portfolio performance. The Realized Conditional Correlation model for convenient covolatility forecasting based on high-frequency data in a high dimensional setting is proposed. It combines the separate modelling of conditional volatility and correlation (an idea of the multivariate GARCH literature) with more recent findings of the realized volatility literature. Furthermore, our model has the advantage that the dynamics of the volatility of each asset can differ and that it retains the coefficients' interpretability, compared to the models used in the empirical literature so far. Our results indicate that in a portfolio context this model outperforms existing models when a trade-off between risk and return is considered and that robustness to jumps yields better results at high sampling frequencies.

CS38 Room 1 ECONOMETRICS OF FINANCIAL DISTRESS AND APPLICATIONS Chair: Andrea Cipollini

C031: **Measuring financial contagion by local Gaussian correlation**
Presenter: **Bard Stove**, Norwegian School of Economics and Business Administration, Norway
Co-authors: Karl Ove Hufthammer, Dag Tjostheim

This paper examines financial contagion, that is, whether the cross-market linkages in financial markets increases after a shock to a country. This effect has been studied in a variety of ways. We introduce the use of a new measure of local dependence to study the contagion effect. The central idea of the new approach is to approximate an arbitrary bivariate return distribution by a family of Gaussian bivariate distributions. At each point of the return distribution there is a Gaussian distribution that gives a good approximation at that point. The correlation of the approximating Gaussian distribution is taken as the local correlation in that neighbourhood. By comparing the local Gaussian correlation for the stable and crisis period, we are able to test whether contagion has occurred. In particular, there are several advantages by using the local Gaussian correlation; the new measure does not suffer from the selection bias of the conditional correlation, and the local Gaussian correlation may be able to detect nonlinear changes in the dependence structure, that a global correlation may mask. Examining several crisis, among others the Asian crisis of 1997 and the financial crisis of 2007-2008, we find evidence of contagion based on our new procedure.

C304: **Asset allocation under trending volatility**
Presenter: **Antonios Antypas**, University of Piraeus, Greece
Co-authors: Nikolaos Kourogenis, Nikitas Pittis

Recent research provides evidence that volatility of stock returns exhibits a time trend. The existence of this volatility trend can be utilized in the estimation of the covariance matrix of the individual stock returns. This task may be achieved by developing

measures of the degree of trending of the unconditional variance of stock returns. We adopt a suggested model where the variance of the error term can change in a polynomial-like fashion and the rate of growth is captured by the polynomial order, k. Hence, by estimating k for individual stock returns, we are able to forecast the future behavior of their volatilities and adjust the sample estimator of the covariance matrix for these dynamics. A positive (negative) value of k is an indicator that the variance around the end of our sample is bigger (smaller) than that near the beginning of the sample. This paper examines the performance of a portfolio selection procedure where the covariance matrix is adjusted for the presence of a polynomial-like trend. We compare this procedure with previous studies that have also focused in correcting the volatility estimation by taking into consideration the dynamic structure of the covariance matrix.

C291: **Testing for contagion: a time scale decomposition**

Presenter: **Andrea Cipollini**, University of Modena and Reggio Emilia, Italy
Co-authors: Iolanda Lo Cascio

This paper tests for shift contagion, which is a temporary shift in the spillover effects of idiosyncratic shocks. Existing studies have concentrated on the time domain and identified a shift in the mutual, contemporaneous interaction between asset returns using identification schemes based upon structural innovations following a GARCH process. The potential endogeneity bias of contagion coefficients can be tackled by using lagged values of a dependent variable as instrumental variables. An alternative approach uses a Granger causality test in the frequency domain. In this paper, wavelets analysis is used to test for contagion. We show how a time scale decomposition of financial returns, and in particular, a decomposition of covariance on a scale by scale basis helps to identify the structural coefficients, including those capturing a potential shift in the contemporaneous interaction among asset returns.

C141: **Useful VAR-VECM representations for real-time data**

Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Jan Jacobs

Before being considered definitive, data currently produced by Statistical Offices undergo a recurrent revision process resulting in different releases of the same phenomenon. The collection of all these vintages is referred to as a real-time data set. Economists and econometricians have realized the importance of this type of information for economic modeling and forecasting. This paper clarifies the link between different representations for modeling the revision process within a multivariate dynamic time series framework. These approaches are the VAR, the VECM and alternative transformed VAR representations. We develop a coherent framework that encompasses seemingly unrelated time series analyses of the revision process that are present in the literature. We show how we can test in these alternative frameworks whether a first-release measure is unbiased. We particularly focus on the presence of co-movements between *diagonal* releases. It emerges that the presence of cointegration and common cyclical features is quite important for studying such process. Moreover we also introduce to this type of data a co-breaking analysis that is seen to be necessary to jointly estimate the revision process and sudden shifts in the measure of series. An example is the introduction of a new base year.

C366: **Downside risk of derivative portfolios with mean-reverting underlyings**

Presenter: **Patrick Leoni**, University of Southern Denmark, Denmark

A Monte-Carlo simulation of a standard portfolio management strategy involving derivatives is performed inn order to estimate the sensitivity of its downside risk to a change of mean-reversion of the underlyings. It is found that the higher the intensity of mean-reversion, the lower the probability of reaching a pre-determined loss level. This phenomenon appears of large statistical significance for large enough loss levels. It is also found that the higher the mean-reversion intensity of the underlyings, the longer the expected time to reach those loss levels. The simulations suggest that selecting underlyings with high mean-reversion effect is a natural way to reduce the downside risk of those widely traded assets.

C256: **The dynamical relation between ad liking and memorial response to advertising**

Presenter: **Marzia Freo**, University of Bologna, Italy
Co-authors: Sergio Brasini, Giorgio Tassinari

The paper addresses the question of carryover effects of advertising (ad) liking on the recall, jointly modeling the patterns of recall, ad pressure and ad liking, by means of the specification of a vector autoregressive model with GRPs acting as exogenous variable. The approach is innovative since marketing literature has mainly investigated until now only the simultaneous relationship between advertising, recall and liking. On the other hand, multiple time series models have been applied to capture dynamical relationships between marketing mix variables and sales, while the relation between the different cognitive and affective facets of response to advertising have never been explored. The analysis is carried out for the markets of small automobiles, deodorants and shampoos. Main empirical findings for the analyzed categories highlight that: carryover effects of ad liking on the recall measures may be detected but not systematically, and the ad liking role of ad likeability on memorial responses varies among product categories classified as approach or avoidance. Moreover a further finding shows that, whereas positive influences

are thoroughly retrievable (in the approach category small car), ad likeability influences more advertising than brand awareness and more total than unaided awareness.

CS47 Room 5 TIME SERIES FINANCIAL ECONOMETRICS 3 Chair: Ana-Maria Fuertes

C252: **Portfolio diversification opportunities in eastern Europe**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Co-authors: Nektarios Aslanidis

The advent of the European Union has decreased the diversification benefits available from country based equity market indices in the region. The increase in stock integration between the three largest new EU members (Hungary, the Czech Republic and Poland who joined in May 2004) and the Euro-zone is measured. A potentially gradual change in correlation between stock markets is allowed. This seems particularly appropriate to analyse the increasing integration between the Eastern European and the Euro-zone stock markets over the recent years. We motivate the idea of the correlation across stock markets by adopting a simple economic model of correlations. At the country market index level all three Eastern European markets show a considerable increase in correlations in 2006. At the industry level the dates and transition periods for the correlations differ, and the correlations are lower although also increasing. The results show that sectoral indices in Eastern European markets may provide larger diversification opportunities than the aggregate market.

