PROGRAMME AND ABSTRACTS

4th CSDA International Conference on

Computational and Financial Econometrics (CFE 10)

http://www.cfe-csda.org/cfe10

and

3rd International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computing & Statistics (ERCIM 10)

http://www.cfe-csda.org/ercim10

Senate House, University of London, UK
10-12 December 2010

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

Birkbeck, University of London
http://www.bbk.ac.uk

London School of Economics
http://www.lse.ac.uk

CFE 10 & ERCIM 10 International Organizing Committee:
A.-M. Fuertes, J. Niland, E. Ruiz, B. Rustem, E.J. Kontoghiorghes and G. Loizou.

CFE 10 Co-chairs:
George Kapetanios, Oliver Linton, Michael McAleer and Esther Ruiz.

CFE 10 Programme Committee:

A. Amendola, F. Audrino, P. Boswijk, M. Billio, M. Binder, C. Chen, D. van Dijk, C. Francq, A.-M. Fuertes, A.B. Galvao, R. Gerlach, D. Guegan, S. Hodges, S.J. Koopman, L. Khalaf, J.F. Kiviet, T. Lux, S. Mittnik, M. Ooms, J. Nankervis, Y. Omori, M. Paolella, D.S.G. Pollock, T. Proietti, M. Reale, M. Riani, B. Rustem, W. Semmler, M. Steel, M. So, E. Tzavalis, D. Veredas, M. Wagner, P. Winker, P.L.H. Yu, P. Zadrozny, P. Zaffaroni, J.-M. Zakoian and A. Zeileis.

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

CSDA Annals of CFE Guest Editors:
Luc Bauwens, Siem Jan Koopman, Oliver Linton and Michael McAleer.

ERCIM 10 Co-chairs:
Ana Colubi, Ana-Maria Fuertes, Irini Moustaki, D. Stephen G. Pollock, Stefan Van Aelst and Peter Winker.

ERCIM 10 Programme Committee:

M. Alfo, W. Bergsma, D. Boehning, E. Ceyhan, A. Christmann, C. Croux, M. Debruyne, M. Deistler, I. Demetriou, T. Denoeux, V. Esposito Vinzi, P. Filzmoser, K. Fokianos, P. Foschi, R. Fried, C. Gatu, M.A. Gil, S.G. Gilmour, J. Godolphin, B. Grun, M. Hubert, D. Kuhn, U. Ligges, A. Luati, D. Maringer, S. Meintanis, A. C. Monti, D. Morales, M.C. Pardo, S. Paterlini, B.U. Park, C. Perna, S. Petrone, I. Poli, I. Prunster, G. Storti, M. Timmerman, K. Triantafyllopoulos, L. Ugarte and A. Zeileis.

Local Organizing Committee:
Diane Loizou, Elizabeth Price, Evgenia Tsinti and Sharon Woolf.

Dear Friends and Colleagues,

We warmly welcome you to London, for the Fourth International Conference on *Computational and Financial Econometrics* (CFE 10) and the Third International Conference of the ERCIM Working Group on *Computing & Statistics* (ERCIM 10). As many of you know, this annual conference has been established as a leading joint international meeting for the interface of computing, empirical finance, econometrics and statistics, and is endorsed by the journal of Computational Statistics & Data Analysis (CSDA). The CFE 10 concurs with the inauguration of the Call for Papers of the CSDA Annals on Computational and Financial Econometrics.

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 10 & ERCIM 10 programme consists of over 160 sessions, 5 plenary talks and over 700 presentations. Peer reviewed papers will be considered for publication in special issues of the journal Computational Statistics & Data Analysis.

The co-chairs have endeavoured to provide a balanced and stimulating programme that will appeal to the diverse interests of the participants. The international organizing committee hopes that the conference venue will provide the appropriate environment to enhance your contacts and to establish new ones. The conference is a collective effort by many individuals and organizations. The Scientific Programme Committee, the Session Organizers, the local hosting universities and many volunteers have contributed substantially to the organization of the conference. We acknowledge their work and the support of our hosts and sponsors, and particularly Queen Mary, University of London, Birkbeck, University of London, London School of Economics, CSDA journal and ERCIM.

Looking forward, the CFE 11 & ERCIM 11 will also take place in the same venue as now (Senate House, London). The first issue of the CSDA Annals on Computational and Financial Econometrics will be available 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 London.

The CFE 10 & ERCIM 10 co-chairs and the International Organizing Committee.

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 application 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, 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.

QF: Quantitative Finance.

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: ercim@cfe-csda.org.

SCHEDULE

CFE 10

Friday, 10th December 2010

08:45 - 08:55	Opening (IoE Logan Hall)
08:55 - 09:45	Plenary Session A (IoE Logan Hall)
09:45 - 10:15	Coffee Break
10:15 - 12:20	Parallel Sessions B
12:20 - 13:40	Lunch Break
13:40 - 15:20	Parallel Sessions C
15:20 - 15:50	Coffee Break
15:50 - 17:05	Parallel Sessions D
17:20 - 18:10	Plenary Session E (IoE Logan Hall)
20:00 - 21:30	Reception (Senate Crush Hall)

Saturday, 11th December 2010

08:40 - 10:45	Parallel Sessions F
10:45 - 11:10	Coffee Break
11:10 - 12:50	Parallel Sessions G
12:50 - 14:15	Lunch Break
14:15 - 15:05	Plenary Session H (Senate Beveridge Hall)
15:15 - 16:55	Parallel Sessions I
16:55 - 17:20	Coffee Break
17:20 - 19:25	Parallel Sessions J
20:30 - 24:00	Conference Dinner

Sunday, 12th December 2010

09:00 - 10:40	Parallel Sessions K
10:40 - 11:10	Coffee Break
11:10 - 13:15	Parallel Sessions L
13:15 - 14:45	Lunch Break
14:45 - 16:50	Parallel Sessions N
16:50 - 17:20	Coffee Break
17:20 - 19:00	Parallel Sessions P
19:05 - 19:15	Closing (Senate Beveridge Hall)
19:15 - 20:00	Closing Drink (Senate Crush Hall)

ERCIM 10

Friday, 10th December 2010

08:45 - 08:55	Opening (IoE Logan Hall)
08:55 - 09:45	Plenary Session A (IoE Logan Hall)
09:45 - 10:15	Coffee Break
10:15 - 12:20	Parallel Sessions B
12:20 - 13:40	Lunch Break
13:40 - 15:20	Parallel Sessions C
15:20 - 15:55	Coffee Break
15:55 - 16:45	Plenary Session D (IoE Logan Hall)
16:55 - 19:00	Parallel Sessions E
20:00 - 21:30	Reception (Senate Crush Hall)

Saturday, 11th December 2010

08:40 - 10:45	Parallel Sessions F
10:45 - 11:10	Coffee Break
11:10 - 13:15	Parallel Sessions G
13:15 - 14:45	Lunch Break
14:45 - 16:50	Parallel Sessions I
16:50 - 17:20	Coffee Break
17:20 - 19:00	Parallel Sessions J
20:30 - 24:00	Conference Dinner

Sunday, 12th December 2010

09:00 - 10:40	Parallel Sessions K
10:40 - 11:10	Coffee Break
11:10 - 12:50	Parallel Sessions L
12:50 - 14:20	Lunch Break
14:20 - 15:10	Plenary Session M (Senate Beveridge Hall)
15:15 - 16:55	Parallel Sessions N
16:55 - 17:20	Coffee Break
17:20 - 19:00	Parallel Sessions P
19:05 - 19:15	Closing (Senate Beveridge Hall)
19:15 - 20:00	Closing Drink (Senate Crush Hall)

MEETINGS AND SOCIAL EVENTS

SPECIAL MEETINGS by invitation to group members

- CSDA Editorial Board meeting, *MacMillan Hall, Senate House*, Thursday 9th of December 2010, 18:00-19:30.
- CSDA Editorial Board dinner, Thursday 9th of December 2010, 20:00 - 23:00.
- Special General Assembly of the IASC, *Beveridge Hall, Senate House*, Sunday 12th of December 2010, 13:30 - 14:00.
- COMPSTAT SPC meeting, *Room MAL G16*, Sunday 12th of December, 17:05-19:00.

SOCIAL EVENTS

- *The coffee breaks* will last 45 minutes each (beginning 15 minutes in advance of the times indicated in the program). These will take place at Birkbeck MAL B02 & MAL B04 rooms, Senate House Crush Hall & MacMillan Halls. Coffee breaks will also be served at the Institute of Education Crush Hall on Friday 10th December (**but only for 30 minutes and limited capacity**). You must have your conference badge in order to attend the coffee breaks.
- *Welcome Reception, Friday 10th of December, 20:00-21:30*. The reception is open to all registrants and accompanying persons who have purchased a reception ticket. It will take place at the Senate House Crush Hall. Conference registrants must bring their conference badge and any accompanying persons should bring their reception tickets in order to to attend the reception.
- The *Working Lunch* will be served at Birkbeck MAL B02 & MAL B04 rooms. You must have your Lunch ticket of the appropriate day in order to attend the lunch. People not registered for lunch can buy lunch on the 5th floor of the Birkbeck extension building and the two cafes at Senate House (limited capacity), and at restaurants and cafes in close walking distance to the conference venue.
- *Conference Dinner, Saturday 11th of December, 20:30*. The Conference Dinner will take place at the *Holiday Inn London Bloomsbury*. The conference dinner is optional and registration is required. You must have your Conference Dinner ticket in order to attend the conference dinner.
- *Closing Drink, Sunday 12th of December, 19:15-20:00*. The Closing Drink will take place in the Senate House Crush Hall.

Addresses of venues:

- Birkbeck, University of London, Malet Street, London WC1E 7HX
- University of London, Senate House, Malet Street, London WC1E 7HU
- Institute of Education, 20 Bedford Way, London WC1H 0AL

Registration and Exhibitors

The registration and exhibitors will be located in the Crush and MacMillan Halls of the Senate House.

Lecture Rooms

The paper presentations will take place at the main building of Birkbeck (Malet Street), Senate House and also on Friday the 10th at the Logan Hall of the Institute of Education (IoE) as well (see map in the next page). The list of rooms and their capacity is listed below. Due to health and safety regulations the maximum capacity of the rooms should be respected. The opening ceremony will take place at the Logan Hall of the IoE. There will be no signs indicating the location of the lecture rooms, and therefore we advise that you visit the venue in advance.
The opening, keynote and closing talks will take place on Friday at the Logan Hall of the IoE and on Saturday and Sunday at the Beveridge Hall of Senate House. The poster sessions will take place at the Crush Halls of the IoE and Senate House. The posters should be displayed only during their assigned session. The authors will be responsible for placing the posters in the poster panel displays. The maximum size of the poster is A0. Chairs are requested to keep the sessions to schedule. Papers should be presented in the order they are listed in the programme for the convenience of attendees who may wish to go to other rooms mid-session to hear particular papers. In the case of a presenter not attending, please use the extra time for a break or a discussion so that the remaining papers stay on schedule.

At Birkbeck all lecture rooms on floors 3, 4 and 5 are accessible via several lifts. The rooms MAL 151, 152, 351 and 352 are in the extension of the main building at Birkbeck. These can be accessed via lifts through the 3rd floor or directly from lift B on the ground floor (right site end of corridor of the main building). In the Senate House, Gordon is Room 34, Bloomsbury is Room 35, and Bedford is Room 37.

Room	Capacity	Floor	Location	Room	Capacity	Floor	Location
MAL B02	35	Basement	Birkbeck Malet St	MAL B04	80	Basement	Birkbeck Malet St
MAL B18	44	Basement	Birkbeck Malet St	MAL B20	80	Basement	Birkbeck Malet St
MAL B29	30	Basement	Birkbeck Malet St	MAL B30	40	Basement	Birkbeck Malet St
MAL B33	133	Basement	Birkbeck Malet St	MAL B34	175	Basement	Birkbeck Malet St
MAL B35	100	Basement	Birkbeck Malet St	MAL B36	100	Basement	Birkbeck Malet St
MAL G16	50	Ground	Birkbeck Malet St	MAL 151	35	First	Birkbeck Malet St
MAL 152	35	First	Birkbeck Malet St	MAL 351	24	Third	Birkbeck Malet St
MAL 355	24	Third	Birkbeck Malet St	MAL 421	130	Fourth	Birkbeck Malet St
MAL 509	40	Fifth	Birkbeck Malet St	MAL 532	70	Fifth	Birkbeck Malet St
MAL 538	45	Fifth	Birkbeck Malet St	MAL 541	45	Fifth	Birkbeck Malet St
Gordon	40	Ground	Senate House	Bloomsbury	40	Ground	Senate House
Bedford	50	Ground	Senate House	Beveridge Hall	500	Ground	Senate House
Logan Hall	933	Basement	Institute of Education				

Presentation Instructions

The lecture rooms will be equipped with a PC and a computer projector. The session chairs should obtain copies of the talks on a USB stick before the session starts (use the lecture room as the meeting place), or obtain the talks by email prior to the start of the conference. Presenters must provide to the session chair with the files for the presentation in PDF (Acrobat) or PPT (Powerpoint) format on a USB memory stick. This must be done ten minutes before each session. The PC in the lecture rooms should be used for presentations. *The session chairs are kindly requested to have a laptop for backup*. Please note that plugs /power outlets of the UK 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.

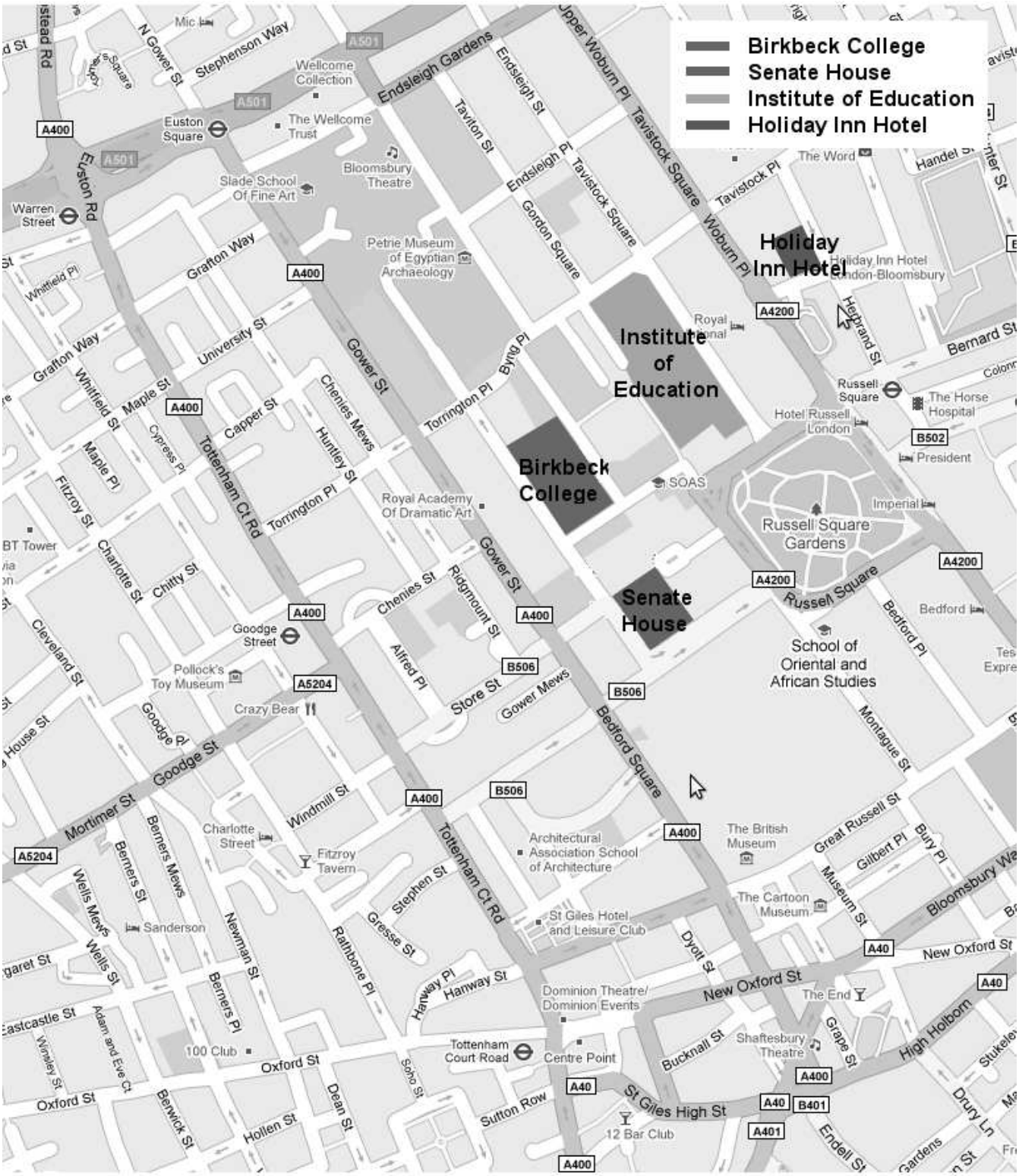
Internet

There will be wireless Internet connection in Senate House. You will need to have your own PC in order to connect to the Internet. The daily login and password will be displayed in the announcement board by the registration desk of the Crush Hall at Senate House.

Information and Messages

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

Map of the Venue and Nearby Area



PUBLICATIONS OUTLETS

Journal of Computational Statistics & Data Analysis (CSDA)

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

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

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

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

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

- CSDA has already planned special issues for 2010-2011 on the following topics:
  - Small Area Estimation (SAE).
  - Robust Analysis of Complex Data.
  - Statistical Algorithms and Software.
  - Optimization Heuristics in Estimation and Modelling.

SPONSORS

Queen Mary, University of London, UK (URL http://www.qmul.ac.uk)

Birkbeck, University of London, UK (URL http://www.bbk.ac.uk)

London School of Economics, UK (URL http://www.lse.ac.uk)

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

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

Commerzbank, UK (URL http://www.commerzbank.com)

EXHIBITORS

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

John Wiley & Sons Ltd (URL http://www.wiley.com)

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

TIMBERLAKE (URL http://www.timberlake.co.uk)

ENDORSED SOCIETIES & GROUPS

Journal of Computational Statistics & Data Analysis, Elsevier

ERCIM Working Group on *Computing & Statistics*

The Society for Computational Economics

International Association for Statistical Computing

Contents

Friday 10.12.2010 08:55-09:45 Room: IoE Logan Hall Chair: George Kapetanios Keynote talk CFE-ERCIM

Factor modelling for time series: a dimension-reduction approach

Speaker: **Qiwei Yao, London School of Economics, UK** Co-authors: Clifford Lam

Following a brief survey on the factor models for multiple time series in econometrics, we introduce a statistical approach from the viewpoint of dimension reduction. Our method can handle nonstationary factors. However under stationary settings, the inference is simple in the sense that the estimation for both the factor dimension and the loadings is resolved by an eigenanalysis for a non-negative definite matrix, and is therefore applicable when the dimension of time series is in the order of a few thousands. Asymptotic properties of the proposed method are investigated. In particular, our estimators for zero-eigenvalues enjoy the faster convergence rates even when the dimension goes to infinity together with the sample size. Numerical illustration with both simulated and real data will also be reported.

Friday 10.12.2010 17:20-18:10 Room: IoE Logan Hall Chair: George Kapetanios Keynote talk CFE 1

Stick-breaking autoregressive processes

Speaker: **Mark Steel, University of Warwick, UK** Co-authors: Jim Griffin

The problem of defining a time-dependent nonparametric prior for use in Bayesian nonparametric modelling of time series is considered. A recursive construction allows the definition of priors whose marginals have a general stick-breaking form. The processes with Poisson-Dirichlet and Dirichlet process marginals are investigated in some detail. We develop a general conditional Markov Chain Monte Carlo (MCMC) method for inference in the wide subclass of these models where the parameters of the marginal stick-breaking process are nondecreasing sequences. We derive a generalized Pólya urn scheme type representation of the Dirichlet process construction, which allows us to develop a marginal MCMC method for this case. We apply the proposed methods to financial data to develop a semi-parametric stochastic volatility model with a time-varying nonparametric returns distribution. Finally, we present two examples concerning the analysis of regional GDP and its growth.

Friday 10.12.2010 15:55-16:45 Room: IoE Logan Hall Chair: Irini Moustaki Keynote talk ERCIM 1

The forward search in robust statistics

Speaker: **Anthony Atkinson, London School of Economics, UK** Co-authors: Marco Riani

The forward search applied to the analysis of data divides the data into a *good* portion that agrees with the model and a set of outliers, if any. This is achieved by fitting subsets of data of increasing size. The talk will focus on multivariate data and will consider the inferential properties of the procedure: how to find and control the size of outlier tests in the face of simultaneous inference. Comparison is made with inferences that come from the plethora of robust methods for this problem, for example the MCD, either unweighted or reweighted in a variety of ways. Aspects of the presentation of power comparisons will be touched on. Our procedure is extended to clustering and compared with established methods such as mclust. The forward search leads naturally to informative graphs of many aspects of the interaction between data and fitted model. These will be illustrated through a variety of dynamic linked plots from the toolbox for the forward search, which also includes regression and data transformation.