C280 : **Forward premium anomaly, realized volatility and jump process in foreign exchange markets**
Presenter: **Young Wook Han**, Hallym University, Korea

By using the daily Dollar-Euro and Yen-Dollar spot and overnight forward exchange rates, this paper reexamines the issue of the forward premium anomaly which has attracted widespread attention in international finance. In particular, this paper investigates the evidence for the role of realized volatility and jump process in the tests of the forward premium anomaly, which have mostly ignored in previous tests. For the purpose, this paper adjusts the usual regression model for the tests of the forward premium anomaly allowing for the realized volatility and the jump process in. After re-estimating the adjusted regression model, this paper finds that the estimated value of the forward premium coefficient in the Yen-Dollar currency market is close to the theoretical value of unit implying that the forward premium anomaly seems to disappear while the coefficient in the Euro-Dollar currency market is still negative so that the anomaly exists but the degree is not as severe as before. Thus, this paper presents the possibility that the empirical phenomenon of the forward premium anomaly in foreign exchange markets may not be as robust as before if the test regression model is specified more appropriately.

C238: **Semiparametric conditional quantile models for financial returns and realized volatility**
Presenter: **Filip Zikes**, Imperial College London, UK

This paper investigates how the conditional quantiles of future returns and volatility of financial assets vary with various measures of ex-post variation in asset prices as well as option-implied volatility. We work in the flexible quantile regression framework and rely on recently developed model-free measures of integrated variance, upside and downside semivariance, and jump variation based on high-frequency data, as covariates. We provide sufficient conditions ensuring that the measurement error associated with the realized measures vanishes in the limit and hence does not affect consistency of the quantile regression estimator. Our empirical results for the SP500 and WTI Crude Oil futures contracts show that simple quantile regressions for returns and heterogenous quantile autoregressions for realized volatility perform very well both in-sample as well as out-of-sample in capturing the dynamics of the respective conditional distributions. The realized downside semivariance is found to be particularly informative for quantiles of future returns as well as realized volatility, while the contribution of jumps seems to be rather small. The models can serve as useful risk managements tools for investors trading the futures contracts themselves or various derivative contracts (e.g. variance swaps) written on the realized volatility corresponding to these contracts.

C096: **The time series properties of annual earnings: new evidence from an ESTAR unit root test**
Presenter: **Andros Gregoriou**, University of East Anglia, UK
Co-authors: Len Skerratt

In this paper we examine the stationarity of earnings, within the context of a panel consisting of 479 firms listed on the London Stock Exchange over the time period 1984-2003. Using standard ADF tests, we find that the annual earnings process is stationary for only 27% of the sample firms. We present new tests of stationarity, which explicitly allow for the possibility that earnings can be characterised by a non-linear mean-reverting process. Once we allow for a non linear adjustment, the earnings process is stationary for 76% of the sample firms. A final contribution is to discuss the implications of our findings for empirical models which involve earnings as an explanatory variable.

C138: **Information flows around the globe: predicting opening gaps from overnight foreign stock price patterns**

Presenter: **Lukasz Gatarek**, Erasmus University Rotterdam/ Tinbergen Institute, Netherlands
Co-authors: Jan De Gooijer, Cees Diks

This paper describes a forecasting exercise of close-to-open returns on major global stock indices, based on price patterns from foreign markets that have become available overnight. As the close-to-open gap is a scalar response variable to a functional variable, it is natural to focus on functional data analysis. Both parametric and non-parametric modeling strategies are considered, and compared with a simple linear benchmark model. The overall best performing model is nonparametric, suggesting the presence of nonlinear relations between the overnight price patterns and the opening gaps. This effect is mainly due to the European and Asian markets. The North-American and Australian markets appear to be informationally more efficient in that linear models using only the last available information perform well.

C289: **Exploiting intra-day prices, jumps and subsampling in daily VaR prediction**

Presenter: **Ana-Maria Fuertes**, City University London, UK
Co-authors: Jose Olmo

This paper illustrates empirically the importance of exploiting high frequency based volatility measures and, relatedly, of acknowledging jumps in returns from the point of view of assessing risk exposure. For this purpose, it compares several VaR models in a univariate context for 14 large NYSE stocks. The benchmark is the standard ARMA-GARCH process augmented by two nonparametric estimators, realized variance and realized bipower variation, which differ in that the former can reflect jumps whereas the latter does not. The loss function used for the comparison of forecasts is the economic one implicit in VaR backtesting. A novel unconditional coverage testing approach is used which is robust to estimation and model risk. We find that neglecting jumps may result in significant risk overestimation and autocorrelation in the sequence of VaR exceptions. Nevertheless, the VaR predictive ability of GARCH is not improved upon by exploiting intraday price information nor by assuming a fat-tailed distribution relative to the standard normal assumption commonly adopted by practitioners.

Saturday 31.10.2009 14:30-16:30 Parallel Session L

ES06 Room 5 STATISTICAL ALGORITHMS AND SOFTWARE Chair: Uwe Ligges

E220: **Missing data imputation by sequential decision trees**
Presenter: **Agostino Di Ciaccio**, University of Roma "La Sapienza", Italy
Co-authors: Giovanni Maria Giorgi

A procedure based on the use of sequential regression and classification trees with bootstrap for the imputation of missing data is proposed. The case of non-monotone patterns of missing data with mixed measurement level of the variables is considered. The aim of the analysis is to obtain a completed data matrix with optimal characteristics (with respect to means, variances and correlations of the variables) which is often the main demand for a statistical office. Moreover we want to obtain a measure of the additional variability due to the presence of missing values. A simulation case with qualitative and quantitative data is analysed and the results compared with other procedures. In particular, the performance of Multiple Imputation, IVEWARE and MIDAS were compared with our procedure using the simulation and the EU-SILC cross-sectional data. Our non-parametric method showed to be very competitive on these data, where the other methods show some faults.

E086: **An application on structural time series using GNU Octave on a cheap optimized Linux computer cluster**
Presenter: **Riccardo Corradini**, ISTAT, Italy

This paper describes how it is possible to estimate unobserved components on Structural Time Series by a popular matrix programming language very similar to MATLAB available for GNU OCTAVE. Unfortunately, this software does not have specific sets of instructions to analyse State-Space Estimation of Econometric Models. Moreover there is no documented effort to achieve parallel computation concerning the Kalman filtering and smoothing equations by a Linux cluster using ATLAS (specialized Fortran libraries written for a specific hardware platform linked with GNU OCTAVE). We shall see how, using some Montecarlo simulations conducted in parallel, it will be possible to verify the numerical accuracy, robustness of parameters when they are compared with real life examples. By further improvements, GNU OCTAVE scripts could be translated into C++ language achieving more speed on execution. Finally it will be demonstrated that the use of sparse matrix algorithms will speed up Kalman filtering and smoothing equations. Any State Space model with a time linear constraint (i.e. with an indicator about quarterly accounts), spatial linear constraint (i.e. a panel data model about GDP of Italian regions subject to a national aggregate value), log-linear time or spatial constraint and nested stochastic regressors could be estimated. If on one hand the computational burden is very heavy, on the other the researcher could neglect the hardware and software cost problem concentrating only on subsequent stages of computational difficulties.

E085: **Maximum likelihood estimation for parameters of stable Paretian distribution: Implementation in R**
Presenter: **Dedi Rosadi**, Gadjah Mada University, Indonesia

Its has been widely known that many large empirical data sets from diverse fields of studies, for instance from telecommunications and network traffics, signal processing, physics and finance, are found to be leptokurtic, i.e., heavy-tailed and peaked around the center. An important and attractive class of distributions for modeling heavy-tailed data is the stable distribution. Unfortunately, the probability distribution function (pdf) of the stable distributions can not be written in a closed form, except for a few cases. Therefore, one has to rely on the numerical method to obtain its pdf. Furthermore, all statistical methods that rely on the existence of closed-form pdf, such as the maximum likelihood estimator, are difficult to implement and time consuming in practical work. Some algorithms have been proposed to calculate the pdf of stable non normal (stable Paretian) such as direct integration of characteristics functions, Fast Fourier Transform, etc. In this paper, we present an implementation of the stable pdf and MLE calculation using R version 2.$\star$ and compare its performance with the existing implementations in literature.

E204: **ROC.Regression: an R package for ROC regression analysis**
Presenter: **Maria Xose Rodriguez-Alvarez**, University of Santiago de Compostela, Spain
Co-authors: Ignacio Lopez de Ullibarri, Carmen Maria Cadarso-Suarez

The receiver operating characteristic (ROC) curve is a widely used tool for characterizing the accuracy of a diagnostic test in distinguishing between two states. Various ROC regression approaches have been proposed in order to account for covariates that might influence the test accuracy. So far, the scarcity of implemented ROC regression software is probably responsible for the lack of popularity of these models in the biomedical community. Therefore, we have developed an user-friendly R package, called *ROC.Regression*. In this software, different approaches to ROC regression analysis have been implemented. This software can be used to fit ROC regression models for a set of continuous and/or categorical covariates, and their possible interactions. From the estimated ROC, other summary measures of the accuracy, such as the area under the curve (AUC) and the generalized Youden index (YI) can be obtained and also the thresholds values based on the YI criterion. Numerical output includes coefficient estimates, standard errors and p-values, whereas graphical output offers ROC, AUC, YI and Threshold curves. To illustrate usage of the program we analyse data from a computer-aided diagnostic (CAD) system dedicated to early detection of breast cancer.

E212: **tuneR - vibration and sound analyses in R**

Presenter: **Uwe Ligges**, TU Dortmund, Germany
Co-authors: Sebastian Krey

The R package tuneR was originally intended for the analysis of music time series. It has been extended with more methods for frequency estimation. It is now capable of analyzing huge data that does not even fit into memory. Moreover, it includes useful methods for a number of different application such as analysis of speech. Besides summarizing the already existing functionality of the package, this talk will focus on new algorithms for transcription, timbre recognition and speech processing as well as vibration detection in technical applications. For speech analysis and timbre recognition, some popular algorithms, e.g. for LPC, PLP, and MFCC calculations, will be implemented at the time this talk is given. The quality of the algorithms will be compared and demonstrated in some brief examples. First example is a classification task where 59 different classes of instruments have to be distinguished given features derived from the data that has been pre-processed by tuneR. A second example is from an application in music psychology. Here we show how to transcribe pieces of music recorded on the streets in noisy environment, sung by amateur singers. Finally, we will present how to analyze a huge 4 Gbytes dataset recorded by a machine's vibration sensor.

ES19 Room 6 ALGORITHMS AND APPLICATIONS OF ROBUST METHODS Chair: Mia Hubert

E015: **Robustified least squares support vector classification**

Presenter: **Tim Verdonck**, University of Antwerp, Belgium
Co-authors: Michiel Debruyne, Sven Serneels

A group of methods which is recently receiving increasing attention in classification problems are Support Vector Machines (SVM). In complex high-dimensional data one often has to deal with outliers. For SVM classification a recent theoretical result shows that such outliers can have an arbitrary large effect on the classifier for certain types of kernels. Also in practice it is easily observed that SVM classification can severely suffer from outlying samples. Here we propose an outlier resistant SVM method for any particular choice of kernel. To obtain robustness we introduced a strategy to specify weights in a general kernel induced feature space. These weights are used to adjust the Least Squares SVM (LS-SVM). We choose the least squares loss since in that case a reweighting step can easily be implemented without significantly increasing the computation time. A specific tuning procedure is proposed to choose the additional parameter controlling robustness. The resulting classifier is called Robustified Least Squares Support Vector Machine and denoted as RLS-SVM. An extensive simulation study shows that the RLS-SVM yields better classification performance for heavily tailed data and data containing outliers. From the real examples, it can be concluded that RLS-SVM is competitive to LS-SVM on clean data.

E023: **Building a robust calibration model for heterogeneous spectral data**

Presenter: **Sabine Verboven**, University of Antwerp, Belgium
Co-authors: Mia Hubert, Peter Goos

Predicting the humidity level from a large set of NIR spectra, measured on different samples of animal feed, is one of the main priorities in one of Belgium's largest animal feed companies. An accurate prediction of the humidity level allows for an effective quality control as well as a substantial reduction in production cost of the feed. Because the data set is high-dimensional and might contain outliers (due to the heterogeneity of the samples) we estimate the regression parameters with a robust Principal Component Regression method and with a robust Partial Least Squares regression method. In order to obtain the optimal calibration model we introduce and compare several robust preprocessing techniques. Also, different robust prediction errors were defined and a desirability index was used to determine the best model with respect to the root mean squared error of prediction (RMSEP). The resulting calibration model yields prediction errors that are 15% smaller than the black-box calibration method applied by the company so far.

E103: **Detecting influential data points in extreme value statistics**

Presenter: **Dina Vanpaemel**, Katholieke Universiteit Leuven, Belgium
Co-authors: Mia Hubert, Goedele Dierckx

In extreme value statistics, the Extreme Value Index (EVI) is used to characterize the tail behavior of a distribution. This real-valued parameter helps to indicate the size and frequency of certain extreme events under a given probability distribution: the larger the EVI is, the heavier the tail of the distribution. The Hill estimator is a well known estimator for the EVI. We present a new method for measuring the influence of the individual observations on the estimation of the EVI. Our approach is based on the empirical influence function (EIF) of the Hill estimator. As the EIF uses an estimate of the EVI, it is important to insert a robust estimate of the EVI. Otherwise the estimate itself will be very biased by highly influential data points. In this presentation, a robust estimator will be introduced based on a robust GLM estimator applied to the log-spacings of the data. Based on the asymptotic normality of this estimator, we can also derive cutoff values for automatically detecting highly influential data points.

E146: **Variable selection for kernel classification**
Presenter: **Sarel Steel**, University of Stellenbosch, South Africa
Co-authors: Nelmarie Louw, Surette Bierman

A variable selection procedure, called surrogate selection, is proposed which can be applied when a support vector machine or kernel Fisher discriminant analysis is used in a binary classification problem. Surrogate selection applies the lasso after substituting the kernel discriminant scores for the binary group labels, as well as values for the input variable observations. Empirical results are reported, showing that surrogate selection performs well. The underlying idea is general enough to make its extension to variable selection in other classification and regression contexts feasible.

E096: **A faster deterministic algorithm for the MCD**
Presenter: **Mia Hubert**, Katholieke Universiteit Leuven, Belgium
Co-authors: Peter Rousseeuw, Tim Verdonck

The Minimum Covariance Determinant estimator is a well-known highly robust estimator of location and scatter of multivariate data, defined as the mean and covariance matrix of the h-subset with smallest covariance determinant. Besides being highly resistant to outliers, the MCD is affine equivariant, which allows for any affine transformation on the data. This is a strong and often desirable property, but it has the disadvantage that the computation of the MCD is very hard. Currently the FAST-MCD algorithm is mostly used. Roughly summarized, this algorithm starts by randomly drawing many elemental subsets and then applies C-steps until convergence. Due to this random sampling mechanism, the algorithm is not deterministic. In this talk we present a new algorithm for the MCD. It is fast, deterministic and permutation invariant, but not fully affine equivariant. It is essentially based on starting from a few h-subsets which are easily computed, and then iterating C-steps until convergence. We discuss the performance of the new algorithm on several data sets and outlier configurations.

ES23 Room 7 PORTFOLIO OPTIMIZATION, HEURISTICS AND RISK MEASURES (ANSET) Chair: Peter Winker

E176: **Cardinality versus q-Norm constraints for index tracking**
Presenter: **Bjoern Fastrich**, University of Giessen, Germany
Co-authors: Sandra Paterlini, Peter Winker

Index tracking aims at replicating a given benchmark with a smaller number of assets. Different quantitative models can be set up to determine the optimal index replicating portfolio. The most common approach is to tackle the problem as an optimization problem with a cardinality constraint, i.e. minimizing a given distance measure between the index and the replicating portfolio that consists of at most K assets. In this work, we propose a possible alternative based on imposing a constraint on the q-norm, $0 < q < 1$, of the replicating portfolios' asset weights: the q-norm constraint regularizes the problem and identifies a sparse model. Both approaches are challenging from an optimization viewpoint due to either the presence of the cardinality constraint or a non-convex q-norm constraint. Furthermore, the problem can become even more complex when non-convex distance measures or other real-world constraints are considered. Optimization heuristics have already shown remarkable performance in index tracking with cardinality constraints. We propose to employ a hybrid heuristic that can provide a flexible tool to tackle both optimization problems. Finally, the empirical analysis on real-world financial data, such as the German stock index DAX100, allows to compare the pros and cons of the two index tracking approaches.