Saturday 11.12.2010 14:15-15:05 Room: Senate Beveridge Hall Chair: D.S.G. Pollock Keynote talk CFE 2

Econometric analysis of high dimensional VARs featuring a dominant unit

Speaker: **M. Hashem Pesaran, Cambridge University, UK and USC, USA** Co-authors: Alexander Chudik

This paper extends the analysis of infinite dimensional vector autoregressive models (IVAR) to the case where one of the variables or the cross section units in the IVAR model is dominant or pervasive. This extension is not straightforward and involves several technical difficulties. The dominant unit influences the rest of the variables in the IVAR model both directly and indirectly, and its effects do not vanish even as the dimension of the model (N) tends to infinity. The dominant unit acts as a dynamic factor in the regressions of the non-dominant units and yields an infinite order distributed lag relationship between the two types of units. Despite this it is shown that the effects of the dominant unit as well as those of the neighborhood units can be consistently estimated by running augmented least squares regressions that include distributed lag functions of the dominant unit. The asymptotic distribution of the estimators is derived and their small sample properties investigated by means of Monte Carlo experiments.

Sunday 12.12.2010 14:20-15:10 Room: Senate Beveridge Hall Chair: George Loizou Keynote talk ERCIM 2

Interdisciplinary research in support of biomedical discoveries: The role of the biostatistician and informaticist

Speaker: **Stanley Azen, University of Southern California, USA**

New biomedical discoveries require the collaboration of biostatisticians and informaticists with multi-disciplinary investigators in conducting interdisciplinary and translational research. Multi-disciplinary collaborations lead not only to creating new knowledge that has biomedical, clinical and public health importance, but also to developing new biostatistical methodology. Examples that will be presented include: 1) translational research in cardiovascular disease leading to drug development; 2) population based studies in cardiovascular disease leading to the development of improved screening strategies; 3) population-based studies in ocular disease in multi-ethnic cohorts and its impact on public health; 4) health promotion and its improvement on quality of life in the aging population; and 5) influence of the neighborhood environment using geographic information systems to determine factors leading to obesity. The presentation also includes a discussion of 1) the challenges associated with developing quality clinical databases using informatics technology; and 2) examples of data mining databases which provide opportunities for junior faculty, post-doctoral fellows and graduate students to identify new and interesting findings leading to new screening tools, and identification of treatment-induced biological markers impacting clinical outcomes.

Friday 10.12.2010 10:15 - 12:20 Parallel Session B – CFE

CI97 Room IoE Logan Hall INVITED SESSION: APPLIED TIME SERIES ECONOMETRICS Chair: Ana-Maria Fuertes

C398: **Supply, demand and monetary policy shocks in a multi-country new Keynesian model**
Presenter: **Ron Smith**, Birkbeck, University of London, UK
Co-authors: Stephane Dees, M. Hashem Pesaran, L. Vanessa Smith

This paper estimates and solves a multi-country version of the standard New Keynesian (NK) model. The country-specific models include a Phillips curve, an IS curve including a real effective exchange rate variable and a country-specific foreign output variable, a Taylor Rule and a real effective exchange rate equation. In accord with theory all variables are measured as deviations from their steady states, estimated as long-horizon forecasts from a reduced-form cointegrating global vector autoregression. The resulting rational expectations model is then estimated for 33 countries on data for 1980Q1-2006Q4, by inequality constrained IV, using lagged and contemporaneous foreign variables as instruments, subject to the restrictions implied by the NK theory. The model is then solved to provide estimates of identified supply, demand and monetary policy shocks. We assume that the within country supply, demand and monetary policy shocks are orthogonal, though shocks of the same type (e.g. supply shocks in different countries) can be correlated. We discuss estimation of impulse response functions and variance decompositions in such large systems, and present estimates allowing for both direct channels of international transmission through regression coefficients and indirect channels through error spillover effects. Bootstrapped error bands are also provided for the US monetary policy shock.

C401: **Inference on stochastic time-varying coefficient models**
Presenter: **George Kapetanios**, Queen Mary University of London, UK
Co-authors: Liudas Giraitis, Tony Yates

Recently there has been considerable work on stochastic time-varying coefficient models as vehicles for modelling structural change in the macroeconomy with a focus on the estimation of the unobserved sample path of time series of coefficient processes. The dominant estimation methods, in this context, are various filters, such as the Kalman filter, that are applicable when the models are cast in state space representations. This paper examines alternative kernel based estimation approaches for such models in a nonparametric framework. The use of such estimation methods for stochastic time-varying coefficient models, or any persistent stochastic process for that matter, is novel and has not been suggested previously in the literature. The proposed estimation methods have desirable properties such as consistency and asymptotic normality. In extensive Monte Carlo and empirical studies, we find that the methods exhibit very good small sample properties and can shed light on important empirical issues such as the evolution of inflation persistence and the PPP hypothesis.

C469: **Financial option pricing in volatile markets**
Presenter: **Christopher Baum**, Boston College, USA
Co-authors: Paola Zerilli

The main objective of this paper is to propose a new model for option pricing that can deliver reliable results even in the event of a major financial market crisis. The basic idea is that when stock return volatility is unobservable, a good indicator for its behaviour can be found in the volatility implied in the corresponding option market. For example, the implied option volatility during 1987 was characterized by very sharp movements in both directions. Existing option pricing models generally allow only for dramatic increases in the volatility. The proposed model is that the stock return volatility is allowed to wonder about its path without any restrictions beyond the zero bound. The model is estimated from daily stock returns and options data using a two-step procedure which combines the Efficient Method of Moments (EMM) with minimization of the mean squared residuals (MSE). The EMM is used to estimate the parameters of the stock return and return volatility processes, while the MSE is used to estimate the risk premia embedded in the option prices using a no-arbitrage argument.

CS12 Room MAL 151 FORECASTING VALUE-AT-RISK Chair: Teodosio Perez Amaral

C510: **The information content of high-frequency data for estimating equity return models and forecasting risk**
Presenter: **Dobrislav Dobrev**, Federal Reserve Board, USA
Co-authors: Pawel Szerszen

We demonstrate that the parameters controlling skewness and kurtosis in popular equity return models estimated at daily frequency can be obtained almost as precisely as if volatility is observable by simply incorporating the strong information content of realized volatility measures extracted from high-frequency data. For this purpose, we introduce asymptotically exact volatility measurement equations in state space form and propose a Bayesian estimation approach. Our highly efficient estimates lead in turn to substantial gains for forecasting various risk measures at horizons ranging from a few days to a few months ahead when taking also into account parameter uncertainty. As a practical rule of thumb, we find that two years of high frequency data often suffice to obtain the same level of precision as twenty years of daily data, thereby making our approach particularly useful in finance applications where only short data samples are available or economically meaningful to use. Moreover, we find that compared to model inference without high-frequency data, our approach largely eliminates underestimation of risk during bad times or overestimation of risk during good times. We assess the attainable improvements in VaR forecast accuracy on simulated data and provide an empirical illustration on stock returns during the financial crisis of 2007-2008.

C521: **Uncertainty of multiple period risk measures**
Presenter: **Carl Lonnbark**, Umea University, Sweden

In general, the properties of the conditional distribution of multiple period returns do not follow easily from the one-period data generating process. This renders computation of Value-at-Risk and Expected Shortfall for multiple period returns a non-trivial task. In this paper we consider some approximation approaches to computing these measures. Based on the results of a simulation experiment we conclude that among the studied analytical approaches the one based on approximating the distribution of the multiple period shocks by a skew-t was the best. It was almost as good as the simulation based alternative. We also found that the uncertainty due to the estimation risk can be quite accurately estimated employing the delta method. In an empirical illustration we computed five day VaR's for the S&P 500 index. The approaches performed about equally well.

C537: **A detailed comparison of Value at Risk estimates**
Presenter: **Sonia Benito**, Universidad Rey Juan Carlos, Spain
Co-authors: Pilar Abad

The performance of different models of Value at Risk is investigated. We include a wider range of methods (Parametric, Historical simulation, Monte Carlo, and Extreme value theory) and several models to compute the conditional variance. We analyse several international stock indexes

and examine two periods: stable and volatile. To choose the best model, we employ a two-stage selection approach. The result indicates that the best model is a Parametric model with conditional variance estimated by asymmetric GARCH model under Student's t-distribution of returns. It is shown that Parametric models can obtain successful VaR measures if conditional variance is estimated properly.

C590: **Minimize capital requirements with a DPOT method**

Presenter: **Paulo Araujo Santos**, Instituto Politucnico de Santarem and CEAUL, Portugal

Co-authors: Isabel Fraga Alves

Threshold methods, based on fitting a stochastic model to the exceedances over a threshold, were developed under the acronym POT (peaks over threshold). In this work we propose a duration based POT (DPOT) method, to forecast one-day-ahead Value-at-Risk (VaR). The proposed method allows for the presence of duration as covariate in order to eliminate or reduce the tendency to clustering of violations that often occurs in the context of heterocedastic financial time series. We compare the out-of-sample performance of the proposed duration based POT method with other methods, using unconditional coverage and independence tests for interval forecasts evaluation and in terms of minimization of daily capital requirements under the Basel II Accord.

C702: **Crisis robust risk management under the Basel accord**

Presenter: **Teodosio Perez Amaral**, Complutense de Madrid, Spain

Co-authors: Juan-Angel Jimenez-Martin, Michael McAleer

A crisis robust risk management strategy is proposed. We suggest selecting a VaR forecast that combines the forecasts of different VaR models. In particular we propose to communicate the median of the point VaR forecast of a set of GARCH-family models (median VaR). This risk management strategy is crisis robust in the sense that maintaining the same rules before, during and after a financial crisis would imply comparatively low capital charges and also numbers of violations that would not jeopardize the bank or institutions that use it. We illustrate this new method using the S&P500 index before, during and after the 2008-09 financial crisis. We investigate the performance of a variety of single and combination VaR forecast models in terms of capital requirements and violations under the Basel II Accord as well as other criteria. It is found that the median VaR strategy is crisis robust because it provides stable results across different periods relative to other models. This new strategy based on combinations of forecasts of single models should be straightforward to incorporate into existing computer software packages used by banks and financial institutions

CS15 Room MAL B29 FINANCIAL RISKS AND THE MACROECONOMY Chair: Monica Billio

C445: **Multifactor pricing models for US real estate**

Presenter: **Andrea Donato Tortora**, Bocconi University, Italy

Co-authors: Massimo Guidolin, Francesco Ravazzolo

We analyze and compare the asset pricing properties and dynamics of publicly traded real estate to general asset classes like stocks and bonds. The aim is to investigate whether there is any evidence of systematic over-pricing of various categories of REITs over the period 2003-2006, as widely believed by many policy-makers and commentators. To this purpose, we estimate a multifactor pricing model with time-varying factor sensitivities and risk premia in a Bayesian dynamic framework and introduce stochastic idiosyncratic volatility. We find no evidence of systematic overpricing of the REIT asset class over the alleged 2003-2006 bubble period. On the contrary, what characterizes the recent years is a growing level of persistent idiosyncratic risk that seems to have been priced in the case of a range of asset classes, including real estate. This evidence is consistent with the notion that the recent financial crisis would have originated more from the poor quality of lending standards than from the presence of obvious upward biases in prices.

C410: **Identifying risks in emerging market sovereign and corporate bond spreads**

Presenter: **Gabriele Zinna**, Bank of England, UK

This study investigates the systematic risk factors driving emerging market (EM) credit risk by jointly modelling sovereign and corporate credit spreads at a global level. We use a multi-regional Bayesian panel VAR model, with time-varying betas and multivariate stochastic volatility. This model allows us to decompose credit spreads and to build indicators of EM risks. We find that indices of EM sovereign and corporate credit spreads differ because of their specific reactions to global risk factors. Following the failure of Lehman Brothers, EM sovereign spreads *decoupled* from the US corporate market. In contrast, EM corporate bond spreads widened in response to higher US corporate default risk. We also find that the response of sovereign bond spreads to the VIX was short-lived. However, both EM sovereign and corporate bond spreads widened in flight-to-liquidity episodes, as proxied by the OIS-Treasury spread. Overall, the model is capable of generating other interesting results about the comovement of sovereign and corporate spreads.

C472: **Econometric measures of systemic risk in the finance and Insurance sectors**

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

Co-authors: Mila Getmansky, Andrew Lo, Loriana Pelizzon

We propose several econometric measures of systemic risk to capture the interconnectedness among the monthly returns of hedge funds, banks, brokers, and insurance companies based on principal components analysis and Granger-causality tests. We find that all four sectors have become highly interrelated over the past decade, increasing the level of systemic risk in the finance and insurance industries. These measures can also identify and quantify financial crisis periods, and seem to contain predictive power for the current financial crisis. Our results suggest that hedge funds can provide early indications of market dislocation, and systemic risk arises from a complex and dynamic network of relationships among hedge funds, banks, insurance companies, and brokers.

C358: **Optimal fiscal policies in booms and in recessions: An econometric case study for Slovenia**

Presenter: **Reinhard Neck**, Klagenfurt University, Austria

Co-authors: Dmitri Blueschke, Viktoria Blueschke-Nikolaeva, Klaus Weyerstrass

Optimal fiscal policies for the next few years are determined for Slovenia under alternative assumptions about the global development. In particular, we distinguish between a scenario of a recession (assuming the recent crisis to exhibit a double-dip) and a scenario of boom (assuming that the crisis is over by the end of 2010). We use the econometric model SLOPOL8 and assume an intertemporal objective function for Slovenian policy maker containing output, unemployment, inflation, the budget deficit, public debt, and the current account as arguments. Using the OPTCON algorithm, approximately optimal policies are calculated under both scenarios. It turns out that the design of fiscal policies is rather similar in both cases, showing the relatively low effectiveness of the fiscal instruments with respect to their influence on the business cycle.

C647: **Real, financial and credit variables in Italian GDP forecasting**

Presenter: **Giulio Nicoletti**, Bank of Italy, Italy

Co-authors: Raffaele Passaro

We probe the role of financial, credit and real variables in forecasting the Italian real output growth. We first recur to classical methods, we then refine our analysis by Bayesian methods, namely Dynamic Model Averaging (DMA). Using quarterly data on 22 indicators (1980–2009) we show that credit and financial spreads forecast GDP dynamics. DMA allows us to describe the over time evolution of the predictive content of our regressors. A preview of our results is as follows: we do find evidence that credit variables have predicting power over the GDP but mostly regarding its medium rather than its short run dynamics. Differently from the US, the forecasting ability of the long end of the government term spread is limited; our predictions improve instead when we use the short end of the yield curve. The forecasting ability of the overnight rates seems also to be rather limited when other variables, such as unemployment, are also included in the regressions. DMA shows that the relation between credit spreads, unemployment and GDP dynamics is more stable than the one with exchange rates and the government term spreads.

CS31 Room MAL G16 HEAVY-TAILED TIME SERIES Chair: Christian Francq

C355: **Prediction in GARCH models under heavy-tailed errors**

Presenter: **Jean-Michel Zakoian**, CREST and University Lille 3, France

Co-authors: Christian Francq

While GARCH models provide an efficient tool for predicting future squared returns, they rely on the assumption that the independent error process has a finite variance. When the errors are heavy-tailed, GARCH models can still be used to predict the future volatility behavior through the prediction of low-order powers of the absolute returns. We propose a new approach for predicting such powers based on a modification of the usual identifiability assumption: instead of a unit variance, the absolute errors are assumed to have a unit r-order moment, with $r < 2$. To estimate a general GARCH-type model based on this assumption, a non-Gaussian QML estimation method is developed, and its asymptotic properties are established. For the sake of comparison we also consider the asymptotic properties of a two-step approach based on the usual identifiability condition and the standard Gaussian QML. In particular, it is shown that for standard GARCH specifications the efficiency comparison only depends on r and the errors distribution. The results are illustrated by numerical experiments. An application to indexes of major stock exchange markets is also proposed.

C185: **Non-parametric modelling of heavy tails and skewness of box-office revenues**

Presenter: **Dimitrios Stasinopoulos**, London Metropolitan University, UK

Co-authors: Robert Rigby, Robert Gilchrist, John Sedgwick, Vlasios Voudouris

Using box-office data for movies released in the US market in the 1990s and 1930s, we establish probabilistic statements for the box-office revenues that the market at these instances dictate. Here, we propose a smooth and non-parametric model of heavy tails and skewness using the GAMLSS (Generalized Additive Models for Location Scale and Shape) framework. In doing so, an understanding of the demand dynamics and adaptive supply arrangements of the motion picture industry is presented. The movie market is particularly interesting due to its skewed and kurtotic macro-regularity, which resulted in the hypothesis that *nobody knows what makes a hit or when it will happen*. This hypothesis is revised here.

C248: **Robust estimation of dynamic conditional correlation GARCH models**

Presenter: **Sebastien Laurent**, Maastricht University, Netherlands

Co-authors: Kris Boudt, Jon Danielsson

The use of dynamic conditional correlation models for the estimation of conditional covariance matrices has now become standard in the financial econometrics. Its estimation is usually performed in two or three steps by Gaussian quasi-maximum likelihood. We show that this method is very sensitive to outliers in the data and propose to use outlier-robust estimators instead. The Monte Carlo study and empirical application document the good properties of this estimation method in absence and presence of outliers.

C443: **Macroeconomic applications of skewed tempered stable distributions**

Presenter: **Svetlana Makarova**, University College London, UK

Co-authors: Wojciech Charemza, Piotr Jelonek

It has been found out that uncertainty related to relevant macroeconomic indicators, most notably inflation, can be best approximated by simple skewed tempered stable (TS) distributions. These distributions account for possible appearance of infrequent extreme events, like hyperinflations and rapid disinflations. The problem is in generating random numbers from the TS distributions. Unlike in most of financial applications, macroeconomic uncertainties can be either positive or negative and the characteristic exponent parameter of the TS distributions is usually greater than one. These make the usual TS random number generators either not applicable, or too slow for practical implementations. We suggest applying the naive TS random number generators, by tilting the stable random numbers and, in the multivariate case, applying Cholesky transformation. Numbers obtained in such a way are biased if skewness is non-zero, but the bias can be accounted for either by introducing an additional adjustment parameter, or by adding a pre-simulated bias correction. The applications include the examination of the predictive power of inflation decompositions for output changes in an inflation-output vector autoregressive model and, in multivariate case, of predicting the expected length of runs and turning points for inflation in OECD countries.

CS35 Room MAL B30 RISK AND COMMODITIES MARKETS Chair: Dominique Guegan

C137: **Event conditional correlation**

Presenter: **Pierre-Andre Maugis**, PSE, France

Co-authors: Dominique Guegan

Given two random variables X and Y and an event $\mathcal{A}$, we are interested in the linear correlation between X and Y conditionally to $\mathcal{A}$. This dependence measure is of interest because even in the Gaussian case, this conditional correlation is significantly higher in the tails as compared to the center of the distribution. This variability of the conditional correlation highlights the importance of a proper study of this type of correlation. Indeed, if for instance the estimation of a correlation parameter is done conditionally to some omitted event, it significantly increases or decreases the estimates. Moreover, we show that conditional correlation offers great flexibility and is a powerful substitute for dynamic correlation and partial correlation, and even permits new uses of correlation. In this work we provide a formula to compute such correlation for any random variables X and Y (with second moments) and any event $\mathcal{A}$ (with a non-null probability of occurring) under weak hypotheses. This formula provides a very efficient estimator of such correlation, as shown in Monte-Carlo simulations.

C138: **On the necessity of five risk measures**
Presenter: **Wayne Tarrant**, Wingate University, USA
Co-authors: Dominique Guegan

The banking systems that deal with risk management depend on underlying risk measures. Following the recommendation of the Basel II Accord, most banks have developed internal models to determine their capital requirement. The Value at Risk measure plays an important role in computing this capital. In this paper we analyze in detail the errors produced by the use of this measure. We then discuss other measures, pointing out their strengths and shortcomings. We give detailed examples, showing the need for five risk measures in order to compute a capital in relation to the risk to which the bank is exposed. In the end we suggest using five different risk measures for computing capital requirements.

C349: **An efficient peak-over-threshold implementation for operational risk capital computation**
Presenter: **Bertrand Hassani**, Universite Paris I Pantheon - Sorbonne - BPCE, France
Co-authors: Dominique Guegan, Cedric Naud

Operational risk quantification requires to deal with data sets which often present extreme values. These values have a tremendous impact on the Capital computation (VaR, CVaR). The calibration of theoretical distributions on these data sets can be considered, but in most cases the goodness-of-fit tests reject them. An alternative, is to use a Peak-Over-Threshold method. This, based on Pickands' theorem, allows to fit a generalized Pareto distribution on the right tail. Unfortunately, Pickands' theorem demands to define a sufficiently high threshold. Therefore, initially we propose to implement slightly modified bootstrap methods to estimate the latter. Then, we use the Expectation-Maximization algorithm to estimate the parameters of the distribution fitted on the data below the threshold, and we discuss several methods to estimate GDP's scale and shape parameters. Finally, using an adapted Monte Carlo method to estimate the Loss Distribution Function, we proceed in the capital computation.

C298: **Volatility transmission between world oil prices and stock markets of the GCC countries**
Presenter: **Amine Lahiani**, Universite d'Orleans and ESC Rennes Business School, France
Co-authors: Mohamed El Hedi Arouri, Duc Khuong Nguyen

The return linkages and volatility transmission between oil and stock markets in the GCC countries over the last turbulent years is investigated. We employ a recent generalized VAR-GARCH approach which allows for transmissions in returns and volatilities. In addition, we analyze the optimal weights and hedge ratios for oil-stock portfolio holdings. Overall, our results point to the existence of substantial return and volatility spillovers between world oil price and GCC stock markets, and appear to be crucial for international portfolio management in the presence of oil price risk.