E115: **Solving a discrete mean-variance-skewness portfolio selection model using multiobjective evolutionary algorithm**
Presenter: **Georgios Mamanis**, Democritus University of Thrace, Greece
Co-authors: Konstantinos Anagnostopoulos

We consider a large scale mean-variance-skewness model with additional real-world constraints which limit the number of assets in the portfolio and the amount invested in these assets. The resulting model is a large scale mixed-integer non-linear multi-objective optimization problem. Furthermore, by incorporating a preference for skewness, the bidimensional efficient frontier is transformed in a three dimensional surface which provides all possible trade-offs between mean, variance and skewness. In this study the three objectives are optimized simultaneously, i.e. expected return and skewness are maximized while variance is minimized. Standard optimization algorithms cannot deal with the problem at hand in reasonable time. On the contrary, multi-objective evolutionary algorithms (MOEA) have already shown remarkable performances in large scale multiobjective optimization problems. Their main advantage is that they generate reasonably good approximations of the efficient surface in a single run and within limited computational effort. Empirical results demonstrate that the MOEA can provide a set of efficient portfolios well capturing the trade-offs among the objectives satisfying different investors' risk preferences.

E064: **On the impact of concordance measures in portfolio selection theory**
Presenter: **Sergio Ortobelli Lozza**, VSB-Technical University Ostrava, Czech Republic
Co-authors: Tomas Tichy

In portfolio several concordance measures (such as Spearman rho, Kendall tau, Gini gamma, Pearson coefficient of linear correlation) have been proposed to value the dependency among random returns. Clearly each measure can have a different impact in valuing the dependence of returns. We discuss and value the impact of different concordance measures in portfolio selection problems. In particular, we propose new concordance measures that capture the tail dependency and asymmetry of returns series. Moreover we compare the ex-post sample paths of wealth obtained adopting portfolio selection strategies that use different concordance measures.

E184: **Exact and heuristic approaches to the index tracking problem with hard real-world constraints**
Presenter: **Andrea Scozzari**, University of Rome "La Sapienza", Italy
Co-authors: Fabio Tardella, Thiemo Krink, Sandra Paterlini

One of the most common problems in asset management is the so-called index tracking problem. The aim is to select a portfolio which replicates an index with a small number of assets while simultaneously has certain attractive characteristics, for instance: low turnover, low transaction and rebalancing costs and so forth. Index tracking can be formulated as a mixed integer linear programming problem. Furthermore, it often happens that there are non-linear constraints that need to be satisfied in practice such as the concentration limits of the UCITS (Undertaking for Collective Investments in Transferable Securities) rules, which are binding for most portfolios in the EU. However, no linear formulation of the UCITS rule compliance seems to have been proposed until now. We propose a mixed integer quadratic programming formulation for the index tracking problem with the UCITS rule compliance and use such formulation to obtain exact solutions for this problem up to a certain size based on real-world datasets. We then compare the results of the exact method to those obtained with a stochastic search heuristic in order to assess its quality, and then present the results of the heuristic on problems based on larger datasets that cannot be solved to optimality with current state-of-the-art mixed integer programming software. The analysis on real-world datasets with small and large number of assets shows the effectiveness and the efficiency of the proposed methods, providing best practice guidelines for asset managers.

E191: **Multimodal optimization for financial portfolio selection with evolutionary algorithms**
Presenter: **Sandra Paterlini**, University of Modena and Reggio E., Italy
Co-authors: Thiemo Krink, Tommaso Minerva, Massimo di Tria

Multiobjective Portfolio Optimization aims at determining the Pareto front of optimal asset allocations with respect to different objective functions, which usually quantify the risk and the return of the desired portfolios. The Markowitz's mean-variance approach is probably the most well-known and widely used in practice. However, the standard Markowitz model cannot handle realistic risk measures and typical real world constraints. We propose a method, based on evolutionary algorithms that can deal with a whole range of optimization problems, without requiring any mathematical properties of the objective functions (i.e. return/risk measures) and/or of the constraints (e.g. convexity, monotonicity). The population of candidate solutions searches simultaneously for a set of solutions that represents the entire Pareto front. We then use multimodal optimization in order to identify more than one optimal portfolio for each risk/return profile. Hence, the new tool does not propose one arbitrary asset allocation for each front point among (possibly) many but a whole variety of alternative optimal investment opportunities. Our analysis with real-world data shows that the proposed approach is not only valuable as a decision support tool in strategic asset allocation but also in tactical asset allocation.

ES30 Room 3 FUZZY STATISTICAL ANALYSIS 3 Chair: Renato Coppi

E048: **Multi-label learning using Dempster-Shafer theory**
Presenter: **Zoulficar Younes**, Universite de Technologie de Compiegne, France
Co-authors: Fahed Abdallah, Thierry Denoeux

Multi-label learning has been increasingly required by modern applications where it is quite natural that some instances belong to several classes at the same time. A first approach, referred to as the indirect multi-label approach, consists in transforming a multi-label problem into a traditional classification task after converting the multi-label datasets in single-labelled ones. A second approach referred to as direct multi-label approach, consists in extending common single-label learning algorithms and making them able to manipulate multi-label data directly. In this paper, we present a comparative study between the two approaches using the *k* nearest neighbor rule. In addition, in order to handle uncertainty for decision making, the Dempster-Shafer theory of belief functions is used. For the first approach where the data are single-labelled, the uncertainty is represented by evidence on multiple hypotheses where each hypothesis is a class from a set of disjoint classes Ω to be assigned or not to an unseen instance. In contrast, when handling multi-labelled data directly, each hypothesis represents a set of labels and the uncertainty is then

expressed by evidence on sets of label sets. In the direct multi-label approach, instead of defining mass functions on the frame 2^{2^Ω}, we restrict the definition on a subset of 2^Ω that is closed under intersection and having much smaller size than 2^{2^Ω} while being rich enough to express evidence in many realistic situations. The performances of different approaches are demonstrated by experiments using several benchmark datasets.

E102: **A linear regression model with LR fuzzy random variables**
Presenter: **Maria Brigida Ferraro**, Sapienza University of Rome, Italy
Co-authors: Paolo Giordani

In standard regression analysis the relationship between one (response) variable and a set of (explanatory) variables is investigated. In a classical framework the response is affected by probabilistic uncertainty (randomness) and, thus, treated as a random variable. However, the data can be also subjected to other kinds of uncertainty, such as imprecision, vagueness, etc. A possible way to manage all of these uncertainties is represented by the concept of fuzzy random variable (FRV). The most common class of FRVs is the LR family, which allows us to express every FRV in terms of three random variables, namely, the center, the left and the right spread. In this work, limiting our attention to the LR FRVs, we address the linear regression problem in presence of one or more imprecise random elements. The procedure for estimating the model parameters is discussed, and the statistical properties of the estimates are analyzed. Furthermore, in order to illustrate how the proposed model works in practice, the results of some case-studies are given.

E111: **Possibilistic clustering for fuzzy data**
Presenter: **Paolo Giordani**, Sapienza University of Rome, Italy
Co-authors: Renato Coppi, Pierpaolo D'Urso

Given a set of objects on which some quantitative features are observed, the fuzzy k-means clustering model (fkm) is a powerful tool for classifying the objects into a set of k homogeneous clusters. This is done by means of the membership degrees of a object in a cluster, which expresses the extent to which an object belongs to each and every cluster. In fkm, for each object, the sum of the membership degrees in the clusters must be equal to 1. Unfortunately, such a constraint may cause meaningless results. To avoid this drawback, it is possible to relax the constraint leading to the so-called possibilistic k-means clustering model (pkm). In the present work we consider the case in which the information at hand is affected by imprecision. Namely, the objects to be clustered are characterized by some imprecise features suitably managed in terms of fuzzy sets. The fkm and pkm procedures are then adapted for coping with such imprecise data.