C380: **Economic value of accounting for large losses in portfolio selection**
Presenter: **Alexandra Dias**, University of Warwick, UK

Financial crises can cause financial portfolios to incur large losses. Methodologies for portfolio selection taking into account the possibility of large losses have existed for decades but their economic value is not established. This article studies if there is actually economic value in reducing the probability of large losses in portfolio selection. We combine mean-variance analysis with semi-parametric estimation of potential portfolio large losses. We find that strategies with reduced probability of large losses outperform efficient minimum variance portfolios, especially when semi-parametric estimation is used. Our results are robust to transaction costs.

CS37 Room MAL B18 MULTIVARIATE DEPENDENCE MODELLING IN FINANCE AND INSURANCE Chair: Vladimir Kaishev

C475: **Beyond fat-tails: A comparison of the methodologies for describing the tail dependence between assets**
Presenter: **Georgi Mitov**, FinAnalytica, Bulgaria
Co-authors: Svetlozar (Zari) Rachev, Boryana Racheva-Yotova

Risk and portfolio managers are finally looking beyond the bell curve and now accept that asset returns are fat-tailed. Indeed, practitioners are ramping up their processes to account for the phenomena that give rise to fat-tails. Most efforts however concentrate on improving the tails of financial instruments. Few have really gone beyond linear correlations and tackled the process of capturing the joint dependencies among risky assets. Continuing blindly down this path will surely lead to future catastrophic losses even with application of distributions that accurately describe the behavior of the individual assets. Modeling joint extreme events is not a simple exercise as it comprises multiple phenomena. It is well known that the dependence between assets is non-linear, being greater during periods of market stress. Furthermore, the dependency is asymmetric: most asset prices become relatively more dependent during significant market downturns than upturns. Finally the modeling process is complicated further by the need to handle large dimensions using relatively small data sets. We review several alternatives for building multivariate models - using the implied dependence of multivariate fat-tailed distributions, Gaussian and Fat-tailed copulas and different variations of empirical dependence models. We discuss their theoretical properties, advantages and disadvantages, practical applicability and compare their behavior in an empirical example which analyses the risk of a portfolio.

C707: **Linear combinations of Gamma, (LG) processes and Dirichlet (B-) splines: Applications in finance and insurance**
Presenter: **Vladimir Kaishev**, Sir John Cass Business School, City University London, UK

We consider a new class of the so called, LG processes, defined as linear combinations of independent gamma processes. Their distributional and path-wise properties are explored by following their relation to polynomial and Dirichlet (B-) splines. In particular, it is shown that the density of an LG process can be expressed in terms of Dirichlet (B-) splines, introduced independently by Ignatov and Kaishev and Karlin, Michelli and Rinott. We further show that the well known variance-gamma (VG) process, and the Bilateral Gamma (BG) process, are special cases of an LG process. Following this LG interpretation, we derive new (alternative) expressions for the VG and BG densities and consider their numerical properties. Multivariate LG processes are also introduced and their relation to multivariate Dirichlet and simplex splines is established. Based on this, expressions for their joint density, the underlying LG-copula and the characteristic function are given. A numerically efficient bridge sampling method for simulating LG sample paths is also developed. Estimation of the parameters of LG processes is further considered and a method of moments is developed. Applications of these new Dirichlet (B-) spline related processes in finance and insurance are also considered.

C791: **Multivariate elliptic processes**
Presenter: **Nick Bingham**, Imperial College London, UK

Multivariate elliptic processes provide a dynamic counterpart to earlier work on multivariate elliptic distributions. There, the aim was to provide a flexible semi-parametric approach to modelling the distribution of a high-dimensional portfolio of assets, avoiding the curse of dimensionality and able to model a variety of tail behaviour. Here, these features are successfully extended to behaviour of such portfolios over time (discrete or continuous). In particular, our approach is able to handle stochastic volatility and volatility clustering very simply and naturally.

C809: **Aggregation and capital allocation for portfolios of dependent risks**

Presenter: **Etienne Marceau**, Universite Laval, Canada

Co-authors: Helene Cossette

We consider an insurance portfolio consisting of several dependent risks and we aim to evaluate not only the capital allocation for the overall portfolio but also the contribution of each risk over their aggregation. We use the Tail Value at Risk (TVaR) as risk measure. We consider the following dependence models: multivariate compound distributions, models based on common mixtures and models based on copulas. We present special cases where exact expressions for the TVaR of the sum of the risks and for the TVaR-based allocations are obtained. We then propose numerical methods applicable for any proposed dependence models in order to approximate the TVaR of the aggregate risk for the portfolio and the contribution of each risk of the portfolio. Several numerical examples are presented in order to illustrate the topics exposed here.

C763: **Asset-liability management for pension funds in a time-varying volatility environment**

Presenter: **Spyridon Vrontos**, University of Piraeus, Greece

Co-authors: Ioannis Vrontos, Loukia Meligkotsidou

The determination of optimal dynamic ALM strategies in pension funds is an important issue. We consider standard approaches to pension fund liability modelling, although our asset management setting can be combined with more sophisticated models for the liabilities. The innovation of this work concerns issues related to asset management. Our aim is to develop modelling devices for asset return series, which take into account the particular characteristics of these series, in order to improve our ability to construct optimal asset portfolios. We address the issue of time-varying variances and covariances/correlations of asset returns and concentrate on the potential impacts in terms of pension fund portfolio construction and risk measurement.

CS50 Room MAL B20 MULTIVARIATE FINANCIAL TIME SERIES Chair: Cathy Chen

C109: **Financial time series clustering with nonparametric Bayes method**

Presenter: **Xiyuan Qian**, East China University of Science and Technology, China

A novel financial time series clustering method based on a novel similarity metric is developed. This similarity metric is defined by the distance between two stocks' log return distribution (PDF). Since the log return distribution of stocks empirically can not be treated as normal distribution, we apply infinite Gaussian mixture modeling, a nonparametric Bayesian method, to model the log return distribution of stocks. As this approach is Bayesian, it allows us to integrate prior knowledge about the distribution in a principled way and since it is non-parametric we are able to avoid model selection, to let the financial time series data "speak directly". Based on the similarity measure, we apply the single linkage agglomerative algorithm to cluster the financial time series. The dendrograms and trees of links among the financial assets are generated, which corresponds to a large extent to the relevant economic sectors. We apply this approach to analyze the $N = 30$ shares composing the Dow Jones industrial average (DJIA) index, collecting the daily closure prices of its stocks for a period of 5 years (2006-2010) which includes the financial crisis period, to find companies that share a similar behavior. We compare this approach with that which clusters the 30 stocks with a Hausdorff distance as a similarity metric. Comparisons are also made with other similarity metrics based clustering results.

C329: **Adaptive estimation of vector autoregressive models with time-varying variance.**

Presenter: **Hamdi Raissi**, INSA Rennes, France

Co-authors: Valentin Patilea

Linear Vector AutoRegressive (VAR) models where the innovations could be unconditionally heteroscedastic and serially dependent are considered. The volatility structure is deterministic and quite general, including breaks or trending variances as special cases. In this framework we propose Ordinary Least Squares (OLS), Generalized Least Squares (GLS) and Adaptive Least Squares (ALS) procedures. The GLS estimator requires the knowledge of the time-varying variance structure while in the ALS approach the unknown variance is estimated by kernel smoothing with the outer product of the OLS residuals vectors. Different bandwidths for the different cells of the time-varying variance matrix are also allowed. We derive the asymptotic distribution of the proposed estimators for the VAR model coefficients and compare their properties. In particular we show that the ALS estimator is asymptotically equivalent to the infeasible GLS estimator. This asymptotic equivalence is obtained uniformly with respect to the bandwidth(s) in a given range and hence justifies data-driven bandwidth rules. Using these results we build Wald tests for the linear Granger causality in mean which are adapted to VAR processes driven by errors with a non stationary volatility. It is also shown that the commonly used standard Wald test for the linear Granger causality in mean is potentially unreliable in our framework. Monte Carlo experiments illustrate the use of the different estimation approaches for the analysis of VAR models with stable innovations.

C546: **Modelling time-varying correlation with parsimonious multivariate linear ARCH**

Presenter: **CY (Chor-yiu) Sin**, National Tsing Hua University, Taiwan

Contrast to a typical volatility model, a linear ARCH (LARCH) model generalizes an RCA model without explicitly specifying the conditional variance. We adopt a parsimonious version of the multivariate LARCH. The salient features of this model include (i) the time-varying conditional variance-covariance matrix is positive semi- definite (and thus all the time-varying conditional correlations lie between -1 and +1); (ii) compared to the general multivariate RCA model or the general multivariate LARCH model, the number of parameters is more manageable; (iii) a factor structure can further be incorporated. With some discussion on the parameter identification, we derive the asymptotic distribution of the QMLE, specifying some primitive assumptions on the multivariate LARCH process. Estimation, order selection and diagnostic checking are suggested and their finite-sample performance is assessed with a small-scale Monte-Carlo experiment.

C664: **Spectral sparsity and the modeling of high dimensional heteroscedastic phenomena**

Presenter: **Juan-Pablo Ortega**, CNRS/Universite de Franche Compte, France

Co-authors: Stephane Chretien

The use of VEC type models in the treatment of high dimensional heteroscedastic phenomena is not free from difficulties due to their lack of parsimony and the complications encountered at the time of formulating sufficient conditions for stationarity and positivity of the resulting covariance matrices. In this talk we will start by providing a global matrix formulation of the calibration problem that will allow us to encode it as a semidefinite programming routine. Moreover, we will consider how recent advances in compressed sensing and first order optimization methods may be used to tackle the calibration problem as the spectral sparsity exhibited by the covariance matrices associated to the components of large portfolios makes particularly appropriate this type of approach.

C467: **A Bayesian nonparametric alternative for multivariate models**

Presenter: **Sjoerd Van den Hauwe**, Erasmus Universiteit Rotterdam, Netherlands

Co-authors: Richard Paap, Dick van Dijk

We propose a robust nonparametric approach to model joint distributions by means of a Dirichlet process prior. As a result, the joint distribution is approximated by a mixture of multivariate normals of ex ante unknown number which also imply the marginal distributions. The focus is on the ability of the model to capture exotic joint behavior, like joint extreme events and changing correlations over the sample space, and marginal properties like skewness and heavy tails, phenomena typically found in financial applications. Though the common stepwise copula approach, make a choice for the marginal distributions and pick a copula function to model the dependence structure, has obvious advantages, empirical results very much rely on the specific (parametric) choices of both copula and marginals. We investigate these two approaches to joint modeling in a comparative study with simulated data sets where we adopt a Bayesian analysis. In order to do so, we also propose a simulation scheme to sample marginal and copula parameters simultaneously. In a real data example that models the dependence between international stock markets we investigate the forecasting implications of both modeling strategies in terms of key measures like value-at-risk and expected shortfall.

CS57 Room MAL B34 BAYESIAN NONPARAMETRIC METHODS IN ECONOMETRICS Chair: Mark Steel

C637: **Bayesian nonparametric models in macroeconomics: an illustration**
Presenter: **Igor Pruenster**, University of Torino & Collegio Carlo Alberto, Italy
Co-authors: Antonio Lijoi, Pietro Muliere, Filippo Taddei

Bayesian nonparametric models are becoming increasingly popular in a variety of application areas. Among them, some interesting economic phenomena have recently been described by resorting to probabilistic tools that are well-established in Bayesian Nonparametrics. In line with these recent advances, the talk focuses on applications of the two parameter Poisson-Dirichlet process. It is shown how this process lends itself to modeling effectively the uncertainty related to some macroeconomic quantities, whose asymptotic behaviour is then thoroughly studied.

C648: **Nonparametric spline regression with portable knots**
Presenter: **Mattias Villani**, Sveriges Riksbank, Sweden
Co-authors: Feng Li

A surface model is a flexible model of the regression surface E(y|x) where there may be interactions between the covariates. It is possible to extend the traditional spline approach with multidimensional basis functions, e.g. radial basis functions, using a fixed set of knots in covariate space, but this is unlikely to work well in moderate to large covariates spaces. We treat the locations of the multidimensional knots as unknown parameters and compute the joint posterior distribution of the knot locations and their regression coefficients using an efficient MCMC algorithm.

C731: **Inferring differences between distributions**
Presenter: **Jim Griffin**, University of Kent, UK
Co-authors: Michalis Kolossiatis, Mark Steel

Observations often naturally fall into groups and understanding the difference between the distributions of the observations for each group is the main focus of modelling. For example, if firms are grouped according to management practices, we might want to understand how the efficiency distribution changes with management practice. Most straightforwardly, we could model the means. However, often in econometrics we are interested in the distribution of unobserved quantities and so it's hard to decide whether such an analysis is adequate. Alternatively, we could model differences across the whole distribution using nonparametric methods. In this talk a Bayesian nonparametric approach will be described including construction of priors, MCMC methods for estimation and a method for deciding where the distributions differ.

C729: **Modelling the conditional distribution of daily stock index returns: an alternative Bayesian semiparametric model.**
Presenter: **Maria Kalli**, University of Kent, UK
Co-authors: Stephen Walker

The aim is to develop a family of Bayesian semi-parametric models for the conditional distribution of daily stock index returns. The proposed models capture critical empirical features of such returns, namely heavy tails, skewness and volatility clustering. Such features have proven difficult to model accurately using parametric models. A Bayesian nonparametric prior is used to generate random density functions that are unimodal and asymmetric. Volatility is modelled parametrically. We apply our model to the daily returns of the S&P 500, EUROSTOXX 50 and FTSE 100 indices and provide a comprehensive empirical analysis comparing our model results to those of GARCH, Stochastic Volatility, and other Bayesian semi-parametric models.

C773: **Bayesian semiparametric instrumental variable regression with non-normal errors**
Presenter: **Manuel Wiesenfarth**, University of Gottingen, Germany
Co-authors: Carlos Matias Hisgen, Thomas Kneib, Carmen Cadarso-Suarez, Daniel Miles Touya

We propose a semiparametric instrumental variable approach that allows us to correct for endogeneity bias in regression models where the covariate effects enter with unknown functional form. Bias correction relies on a simultaneous equations specification with flexible modeling of the joint error distribution. We specify both the structural and instrumental variable equation based on geoadditive predictors comprising penalised splines for nonlinear effects of continuous covariates, Gaussian Markov random fields for spatial effects, and a Dirichlet process prior to deal with flexible estimation of the distribution of the error terms. Inference is fully Bayesian, employing efficient Markov Chain Monte Carlo simulation techniques. The resulting posterior samples do not only provide us with point estimates, but allow us to construct uniform confidence bands for the nonparametric effects, which was not discussed so far in the simultaneous equations literature. In simulations, we show that the Bayesian model formulation is advantageous in small sample scenarios in comparison to classical procedures, particularly in the case of weak instruments and non-normal errors. The investigation of the effect of class size on student performance in Uruguay provides an illustration of the proposed approach, which is implemented in an easy-to-use R package.

CP01 Room IoE Crush Hall POSTERS I Chair: Christos Savva

C855: **Multi-scale correlation analysis of coherent variability of economical time series**
Presenter: **Anton Chernenko**, Space Research Institute (IKI), Russia

Alongside with fundamental trends, economical variables, such as stock prices, are driven by large number of variable or even transient factors of different durations and at unpredictable moments of time. A natural framework for identification of such hidden factors has been factor analysis which derives significant factors from the investigation of the correlation matrix of the time series. For long periods of observations, it is obvious that, the observed time series should be superposition of a large number of independent factors. However, for certain short time intervals one may expect that a single important factor may be the dominant reason for variability of all observed variables. On the other hand, for short time series large number of variables, correlation matrices become singular and direct analysis of these matrices becomes more relevant then any technique

based on their eigenvalues. In this paper we present statistical properties of sample correlation matrices for end-of-day prices of 100 randomly selected stocks from NYSE on multiple times scales from 1 week to 1 year from 1990 till 2010. We show how periods of coherent behaviour could be extracted and used to detect and understand unusual events on market, such as insider trades.

C614: **Further developments on time series clustering**

Presenter: **Margherita Gerolimetto**, Ca' Foscari University of Venice, Italy

Co-authors: Isabella Procidano

Clustering time series has recently received a lot of attention. It can be very useful, for example, to classify the economic situation of a country by studying some time series indicators. There are two broad categories of clustering time series methods, depending upon whether they work in the frequency domain or in the time domain. Moreover, within each category, methods may i) work directly with raw data, ii) work indirectly with features extracted from the data, iii) work indirectly with models built from the raw data. Unsurprisingly, a very relevant issue in this framework, is the choice of the measure of the dissimilarity between time series, of which there is a number of proposals in literature. In this work we conduct, firstly, a Monte Carlo study to investigate how much the results are affected by the chosen dissimilarity measure and, possibly, to find out if there are specific situations where some dissimilarity measures may be more recommended than others. Then we discuss the issue of classifying time series from a different point of view using Kohonen Self Organizing Maps (typically employed for cross-sectional data classification) and present an application of this to the Italian Industrial Production Index series.

C740: **Comparison of different misspecification tests designed for nonlinear time series models**

Presenter: **Leena Kalliovirta**, University of Helsinki, Finland

We simulate nonlinear time series models based on mixture distributions and use these models to compare the size performance of different recently suggested misspecification tests. The uncertainty caused by parameter estimation is properly taken into account in all these tests. However, one employs different methodologies to achieve such a correction. We find that tests based on moments have more reliable size properties than tests that employ the empirical distribution function or a nonparametric estimate of the generalized characteristic function.

C598: **Capturing correlated effects in adoption and diffusion in large virtual networks**

Presenter: **Elenna Dugundji**, Universiteit van Amsterdam, Netherlands

Co-authors: Ate Poorthuis, Michiel van Meeteren

With the onset of Web 2.0, people leave numerous traces of their behavior in – often publicly available – data sets. In this paper we study a virtual community of independent software developers that use the social networking site Twitter. We collect longitudinal data on network connections among the developers, their friends and followers (approximately 15,000 nodes) and their Twitter behavior over a period of five weeks (more than 600,000 Tweets). We use this large dynamic panel data to analyze the adoption and diffusion of Twitter client software. We estimate individual preferences, social contagion, opinion-maker influence and other contextual effects, behavioral characteristics, socio-centric network measures and ego-centric network measures. Furthermore, we simulate the size of the error due to unobserved correlated effects. This is critical to test for in any application of multinomial logistic regression where social influence variables and/or other network measures are used as explanatory variables, since their use poses a classic case of endogeneity. We show that even in a seemingly saturated model, the model fit can increase significantly by accounting for unobserved correlated effects, and the estimated coefficients in the uncorrected model can be significantly biased. Failing to account for correlated effects can yield misleading policy interpretations.

C282: **A new correlation coefficient for bivariate time-series data**

Presenter: **Yusuf Varli**, Istanbul Bilgi University, Turkey

Co-authors: Elvan Ceyhan, Orhan Erdem

Various financial models such as pairs trading etc. are interested in the correlation between two different time series data, e.g. stock prices, returns etc. Pearson's correlation coefficient is one of the most commonly used formulas to measure this correlation. However there are many assumptions for the usage of this formula that people ignore. In this article we propose a new correlation coefficient that measures the distance between two subsequent data points by taking the lag difference. Even though the very-first data point is lost, we demonstrate that the new correlation coefficient captures the direction of the movements of the two variables over time. We also propose various extensions of this coefficient to obtain more reasonable and reliable results at the expense of having more complex formulas

C658: **Bayesian stochastic frontier models with heterogeneity**

Presenter: **Helena Veiga**, Universidad Carlos III de Madrid, Spain

Co-authors: Jorge Galan, Michael Wiper

Frontier models are a wide research area with very relevant theoretical developments and plenty of applications of increasing interest for managers and policy makers. There exist two main approaches in this literature, the parametric and the nonparametric approach. The former relies on the specification of a functional form while the latter is more flexible. More recently, a Bayesian approach to Stochastic Frontier Analysis (SFA) has become very influential given its interesting properties. However, a general framework for the treatment of heterogeneity is missing under Bayesian SFA. Therefore, we propose to define a complete framework to study the effects of heterogeneity as a contribution to Bayesian SFA literature. On this work, we present basic heterogeneity models with effects on the variance and mean of inefficiencies for measuring technical efficiency and study their estimation properties via Monte Carlo simulations. Finally, we perform an empirical application, whose results evidence a relevant effect of heterogeneity on the variance of inefficiencies. These findings motivate further research on a full heterogeneity framework for Bayesian SFA with extensions to dynamic effects and semiparametric models.

C780: **A comparison of approximations for compound Poisson processes**

Presenter: **Raffaello Seri**, Universita degli Studi dell Insubria, Italy

Co-authors: Christine Choirat

The aim is to provide a comparison of the error in several approximation methods for the cumulative aggregate claim distribution customarily used in the collective model of insurance theory and in quantitative risk management. If the number of claims in a time period follows a Poisson process, the cumulative aggregate claim distribution is distributed as a compound Poisson process. Several approximations have been proposed for this distribution: here we only consider the ones that use information on the lower order moments of the involved distributions. We consider the normal, Edgeworth, NP2, NP2a, adjusted NP2, NP3, Gamma, Inverse Gaussian and Gamma-IG approximations and we provide a comparison of the error in these methods as the Poisson intensity diverges to infinity. Several statements in the literature concerning the quality of approximations for the distribution of the aggregate claim process can find theoretical support through our results. Other statements can be disproved on the same grounds.