E167: **A dependent multi-sample test for fuzzy means**
Presenter: **Manuel F. Montenegro**, University of Oviedo, Spain
Co-authors: Teresa Lopez-Garcia, Asuncion Lubiano, Gil Gonzalez-Rodriguez

An asymptotic procedure to test equality of means of k fuzzy random variables (FRVs) measured on the same population (k-dependent sample test) is presented. In previous works, the independent multi-sample test and the paired two-sample test for the fuzzy mean have been developed by taking advantage of well-known results in Hilbert spaces. On the basis of such results, we find the null distribution of the natural statistics based on the sum of the distances of the sample mean of each population to the overall average. As expected, the limit distribution of the suitably normalized statistic under the null hypothesis is a function of a Gaussian process, and can be used to develop a testing procedure for large samples. For illustrative purposes we compare the approach with the analogous obtained for both paired two-sample and independent multi-sample test. Since there are no realistic parametric models for multi-dimensional FRVs, the dependence between the different FRVs in the simulations is fixed by copulas. Empirical results for different degrees of dependence are shown. Finally, we apply the approach to evaluate if the expected value of different fuzzy random variables employed to model the answers in a survey is the same on a given population.

E024: **Learning fuzzy rules with arbitrary reference functions using GSVM**
Presenter: **Christian Moewes**, University of Magdeburg, Germany
Co-authors: Rudolf Kruse

How to learn fuzzy rules from observations is still an open question in machine learning. In the last years, some researchers tried to generate them from Support Vector Machines (SVMs). Using the SVM formulation, however, only a certain subclass of reference functions is theoretically suitable to obtain fuzzy rules. We suggest to use the Generalized SVM since it enables to use arbitrary reference functions. Experiments on benchmark datasets show the manifoldness of the proposed method.

CS03 Room 4 MULTIFRACTAL VOLATILITY Chair: Laurent E. Calvet

C174: **Multifractal scaling in the interest rate term structure**
Presenter: **Adlai Fisher**, University of British Columbia, Canada
Co-authors: Laurent Calvet, Liuren Wu

We develop a flexible yet parsimonious no-arbitrage model of the interest rate term structure using restrictions derived from multifractal scaling laws. Using as few as five parameters and a high-dimensional state space, we represent the term structure of interest rates at any point in time to near perfection. The parameters of the model are precisely estimated, and out-of-sample forecasts show substantial improvement over prior specifications.

C194: **Relative forecasting performance of volatility models: Monte Carlo evidence**
Presenter: **Leonardo Morales-Arias**, University of Kiel / Kiel Institute for the World Economy, Germany
Co-authors: Helmut Herwartz, Thomas Lux

We conduct a Monte Carlo (MC) experiment to study the relative performance of the Markov-Switching Multifractal Model (MSM), the (Fractionally Integrated) Generalized Autoregressive Conditional Heteroskedasticity models ((FI)GARCH), and the Stochastic Volatility model (SV) for in-sample fitting and out-of-sample forecasting. The MC experiment consists on simulating data from the various models at hand (MSM, GARCH, FIGARCH, SV) and then estimating the models' parameters and forecasting volatility based on the estimated parameters under the alternative data generating processes (DGPs). The MC study enables to compare the relative forecasting performance of various models, which account for different characterizations of the latent volatility: specifications which incorporate short/long memory, autoregressive components and stochastic shocks (GARCH, FIGARCH, SVOL) against specifications which incorporate Markov-switching, multifractality and (apparent) long memory (MSM). Moreover, we may compare three popular volatility models (GARCH, FIGARCH, SVOL) vis-a-vis. MC complementarities between models are explored via forecast combinations. Forecasts are evaluated by means of Mean Squared Errors (MSE), Mean Absolute Errors (MAE) and Value-at-Risk (VaR) diagnostics. We find that: (i) the MSM model best forecasts volatility under any other alternative characterization of the latent volatility and (ii) forecast combinations provide an improvement upon single models' forecasts.

C203: **Equity skew and the Markov switching multifractal**
Presenter: **Marcus Fearnley**, HEC Paris, France
Co-authors: Laurent Calvet, Adlai Fisher, Markus Leippold

The MSM model parsimoniously captures the persistence of asset return volatility over multiple frequencies. It has been shown to out-perform alternative long-memory models including FI-GARCH and MS-GARCH, both in– and out–of–sample. We build upon the econometric performance of the MSM model by specifying several extensions that are able to capture the skewness of equity returns and the 'leverage effect'. The new specifications are tightly parameterized and the likelihood functions are available in closed form. Likelihood-based comparisons and VaR forecasts indicate that the new specifications significantly outperform other leading models, including affine and non-affine stochastic volatility and GARCH type models.

C136: **(Re)correlation: a Markov switching multifractal model with time varying correlations**
Presenter: **Julien Idier**, Banque de France, France

The paper develops a Markov switching multifractal model with dynamic conditional correlations. The aim is to give more flexibility to the initial bivariate Markov switching multifractal [MSM] model by introducing some time dependency in the comovement structure. The new defined model is applied to stock index data (CAC, DAX, FTSE, NYSE) between 1996 and 2008 and compared to both the standard MSM and the DCC model. The MSMDCC models present better fit than the MSM and DCC models, and the formulation of the correlations gives good insight in its economic interpretation.

C171: **Multifractal volatility: theory, forecasting and pricing**
Presenter: **Laurent Calvet**, HEC Paris, France
Co-authors: Adlai Fisher

Multifractal volatility models efficiently capture seemingly disparate aspect of asset returns and strongly outperform the forecasting accuracy of traditional approaches. The Markov-switching multifractal (MSM) is based on the observation that financial markets are affected by shocks with highly heterogeneous degrees of persistence, ranging from intraday intervals to decades. While traditional approaches such as GARCH consider one or sometimes two persistence levels, MSM proposes a pure regime-switching approach that incorporates volatility fluctuations on an arbitrarily large range of frequencies. The model has only four parameters even if it has many volatility components and a very large number of states. The parsimony makes MSM a strong performer in forecasting return volatility out of sample. The next step is to consider the equilibrium implications of multifrequency risk in fundamentals, a highly pervasive but hitherto little studied feature of market economies. Multifrequency regime-switching

is conveniently embedded into a Lucas tree economy. The MSM equilibrium model captures the extreme realizations of actual equity returns, as well as the excess volatility of returns compared with the volatility fundamentals. In continuous time, we find that exogenous jumps in fundamental volatility causes endogenous jumps in asset prices, providing an attractive economic explanation of the jump correlations.

CS26 Room 1 BAYESIAN ECONOMETRICS 2 Chair: Yasuhiro Omori

C051: **Bayesian estimation of the cost of equity with a hierarchical prior**
Presenter: **Teruo Nakatsuma**, Keio University, Japan

We propose a new hierarchical Bayes approach for estimating the cost of equity in a specific industry. Our approach utilizes the well known relationship among firm-specific levered betas and the industry-specific unlevered beta to construct a hierarchical prior of parameters in a multi-factor model of stock returns. The Bayesian estimation procedure is conducted with the standard Gibbs sampling. We apply the new approach to the Japanese automobile industry and perform sensitivity analysis on the subjective belief about the future market performance and other prior information.