Friday 10.12.2010 10:15 - 12:20 Parallel Session B – ERCIM

EI80 Room Senate Beveridge Hall INVITED SESSION: ROBUST METHODS Chair: Stefan Van Aelst

E111: **Accurate and robust inference**

Presenter: **Elvezio Ronchetti**, University of Geneva, Switzerland

Classical statistics and econometrics typically rely on assumptions made on the structural and the stochastic parts of the model and on optimal procedures derived under these assumptions. Standard examples are least squares estimators in linear models and their extensions, maximum likelihood estimators and the corresponding likelihood-based tests, and GMM techniques in econometrics. Inference is typically based on approximations obtained by standard first-order asymptotic theory. However, in the presence of small deviations from the assumed model, this can lead to inaccurate p-values and confidence intervals. Moreover, when the sample size is moderate to small or when probabilities in the tails are required, first-order asymptotic analysis often fails for classical and even for robust procedures. In this talk we review some techniques which combine robustness and good accuracy in finite samples. They are derived using saddlepoint methods and provide robust and accurate tests and confidence intervals with relative errors for coverage probabilities of order O(1/n). The theory is discussed in detail in several important classes of models, including linear and generalized linear models, functional measurement error models, and indirect inference in diffusion models.

E121: **High breakdown point and efficient estimates for generalized linear models**

Presenter: **Victor Yohai**, Universidad de Buenos Aires-CONICET, Argentina

Co-authors: Andrea Bergesio

We present a class of estimators for generalized linear models which combine high breakdown point and high efficiency. The proposed estimators are one step weighted M estimators, where the weights penalize leverage points. The initial estimator is a projection estimator similar to those defined for the linear model. The final estimator is obtained by performing one step of the scores algorithm for computing weighted M-estimates. We prove that the proposed estimators may have a breakdown point as close to 0.5 as desired. Moreover, we also prove that they are asymptotically normal with the same asymptotic covariance matrix as the fully iterated weighted M-estimate. A Monte Carlo simulation shows that these estimators perform favorably for finite sample size compared with other robust estimators.

E350: **Robust Sparse K-Means**

Presenter: **Ruben Zamar**, UBC, Canada

We consider the problem of clustering using a large set of variables. The true underlying clusters present in the data may be determined by a relatively small fraction of the variables, and may be missed if one clusters the observations using the full set of features. A novel framework for sparse clustering in which the clusters are found using an adaptively chosen subset of the features has been previously proposed. This method for sparse k-means uses a lasso-type penalty to select the features. Unfortunately this procedure cannot handle outliers nor missing data, which are prevalent in many applications. We propose a modification of sparse k-means which we call robust-sparse k-means, that meets these two requirements. We show some simulation results and an application to a large dataset.

ES21 Room MAL 538 CLASSIFICATION AND DISCRIMINANT PROCEDURES FOR DEPENDENT DATA Chair: Andres M. Alonso

E224: **A comparative study of dissimilarity measures for time series classification**

Presenter: **Jose A. Vilar**, University of A Coruna, Spain

Co-authors: Sonia Pertega, Juan Vilar

One key point in cluster analysis is to determine a similarity or dissimilarity measure between data objects. When working with time series, the concept of similarity can be established in different ways. In this work, several nonparametric statistics originally designed to test the equality of the log-spectra of two stochastic processes are proposed as dissimilarity measures between time series data. Their behavior in time series clustering is analyzed throughout a simulation study, and compared with the performance of several model-free and model-based dissimilarity measures. Up to three different classification settings are considered: (i) to distinguish between stationary and non-stationary time series, (ii) to classify different ARMA processes and (iii) to classify several non-linear time series models. As expected, the performance of a particular dissimilarity metric strongly depended on the type of processes subjected to clustering. Among all the measures studied, the non-parametric distances showed the most robust behavior.

E297: **The factor structure of international stock market returns**

Presenter: **Jorge Caiado**, CEMAPRE/University of Lisbon, Portugal

Co-authors: Joao Bastos

The behavior of international stock market returns in terms of rate of return, unconditional volatility, skewness, excess kurtosis, serial dependence and long-memory is examined. A factor analysis approach is employed to identify the underlying dimensions of stock market returns. In our approach, the factors are estimated not from the observed historical returns but from their empirical properties, without imposing any restriction about the time dependence of the observations. To identify clusters of markets and multivariate outliers, factor analysis is then used to generate factor scores. The findings suggest the existence of meaningful factors which determine the differences in terms of the dependence structure between developed and emerging market returns.

E386: **Mixture models for multiple time series**

Presenter: **Nicholas Longford**, Universitat Pompeu Fabra, Spain

Co-authors: Pierpaolo D'Urso

In the analysis of multiple time-series or panel data, we come across problems similar to the analysis of other multivariate data – a particular simple model fits well to a subset of the cases, but not to all of them. We apply mixture models to such a setting, assuming that the cases in the study represent a small number of distinct subpopulations (or clusters) to which different models apply. In particular, we consider mixtures of autoregressive models of possibly different orders. Further, we consider a subset of the cases to which no model applies, and associate them with an improper (black-hole) distribution. An example with monthly income from outlets of a retail company is presented.

E563: **Cluster identification for Gaussian processes by projection pursuit**

Presenter: **Pedro Galeano**, Universidad Carlos III de Madrid, Spain

A procedure to identify clusters in a sample of Gaussian processes by projection pursuit is proposed. The main idea of the procedure is to project the

sample onto directions that minimize and maximize the kurtosis coefficient of the projected dataset. The projected univariate data obtained allow us to group the observations according to the values of the gaps between consecutive ordered observations. The proposed algorithm is iterative, flexible, robust to outliers, fast to implement, and appears to work well in practice, as shown in several simulation studies and a real data example.

E665: **Cluster analysis of European daily temperature series: an extreme value approach**

Presenter: **Andres M. Alonso**, Universidad Carlos III de Madrid, Spain

Co-authors: Susana Barbosa, Manuel Scotto

Time series of daily mean temperature obtained from the European Climate Assessment dataset are analyzed with respect to their extremal properties. A time series clustering approach which combines Bayesian methodology, extreme value theory, and classification techniques is adopted for the analysis of the regional variability of temperature extremes. The daily mean temperatures records are clustered on the basis of their corresponding predictive distributions for 25-, 50- and 100-years return values. The results of the cluster analysis show a clear distinction between the highest-altitude stations, for which the return values are lowest, and the remaining stations. Furthermore, a clear distinction is also found between the northernmost stations in Scandinavia, and the stations in central and southern Europe. This spatial structure of the return periods distributions for 25-, 50- and 100-years seems to be consistent with projected changes in the variability of temperature extremes over Europe pointing to a different behaviour (increased variability) in central Europe than in northern Europe and the Mediterranean area, possibly related to the effect of soil moisture and land-atmosphere coupling.

ES33 Room Senate Bloomsbury Room OPTIMAL DESIGN AND MODEL CHOICE Chair: Jesus Lopez-Fidalgo

E194: **T-optimality and observational data: some examples**

Presenter: **Rossella Berni**, University of Florence, Italy

The use of observational data to implement an experimental design by applying optimal experimental design criteria is considered. The aim is to show an efficient use of the available data by exploiting the T-optimality criteria and building an experimental design which may be assumed optimal when considering a sequential optimality. Undoubtedly, an experimental design is characterized by a specific context; sources of variability and variables may be studied and planned by considering the a-priori information. The system is under the control of the experimenter: elements characterizing this difference include level setting of variables under study, methods and measurement procedures, random allocation of trials to experimental units. Therefore, the proposed procedure is based on several steps, aimed at avoiding some of the deficiencies in observational data; the final result is the implementation of an algorithm where the selection of each trial is performed by considering the factor levels and the corresponding real response value. Furthermore, it is quite common in industry to collect large data-sets over time and our proposal is specifically directed to improve this situation. Empirical and simulated examples applied in the technological field are shown, in order to support our proposal.

E265: **Staggered designs for experiments with several hard-to-change factors**

Presenter: **Peter Goos**, Universiteit Antwerpen, Belgium

Co-authors: Heidi Arnouts

In many industrial experiments, some of the factors are not independently reset for each run. This is due to time and/or cost constraints and to the hard-to-change nature of these factors. Most of the literature restricts the attention to split-plot designs in which all the hard-to-change factors are independently reset at the same points in time. This constraint is to some extent relaxed in split-split-plot designs because these require the least hard-to-change factors to be reset more often than the most hard-to-change factors. A key feature of the split-split-plot designs, however, is that the least hard-to-change factors are reset whenever the most hard-to-change factors are reset. In this article, we relax this constraint and present a new type of design which allows the hard-to-change factor levels to be reset at entirely different points in time. We show that the new designs are cost-efficient and that they outperform split-plot and split-split-plot designs in terms of statistical efficiency. Because of the fact that the hard-to-change factors are independently reset alternatingly, we name the new design a staggered design.

E750: **Numerical methods in optimal design**

Presenter: **Ben Torsney**, University of Glasgow, UK

In the approximate approach to optimal design determination, the objective is to determine optimal weights, the weights or proportions of observations which will be taken at each design point. When the design space is continuous, observations will only be taken at a small number of design points, known as the support of the design. A useful initial task is to attempt to narrow down what this set is. Empirical results can easily demonstrate that the Wynn-Fedorov-Vuchkov algorithm, which changes one weight at a time, can quickly identify most non-support points. We are then left with support points and a few neighbouring unidentified non-support points. Multiplicative algorithms, which update all weights simultaneously can then perform well in finding optimal weights. I will briefly review this progression and report on a new approach which is a kind of fusion of the twin problems of support identification and optimal weight determination.

E828: **Design efficiency for component amount mixture models homogeneous of degree zero and one**

Presenter: **Ralf-Dieter Hilgers**, RWTH Aachen, Germany

Component amount mixture models are extensions of the classical mixture experiments, where the response depends not only on the proportions of the components but also on the total amount. If the maximal amount of the individual q components is restricted, the sample space is given by a q dimensional simplex. As an extension of the classical models to describe mixture experiments various types of regression models were introduced to describe the response surface in this setting. A special class are models, which are homogenous of degree zero or one. Only some results concerning optimal designs for these special classes of regression models are available. Thus efficiency considerations are necessary. In the talk an upper bound for the G-efficiency in the non full saturated regression models, which are homogeneous of degree zero and one, will be derived. It is shown that the designs not supported by all barycenters have bad G-efficiency. The results indicate that the optimal designs are supported on all barycenters even in the non full saturated cases. Optimal designs for the component amount experiments are interesting from the theoretical point, because they show how optimality changes, when an intercept is included in the regression model and the design space is extended.

E868: **Design with smooth polynomial methods**

Presenter: **Hugo Maruri Aguilar**, Queen Mary, University of London, UK

Co-authors: Henry Wynn, Ron Bates

Smooth supersaturated polynomial interpolators are an alternative to modelling computer simulations. They have the flexibility of polynomial modeling, while avoiding the inconvenience of undesired polynomial oscillations (i.e. Runge's phenomenon). Smooth polynomials have been observed to be most effective for small sample sizes, although their use is not restricted in this respect. The talk will survey the smooth polynomial technique, and then concentrate on the implications these models have for designing experiments. Extensions and examples will be presented.

ES35 Room MAL 541 VALIDATION AND SPECIFICATION TESTS WITH STRUCTURED DATA Chair: M. Dolores Jimenez-Gamero

E168: **Testing for conditional symmetry in absolutely regular and possibly nonstationary dynamical models**

Presenter: **Joseph Ngatchou-Wandji**, EHESP de Rennes and Universite Nancy 1, France

Co-authors: Michel Harel

We propose a symmetry test for the errors distribution in a class of heteroscedastic models. The observations as well as the errors are not necessarily stationary, but are required to be absolutely regular. The convergence of the residual-based empirical distribution function is established, as well as other functional limit theorems. The null cumulative distribution function of the test statistic is approximated. A simulation experiment shows that the test performs well on the example tested.

E309: **Nonparametric tests for risk-return relationships**

Presenter: **Juan Carlos Pardo-Fernandez**, Universidade de Vigo, Spain

Co-authors: Juan Carlos Escanciano, Ingrid Van Keilegom

Nonparametric specification tests for checking parametric risk-return relationships in a dependent data context are presented. Specifically, given a response variable and a vector of covariates, we present tests for possible parametric specifications of the relationship between the conditional mean and the conditional variance of the response given the covariate. We allow for unobserved, but parametrically generated, covariates. A distinctive feature of our setup is that it does not require parametric models for the conditional mean and conditional variance, while it allows for flexible parametric risk-return relationships. This setup covers many interesting problems in finance, such as testing for the form of the Sharpe ratio. The tests are based on the difference of the estimated restricted and unrestricted pricing error distributions. A suitable transformation of this difference renders the tests asymptotically distribution-free, with limits that are functionals of a standard normal variable. Hence, the tests are easy to implement. Several examples will be studied in detail, and simulations will show the practical performance of the proposed methodology.

E583: **Goodness-of-fit tests in semi-linear models**

Presenter: **Jochen Einbeck**, Durham University, UK

Co-authors: Simos Meintanis

We introduce specification tests for the semi-linear regression model $y = Xb + g(z) + e$, where Xb is a parametric term and $g(z)$ a smooth, unspecified, uni-or multivariate, function. Specifically, tests are proposed for the hypotheses : 1) The distribution function of the residuals, e, belongs to a specific parametric family of distributions (in particular, we consider the Gaussian and Laplace distribution). 2) The distribution of the residuals is symmetric about 0. The test statistic is (in both cases) based on the empirical characteristic function of the residuals. Due to difficult asymptotics, a bootstrap technique is utilized for the determination of p-values. It is demonstrated that, under the null hypothesis, the bootstrap tests reproduce closely the target significance level, while they achieve higher test powers than competing (bootstrapped or non-bootstrapped) tests based on the empirical distribution function of the residuals. The method is illustrated through real data examples.

E602: **Two classes of divergence statistics for testing uniform association.**

Presenter: **Virtudes Alba-Fernandez**, University of Jaen, Spain

Co-authors: M. Dolores Jimenez-Gamero

The problem of testing uniform association in cross-classifications having ordered categories is considered. Specifically, two families of test statistics, both based on divergences between certain functions of the observed data, are studied and compared. Our theoretical study is based on asymptotic properties. For each family, two consistent approximations to the null distribution of the test statistic are studied: the asymptotic null distribution and a bootstrap estimator; we also show that the tests considered are all of them consistent against fixed alternatives; finally, we do a local power study. The finite sample performance of the tests in these two classes is numerically investigated through some simulation experiments.

E715: **A simple misspecification test for regression models**

Presenter: **Jean-Baptiste Aubin**, Insa-Lyon, France

Co-authors: Samuela Leoni-Aubin

We consider the following model in the context of univariate regression $Y_i = m_0(x_i) + \varepsilon_i$, where the response variable Y is observed at n fixed design points x_i of a covariate x and m_0 is the true regression function. Moreover, the ε_i are independent and centered random variables such that $P(\varepsilon_i > 0) = P(\varepsilon_i < 0) = 1/2$ for all i. A simple test is proposed for examining the correctness of a given completely specified response function m against unspecified general alternatives. Technically, the test statistic is the maximum length of the sequences of ordered (with respect to the covariate) observations that are consecutively overestimated or underestimated by the candidate regression function. This test is very simple and can be computed visually if the sample size is small enough and is a modification of a nonrandomness test. Note that the testing procedure can cope with heteroscedastic errors and that, contrary to other classical tests, no replicates are needed to compute the test statistic. Recursive formulae to calculate the exact distribution of the test statistic under the null hypothesis and under a class of alternative hypotheses are given here. The relative performances of the longest run test and other classical tests (the Rainbow test, the RESET test and the CUSUM test) are compared by means of a simulation study.

ES17 Room MAL 532 TIME SERIES Chair: Claudia Kirch

E103: **Early warning with calibrated and sharper probabilistic forecasts**

Presenter: **Reason Machete**, University of Reading, UK

Given a non-linear deterministic model, a density forecast is obtained by evolving forward an ensemble of starting values and doing density estimation with the final ensemble. The density forecasts will inevitably be downgraded by model misspecification. To mitigate model misspecification and enhance the quality of the predictive densities, one can mix them with the system's climatology (or time series density). This paper examines the effect of including the climatology on the sharpness and calibration of density forecasts at various time horizons. A new theorem is also presented that helps explain the findings. The density forecasts are estimated using a non-parametric approach. The findings have positive implications to issuing early warnings in different disciplines including economic applications and weather forecasting.

E801: **Robust online monitoring**

Presenter: **Roland Fried**, TU Dortmund University, Germany

Co-authors: Ursula Gather, Herold Dehling

Automatic online monitoring time series are often disturbed by a high level of noise and different types of measurement artefacts. A basic task

is the reliable detection of abrupt changes with only short time delays. For a robust statistical analysis of this problem we need to distinguish relevant changes from short-term fluctuations and (patches of a few) outlying observations. Reliable methods for sequential (online) detection and classification of changes in the data-generating mechanism are an essential component of decision support. 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 median of all pairwise differences between the observations in subsequent moving time windows, which is the Hodges-Lehmann two-sample estimator for a difference in location and can be derived from the Wilcoxon test. For online monitoring of time series with slowly varying trends, we can additionally estimate the underlying trend applying Siegel's repeated median to a moving time window. The usefulness of the approaches is investigated via simulations and applications to some real time series.

E341: **CloCK-TiME: cloud computing kernel for time-series modeling engine**

Presenter: **Hiromichi Nagao**, The Institute of Statistical Mathematics, Japan

Co-authors: Tomoyuki Higuchi

Bayesian multivariate time-series analysis is recognized to be a powerful method in various fields of science in order to discover new knowledge buried commonly in different degrees of observation. However, this type of analysis usually requires a complex programming and costs immeasurable computation time especially in the case that non-Gaussian distributions are included in its state space model. For the purpose to provide a platform of multivariate analysis available to everyone, we have developed web application *CloCK-TiME* (Cloud Computing Kernel for Time-series Modeling Engine). CloCK-TiME applies a multivariate analysis based on the particle filter algorithm to uploaded time-series data by utilizing network distributed PC clusters in a cloud computing system. The details of analysis can be controlled interactively through a Flash-based user interface, e.g., select trend, seasonal and/or autoregressive (AR) components to be included in the observation model, set trend order, seasonal period, AR order, distribution form for each system noise, and the number of candidate hyperparameters. The user interface draws the results on display such as a list of optimum value and prior/posterior distribution for each model parameter. CloCK-TiME is now open to public via our website http://sheep.ism.ac.jp/CloCK-TiME/index.html.

E623: **Dynamic mixture models with covariate interventions**

Presenter: **Juan-Carlos Martinez-Ovando**, Universita degli Studi di Torino & University of Kent, UK

Co-authors: Stephen G. Walker

The construction and (Bayesian) inference of a novel class of semi-parametric dynamic mixture models with transition dynamics affected by covariate interventions will be presented. It will be shown that the model is coherent to marginalization and that it has a unique invariant distribution. By construction, the covariate-intervention appears only in the transition dynamics, while leaving the invariant distribution unaffected. Illustrations will be given to the study of event-count data and stochastic volatility modeling.

E289: **Assessing the relationship between shadow economy and the unemployment rate in U.S.**

Presenter: **Adriana AnaMaria Alexandru**, Academy of Economic Studies, Romania

Co-authors: Ion Dobre, Catalin Ghinararu

The aim is to estimate the size of the U.S. shadow economy (SE) using a structural equation approach with quarterly data for 1980-2009. For this purpose, the shadow economy is modeled like a latent variable using a particular type of the structural equation models-the MIMIC model. Its dimensions (i.e. SE) are estimated to be decreasing over the said period of time. Furthermore, we also investigate the long-run equilibrium relationship and the direction of causality between U.S. shadow economy (SE) and the unemployment rate (UR) using bounds test for co-integration and Granger causality approach. The empirical results reveal the existence of a long run relationship between the two variables with Granger causality tests identifying a unique direction of causality that runs from unemployment rate to shadow economy though only in the short term.

ES19 Room MAL 351 COMPUTATIONAL STATISTICS FOR CLINICAL RESEARCH Chair: Joyce Niland

E378: **A model selection approach to genome wide association studies**

Presenter: **Florian Frommlet**, University Vienna, Austria

Co-authors: Malgorzata Bogdan, Felix Ruhaltinger

For the vast majority of genome wide association studies (GWAS) published so far, statistical analysis is performed by testing markers individually. Elementary statistical considerations show that in case of complex traits model selection is preferable to multiple testing. We will introduce a model selection approach to GWAS based on modifications of BIC and develop some simple but efficient search strategies to deal with the huge number of potential models. Comprehensive simulations based on real SNP data confirm that model selection has larger power than multiple testing to detect causal SNPs in complex models. On the other hand multiple testing tends to detect systematically particular false positive SNPs. Finally we discuss the advantages of our model selection approach in the context of real GWAS, where we consider publicly available gene expression data as traits for individuals from the HapMap project.

E226: **Stereotype models for joint analysis of quality of life and survival**

Presenter: **Kemmawadee Preedalikit**, Victoria University of Wellington, New Zealand

Co-authors: Ivy (I-Ming) Liu, Nokuthaba Sibanda

The joint models for longitudinal and survival data have been widely discussed in the literature. This talk proposes a joint model using a stereotype model for the longitudinal ordinal responses. The stereotype model can improve the fit by adding extra score parameters, but it still has the advantage of requiring only a single parameter to describe the effect of a predictor. We give an example to model quality of life and survival time for patients being followed up after treatments. The ordinal quality of life measures are modelled using the stereotype model, while the survival time is modelled using Cox's proportional hazards model. They are linked through a continuous latent variable that is assumed to underlie the quality of life and the hazard rate. We use a Bayesian approach to obtain the estimates of the parameters. Furthermore, we illustrate our methodology by analyzing data from Staccato study, a randomized trial to compare two treatments, for HIV infection of Thai patients on highly active antiretroviral therapy, in which quality of life was assessed with HIV Medical Outcome Study (MOS-HIV) questionnaires.

E520: **Evaluation of a 2x2 study design under non-proportional hazards addressing long-term outcomes**

Presenter: **Ulrike Poetschger**, CCRI, Austria

Co-authors: Harald Heinzl, Milen Minkov, Helmut Gadner, Ruth Ladenstein, Martina Mittlboeck

The aim is to describe the statistical design of the LCH-IV study, an international randomised trial in children with the Langerhans Cell Histiocytosis (LCH). The intended design is 2x2 factorial, aiming to investigate 1) the role of the prolongation of therapy from 12 to 24 months and 2) the role of the addition of mercaptopurine to standard treatment on the long-term reactivation rate. When comparing identical treatments of different duration, identical hazards are anticipated for the initial period. Furthermore, the prolongation of treatment duration may solely delay but not stop

the occurrence of reactivations. Both result in time-dependent hazard ratios. Therefore, standard methods of survival analysis are inappropriate. Instead, mixture cure-models are used to determine treatment differences. Without relying on the proportional hazard assumption, cure-models explicitly allow estimation and modelling of the proportion of individuals who will not experience an event. Since early differences in event rates may not translate to a different long-term outcome, early stopping is implemented in favour of the null-hypothesis only. Monte Carlo methods are used to evaluate the statistical power and interim-analysis. Appropriateness of Cox-regression and Cure models are compared and discussed.

E565: **Statistical implications of survival endpoint options in randomised clinical trials**

Presenter: **Martina Mittlboeck**, Medical University of Vienna, Austria

Co-authors: Harald Heinzl

Survival time is often chosen as primary endpoint in many randomized clinical trials. However, the definition of the survival time is based on interesting events and has to be discussed in advance. In cancer studies, overall-survival, event-free survival or recurrence-free survival are frequently chosen candidates. If there are several candidates under discussion for the primary outcome, then often the outcome with the shortest expected study duration will be chosen. Consequences of such an approach will be compared and discussed. Effects on the planning phase, analysis and interpretation of results are investigated. Pitfalls when comparing treatment effects between different studies are mentioned. Special attention will be given to the comparison of hazard ratios from studies with unequal proportions of non-disease related events. Different choices and definitions of endpoints are not interchangeable, and sample size calculations and interpretations of resulting hazard ratios should be approached with caution. Furthermore, hazard ratios among studies with unequal proportions of non-disease related events cannot be compared in a straightforward manner. Systematic reviews or meta-analyses of treatment comparisons become more complicated when either the primary endpoints differ across studies or when they are identical, but the hazards for the non-disease related events differ.

E788: **Evaluating seamless adaptive designs using simulation**

Presenter: **Susan Todd**, University of Reading, UK

Co-authors: Nick Parsons, Tim Friede, Nigel Stallard, Elsa Valdes-Marquez, Richard Nicholas, Jeremy Chataway

Recent advances in statistical methodology for clinical trials have led to the development of seamless adaptive designs, which allow modifications during the course of a trial whilst maintaining statistical and scientific integrity through the control of the family-wise type I error rate. Implementation of adaptive design methods usually requires detailed simulations to be evaluated examining a wide range of realistic clinical scenarios. As part of a project considering the implementation of an adaptive trial in Multiple Sclerosis, code was developed in the R statistical programming environment (http://www.r-project.org/) to implement adaptive design methodology incorporating the following three techniques; (i) Dunnett's test, (ii) the inverse Normal combination test and (iii) closure test principles. In this talk, we report our work in developing both methodology and the software tools and use an example application in progressive Multiple Sclerosis to demonstrate how a wide range of design options and assumptions can be explored using simulation.

ES54 Room Senate Gordon Room MODEL SELECTION AND REGULARIZATION/ESTIMATION STRATEGIES Chair: Ejaz Ahmed

E196: **Information ratio test for model misspecification on volatility structures in stochastic diffusion models**

Presenter: **Peter Song**, University of Michigan, USA

Co-authors: Shulin Zhang, Daimin Shi

Model misspecification may lead to serious consequences in data analysis, and thus model diagnostics become one of critical tasks in model building. In this paper, we propose a powerful testing procedure to detect possibly model misspecification on the variance or volatility structure in stochastic diffusion models for time series data. The test statistic is constructed via a contrast of sensitivity matrix and variability matrix in the context of martingale estimating equations, in which no parametric marginal distributions are assumed. We discuss large sample properties for the proposed test statistic. We illustrate our method using both simulation studies and real world data analysis. In simulation studies we compare our method with the classic test method for heteroskedasticity, and in many occasions, our method outperforms White's method in terms of controlling the type I error and power.

E531: **Censored linear regression when p>n and p<n**

Presenter: **Yi Li**, Harvard, USA

We introduce a class of Dantzig variable selectors for linear regression models for right-censored outcomes. We first establish the finite sample error bound for the estimator and show the proposed selector is nearly optimal in the ℓ_2 sense. To improve model selection performance, we further propose an adaptive Dantzig variable selector and discuss its large sample properties, namely, consistency in model selection and asymptotic normality of the estimator. The practical utility of the proposed adaptive Dantzig selectors is verified via extensive simulations. We apply the proposed methods to a myeloma clinical trial and identify important predictive genes for patients' survival.

E578: **Estimating the correlation coefficient from several elliptical distributions**

Presenter: **Dietrich von Rosen**, Swedish University of Agricultural Sciences, Sweden

Co-authors: Ejaz Ahmed

The problem in focus is how to estimate/test appropriately chosen functions of the correlation coefficient in a multivariate p-dimensional elliptical distribution with intra class dispersion structure under the assumptions that there exists $q > 2$ populations. In particular one has to take into account the estimation of the kurtosis, which among others appears in the asymptotic distribution of the correlation coefficient. For each of the q population there will be a baseline estimator which is constructed via moment estimators of the *variance* and *covariance* and then averaged according to the structure in the original covariance matrix. Thereafter a variance stabilizing transformation is performed. An asymptotic test for the homogeneity of the kurtosis parameters is suggested. The proposed test is used to construct some improved estimators. More specifically we suggest pretest and shrinkage estimation strategy for the problem at hand. The asymptotic properties of suggested strategy will be developed. Monte Carlo simulation experiments will be conducted and the performance of each procedure is evaluated in terms of simulated mean squared error. For illustration, data examples will be showcased.

E550: **Model averaging for varying-coefficient partially linear measurement error models**

Presenter: **Sherry Zhou**, City University of Hong Kong, Hong Kong

Co-authors: Haiying Wang, Alan Wan

A local mis-specification framework has been recently proposed for studying the limiting distributions and asymptotic risk properties of model selection and model average estimators in parametric models. Since the asymptotic distribution under such framework is not normal, they suggested a method to construct precise confidence intervals for the parameters of interest. However, in real data analysis, it is common that the variables cannot be observed precisely. If estimations are conducted without considering the presence of measurement error, the obtained estimators are often

biased. In this paper, we extend the previous analysis on model averaging estimation for the semi-parametric varying-coefficient partially linear models when the measurement error is taken into account. We derive the asymptotic distribution of the estimator for the interested parameters and construct confidence intervals that cover the true parameters with probability tending to one. It is found that the resultant intervals are asymptotically equivalent to those developed based on the full model. The finite sample performance of the model average estimators is also examined through Monte Carlo simulations.

E619: **To homogenize or not to homogenize: The case of generalized linear mixed models**

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

Co-authors: Ejaz Ahmed, Marwan Al-Momani

The problem of whether a given data supports heterogeneous or homogeneous models has a long history and perhaps its major manifestation is in the form of generalized linear mixed models. By heterogeneous models we mean models where diversity among possible subpopulations is accommodated by using variance components. Among other areas, this problem arises in economics, finance, and Biostatistics under various names such as panel, longitudinal or cluster correlated data. Homogeneity is a desired property while heterogeneity is often a fact of life. We propose and explore the shrinkage-type estimators for both regression coefficient parameters as well as the variance components in order to reconcile these two types of models and seek unity in diversity. The base estimators used in this study will include penelized likelihood estimators.

ES71 Room Senate Bedford Room STATISTICAL ALGORITHMS AND SOFTWARE I Chair: Cristian Gatu

E068: **Logistic mixed-effects models with ambiguous cluster membership**

Presenter: **Recai Yucel**, State University of New York at Albany, USA

A novel likelihood-based approach to estimating logistic mixed-effects regression model is presented in settings in which cluster indicators can only be determined with a given uncertainty. In haplotype-based investigations of unrelated individuals, cluster assignments are unobservable since the alignment of alleles within chromosomal copies is not generally observed, but rather can be determined under a probabilistic approach. In social studies, where individuals are subject to unobservable effects of multiple social context, these probabilities can indicate the multiple-membership weights. I develop a maximum likelihood (ML) estimation utilizing multidimensional adaptive Gauss-Hermite quadrature for integration of the random-effects. Our models and accompanying computational algorithms are illustrated by a case-control study of colorectal adenoma.

E704: **PLS-PM segmentation. The PATHMOX approach**

Presenter: **Tomas Aluja-Banet**, Universitat Politecnica de Catalunya - Barcelona Tech, Spain

Co-authors: Gaston Sanchez

One common assumption when estimating PLS-PM models is to suppose homogeneity over the entire set of individuals, all individuals are treated alike as if they belonged to the same population. This assumption is unrealistic in many cases. Consider for example survey research studies, where potential sources of heterogeneity can be expected among different subgroups defined by gender, groups of age, ethnicity, education level, etc. In those cases having a unique model when heterogeneity is present can be misleading and the analyst runs the risk of drawing erroneous or poor conclusions. The PATHMOX approach overcomes this problem by building a binary tree of PLS-PM models, revealing groups of individuals following different patterns of relationships. To that purpose auxiliary information must be available on individuals, regarding their socio-economic condition, psychographic profile, consumption habits, etc. like it is the case in most surveys. In this sense PATHMOX entails a data mining approach for identifying unexpected models of segments in the population. It uses a F statistic as split criterion and maximum deep and minimum size as rules to limit the algorithm. We will present the R function pathmox, which implements the aforementioned approach and its application to detect different groups of employee motivation in a Spanish banking institution.

E813: **Combining classifiers in credit scoring**

Presenter: **Farid Beninel**, CREST- ENSAI, France

Co-authors: Marian Hristach

The main problem of credit-scoring is the prediction of borrowers' behavior. This problem could be solved using supervised classification methods, where the behavior is given by the target variable and creditworthy and not-creditworthy are the pre-specified classes. Until recently the methodology involved (with a suitable encoding of data) finding the best rule or classifier among usual classifiers: LDC, QDC, WDC, DDR, kNN, Parzen classification. This approach is unsatisfactory as it assumes that the population of candidates for a loan is homogeneous, which is rarely the case. Indeed implementation of an allocation rule devoted to the prediction of the behavior classes requires stability in the studied population and in the distribution of the available covariates. For discriminating on a mixture of sub-populations, we distinguish the two following situations. In the first case, the partition into homogeneous sub-populations is known and a set of classifiers is estimated. We construct the classifier for each subpopulation using the corresponding learning sub-sample. In the second case, the population partition is unknown and we look for a set of diverse classifiers using the global training sample to estimate each of them. The implemented methods are compared by simulation using Deutsche Bank data.

E832: **Applying sequential parameter optimization to machine learning models and some results regarding qualitative factors**

Presenter: **Bernd Bischl**, TU Dortmund, Germany

Co-authors: Tobias Wagner, Simon Wessing

Using kriging as a meta-model in sequential optimization methods has become one of the most popular methods for nonlinear black-box functions in computer experiments; especially, when function evaluations are expensive and therefore only a low number of them are feasible. Still, the efficient modeling of problems containing quantitative as well as qualitative factors is not completely solved. As we plan tune algorithms like machine learning systems and evolution strategies, we are especially interested in simpler methods for these problems, which can easily be integrated in more complex software environments. Therefore we will take a largely practical approach in our contribution and: (a) demonstrate the integration of sequential parameter optimization for noisy target functions based on the DiceOptim package into the machine learning package mlr. (b) discuss the theoretical and practical approach of parallelizing multiple evaluations of the target function (e.g. maximum expected improvement criterion) on a number of available cores. (c) review the proposed approaches to include qualitative factors into a kriging model, compare some simpler methods and discuss possible improvements. (d) present results from optimizing chains of machine learning operations (preprocessing, filtering, models with hyperparameters).

E806: **An exact algorithm for weighted-mean trimmed regions**

Presenter: **Karl Mosler**, University of Cologne, Germany

Co-authors: Pavel Bazovkin

Trimmed regions are a powerful tool of multivariate data analysis. They describe a probability distribution in Euclidean d-space regarding location,

dispersion and shape, and they order multivariate data with respect to their centrality. Dyckerhoff and Mosler have recently introduced the class of weighted-mean trimmed regions that possesses attractive properties regarding continuity, subadditivity, and monotonicity. We present an exact algorithm to compute the weighted-mean trimmed regions of a given data cloud in arbitrary dimension d. These trimmed regions are convex polytopes in R^d. To calculate them, the algorithm builds on methods from computational geometry. A characterization of a region's facets is used, and information about the adjacency of the facets is extracted from the data. A key problem consists in ordering the facets. It is solved by the introduction of a tree-based order. The algorithm has been programmed in C++ and is available as an R package.

Friday 10.12.2010 13:40 - 15:20 Parallel Session C – CFE

CI99 Room IoE Logan Hall INVITED SESSION: FACTOR MODELS Chair: Tommaso Proietti

C370: **The unrestricted generalized dynamic factor model**

Presenter: **Marco Lippi**, Universita di Roma La Sapienza, Italy

Co-authors: Mario Forni, Marc Hallin, Paolo Zaffaroni

With a few exceptions generalized dynamic factor models have been studied under the assumption that there exists a static representation with a finite number of static factors. We argue that this is quite a severe restriction, ruling out very important cases, and go back to the general model. Assuming that the spectral density of the common components is rational, we derive a one-sided, actually a finite autoregressive, representation for the common components, thus solving a problem that had remained unsolved within the frequency-domain approach in previous literature. The paper builds upon results by the authors themselves and one recent work on stochastic processes whose spectral density is rational and singular. We produce population results and consistency results for the corresponding estimator. An empirical analysis, based on USA data, shows that relaxing the common assumption of a static representation is a considerable advantage.

C460: **Generalized factor models-a structure theory**

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

Co-authors: Brian D O Anderson, Alexander Filler, Weitian Chen

We consider generalized linear dynamic factor models. These models have been developed recently and they are used for forecasting and analysis of high dimensional time series in order to overcome the curse of dimensionality plaguing traditional multivariate time series analysis. We consider a stationary framework; the observations are represented as the sum of two uncorrelated component processes: The so called latent process, which is obtained from a dynamic linear transformation of a low dimensional factor process and which shows strong dependence of its components, and the noise process, which shows weak dependence of the components. The latent process is assumed to have a singular rational spectral density. For the analysis, the cross sectional dimension n, i.e. the number of single time series is going to infinity; the decomposition of the observations into these two components is unique only for n tending to infinity. We present a structure theory giving a state space or ARMA realization for the latent process, commencing from the second moments of the observations. The emphasis is 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 necessarily have a unique solution in the singular case, and the resulting complexities are examined with a view to find a stable and coprime system.

C889: **Dynamic sparse factor model**

Presenter: **Christian Schumacher**, Deutsche Bundesbank, Germany

Co-authors: Sylvia Kaufmann

Factor models based on large datasets have received increasing attention in the recent macroeconomic literature, in particular for forecasting and analyzing the transmission of shocks. An open issue in the literature is the composition of the data used for estimating factors. For example, in many empirical papers it can be found that groups of data depend differently on the factors, and that the inclusion of uninformative data can distort factor estimates. In the present paper, we make an attempt to estimate a factor model that accounts for this heterogeneity in the data. The factors are estimated within a state-space framework with Bayesian techniques. To achieve a sparse, parsimonious parametrization, the factor loading matrix is estimated under a sparse prior, which assumes that either many zeros may be present in each column of the matrix, or many rows may contain zeros. The paper also tackles identification in a novel way, related to the recent Bayesian cointegration literature. We investigate the properties of the sparse factor model by simulations, where the DGP allows for different degrees of sparsity. Compared to a benchmark normal prior for the loadings, the sparse prior detects sparsity in the data well. We also provide empirical applications to large datasets from the literature. The results indicate that sparsity matters a lot in the data.

CS14 Room MAL 151 FINANCIAL VOLATILITY ESTIMATION AND FORECASTING I Chair: Francesco Audrino

C197: **Forecasting covariance matrices: a mixed frequency approach**

Presenter: **Roxana Halbleib**, University of Konstanz, Germany

Co-authors: Valeri Voev

A new methodology of accurately forecasting high dimensional covariance matrices of financial returns is proposed. It consists of mixing forecasts computed on different information sets within a parsimonious framework, which automatically assures the positivity and symmetry of the matrix forecasts. More precisely it consists of efficiently combining volatility forecasts stemming from the dynamic modelling of daily volatilities consistently estimated on high-frequency data (realized volatilities) with correlation forecasts stemming from parsimonious specifications estimated on daily data. This new approach allows for flexible dependence patterns for volatilities and correlations, which can be easily applied to covariance matrices of large dimensions without further parameter restrictions. The fact that the new approach builds on separate volatility and correlation forecasts considerably reduces the estimation and measurement error implied by the joint estimation and modelling of daily variance and covariance dynamics. Our empirical results show that the new mixing approach provides the best forecasts (smallest MSE) when compared to other multivariate volatility specifications computed on single sources of information.

C241: **Improving volatility forecasts by combining information**

Presenter: **Marcelo Medeiros**, PUC-Rio, Brazil

Co-authors: Thiago Ferreira

Evidence on whether (common) market information can bring gains in the daily forecasts of firm-specific realized volatility are provided. To achieve this goal we construct a benchmark model where only asset-specific information is included, and, then, we add regressors with a wider market information, typically common factors from many assets' volatility and returns. The benchmark specification is a Heterogeneous Autoregressive (HAR) model with one, five, and twenty two lags. Cumulated returns over different horizons are also considered in some specifications. To test whether market information is relevant or not for volatility forecasting, we construct factors from realized volatility and returns of 23 firms. Our results strongly suggest that adding factors to the basic HAR model can bring gains in volatility forecasting.

C310: **Option trading strategies based on semi-parametric implied volatility surface prediction**

Presenter: **Dominik Colangelo**, University of St. Gallen, Switzerland

Co-authors: Francesco Audrino

We construct a set of trading strategies using predicted option returns for a forecasting period of ten trading days and form profitable hold-to-expiration, equally weighted, zero-cost portfolios with one month at-the-money options. A statistical machine learning procedure based on regression trees accurately predicts future implied volatility surfaces. These forecasts assist in obtaining reliable option returns used as trading signals in our strategies. We test the performance of the proposed strategies on options on the S&P100 and on its constituents between 2002 and 2006. Positive annualized returns of up to more than 50% are achieved.

C566: **Robust volatility forecasts in the presence of structural breaks**
Presenter: **Constantinos Kourouyiannis**, University of Cyprus, Cyprus
Co-authors: Elena Andreou, Eric Ghysels

Financial time series often undergo periods of structural change that yield biased estimates or forecasts of volatility and thereby risk management measures. We show that in the context of GARCH diffusion or Two Factor Affine models ignoring structural breaks leads to biased GARCH-type volatility estimates. Realized Volatility, Realized Bipower Variation and Rolling Volatility provide unbiased estimates even in the presence of structural breaks. However, obtaining unbiased and efficient ex-post forecasts of volatility when there are structural changes in volatility, is much more difficult. Hence, we propose forecast combinations that take into account possible structural breaks and provide robust predictions of volatility. We show how forecast combinations based on alternative loss functions can be useful for risk management and particularly in forecasting Value at Risk and Expected Shortfall. In addition, we compare the performance of forecast combinations with individual models, especially those based on non-parametric methods. An empirical application illustrates our methods using both stock market and foreign exchange high frequency series.

CS19 Room MAL B29 MULTIFREQUENCY MODELLING Chair: Laurent Calvet

C286: **Day-varying weights in mixtures of stochastic volatility diffusion processes**
Presenter: **Christos Ntantamis**, CREATES, Aarhus University, Denmark

Current econometrics literature assumes the intraday asset prices to be driven by both a continuous process, and a jump process. Nevertheless, the effect of the jump process may not be the same across all days, e.g. it may be more important during days that changes in macroeconomic conditions occur. In order to investigate this hypothesis, this paper introduces a model for the logarithmic price of a speculative asset that is a mixture of two underlying stochastic volatility diffusion processes, one including jumps the other without, that share the same stochastic volatility process assumed to follow a square root diffusion process. The mixing probability is considered to be day-varying; it is given by a logistic model estimated on a set of explanatory variables. The estimation of the model parameters is performed by considering a 'mixed-frequency' dataset in a Bayesian framework: high frequency (intraday) data are used to determine the parameters of the diffusion processes, whereas low frequency (daily) data are used to determine the mixing probabilities. The model is estimated for the S&P500 index, and some US equity data, using the short-term interest rate, the interest rate spread and the dates of macroeconomic announcements as explanatory variables.