C144: **Uncertainty in asset correlation for portfolio credit risk: the shortcomings of the Basel II framework**
Presenter: **Carlos Castro**, Universite Libre de Bruxelles, Belgium

Moody's databases of corporate issuers of long term bonds and structured products are used to estimate asset correlations across a group of sectors, world regions and products. The estimation of a dynamic factor model for default risk is performed using Bayesian methods. Results indicate that a two factor model rather than the one factor model, as proposed by the Basel II framework, better represents the historical default data. Furthermore, the results reinforce the importance of unobserved factors in this type of models and point out that the levels of the implied asset correlations critically depend on the latent state variable used to capture the dynamics of default, as well as other assumptions on the statistical model. Finally, the posterior distributions of the asset correlations show that the Basel recommended bounds, for this parameter, undermine the level of systemic risk.

C307: **The effects of parameter uncertainty and model risk in interest rate models**
Presenter: **Paul Koerbitz**, Ulm University, Germany

Model risk arising from uncertainty in model and parameter choice is an important risk faced by financial institutions. However, in historically estimated short-rate models, the amount of model and parameter uncertainty and the risks implicated by it are not well understood. A Bayesian methodology is applied in order to derive a framework in which the risks arising from parameter and model uncertainty can be assessed. Specifically, joint posterior distributions of the model parameters for the Vasicek and Chan-Karolyi-Longstaff-Sanders model are derived via Gibbs-Sampling. The extent of parameter and model uncertainty, its implications for model results, and the validity of the method are demonstrated in a large simulation study. Furthermore, the method is applied to the forecasting of price distributions of interest rate derivatives and to fixed-income portfolio optimization. In both cases the impact of parameter and model uncertainty is demonstrated and measures to counter these effects are discussed. The results suggest that parameter and model uncertainty is an important source of risk and that our method provides ways to remedy the effects of such uncertainty.

C050: **Spatio-temporal dynamics in economic growth**
Presenter: **Kazuhiko Kakamu**, Chiba University, Japan
Co-authors: Hideo Kozumi

This paper considers the convergence hypothesis in economic growth by Markov transition approach. We propose a spatio-temporal varying transition probability model and construct a Markov chain Monte Carlo method to estimate the parameters of the model. Finally, using the Penn World Tables data, we examine the dynamics of per capita income in the world.

C324: **Assessing the efficiency of local government in Italy: a spatial productivity analysis**
Presenter: **Gianfranco Di Vaio**, University of Perugia and LUISS, Italy
Co-authors: Carlo Andrea Bollino, Paolo Polinori

We make a first attempt to analyze the spatial dynamics of local government efficiency in Italy. In doing this, we propose a two–step methodology. In the first step, we estimate the (bias–corrected) efficiency scores by means of a non-parametric Data Envelopment Analysis (DEA). In the second step, the efficiency scores are inserted in a Bayesian spatial tobit model, which allow us to identify the existence of spatial externalities. Results show that the efficiency performance is heterogeneous and positively correlated across space. In particular, efficient municipalities tend to be surrounded by efficient ones, and vice versa, probably due to knowledge diffusion across space. The paper is one of the first attempt to deal jointly with productivity and spatial analysis. Notably, the data employed in the empirical exercise are taken from a unique database which includes 341 municipalities from Emilia-Romagna.

CS27 Room 2 FINANCIAL MARKETS 3 **Chair: Elena Kalotychou**

C311: **Intraday price and volume information for volatility-based trading**
Presenter: **Natasa Todorovic**, Cass Business School, UK
Co-authors: Ana-Maria Fuertes, Elena Kalotychou

The role of intraday prices and daily volume in providing daily conditional volatility forecasts that are effective for individual equity trading is investigated. The analysis is applied to 14 SP500 stocks over 7 years. The benchmark is a GARCH equation fitted to daily returns which is augmented with several realised volatility measures or volume. Our findings document very low correlation between statistical accuracy and profitability of volatility forecasts. Profitability-type loss functions indicate that intraday prices do not add any valuable information that is not already incorporated in the baseline GARCH forecasts or in those augmented with volume, whereas statistical loss functions suggest the opposite. The best performing strategy involves buying the stock when its forecasted volatility is extremely high, suggesting a stronger volatility-return relationship in turbulent periods.

C065: **The impact of manager changes on UK fund performance and flows**
Presenter: **Svetlana Sapuric**, Cass Business School City University London, UK
Co-authors: Andrew Clare, Natasa Todorovic

Using our unique database of UK fund manager changes and event study methodology, we examine the impact of such changes on fund performance and fund flows. This allows us to establish whether this impact varies depending upon whether the fund manager is male or female; whether the fund is a developed or emerging market; and depending upon the fund's style, that is, growth, value or small capitalisation. Our results show clearly across different categories of funds that a change in fund manager can have a significant impact on fund performance, at least in the first year following the event. Additionally, we find greater persistence in performance of the bottom performing funds compared with the top performing funds. Finally, our evidence proves that managers' gender, the market or the type of asset they invest in do not influence the level of fund flows, however the change of the fund manager and the past performance do.

C318: **Modeling volatility and conditional correlations**
Presenter: **Malgorzata Doman**, Poznan University of Economics, Poland

Multivariate volatility models are usually based on the assumption that the observations are equally spaced. However, in practice of modeling the daily returns, we deal with unequally spaced data due to weekends and holidays. Moreover, the irregularities in the observations from different markets are often incompatible. This is because of differences in the placing of non-trade days during a year in different countries. Here the aim is to evaluate how these discrepancies affect the outcome of the dependencies analysis. The results are based on an extensive simulation analysis. Assuming that the data generating process is a multi-GARCH or multivariate SV process, we try to examine to what extent the irregular gaps caused by holidays influence the results of modeling, and distort the structure of dependencies. Moreover, we investigate different data pre-treatments. There exist two main possibilities. The first is to remove all the incompatible observations. The second is to fill the existing gaps using some function of observed returns. We compare the modeling and forecasting results for a few of the possible approaches. The Model Confidence Set methodology is applied to comparing the forecasts, and the sensitivity analysis is provided.

C148: **On the estimation of dynamic conditional correlation models**
Presenter: **Olga Reznikova**, Universite Catholique de Louvain, Belgium
Co-authors: Christian Hafner

Financial time series models usually deal with high dimensional data. The time varying correlation model for vast dimensions estimates the parameters with a bias which increases considerably with the number of dimensions. A covariance matrix estimator, which is essential for the DCC model, is not well-conditioned for high dimensions. In this paper we propose to reduce the bias of parameters of the DCC model by using shrinkage to target method for the sample covariance matrix. As targets we use identity matrix, factor model matrix, equicorrelation matrix. Such an estimator of the covariance matrix is asymptotically optimal and has a simple interpretation. The superiority of this approach is demonstrated through a Monte Carlo study. Finally, we provide an illustrating example by applying the method on a financial data series.

C293: **Structured financial products and investor decision making**
Presenter: **Rosella Castellano**, University of Macerata, Italy
Co-authors: Roy Cerqueti

The recent financial crisis, among other things, emphasized the role of retail structured products since many buyers have experienced huge losses and, in some cases, have seen the investments completely wiped out. In both theory and practice there is no unique definition of the term structured product but some characterizing elements can be identified: they are issued by financial institutions and combine at least two different financial assets of which at least one is a derivative written on stocks, basket of

stocks, indices, multiple indices or other investment opportunities (some of which have been blamed for the current financial crisis). Structured equity products represent liabilities of the issuing institutions that select pay-off patterns appealing for the retail investors. As a consequence, the pay-off function of the structured products reveals information about the pay-off pattern demanded by investors. Evidences suggest that structured products tend to be overpriced and this is consistent with models where producers of financial products take advantage of uniformed investors by making these product complicated and over weighting the probabilities of attaining the maximum possible return. We show that simpler products should be preferred by investors and the decision to invest in structured products is puzzling from the stand point of decision theory.