C364: **The oil price-dollar link : A wavelet based approach**
Presenter: **Francois Benhmad**, Universite Montpellier 1, France
Co-authors: Anne Peguin-Feissolle

In this paper, we use a wavelet filtering based approach to study the linear and nonlinear Granger causality between the oil price and the U.S. Dollar exchange rate. We thus use different non-causality tests. Moreover, the much-studied relationship between these two fundamental macroeconomic variables is explored with the help of the wavelet multi-resolution filtering technique. Instead of an analysis at the original series level, as is usually done, we first decompose the variables using a wavelet decomposition technique at various scales of resolution and obtain a relationship among components of the decomposed series matched to its scale. A major finding of this paper is that the linear and nonlinear causal relationship between the oil price and the U.S. Dollar exchange rate is not constant over time: it varies over time, depending on the wavelet time scale. There is a strong bidirectional causal relationship between the real oil and the real dollar for large horizons, i.e. corresponding to fundamentalist traders, especially fund managers and institutional investors. But, for the first frequency band which corresponds to a class of traders whom investment horizon is about 3-months and who correspond principally to speculative trading (noise trading), the causality runs only from the real oil prices to real dollar effective exchange rate.

C765: **Monetary policy effectiveness in times of crisis: Evidence from the Euro area money market**
Presenter: **Paola Donati**, European Central Bank, Germany

The effectiveness of the measures adopted by the ECB during the recent financial crisis in reducing the risk premia, and ultimately the level, of the Euribor with three-month maturity is assessed. This money market rate, which is a benchmark for pricing euro-denominated securities, sizeably departed from the monetary policy rate during the crisis hindering the transmission mechanism of monetary policy. The empirical evaluation of whether the extraordinary programmes pursued by central banks have worked is notoriously difficult because, over the short period of interest, data displayed record high volatility and several factors interacted with central banks' interventions. To account for various spillover effects gauging, at the same time, both the overall effect and the various degrees of persistency of the ECB' s corrective measures on the Euribor, we use a multivariate frequency decomposition approach cast in the unobserved component framework. We explicitly model, in the time domain, each frequency component of each variable of interest with a time-varying state space model. We perform, in real time, the required spectral decompositions while estimating the resulting multiple-input multiple-out (MIMO) system using a dynamic filter which exploits the time and frequency domain interpretation of the eigenvalues of state transition matrices of dynamic systems.

C465: **Equity skew and the Markov-switching multifractal: estimation and option pricing**
Presenter: **Marcus Fearnley**, HEC Paris, France
Co-authors: Laurent Calvet, Adlai Fisher, Markus Leippold

We propose a parsimonious multifrequency model which captures the leverage effect, volatility dependent jumps in returns, and the hyperbolic decay of the variance autocorrelation function. Our specification extends the conditionally Gaussian regime-switching framework of the MSM process to conditionally bivariate Wiener processes. We develop a smooth particle filter-based approximation of the likelihood function which allows us to estimate the model on S&P 500 index returns, and we introduce a method for efficiently calculating the characteristic function of the return which allows option prices to be computed quickly using FFT methods. Likelihood-based comparisons and VaR forecasts indicate that the new specification significantly outperforms other leading models, including affine and non-affine jump diffusion and GARCH type models.

CS30 Room MAL B34 COPULAS IN FINANCIAL ECONOMETRICS: RECENT DEVELOPMENTS Chair: Dick van Dijk

C205: **Tails of correlation mixtures of elliptical copulas**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Johan Segers

Correlation mixtures of elliptical copulas arise when the correlation parameter is driven itself by a latent random process. For such copulas, both penultimate and asymptotic tail dependence are much larger than for ordinary elliptical copulas with the same unconditional correlation. Furthermore, for Gaussian and Student t-copulas, tail dependence at sub-asymptotic levels is generally larger than in the limit, which can have serious consequences for estimation and evaluation of extreme risk. Finally, although correlation mixtures of Gaussian copulas inherit the property of asymptotic independence, at the same time they fall in the newly defined category of *near asymptotic dependence*. The consequences of these findings for modeling are assessed by means of a simulation study and a case study involving financial time series.

C477: **Comparing the accuracy of copula-based multivariate density forecasts in selected regions of support**
Presenter: **Oleg Sokolinskiy**, Tinbergen Institute, Erasmus University Rotterdam, Netherlands
Co-authors: Cees Diks, Valentyn Panchenko, Dick van Dijk

We develop a testing framework for evaluating and comparing the accuracy of copula-based multivariate density forecasts. The unique feature of our newly developed predictive accuracy test is that it allows us to focus on a specific part of the joint distribution. The test has a clear intuition as it makes use of likelihood based scoring rules, which can be interpreted in terms of the Kullback-Leibler Information Criterion (KLIC). We show that a similar test based on the conventional weighted likelihood scoring rule is biased towards fat-tailed distributions. Our proposed scoring rules based on conditional likelihood and censored likelihood do not suffer from this bias. This is confirmed by extensive Monte Carlo simulations, documenting the size and power properties of tests using these two scoring rules. In an empirical application to daily exchange rate returns, the Student's t copula is found to deliver superior forecasts, especially for regions in the lower and upper tails of the distribution. This indicates the importance of allowing for lower and upper tail dependence for accurate forecasting of common extreme appreciation and depreciation.

C487: **Measuring asymmetric tail dependences between the returns on equity style portfolios**
Presenter: **Nikos Thomaidis**, University of the Aegean, Greece
Co-authors: Efthimios Roumpis

Financial practitioners have long recognized persistent differences among various groups of stocks, based upon market capitalisation, industry classification, value or liquidity. There is though ample evidence that none of the predefined equity styles persistently outperforms the market, while their risk varies considerably with time. Recently, many authors have underlined the importance of tail behaviour and asymmetric dependence in the determination of equity premia. Hence, the use of advanced econometric models that capture these features might help to explain pricing anomalies and also bring out important risk components. This paper investigates asymmetric dependences between various equity styles using copula techniques. Our objective is to test whether these flexible parametric families lead to superior forecasts for the distribution of style returns, which can be further exploited in refining portfolio allocation decisions. We also investigate the importance of macroeconomic, business-cycle and other market-wide indicators in predicting shifts in the dependence structure.

C724: **Is the potential for international diversification disappearing?**
Presenter: **Kris Jacobs**, University of Houston, USA

Quantifying the evolution of security co-movements is critical for asset pricing and portfolio allocation, and so we investigate patterns and trends in correlations and tail dependence over time using weekly returns for developed markets (DMs) and emerging markets (EMs) during the period 1973-2009. We use the DECO, DCC, and BEKK correlation models, and develop a novel dynamic t-copula to allow for dynamic tail dependence. We show that it is possible to characterize co-movements for many countries simultaneously. Correlations have significantly trended upward for both DMs and EMs, but correlations between EMs are lower than between DMs. Further, our evidence clearly contradicts the decoupling hypothesis. The tail dependence has also increased for both EMs and DMs, but its level is still very low for EMs as compared to DMs. Thus, while our correlation analysis suggests that the diversification potential of EMs has reduced over time, the tail dependence analysis suggests that EMs offer diversification benefits during large market moves.

CS43 Room MAL B30 NONLINEAR MODELLING IN MACROECONOMICS Chair: Costas Milas

C141: **Macroeconomic uncertainty, inflation and growth: Regime-dependent effects in the G7**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Co-authors: Kyriakos Neanidis

We analyze the causal effects of real and nominal macroeconomic uncertainty on inflation and output growth and examine whether these effects vary with the level of inflation and location on the business cycle. Employing a bivariate Smooth Transition VAR GARCH-M model for the G7 countries during the period 1957-2009, we find strong nonlinearities in these effects. First, uncertainty regarding the output growth rate is related with a higher average growth rate mostly in the low-growth regime, supporting the theory of "creative destruction". Second, higher inflation uncertainty induces lower growth rates, increasingly so at the high-inflation regime. Third, real and nominal uncertainties have mixed effects on average inflation. Nevertheless, there is a trend in favour of the Cukierman-Meltzer hypothesis in the high-inflation regime. Our results can be viewed as offering an explanation for the often mixed and ambiguous findings in the literature.

C539: **Real exchange rates and consumption: A nonlinear perspective**
Presenter: **Efthymios Pavlidis**, Lancaster University, UK
Co-authors: Ivan Paya, David Peel

We adopt a nonlinear framework to model the deviations of the real exchange rate from its fundamental value implied by International Real Business Cycle models. By focusing on the post-Bretton Woods era, we find that in several cases there is a long run relationship between real exchange rates and consumption series in line with international risk sharing. Further, linearity tests indicate that the majority of the deviation processes exhibit significant smooth transition nonlinearity. ESTAR models appear parsimoniously to capture the nonlinear adjustment. These findings provide an explanation for the empirical regularities noted in the literature on the relation between the real exchange rate and consumption. Finally, GIRFs show that shock absorption is significantly faster than suggested in the PPP puzzle.

C295: **Global CO2 and temperature over the last five centuries: Change in persistence and cointegration**
Presenter: **Theodore Panagiotidis**, University of Macedonia, Greece
Co-authors: Theologos Dergiades

The order of integration of global CO2 and temperature is investigated for a period of five centuries. Rolling unit root (ADF and DF-GLS) and stationarity (KPSS) tests reveal that a change is taking place in the last century. Formal testing for a single change in persistence confirms that CO2 changes from I(1) to I(2) in 1934 and temperature from I(0) to I(1) in 1926. In the long-run we establish a long-run relationship between the two variables over the sample with a long-run elasticity that increased for close to 0 to 120 in the last window.

C207: **ECB policy and financial stability**
Presenter: **Costas Milas**, Keele University, UK
Co-authors: Ruthira Naraidoo

We explore how the ECB sets interest rates in the context of policy reaction functions. Using both real-time and revised information, we consider linear and nonlinear policy functions in inflation, output and a measure of financial conditions. We find that amongst Taylor rule models, linear and nonlinear models are empirically indistinguishable within sample and that model specifications with real-time data provide the best description of in-sample ECB interest rate setting behavior. The 2007-2009 financial crisis witnesses a shift from inflation targeting to output stabilisation and a shift, from an asymmetric policy response to financial conditions at high inflation rates, to a more symmetric response irrespectively of the state of inflation. Finally, without imposing an a priori choice of parametric functional form, semiparametric models forecast out-of-sample better than linear and nonlinear Taylor rule models.

CS53 Room MAL B20 FINANCIAL TIME SERIES Chair: Mike So

C162: **Bayesian panel data analysis for exploring the impact of recent financial crisis on the U.S stock market**
Presenter: **Tomohiro Ando**, Keio university, Japan
Co-authors: Ruey Tsay

The effects of recent subprime financial crisis on the U.S. stock market are analyzed. In order to investigate this problem, we develop a Bayesian panel data analysis to identify common factors that explain the movement of stock returns when the dimension is high. For high dimensional panel data, it is known that previously proposed approaches cannot estimate accurately the variance-covariance matrix. One advantage of the proposed method is that it considers parameter uncertainty in variance-covariance estimation and factor selection. We develop two new criteria for determining the number of factors in the data and establish the consistency of the selection criteria as both the number of observations and the cross-section dimension tend to infinity. We find that the U.S. stock market was subject to 8 common factors before the outbreak of the subprime crisis, but the number of factors reduced substantially after the outbreak. In particular, a small number of common factors govern the fluctuations of the stock market after the failure of Lehamn Brothers. That is, we obtain empirical evidence that the structure of U.S. stock market has changed drastically after the subprime crisis. We also show that the factor models selected by the proposed criteria work well in out-of-sample forecasting of asset returns.

C191: **Stochastic covariance models**
Presenter: **Manabu Asai**, Soka University, Japan
Co-authors: Mike So

We propose a new class of stochastic covariance models based on Wishart distribution. Three categories of dynamic correlation models are introduced depending on how we formulate the time-varying covariance matrix and whether it is a latent variable. We also develop a stochastic covariance filter for filtering and prediction of covariances. Extensions of basic models enable us to study the long memory properties in dynamic correlations, threshold correlation effects and the portfolio analysis. Suitable parameterization in our stochastic covariance models and the stochastic covariance filter facilitate efficient calculation of the likelihood function in high-dimensional problems, no matter the covariance matrix is observable or latent. We conduct Monte Carlo experiments to investigate finite sample properties of the maximum likelihood estimator. Two empirical examples are presented; one deals with the realized covariance using the high frequency data for exchange rates, while the other examines daily stock returns.

C276: **On a Student t-mixture autoregressive conditional heteroscedastic model**
Presenter: **Chun Shan Wong**, The Chinese University of Hong Kong, Hong Kong

A Student t-mixture autoregressive conditional heteroscedastic model is proposed for nonlinear time series modelling. The model combines the nice features of two recently proposed models, namely, the mixture autoregressive conditional heteroscedastic and Student t-mixture autoregressive models. The model is able to capture serial correlations, time-varies means and volatilities, and the shape of the conditional distributions can be time varied from short-tailed to long-tailed, or from unimodal to multimodal. The use of t-distributed errors in each component of the model allows conditional leptokurtic distributions. The statistical properties, parameter estimation, and approximation of standard error of estimates will be discussed. Simulation studies for verifying the small-sample estimation performance will be reported. The proposed modelling procedure is illustrated through a real example.

C172: **On a class of GARCH-M models**
Presenter: **Heung Wong**, The Hong Kong Polytechnic University, China
Co-authors: Xingfa Zhang

Motivated by the time-varying property of risk aversion and the psychology of investors in the stock market, we propose a class of functional coefficient GARCH-M models to study the effect of past returns on the investors' behavior of risk aversion. Unlike previous works in the literature, the new GARCH-M model considers the risk aversion as an unknown smooth function of the past returns. This helps us understand how the investors' attitude towards risk changes following the variation of the previous returns. Compared to the classic GARCH-M model, another novelty of the proposed model lies in adopting a re-specified GARCH-process which makes the model estimable. Due to its nice properties, this new specification has attracted the attention of a number of researchers. By incorporating the idea of the functional coefficient model and the quasi maximum likelihood method, approaches are given to estimate the model and develop a goodness of fit test. Simulations show that our method works well. To examine the proposed model, we apply it to both monthly and weekly excess return data that are calculated from the CRSP value weighted index, which includes the NYSE, the AMEX and NASDAQ. It is found that the proposed model can better fit the excess return series as compared to the purely parametric models and it also can provide insights on some potential parametric models. Moreover, some interesting results about relationship between risk aversion and the excess return are acquired.

CS56 Room MAL B18 NONPARAMETRIC VOLATILITY ESTIMATION Chair: Simona Sanfelici

C330: **Composite loss and realised kernels**
Presenter: **Asger Lunde**, Aarhus University, Denmark
Co-authors: Kevin Sheppard, Niel Shephard

The last seven years has seen dramatic improvements in the way econometricians think about time-varying volatility, first brought about by harnessing high frequency data and then by mitigating the effects of market microstructure effects. Extending this work to the multivariate case is challenging as this needs to additionally remove the effects of non-synchronous trading while simultaneously requiring that the covariance matrix estimator must be guaranteed to be positive semi-definite. The problem has partly been solved by the multivariate realised kernel estimator. The realised kernel estimator is pretty efficient in the bivariate case, but as the dimension of the problem increases it becomes less so due to its use of a synchronisation technique called refresh time. We propose an alternative which takes the bivariate method, apply it to all possible pairs and then put the results together to form a large dimensional covariance matrix. The bivariate submatrices are not necessarily compatible in finite samples and so we need to squash them together. We do this using an average of James and Stein loss functions over all possible pairs. We show this new estimator has better properties than that based on high dimensional realised kernels.

C862: **Nonparametric spot volatility estimation in microstructure noise models**
Presenter: **Johannes Schmidt-Hieber**, Universitat Gottingen, Germany
Co-authors: Marc Hoffmann, Axel Munk

Some results on adaptive wavelet estimation of the spot volatility are presented. We assume to have high-frequency data obtained from a continuous semimartingale and additionally corrupted by general microstructure noise. This model is widely used to describe financial data which result from trading on very short time intervals. The estimation technique is based on pre-averaging combined with thresholding. The proposed spot volatility estimator is rate-optimal and adapts to the correct smoothness over wide ranges of Besov spaces. We illustrate our method by real data applications.

C518: **Jumps at ultra high frequency**
Presenter: **Kim Christensen**, Deutsche Bank, UK
Co-authors: Mark Podolskij, Roel Oomen

We provide an in-depth investigation of the relative contribution of jumps to total asset return variation. We formulate a new non-parametric framework for measuring the importance of jumps and for performing empirical tests for jumps on financial high-frequency data, which are potentially contaminated by both microstructure noise and outliers. Using an extensive unique set of high-frequency data recorded at milli-second precision and a sample period, which includes the exceptionally turbulent recent period of financial turmoil, we find strong empirical evidence that the jump component is relatively insignificant as a fraction of total return variation. In line with these findings, our feasible jump test detects far less jumps than previously reported in the literature. Overall, our results suggest that jumps are not as common as generally perceived, which has a number of important implications for asset pricing and risk management. We demonstrate this with a delta hedging example of an option trader, who is short gamma.

C254: **Nonparametric stochastic volatility**
Presenter: **Bandi Federico**, University of Siena, Italy
Co-authors: Roberto Reno

We provide nonparametric methods for stochastic volatility modelling. Our methods allow for the joint evaluation of return and volatility dynamics with nonlinear drift and diffusion functions, nonlinear leverage effects, and jumps in returns and volatility with possibly state-dependent jump intensities. In the first stage, we identify spot volatility by virtue of jump-robust nonparametric estimates. Using observed prices and estimated spot volatilities, the second stage extracts the functions and parameters driving price and volatility dynamics from nonparametric estimates of the bivariate process' infinitesimal moments. We present a complete asymptotic theory under recurrence, thereby accommodating the persistence properties of volatility in finite samples.

C292: **Assessing the quality of volatility estimators via option pricing**
Presenter: **Simona Sanfelici**, University of Parma, Italy
Co-authors: Adamo Uboldi

The aim is to measure and assess the accuracy of different volatility estimators based on high frequency data in an option pricing context. For this we use a stochastic volatility model based on Auto-Regressive-Gamma (ARG) dynamics for the volatility. First, ARG processes are presented both under historical and risk-neutral measure, in an affine stochastic discount factor framework. The model parameters are estimated exploiting the informative content of historical high frequency data. Second, option pricing is performed via Monte Carlo techniques. This framework allows us to measure the quality of different volatility estimators in terms of mispricing with respect to real option data, leaving to the ARG volatility model the role of a tool.

CS58 Room MAL G16 CONTINUOUS TIME ASSET PRICING MODELS Chair: Leopold Soegner

C246: **Continuous time Markov switching models: Estimation and discretization**
Presenter: **Jorn Sass**, University of Kaiserslautern, Germany

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. That is, we deal with a linear, non-autoregressive Markov switching model. We are interested 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, we compare different approaches to estimate parameters. In particular, for medium sizes of observations, a Markov chain Monte Carlo method is suitable. We discuss problems for estimation, pricing and optimization arising if discrete time methods are applied to the continuous time model.

C504: **The expectations hypothesis and properties of bond risk premia**
Presenter: **Christian Wagner**, Vienna University of Economics and Business, Austria
Co-authors: Lucio Sarno, Paul Schneider

We study the validity of the expectations hypothesis (EH) of the term structure using a novel testing procedure based on affine term structure

models (ATSMs). We analyze whether an ATSM on which we impose the EH postulate of constant risk premia matches the data equally well as compared to an unrestricted model which allows for time-variation in risk premia. The unrestricted model produces unbiased predictions and has high explanatory power for realized risk premia. In the full sample analysis we find that the model significantly dominates the EH-restricted model as judged by statistical criteria and, for horizons up to two years, also in terms of economic value. The out-of-sample analysis, however, weakens the case against the EH. While the unrestricted dominates the pure EH predicting that risk premia follow a random-walk, it neither has higher predictive ability as compared to the EH-restricted model, nor does conditioning on time-varying risk and premium forecasts add economic value. These findings, which apply uniformly to the United States and five other countries, suggest that presuming the EH can still be viewed as useful for out-of-sample purposes.

C406: **Dynamic estimation of implied recovery rates from CDS spreads**

Presenter: **Marcin Jaskowski**, Vienna Graduate School of Finance, Austria

An important step in the credit analysis is the assessment of the value of collateral, that can be seized in case of default - the so called recovery rate value. Credit risk models should reflect the observation that the relevant value of collateral is generally not the average value of the asset over all possible states of nature. In most of the cases the relevant value of collateral for the lender is its secondary market value in bad states of nature. Although the negative relation between recovery rates and default probabilities is well documented, most of the pricing models do not allow for the correlation between the two. In this paper, we propose a relatively parsimonious reduced-form continuous time model that allows to estimate expected recovery rates and default probabilities from CDSs. Parameters of the model and latent factors driving recovery risk and default risk are estimated with a Bayesian MCMC algorithm. Finally, we show how to link those estimates with the structural model of debt where default is chosen endogenously.