CS43 Room 9 COMPUTATIONAL ECONOMETRICS: SIMULATION AND DYNAMICS Chair: Cathy Chen

C268: **The Yuima Project: a computational framework for simulation and inference of SDEs with jumps**
Presenter: **Stefano Iacus**, University of Milan, Italy

We will introduce a new object oriented framework, the Yuima Project, for simulation and inference of general stochastic differential equations. Most of the theoretical results in finance rely on the assumption that the underlying dynamics of asset prices, exchange or interest rates, etc are continuous time stochastic processes. These models are also at the basis of option pricing and option pricing often requires Monte Carlo methods. In turn, Monte Carlo requires a preliminary estimation of the model to simulate. The yuima package is an S4 framework for simulation and inference for multidimensional SDEs driven by either Wiener, Levy, etc. noise. The building blocks are the definition of the S4 class which describes the model in an abstract way and a series of S4 methods for simulation and inference. The definition of the model does not include the description of the data which may come in many forms (tseries, zoo, xts). The combination of the Model (jump diffusion, stochastic volatility, fractional BM), the Data and the sampling scheme (high frequency, non small Delta, tick data) and the inference problem (parametric/nonparametric estimation; change point analysis; covariance estimation; asymptotic expansion; estimation of functionals; etc) produces the optimal solution.

C153: **Explaining macroeconomic and term structure dynamics jointly in a non-linear DSGE Model**
Presenter: **Martin M. Andreasen**, Bank of England, UK

This paper estimates a fairly standard DSGE model extended with recursive preferences and long-run risk on post-war US data. We show that this model can reproduce the dynamics in the 10 year yield curve with almost the same degree of precision as reduced form term structure models. The model also generates a 10 year term premium with a mean value of 69 basis points and reproduces the dynamics of four key macro variables. A decomposition of the variation in the yield curve shows that the level factor is explained by a combination of stationary labour productivity shocks and shocks to firms' fixed costs. The slope factor and the curvature factor are explained by a combination of shocks to the inflation rate target and firms' fixed costs.

C337: **Cartesian genetic programming approach to find a best regression model between credit default swap spreads and bond yields**
Presenter: **Laleh Zangeneh**, University College London, UK
Co-authors: Peter Bentley

Credit default swap (CDS) has been blamed globally for the recent credit crisis but more research is vitally needed to analyze and define its impact more precisely in financial market. CDS is a swap contract in which the buyer of the CDS makes a series of payments to the seller and, in return, receives a payoff if a credit instrument goes into default (fails to pay). The relationship between CDS spreads and bond yields is examined. The CDS spreads and related company bond prices are collected from financial market data provider (Reuters). We propose a Cartesian genetic programming (CGP) approach to find a best regression model among all candidates in order to model the relationship between credit default swap spreads and bond yields. CGP represents a program as a directed graph. The advantage of this type of representation over tree-based genetic programming representations is that it allows the implicit re-use of nodes, as a node can be connected to the output of any previous node in the graph. Thereby allowing the repeated re-use of sub-graphs. This paper compares the new proposed model results with the current statistical approaches and evaluates its performance.

C176: **Stochastic relaxation algorithms for the analysis of regional economic growth**
Presenter: **Paolo Postiglione**, Chieti-Pescara, Italy
Co-authors: M. Simona Andreano, Roberto Benedetti

The methods for spatial data analysis are often based on the assumption of stationarity of the estimated parameters. This hypothesis is patently violated when the data are characterized by information relative to predefined but unknown sub-groups of the reference population. It is clear that for spatial data which follow this hypothesis, the main analytic issue is not to estimate the model parameters or to introduce a structural dependence among observations, but to identify the geographical units where the parameters model are homogeneous. In this paper, the aim is the analysis of the concept of convergence clubs. Different strate-

gies are adopted in the definition of the model used in the objective function of the algorithm. The first is the classical non-spatial conditional beta-convergence model. The others are modified beta-convergence models which take into account the dependence showed by spatially distributed data. Three different stochastic relaxation algorithms for the identification of convergence clubs are proposed and compared: Simulated Annealing, Generalized Simulated Annealing and Iterated Conditional Model (ICM). The algorithm is then applied to 187 European regions for the period 1981-2004. Given the adaptability of the algorithms, they can be applied to any regression model in the analysis of heterogeneous spatial data.

E228: **Sequential parameter optimization applied to evolutionary strategies for portfolio optimization**
Presenter: **Oliver Flasch**, Cologne University of Applied Sciences, Germany
Co-authors: Thomas Bartz-Beielstein

Sequential parameter optimization (SPO) is a heuristic that combines classical and modern statistical techniques to improve the performance of search algorithms. Although sequential parameter optimization relies on enhanced statistical techniques such as design and analysis of computer experiments, it can be performed algorithmically and requires only the specification of the relevant algorithm's parameters. We apply SPO to tune the parameters of an evolutionary strategy (ES) used commercially for the optimization of financial portfolios. This provides a real-world example on using SPO in a commercial setting, as well as on the practical aspects of applying the SPO toolbox (SPOT). We show that SPO can improve on the results of an hand-tuned ES on this real-world problem.

CS53 Room 10 ECONOMIC AND FINANCIAL APPLICATIONS Chair: Alessandra Amendola

C331: **Determinants of liquidity holdings of Turkish commercial banks**
Presenter: **Esma Gaygisiz**, Middle East Technical University, Turkey
Co-authors: Didem Pekkurnaz, Hande Ayaydin

The Turkish economy has experienced many financial instabilities and crises arising from its banking structure since the introduction of the financial liberalization process in 1980s. Regulating and managing banks' liquidity have become essential for Turkey as well as many other countries after experiencing severe bank failures. This study investigates the key determinants of liquidity holdings of Turkish commercial banks during period 2003Q1 to 2007Q2. For this purpose panel data on quarterly unconsolidated balance sheets, income statements and independent supervision reports of private commercial banks are used. The macroeconomic and bank specific determinants of bank liquidity are analyzed using dynamic panel data methods. The observation that profitable as well as large banks in Turkey hold less liquid assets over both total deposits and total assets is interesting. Also, in cases of the loan growth and GDP growth, banks tend to hold smaller liquid assets over total deposits. Interest rate negatively affects the liquidity holdings of banks while for foreign owned banks its effect is positive.

C332: **Performance and cost efficiency of Russian small-sized banks**
Presenter: **Veronika Belousova**, University - Higher School of Economics, Russia

The aim of this study is two-fold. First, cost efficiency of Russian commercial banks is estimated for the period from 1q/2006 to 1q/2009. Stochastic frontier analysis with output quality and risk factors is applied to the data of Russian banks. In the case of single frontier usage it is necessary to assume that all banks use the same production technology, and so provide banking services under the same frontier. In practice not all banks may use similar transformation functions and so do not have equal access to the single cost frontier. To divide the Russian commercial banks into homogeneous groups, the Zipf-Pareto distribution of banks' total assets value is used. Second, determinants of cost efficiency variations among homogeneous group of banks are analyzed. We found that Russian small-sized banks were quite small and risky; however, they acted as a financial intermediary. In order to raise the minimum level of capital for this category of banks, the supervisor should pursue M&A of small-sized banks mostly on the base of medium-sized banks due to existing room for their efficiency improvement by means of exploiting economy of scales.

C052: **On stationary distribution of heteroskedastic conditional variance**
Presenter: **Jevgenijs Carkovs**, Riga Technical University, Latvia
Co-authors: Viktorija Carkova

This paper proposes an algorithm for diffusion approximation of a discrete ARCH process. In contrast to classical autoregressive models with independent random perturbations, this paper deals with uncertainty given as a stationary ergodic Markov chain. The method is based on a stochastic analysis approach to finite dimensional difference equations. By deriving a point-form solution of this difference equation, probabilistic limit theorems for dynamical systems with rapid Markov switching can be applied. The distribution of the stationary solution of the resulting stochastic equation may be successfully used for analysis of the initial discrete model. This method permits the analysis of a correlation effect on the log of cumulative excess returns with stochastic volatility. A model-based analysis shows that it is important to take into account possible serial residual correlation

in the conditional variance process. The proposed method is applied to GARCH(1,1) processes and to stocks with stochastic volatility.