C322: **Bayesian parameter estimation and identification in affine term structure models**

Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria

Problems arising with near unit root behavior and different market micro-structure noise assumptions for affine term structure models are investigated. Since interest rate data exhibit a high degree of serial correlation, near unit-root behavior becomes an important aspect. We show that with increasing serial correlation the Fisher information matrix approaches a singularity, such that the standard deviation of the level parameter is going to explode. We apply Markov Chain Monte Carlo simulation techniques in connection with regularized priors to simulate the joint posterior distribution of the model parameters.

CS77 Room MAL 351 RISK MANAGEMENT Chair: Ronald Hochreiter

C368: **Worst-case Omega ratio**

Presenter: **Michalis Kapsos**, Imperial College, UK

Co-authors: Daniel Kuhn, Berc Rustem

The Omega Ratio is a new performance measure that aims to surpass the known drawbacks of the Sharpe Ratio. Until recently, it was believed that optimizing the Omega Ratio is computationally intractable, and research was focused on heuristic optimization procedures. However, today it is known that the Omega Ratio optimization problem is equivalent to a linear program that can be solved exactly in polynomial time. Thus, more complex and realistic variants of the problem can now be investigated. In portfolio optimization it is commonly assumed that the probability distribution of the asset returns is precisely known. This distribution is usually estimated from historical data or prescribed by an expert. Unfortunately, solutions of the Omega Ratio optimization problem are severely biased and overoptimistic when the probability distribution is only partially known or corrupted by estimation errors. Therefore, we introduce in this talk the Worst-Case Omega Ratio, a robust variant of the conventional Omega Ratio which hedges against uncertainty in the probability distribution. We examine the Worst-Case Omega Ratio optimization problem under mixture distributions with uncertain mixing probabilities. These mixing probabilities are confined to an ellipsoidal or polyhedral uncertainty set. We show that under these assumptions the Worst-Case Omega Ratio optimization problems can still be solved efficiently as a tractable linear- or second-order cone program.

C684: **Mean-variance hedging in a Levy-Ito framework**

Presenter: **Florian Locker**, WU Vienna, Austria

In this work, we take a fresh look onto hedging methods arising from the well-known mean-variance criterion in cases where the classical delta hedging method becomes inaccurate due to the presence of large jumps generated by a Levy-type driving process. In particular, this paper includes some novel formulae for the optimal weights and errors in the multivariate pure jump case. We show that multivariate hedging weights in this framework can, under mild conditions, be reduced to a linear combination of univariate hedging strategies, which considerably facilitates numerical computations. As a first application of these equations, we address some interesting problems such as hedging derivatives of one underlying by several assets and in that respect review different choices of volatility matrices (Cholesky and spectral decompositions) and their implications. Hedging errors as tested on empirical markets stem from three, hard to disentangle error sources, namely misspecification of model, discretization and pure error due to incompleteness of the market. However, the latter two are amenable to numerical experimentation and thus, as a first step to a full-blown empirical study, this paper aims to determine their magnitude by simulation studies. We apply mean-variance hedging to several configurations of geometric pure jump processes, ranging from compound Poisson over Variance Gamma to the infinite variation Meixner processes, and separate the pure from the discretization hedging error by comparison with appropriate complete market models.

C675: **Corporate credit default models: a mixed logit approach**

Presenter: **Martin Kukuk**, University of Wuerzburg, Germany

Co-authors: Michael Roennberg

Analysing a cross-sectional sample of German corporate credit defaults, the popular logit model is extended to allow for varying stochastic parameters (mixed logit) and non-linearities of regressor variables. With respect to economic interpretability and goodness of probability forecasts according to discriminatory power and calibration, empirical results favor the extended specifications. The mixed logit model is particularly useful with respect to interpretability. However, probability forecasts based on the mixed logit model are not distinctively preferred to extended logit models allowing for non-linearities in variables. Further potential improvements with the help of the mixed logit approach for panel data are shown in a Monte Carlo study.

C819: **A proof system for pricing securities**

Presenter: **Erik Mathiesen**, , UK

We present a tool for approximating the solution of systems of differential equations. The tool is based on a proof system which describes the syntactical structure of the systems. Technically the proof system arises as the internal logic of a certain type of category, namely a traced monoidal

category. By defining a semantical interpretation of the proof system we give systems described by the proof system a semantical meaning. With one such interpretation we can approximate the solution of an underlying model and price a path-independent contingent claim. What we essentially achieve is an abstract description of the Taylor expansion of the solution. We can use this to price using Monte Carlo techniques in a way similar to standard schemes but with a variable degree of expansion. More importantly we can use it to approximate the characteristic function of the underlying process. This allows us to price using Fourier transform techniques on models for which the characteristic function is not readily available. The main contribution is to shed a different light on a well-known and often used technique, by providing a nice functional diagrammatical language, a "systems-as-proofs" idiom and a straightforward functional implementation making it useful for prototyping.

CP02 Room IoE Crush Hall POSTERS II Chair: Panayiotis Andreou

C796: **Evaluate the one-trading-day-ahead predictive ability of intra-day models for the CAC40 realized volatility**
Presenter: **Christos Floros**, University of Portsmouth, UK
Co-authors: Stavros Degiannakis

Autoregressive fractionally integrated moving average (ARFIMA) and heterogeneous autoregressive models are estimated and their ability to predict the one-trading-day-ahead CAC40 realized volatility is investigated. A method of model comparison, named standardized prediction error criterion (SPEC), which is based on a standardized version of the one-step-ahead forecast errors, is explored. At each point in time that we are interested in selecting one model among a set of candidate models, we could naturally pick the model with the minimum value of the loss function which is computed as the half sum of the squared standardized one-step-ahead forecast errors. In a Monte Carlo simulation, as well as, in an empirical application, the proposed method of model selection appears to have picked models that provide more accurate forecasts (minimize the loss function) than any model independently. According to the proposed loss function, a simple ARFIMA model provides the most accurate one-trading day-ahead forecasts of the logarithmic realized volatility of the CAC40 index.

C533: **Interdependence of CAC40, DAX30 and WIG20 - evidence from intraday data**
Presenter: **Barbara Bedowska-Sojka**, Poznan University of Economics, Poland

This paper seeks to find out if shocks observed on one of three financial markets, French, German and Polish, are observed at the same moment or are transmitted to the other markets within the very short time period. The bivariate GARCH models are estimated with variables standing for shocks in order to examine if there exists interdependence or contagion effect. The sample consists of 5 minute returns of CAC40, DAX30 and WIG20 indices from January 2009 to April 2010. We filter intraday series from periodicity with Weighted Standard Deviation filter and detect shocks with non-parametric Lee-Mykland jump detection test. Using high-frequency data and within the framework of dynamic conditional correlation models, we show that volatility increases at the same time in all series and rarely in the nearest intraday interval. Our results indicate co-movement in 3 pairs of indices that suggest common exogenous sources of volatility shocks.

C870: **Stocks splits and herding**
Presenter: **Maria Chiara Iannino**, Queen Mary University of London, UK

The relation between institutional herding and stock splits are examined. We use data on the buying and selling activity of US institutional investors, from 1994 to 2005. We compute the abnormal correlation of trades among institutional investors in companies that have announced a stock split, compared to a sample of non-splitting companies. The results show a significant level of convergence for both samples, slightly higher for splitting companies between 1998 and 2001. Moreover, we observe that herding has a stabilizing effect in the future returns especially for splitting companies. We also decompose the correlation of trades into the contributions of several types of herding. We observe the significant impact of informational-based herding in the splitting companies sample, that motivates the difference in herding between splitting and non-splitting stocks.

C253: **The forecasting performance of regime-switching models of speculative behaviour for exchange rates**
Presenter: **Ekaterini Panopoulou**, University of Piraeus, Greece
Co-authors: Theologos Pantelidis

The aim is to test whether a regime-switching model implied by the theory of periodically collapsing speculative bubbles is able to capture the observed dynamics of the exchange rates of the major currencies against the US dollar. To achieve our goal, we initially estimate alternative bubble measures for the exchange rates under scrutiny by means of typical models for the fundamental values of exchange rates. We assume that the bubble can be in either the survival state (a period when the size of the bubble increases) or the collapse state (a period when the size of the bubble declines towards zero). Moreover, the probability equation that governs the classification of regimes is allowed to be a function of the bubble. Our next step is to determine the optimal regime-switching model that captures the stochastic properties of our data which will be tested against the benchmark random walk model in an out of sample mode by employing typical tests for equal predictive ability. This model gives rise to the development of suitable trading strategies that will facilitate international investors to achieve profit maximization. Finally, we evaluate the reliability of the forecasts of our regime-switching model by means of Monte Carlo simulations.

C220: **On the impossibility of cointegration of international financial markets**
Presenter: **Thomas Dimpfl**, University of Tubingen, Germany

Cointegration has frequently been used in the financial econometrics literature to assess the degree of interdependence of financial markets. We show that if individual stock prices are generated by random walks with possibly contemporaneously correlated innovations, the resulting indices cannot be cointegrated as they are a combination of n random walks which itself is non-stationary by construction. This result holds if (as in factor models) an additional common global or local random walk is allowed for. There will, however, never be less than n random walk components, as otherwise company specific characteristics would be ruled out to affect the stock price permanently. To substantiate the theoretical propositions we simulate stock prices (allowing for heteroskedasticity, correlated innovations and common factors), construct indices and test whether these indices are cointegrated. We show that while heteroskedasticity alone is able to mislead cointegration tests, it is not sufficient to explain at the same time the empirically found high correlation between stock market indices. A common stochastic factor as well as correlated price innovations are necessary to reproduce the empirical characteristic features. We conclude that cointegration is not a suitable method to analyze stock market interdependence.

C711: **Sign RCA GARCH models in modeling Polish financial time series.**
Presenter: **Joanna Gorka**, Nicolaus Copernicus University, Poland

In the econometric literature, empirical models from family of Sign RCA models occur very seldom. One can find some empirical examples of these models but they concern mainly random coefficient autoregressive models (RCA) and there is a lack of empirical examples of other models of this family because the modifications of the RCA model were proposed relatively not long ago. Therefore, it is worth to raise the question whether

financial time series can be described by this family of models and whether they can be useful. In this paper, empirical models from family of sign RCA models with regard to Polish financial time series are presented. Moreover, these models are applied to calculate the forecasts and risk measures. Finally, the accuracy measures are obtained.

C622: **Heterogeneity and strategic choices: The case of stock repurchases**

Presenter: **Bernardo da Veiga**, Curtin University, Australia

Co-authors: Anup Nandialath

Strategic decisions are fundamentally tough choices. Although theory suggests that managers are likely to display bounded rationality, empirics assume rationality in choice behavior. Recognizing this inherent disconnect, we try to account for behavioral biases using a theoretically consistent choice model. We argue that the traditional approach to modeling strategic choice using discrete choice models and making inference on the conditional mean effects does not capture behavioral biases. By focusing on the conditional variance and explicitly modeling the conditional variance (in the discrete choice framework) we derive valuable information on individual level variation in decision-making. We demonstrate the effect of ignoring the role of variance in choice modeling in the context of firms' decisions to conduct open market repurchases. We show that when taking into account the heterogeneity in choices, a manager's choices of conducting open market repurchases displays considerable heterogeneity and that not accounting for such heterogeneity might lead to inappropriate conclusions regarding the mean effects.

C140: **World steam coal model mixing fundamental and statistic modeling**

Presenter: **Sebastien Lepaul**, EDF, France

We present a World Coal Model (WCM) computing the marginal costs and the trade steam coal flows from 2000 to 2035. WCM is based on a bottom-up approach (coal industry costs). The production, the transport and the investments are organized to response to the demand at the least cost in a perfectly competitive market. The originality of the paper is the adaptive rational price expectations framework. WCM is made of several modules modeling the market, the investors, the construction capacity lead times and the forecast of future price. In a loop, every year, a first module calculates a marginal cost (Price) by minimizing a cost function under capacity constraints, demand and production equilibrium. The marginal cost is then added to an historical chronic of price and the price is forecast for the next five years. With the knowledge of the future price a module mimicking the investor (by maximizing a profit) computes the new mines needed to face the future demand. The investments are delayed in a planning module taking into account the construction lead times before being sent to the market module. The aim of the paper is to focus on the price forecast through optimization or statistic solutions.

C767: **Survival analysis approach in Basel2 credit risk management: modelling danger rates in loss given default parameter**

Presenter: **Stefano Bonini**, Deloitte Consulting, Italy

Co-authors: Giuliana Caivano

In the last three years the majority of European banking groups has chosen to adopt the advance status under Basel II. This has required banks to develop statistical models for estimating Probability of Default, Loss Given Default and Exposure at Default, within a horizon time of 1 year. Such models make no attempt at describing the exact timing of default. In particular, beside an extensive academic and practitioner's literature on PD, LGD studies are in a less advance status. One of the main reasons could be due to the difficulties in modeling and forecasting the danger rates. The aim of this paper is to show the results of the first application on an Italian Bank Retail portfolio of survival analysis technique for estimating LGD, by modeling the danger rates. Two issues arise from the forecasting of danger rates: dealing positions that change, or not, their status towards charge off and obtaining a certain level of accuracy across time, thus resulting more difficult than in simpler classification methods. This paper analyzes the use of a parametric survival model, where time is assumed to follow some distributions whose PDF can be expressed in terms of unknown parameters: hazard and shape.

Friday 10.12.2010 13:40 - 15:20 Parallel Session C – ERCIM

EI81 Room Senate Beveridge Hall INVITED SESSION: NONPARAMETRIC STATISTICS Chair: Igor Pruenster

E127: **Rank-based tests for unit roots**

Presenter: **Marc Hallin**, Universite libre de Bruxelles, Belgium

Co-authors: Ramon van den Akker, Bas Werker

We propose a class of distribution-free rank-based tests for the null hypothesis of a unit root. This class is indexed by the choice of a *reference* density g, which need not coincide with the unknown actual innovation density f. The validity of these tests, in terms of exact finite sample size, is guaranteed, irrespective of the actual underlying density, by distribution-freeness. Those tests are locally and asymptotically optimal under a particular asymptotic scheme, for which we provide a complete analysis of asymptotic relative efficiencies. Rather than asymptotic optimality, however, we emphasize finite-sample performances, and show that, our rank-based tests perform significantly better than the traditional Dickey-Fuller tests.

E452: **Local dependence and tests of independence**

Presenter: **Dag Tjostheim**, University of Bergen, Norway

Co-authors: Geir Berentsen, Karl Ove Hufthammer

A bivariate density function f can be approximated by a family of bivariate Gaussian distributions. At each point (x,y), $f(x,y)$ can be approximated by a Gaussian bivariate distribution having mean $\mu(x,y)$ and covariance matrix $\Lambda(x,y)$. The corresponding correlation $\rho(x,y)$ can be thought of a a local measure of dependence. It is shown how $\rho(x,y)$ can be estimated by local likelihood, and theoretical properties of the estimate are given. The theory is exemplified by simulation and financial return data. Emphasis will be on measuring asymmetry of local Gaussian correlation with increased local dependence in bear markets, and on testing of independence by testing of $\rho(x,y) = 0$.

E853: **Semiparametric estimation of locally stationary diffusion models**

Presenter: **Oliver Linton**, London School of Economics and Political Science, UK

Co-authors: Bonsoo Koo

A class of locally stationary diffusion processes is proposed. The model has a time varying but locally linear drift and a volatility coefficient that is allowed to vary over time and space. We propose estimators of all the unknown quantities based on long span data. Our estimation method makes use of the local stationarity. We establish asymptotic theory for the proposed estimators as the time span increases. We apply this method to real financial data to illustrate the validity of our model. Finally, we present a simulation study to provide the finite-sample performance of the proposed estimators.

ES29 Room Senate Bedford Room ADVANCES IN ROBUST DATA ANALYSIS Chair: L.A. Garcia-Escudero

E107: **A trimming approach to cluster analysis**

Presenter: **Heinrich Fritz**, Vienna University of Technology, Austria

Co-authors: Luis Angel Garcia-Escudero, Agustin Mayo-Iscar

The introduced algorithm "tclust" expands the principle idea of the well known cluster algorithm "tkmeans": On the one hand the approach is robustified by applying certain trimming criteria to outlying observations, on the other hand the restrictions on the clusters' scatter structures are defined more flexibly and thus more heterogeneous datasets can be computed. In fact "tkmeans" is able to handle spherical clusters of same size and scatter. However, "tclust" handles clusters of different sizes and provides three different basic types of scatter constraints: Two restrictions are quite similar as either the relative range of the cluster's covariance matrix determinants, or the relative range of the cluster's covariance matrix eigenvalues is limited. The third restriction assigns the same scatter structure to all clusters which itself is not restricted anymore. As with certain parameter settings our method is able to cover further algorithms which appear in literature; they are discussed briefly. Apart from the method itself some exploratory tools are introduced for choosing the method's parameters as the number of clusters and the trimming level.

E112: **Asymptotic behavior of some robust estimators under long-range dependence.**

Presenter: **Helene Boistard**, Toulouse School of Economics (University of Toulouse 1), France

Co-authors: Celine Levy-Leduc, Eric Moulines, Valderio Reisen, Murad Taqqu

The asymptotic properties of some robust location and scale estimators under long-range dependence will be considered. These estimators can be expressed as empirical quantiles of a U-process. The main result of this talk deals with the asymptotics of U-processes under long-range dependence. It can be used to derive the asymptotic distribution and efficiency of the estimators. Some Monte-Carlo experiments will illustrate the large sample behavior of the estimators and the robustness properties when the data are contaminated with outliers.

E118: **Robust forward search clustering**

Presenter: **Domenico Perrotta**, European Commission, Italy

Co-authors: Francesca Torti

Deviations from model assumptions and the presence of outlying observations are common in many practical statistical applications. These problems may lead to unreasonable results even with cluster analysis methods. In this paper we extend the random starts Forward Search approach coupled with envelopes to the context of cluster analysis. The output is an automatic procedure for classification of the units into different clusters which is robust in the presence of atypical observations, does not require optimization tools which can lead to local minima and does not force all units to be classified.

E128: **Finding curves in noisy data**

Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain

Co-authors: Luis A. Garcia-Escudero, Carlos Matran, Agustin Mayo-Iscar

The problem of finding structures in data has been frequently tackled in the statistical literature. Regression models, cluster analysis, principal curves, etc. are examples of approaches to this problem. We intend to address the problem of finding curves in noisy data through a parametric approach. The statistical model is made of a parametric curve plus orthogonal normal errors and random noisy points. The curves belong to a parametric family obtained from a canonical curve through different families of transformations. We will focus on the case of conics, but the methodology extends to other parametric curves. To avoid the influence of noise or outliers, we use the impartial trimming methodology, already used successfully in other settings. The estimation of the parameters is done by solving a double minimization problem, looking for the best sub-

sample (non trimmed points) and the best curve in the parametric family, which minimize the sum of squared orthogonal residuals of non trimmed data. The minimization is solved through an EM-type algorithm. Finally, our methodology is tested on both simulated and real data sets coming from Computer Vision and Retinography. We also explore the extension of the methodology to the more general setting of searching for k curves.

ES32 Room MAL 538 FUZZY STATISTICS AND DECISION MAKING Chair: Teresa Leon

E083: **GATSH: a decision support system for portfolio**

Presenter: **Juan V. Carrillo**, University of Murcia, Spain

Co-authors: Jose Manuel Cadenas, M. Carmen Garrido, Vicente Liern

Markowitz stated in his work the evolution of the market can be modeled with basic statistics; the mean/covariance matrix. Using this model, an investor can choose a portfolio minimizing the risk and maximizing the expected return. As the investor preferences can be different from the optimum, he will prefer a decision support system which gives him more than one possibility. Nevertheless, this portfolio is restricted to just one possibility: the (expected) optimum. The market is a very uncertain environment, so a fuzzy model would help us to increase the degree of freedom to choose a different solution from min. risk/max. profit in order to improve the portfolio return, by means of relaxing the constraint about the risk. We propose a hybrid strategy that acts as a decision support system. The strategy "Genetic Algorithm + Tabu Search (GA+TS)" uses knowledge about the assets to find the best set of solutions for the portfolio. Using the higher return order of the assets, TS guides the GA mutation limiting the randomness factor of the operator to quickly reach the optimum. Results are promising, as tests show the technique is very fast and improves previous results of the literature.

E163: **Decision making for selection of socially responsible portfolios using fuzzy modeling**

Presenter: **Veronica Canal-Fernandez**, University of Oviedo, Spain

Co-authors: Amelia Bilbao-Terol, Mar Arenas-Parra, Mariano Jimenez-Lopez, M.Victoria Rodriguez-Uria

An approach is introduced for selecting portfolios with financial products by verifying the so-called *socially responsible* criteria. We have used an index, called Sri-Attractiveness, for evaluating the socially responsible performance of different kinds of mutual fund in accordance with a Socially Responsible Investing approach. This index summarizes the *social, environmental and ethical performance* of each mutual fund for a particular investor and is constructed via a process of aggregation and using Fuzzy Multiple Criteria Decision Making techniques. Indeed, the Socially Responsible Investing approach also considers conventional financial objectives such as final wealth, or net gains, or wealth relative to some benchmark. Here, the evaluation of the financial criteria has been handled by the expected value of final wealth and the Conditional Value-at-Risk. We use a mathematical programming technique that allows us to work with multiple criteria: Goal Programming models where the importance relations among the goals are established in an imprecise way. Our method is based on an approach for solving fuzzy goal programming with imprecise goal hierarchy. A case study on a set of Socially Responsible mutual funds domiciled in UK is presented to illustrate the proposed method.