C117: **Progressive stress accelerated life tests under progressive type-II censoring**
Presenter: **Alaa Abdel-Hamid**, Beni-Suef, Egypt
Co-authors: Essam AL-Hussaini

Progressive stress accelerated life tests based on progressively type-II censored samples is considered. The lifetime of an item under use condition follows the Weibull distribution. The inverse power law holds between the Weibull scale parameter and the constant stress. It is assumed that the progressive stress is directly proportional to time and the cumulative exposure model for the effect of changing stress holds. The model parameters are estimated graphically using a Weibull probability paper plot that serves as a tool for model identification. The likelihood equations are derived and solved numerically to obtain the maximum likelihood estimates (MLEs) of the parameters involved. The observed Fisher information matrix and the asymptotic variance-covariance matrix of the MLEs are derived. Approximate confidence intervals (CIs) for the parameters as well as the studentized-t and percentile bootstrap CIs are constructed. Monte Carlo simulation study is carried out to investigate the precision of the MLEs and to compare the performance of the CIs considered. Finally, some examples are presented to illustrate our results.

C223: **Decision support system for a project management application**
Presenter: **Natasa Glisovic**, Mathematical Faculty, Serbia/Montenegro
Co-authors: Nebojsa Bojovic, Milos Milenkovic, Nikola Knezevic

Modern organizations and companies exist in time of progressiveness in almost all segments of life and work influenced by large, rapid and dynamic changes. In such conditions, business world is enforced to adapt to these changes. Character of resulting changes is such that leads to question the existing social theories, especially in the macroeconomic field. Project management is a new and specialized management discipline. The main goal of project management is an efficient project realization within the planned time and planned costs in the conditions of significant external and internal changes. In this paper we developed a decision support system based on fuzzy Delphi method for assisting during the project activity time forecasting. We applied this software tool on a project of Serbian postal company restructuring and obtained results appeared as very consistent.

C181: **The game playing in knock-out discount accumulator**
Presenter: **Tian Wang**, University of Shanghai for Science and Technology, China
Co-authors: Yi Zhang, Jihuai Wang

In the past economic crisis in 2008, derivatives, such as accumulator, have caused great loss to many investors due to the plunge in all kinds of market. Therefore many economists claim that accumulator is unfair and cannot be hedged. The contention has been approached by modeling and other interpretation has been given, i.e. Knock-out Discount Accumulator (KODA) has been proposed fairly. It is regarded as impartial because of its three characteristics: 1) to make profits in accumulator is possible; 2) the profit is proportional to the risk that investors undertake; 3) the hedge is theoretically possible. In the simulation process, the Black-Scholes Model has been applied to calculate how many days of the bull market at least an investor may need and to prove KODA is fair enough to make the reasonable breakeven point exist. The results show that 1) the longer the bull run lasts, the higher probability that KODA makes profit; 2) the breakeven point is where the bull run and bear run divide each other, and the average breakeven point stays at the around 30 days, which is comparatively easy to realize in the real market. 3) 252 trading-days' price by simulation follows a Beta distribution.

CS62 Room 8 FINANCIAL ECONOMETRICS: PORTFOLIO, RISK, AND GARCH MODELS Chair: Arco van Oord

C069: **Dominance results of shrinkage estimators for the mean-variance optimal portfolio weights and their applications**
Presenter: **Takuya Kinkawa**, Keio University, Japan
Co-authors: Nobuo Shinozaki

In the estimation problem of the mean-variance optimal portfolio weights, some previous studies have proposed applying shrinkage estimators. However, only a few studies have addressed this problem analytically. Since the form of the loss function used in this problem is not the quadratic one used in statistical literature, there have been some difficulties in showing analytically the general dominance results. In this study, the dominance of a broader class of Stein type estimators is shown analytically when the covariance matrix is unknown and is estimated. The obtained results enable us to clarify the conditions for some previously proposed estimators in finance to have smaller risks than the classical estimator which we obtain by plugging in the sample estimates. We also show the dominance when there are linear constraints on portfolio weights.

C079: **A hierarchical Bayesian dynamic latent variable model for credit rating**

Presenter: **Saebom Jeon**, Korea University, Korea (ROK)
Co-authors: Yousung Park

The New Basel Capital Accord (Basel II) allows financial institutions to establish internal ratings based (IRB) approach to measure their capital requirement for credit risk, or the probability of default. We propose a hierarchical Bayesian latent variable model that enables us to calculate the default probability and transition matrix for individual obligors. First, using the multivariate latent variables for credit rating, we can evaluate the credit worthiness of an obligor. This model reflects not only the pattern of cross-sectional default correlation among obligors, but also serial correlation over time for each obligor. By introducing the dynamic structure with unit root to the latent variable model, we can test if the probability of default would promptly change or would last some while in any unexpected circumstances. It allows us to more precisely predict the default probability of obligor under unusual mechanism such as stress situation. Then, we estimate the transition probability between rating through the evolving dependences of latent variable over time. We handle missing values in covariates by multiple imputations. We illustrate findings of the external audit data from the Korean domestic corporations during 2000-2006 with above model.

C116: **A particle filter approach for money market yield curve estimation**

Presenter: **Josef Hayden**, University of Regensburg, Germany
Co-authors: Robert Ferstl

Non-linear filtering techniques like the unscented Kalman filter were recently successfully applied in estimating the term structure of interest rates where the underlying instruments typically have non-linear pricing equations. It is well known from the finance literature, that – in addition to the non-linearities – non-Gaussian effects should also be considered when modelling short-rate processes. Therefore, we first compare the term structure estimation performance of several non-linear non-Gaussian state-space models with various non-linear filtering approaches. Because of the ability to produce more accurate forecasts, we consider the common class of multivariate essentially affine term structure models. Second, we estimate and forecast the money market and swap yield curve on a weekly basis for a cross-section of European and US money market and swap yields. We evaluate the out-of-sample forecasts based on the pricing performance of the underlying fixed income securities and test for superior predictive ability over naive forecast strategies.

C214: **Bootstrap prediction intervals for risk measures in the context of GARCH models**

Presenter: **Maria Rosa Nieto Delfin**, Universidad Carlos III de Madrid, Spain
Co-authors: Esther Ruiz Ortega

A new bootstrap procedure to obtain prediction intervals of future Value at Risk (VaR) and Expected Shortfall (ES) in the context of univariate GARCH models is proposed. Alternative bootstrap intervals previously proposed in the literature incorporate the parameter uncertainty in the computation of the conditional standard deviation. We propose to use a second bootstrap step to incorporate the uncertainty attributable to deviations from the assumed error distribution into the computation of the corresponding quantile. Furthermore, in this second bootstrap step, we propose using an iterated smoothed bootstrap with better properties than the traditional one when computing prediction intervals for quantiles. We show with simulated data that the coverage of our proposed procedure are closer to the nominal than those of the alternatives. All the results are illustrated by obtaining one-step-ahead prediction intervals of the VaR and ES of several real time series of financial returns.

C237: **A comprehensive comparison of alternative tests for jumps in asset prices**

Presenter: **Marina Theodosiou**, Imperial College London, UK
Co-authors: Filip Zikes

A comprehensive comparison of the various existing jump tests for disentangling the continuous and jump components in discretely observed prices of financial assets is presented. The relative performance of the tests is examined in a Monte Carlo simulation, covering scenarios of both finite and infinite activity jumps and stochastic volatility models with continuous and discontinuous volatility sample paths. The impact of microstructure noise and infrequent trading (zero intraday returns) on the size and power properties of these tests is also investigated. An empirical application to high-frequency data from the index futures, equity and foreign exchange markets complements the analysis. The simulation results reveal important differences in terms of size and power across the different data generating processes considered. The conclusions from the empirical application conform to those of the Monte Carlo simulation showing that the test statistics are very sensitive to the presence of zero returns and microstructure frictions in the data. The aim is to provide practical guidance to empirical researchers studying the fine structure of asset returns, which has so far not been considered.

Author index