E164: **Fuzzy measures of return and risk of a portfolio**

Presenter: **Enriqueta Vercher**, Universitat de Valencia, Spain

Co-authors: Jose D. Bermudez, Jose V. Segura

Modern portfolio selection theory is derived from the mean-variance probabilistic model due to Markowitz which considered trade-off between return and risk. From that some other measures have been considered in portfolio selection. We assume that the uncertainty on the future returns of an investment can be modeled by means of fuzzy logic and then we will approach the portfolio selection problem under fuzziness. Here we analyze different possible mean values and variances; other central models and values are defined and their mathematical properties are studied. Finally, the models are illustrated with numerical examples.

E102: **Human resource management based on soft-computing: evaluation of skills.**

Presenter: **Teresa Leon**, University of Valencia, Spain

Co-authors: Teresa Lamata, Vicente Liern

Decision-making management in the human resources department of a company is often based on the assessment of certain skills and their aggregation. Performing these tasks using real (crisp) numbers generates several problems; mainly it is not natural for the evaluators and does not allow taking into account the attitude of the candidate in different situations. Therefore, we propose the use of fuzzy logic in two different situations: selection and reassignment of personnel to jobs. The assessment of skills is made with intervals and the selection of candidates by using LR-fuzzy numbers. Such numbers arise from the skill aggregation, possibly considering different managers' opinions (for the selection problem) or from the historical information on the candidates (for reallocation). In general, decisions are made by measuring the distance of each candidate to an ideal one set by the company, by using for example, the Hamming distance, or by calculating a degree of "likeness" between each candidate and the ideal one. However, in conflict situations in which one single ideal profile does not exist, the selection can be made through LR-fuzzy number ranking.

ES34 Room Senate Bloomsbury Room APPLIED STATISTICS I Chair: Marc Lavielle

E395: **Simultaneous confidence bands in curve prediction applied to load curves**

Presenter: **Jean-Claude Fort**, Universite Paris Decartes, France

Co-authors: Jean-Marc Azais, Sophie Bercu, Agnes Lagnoux, Pierre Le

Starting with the problem of predicting the whole annual load curve of some Electricite de France (EDF) clients from easily available explanatory variables, we derived simultaneous confidence bands for these predictions. This may apply in numerous contexts. Thus we construct simultaneous confidence bands in a general setting by using theoretical and numerical results on the maximum of Gaussian sequences. Then we adapt the developed method in the context of the prediction of the electricity annual load curves.

E416: **On the asymptotic behaviour of bifurcating autoregressive processes**

Presenter: **Bernard Bercu**, University Bordeaux 1, France

Co-authors: Benoite De Saporta, Anne Gegout-Petit

Recent results on the asymptotic behaviour of the least squares estimators of the unknown parameters of general pth-order bifurcating autoregressive processes are considered. Under very weak assumptions on the driven noise of the process, we establish the almost sure convergence of our estimators together with the quadratic strong law and the asymptotic normality. The analysis relies on non standard asymptotic results for martingales.

E555: **Statistical tools for road traffic prediction**

Presenter: **Jean-Noel Kien**, Universite Paul Sabatier, France

Co-authors: Guillaume Allain, Philippe Goudal, Jean-Michel Loubes, Fabrice Gamboa

Some recent advances in road information processing will be addressed. We will explain how modern statistical techniques are used both to filter and predict vehicle speeds on road networks. For example, these techniques include classification, specialised statistical modelling and process filtering. This is a joint research work between academia and industrial labs.

E687: **The SAEM algorithm: a powerful stochastic algorithm for population pharmacology modeling**

Presenter: **Marc Lavielle**, INRIA, France

In order to evaluate a drug's impact, trial researchers must confront the need to model biological phenomena that evolve over time and present a high level of inter-individual variability. These highly variable systems are depicted by so called non-linear mixed effect models, often defined by systems of differential equations and for which it is particularly hard to estimate their numerous parameters. For 30 years now, pharmacologists wanting to interpret their clinical trials have been using software programs that employ model linearization. This solution, however, raises major practical problems. The SAEM algorithm is a stochastic approximation version of the EM algorithm for maximum likelihood estimation in the context of incomplete data models. This stochastic algorithm has demonstrated very good practical properties for many complex applications in the framework of non-linear mixed effect models: pharmacokinetics-pharmacodynamics, viral dynamics, glucose-insulin, epilepsy etc. SAEM combined with MCMC and Simulated Annealing is less sensitive to the initial guess and it is fast. SAEM was first implemented in the MONOLIX software (http://software.monolix.org). This stochastic algorithm is now available in NONMEM and was recently implemented in the statistics toolbox of MATLAB. Several extensions of the SAEM algorithm for complex models will be presented: mixed models defined with stochastic differential equations, mixed hidden Markov models.

ES42 Room MAL 541 STOCHASTIC PROCESSES Chair: Paula Rodriguez Bouzas

E256: **Polya-Aeppli of order k risk process**

Presenter: **Leda Minkova**, Sofia University, Bulgaria

Co-authors: Stefanka Chukova

The Polya - Aeppli process of order k is defined as a compound Poisson process with truncated geometric compounding distribution and some of its basic properties are investigated. Using simulation we provide a comparison between the sample paths of the Polya - Aeppli of order k process and the Poisson process. Also, we consider a risk model in which the claim counting process is Polya - Aeppli process of order k, and call it a Polya - Aeppli of order k risk model. For the Polya - Aeppli of order k risk model we derive the joint distribution of the time to ruin and the deficit at ruin as well as the ruin probability. We discuss in detail the particular case of exponentially distributed claims.

E136: **A smoothing algorithm for widely factorizable signals**

Presenter: **Rosa Maria Fernandez Alcala**, University of Jaen, Spain

Co-authors: Jesus Navarro-Moreno, Juan Carlos Ruiz-Molina

The smoothing estimation problem is analyzed for a class of improper complex-valued signals, called widely factorizable, characterized because the covariance of the augmented vector formed by the signal and its conjugate is a factorizable kernel. For this type of signal, widely lineal processing is the most suitable approach by considering the complete information of the augmented covariance matrix. Then, from only the knowledge of the second order properties of the augmented vectors involved, a recursive algorithm is provided for computing the widely linear smoother of this type of signals observed in presence of noise as well as its associated error. Observe that the proposed methodology is valid for both stationary and nonstationary signal.

E417: **Stochastic innovation diffusion models: closed form solutions and related simulations**

Presenter: **Christos Skiadas**, Technical University of Crete, Greece

Co-authors: Yiannis Dimotikalis, Charilaos Skiadas

The innovation diffusion models are applied in various fields. The stochastic versions of these models with an additive or a multiplicative error term are of particular importance in Economics, Marketing, Ecology or Sociology. This paper explores the main stochastic innovation diffusion models and gives closed form solutions for the related stochastic differential equations. Furthermore, the first and second moments are estimated and simulations are done. The related figures show the stochastic paths, the graphs of the mean values and the confidence intervals.

E758: **Y-linked bisexual branching processes with blind choice: parametric Bayesian inference using MCMC methods**

Presenter: **Rodrigo Martinez**, University of Extremadura, Spain

Co-authors: Miguel Gonzalez, Cristina Gutierrez

Y-linked bisexual branching processes with blind choice are two-sex stochastic models useful to analyze the evolution of the number of carriers of Y-linked genes which are not expressed in males or if they are, they do not play any role in the mating of these males (so that females choose blindly their mates with respect to these genes). The behaviour of these kinds of genes is strongly related to the reproduction law of each of their genotypes. In practice, these offspring distributions are usually unknown and their estimation is necessary. In this paper, we deal with the problem of estimating the offspring distribution of each genotype of such a Y-linked gene when the only data we assume available are the total number of males of each genotype and the total number of females up to some generation. We set out the problem in a parametric Bayesian framework and provide an algorithm, based on the Gibbs sampler, to approximate the posterior distribution of the reproduction law of each genotype. From these estimated posterior distributions, we also approximate the posterior predictive distributions of future population sizes. By way of simulated examples, developed with the statistical software R, we illustrate the accuracy of the proposed algorithm.

ES48 Room Senate Gordon Room RECENT ADVANCES IN MULTI-STATE MODELS Chair: Jacobo de Una-Alvzrez

E072: **A simple simulation technique for nonparametric analysis of time-to-event data arising in multistate models**

Presenter: **Jan Beyersmann**, University Medical Centre Freiburg, Germany

Co-authors: Susanna di Termini, Martin Schumacher

Multistate models consider complex time-to-event data as transitions between multiple states. Nonparametric analysis of the transition probabilities based on asymptotic theory is hampered by the complicated structure of the limiting covariance process. Regression analysis of survival hazards and nonparametric estimation of cumulative event probabilities in competing risks have approached such problems via a simple resampling technique. The resampling uses conditional multiplier approximation, which keeps the data fixed and introduces computer generated normal variates in

martingale representations of the statistics of interest. We propose to extend this resampling technique to multistate situations where the aim is to estimate the matrix of transition probabilities. We consider multistate models which are realized as a series of nested competing risks experiments. Observation of the data may be subject to independent left-truncation and right-censoring. Our proposal relies on approximating the asymptotic distribution corresponding to the multiple transitions hazards first. Subsequently, the asymptotic distribution corresponding to the matrix of transition probabilities is approximated using product integration and the functional delta method. From that, standard errors and simultaneous confidence bands can easily be obtained. The practical utility of the method is illustrated with data from drug exposed pregnancies and with data on hospital complications.

E077: **Multi-state models and length biased sampling**

Presenter: **Micha Mandel**, Hebrew University of Jerusalem, Israel

Non-parametric estimation of multi-state models where data are subject to length biased sampling is addressed. The problem occurs frequently in Biostatistics and other areas where data are obtained by sampling individuals who are available at a given place and time window (prevalence designs). In addition to bias, data are often subject to censoring and truncation, or in general to problems of missing data. When these distortions are not taken into account, one may obtain highly biased and inconsistent estimators. The motivating example is of a cross-sectional study of patients hospitalized in intensive care units, where the total hospitalization time is subject to length bias. Additional biases occur in other time-to-event variables, such as time to infection, in a way that depends on the joint distribution and the dependence structure involved. The problem is presented in the framework of multi-state processes, non-parametric estimators for different functionals of the underline process are developed, and their asymptotic properties are derived.

E090: **Presmoothing the transition probabilities in the illness-death model**

Presenter: **Ana Paula Amorim**, University of Minho, Portugal

Co-authors: Jacobo de Una-Alvarez, Luis Meira-Machado

Multi-state models are the most common models for describing longitudinal survival data. A multi-state model is a model for a stochastic process, which is characterized by a set of states and the possible transitions among them. The states represent different situations of the individual (healthy, diseased, etc) along a follow-up. Let $X(t)$ represent the state occupied by the process at time $t \geq 0$. For two states i, j and $s < t$, let introduce the transition probability $p_{ij}(s,t) = P(X(t) = j|X(s) = i)$. There has been much interest in the estimation of $p_{ij}(s,t)$ since it allows for long-term predictions of the process. We propose here an estimator in the case of a non-Markov illness-death model based on presmoothing ideas, which allows for a variance reduction in the presence of censoring. In this work we propose presmoothed estimators of the transition probabilities $p_{ij}(s,t)$ in the scope of the illness-death model, and their consistency is established. We investigate the performance of the proposed estimators by simulations. An illustration through real data analysis is included.

E125: **Advances in nonparametric estimation and testing problems: the illness-death model**

Presenter: **Jacobo de Una-Alvarez**, University of Vigo, Spain

Co-authors: Mar Rodriguez-Girondo

Traditionally, nonparametric estimation of transition probabilities for multi-state survival data is performed via the time honoured Aalen-Johansen estimator. However, this estimator may be inconsistent, showing a large bias when the Markov assumption is violated. Alternative nonparametric estimators which do not rely on such an assumption have been recently introduced and investigated. However, these estimators suffer from large standard errors which make difficult the inferences. For these reasons, it is important to provide reliable testing methods for markovianity which may indicate if Aalen-Johansen estimation is permitted, or recommend using other estimators. In this talk we consider a number of issues in this setup, namely: (1) Existing methods for nonparametric estimation of transition probabilities are reviewed; (2) Goodness-of-fit tests for the Markov assumption are introduced and explored; and, as a relevant by-product, (3) Resampling methods under markovianity are discussed. We focus on the illness-death model and we give some ideas for possible extension to other multi-state models.

ES59 Room MAL 532 MIXTURE MODELLING FOR CLUSTERING Chair: Christian Hennig

E381: **Model-based clustering of probability density functions**

Presenter: **Angela Montanari**, University of Bologna, Italy

Co-authors: Daniela G. Calo', Matteo Farne'

Probability density functions (pdfs) are usually employed to model complex datasets in several research fields, from frequency domain methods in time series analysis to computer vision and image analysis. This has led to the development of various metrics to measure the similarities between different pdfs, and to address classification problems on a given set of observed pdfs. In this paper, we present a method for clustering a set of pdfs which have been estimated from data using nonparametric methods. The idea is to "learn" a probability model on the space of pdfs: the aim is to define and to estimate a proper mixture model, since elements belonging to the same cluster are naturally thought of as samples from the same probability model. The task is rather challenging because of the infinite-dimensionality of the domain space. Moreover, this space has an intrinsic non-Euclidean geometry, since it forms a nonlinear manifold, namely the unit Hilbert sphere (under the square-root parametrization). We tackle the problems by approximating each pdf by a finite-dimensional vector representation and by introducing suitable mixture models for hyper-spherical data. The performance of the proposed solution is evaluated both on simulated and real data and compared with alternative clustering methods.

E425: **Gaussian mixture models: constrained and penalized approaches**

Presenter: **Roberto Rocci**, University of Tor Vergata, Italy

Finite mixture of Gaussians is a well known model frequently used to classify a sample of observations. The idea is to consider the sample as drawn by an heterogeneous population where each sub-population is described by a component of the mixture. The unknown parameters are usually estimated by maximizing the likelihood via an EM algorithm. In practice, this poses several problems, especially if the number of observations is small compared to the number of variables, because the likelihood function is unbounded from above and may present many local spurious maxima. Several methods have been proposed to circumvent this problem. They are based on the fact that unboundedness is avoided if the covariance matrices are not allowed to become singular. The problem is then solved by imposing some constraints on the eigenvalues or a convenient prior on the covariance matrices. In both cases, the solution is obtained by incorporating a priori information about the covariance structure of the components. The effectiveness of the method then depends on the quality and appropriateness of this information. The subject of this work is to reconsider the two approaches into a unified framework where the class conditional covariance matrices are shrink toward a pre-specified matrix Ψ. Our purpose is to investigate two closely related problems. First, how to estimate the matrix Ψ from the data when a priori information is not available. Second, how to determine the amount of shrinking needed.

E459: **A new approach to the computation of a robust estimator for Gaussian mixtures**

Presenter: **Pietro Coretto**, Universita degli Studi di Salerno, Italy

Co-authors: Christian Hennig

Gaussian mixtures (GM) are a popular tool used to represent elliptical-shaped clusters. The parameters of a GM can be estimated by maximum likelihood (ML). However, a small proportion of outliers (or noisy points), i.e. points not consistent with the GM model, can affect the ML estimator and the corresponding clustering severely. A possible solution to this problem is to add a noise component. This can be performed by adding a uniform component. Another approach is to represent the noise by an improper constant density, the latter corresponds to the Robust Improper Maximum Likelihood Estimator (RIMLE). ML estimates are usually computed by the EM algorithm which heavily depends on the choice of the initial values; moreover noise makes this dependence even stronger. We propose a method where initial values are estimated using a mix of the k-medoids and the Minimum Covariance Matrix (MCD) estimator. Within each mixture sub-population, outlying observations are identified using the MCD-based testing strategy. The computing load is feasible even in the case of high-dimensional datasets.

E433: **Identifying finite mixture models under model uncertainty**

Presenter: **Sylvia Fruhwirth-Schnatter**, Johannes Kepler University Linz, Austria

The identification of finite mixture models when the number of components is unknown is considered. The first part of the talk sheds some light on the role the prior of the weight distribution $p(\eta)$ plays when the true number of components is unknown. It is shown that the very popular uniform prior is usually a poor choice for overfitting models. A prior decision has to be made through the choice of $p(\eta)$ whether for overfitting mixture models empty components or identical, non-empty components should be introduced. As a consequence of this choice, either the number of non-empty components or the total of components is a better estimator of the true number of components. In the second part of the talk identification of finite mixture models that are strongly overfitting heterogeneity in the component-specific parameters is discussed. While standard priors lead to underfitting the true number of components, shrinkage priors well-known from variable selection are applied to handle overfitting heterogeneity. Such priors are able to discriminate between coefficients which are more or less homogenous and coefficients which are heterogeneous and avoid underfitting of the number of components by reducing automatically the prior variance of homogeneous components.

EP01 Room Senate Crush Hall POSTERS I Chair: Marios Fyrillas

E271: **Statistics on ECGs: wavelet smoothing, registration and classification**

Presenter: **Anna Maria Paganoni**, Politecnico di Milano, Italy

Co-authors: Francesca Ieva, Valeria Vitelli, Davide Pigoli

Cardiovascular diseases are one of the main causes of death all over the world. In this kind of pathologies, it is fundamental to be well-timed in order to obtain good prognosis in reperfusive treatment. PROMETEO project, realized by 118 Operating Center of Milan and funded by Mortara Rangoni Europe s.r.l., aims at anticipating diagnostic times. In particular, an automatic classification procedure based on statistical analyses of tele-transmitted ECG traces would be very helpful for an early diagnosis. The case study consists of 10992 ECG traces; we select a training set (96 signals) containing both normal and pathological (Atrial Fibrillation, STEMI and Non STEMI) traces. First we reconstruct signals, denoising measurements with a Daubechies wavelet basis. Then we remove the biological variability in the signals that induces a phase variability irrelevant to the classification's scope, through data registration, based on suitable landmarks provided by the measurement system. Finally, we perform a functional principal component analysis of reconstructed and registered ECGs. Some of the detected principal components are clearly associated with variability sources due to different cardiovascular pathologies; we can thus perform patients classification and pathologies prediction projecting the remaining 10896 curves in the functional space generated by these relevant principal components.

E438: **ROCtest: R package for flexible ROC curve comparison**

Presenter: **Carlos Carleos**, Universidad Oviedo, Spain

Co-authors: Pablo Martinez-Camblor, Norberto Corral

Receiver operating characteristic curves (ROCs) are a widespread method to represent the error rates of binary classifiers as a discrimination threshold varies. A typical problem in ROC analysis is to test whether several ROCs are significantly different, in order to compare the corresponding classification models. In this context, nonparametric methods are particularly useful, because of usual lack of knowledge on the underlying stochastic behaviour. The authors have recently introduced a novel nonparametric method, based on handling ROCs as cumulative distribution functions, in order to test for equality of several ROCs, both in paired and independent samples. This technique is quite flexible in that it can incorporate a plethora of distance measures, weights, aggregation type (supremum or average), and resampling regime (permutations or bootstrap). Here an R package is presented that implements the new method, besides omnibus versions of other traditional nonparametric ones (AUC, Venkatraman's, etc). A comparison is done with respect to other ROC-related R packages (ROCR, Bioconductor-ROC) and conversion tools to transfer data between them is provided.

E470: **Repeated measures analysis for functional data**

Presenter: **Pablo Martinez-Camblor**, Universidad de Oviedo, Spain

Co-authors: Norberto Corral

In an increasing number of fields, the advances of the computational techniques allow to obtain (almost) continuous information. Functional data analysis (FDA) is devoted to analyze data providing information about curves. Most of the (more or less) traditional statistical methods are being adapted to the FDA context. The repeated measures analysis which deals with the k-sample problem when the data are from the same subjects is investigated. Both parametric and nonparametric approaches are considered. A Monte Carlo simulation study is carried out in order to explore the statistical power of the proposed method in different scenarios.

E871: **Rank methods for microarray data**

Presenter: **Martin Schindler**, Technical University of Liberec, Czech Republic

Co-authors: Daniel Fischer, Hannu Oja

Our study was motivated by lymphoblastoid microRNA (miRNA) microarray data from more than 100 prostate cancer and healthy brothers. Due to nature of the data we use rank methods. After mentioning preprocessing, quality control and filtering procedures we first describe miRNA-wise analysis based on regression rank score test taking into account family effects. In the second part we show that using Ward's hierarchical clustering method with a distance based on Kendall's τ between individuals leads to very good results. This way we are also able to find the distance between the cluster of healthy and the cluster of cancer patients and by decomposition of τ to find the effects of individual miRNAs. Finally, we use a suitable permutation test to find the miRNAs with a significant effect. We propose a similar procedure in case that a family effect is taken into account. Similar approach is also used to show the level of dependency between miRNAs. This procedure is, thanks to Kendall's τ, very robust against potential outliers and preprocessing and normalization free.

E482: **Linear estimation of signals from observations with Markovian delays**

Presenter: **Maria Jesus Garcia-Ligero**, Granada, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez

Traditionally, the estimation problem of signals was addressed assuming that the measurement data are transmitted over perfect communications channels. Unfortunately, in practice, an unavoidable problem in wireless communication networks is the existence of errors during the transmission such as that the measurements can arrive with delay. For random-delay models, the signal estimation problem has been usually addressed assuming that the delay is modelled by a sequence of independent Bernoulli random variables that takes on values 0 and 1 depending on the real observation is received on time or it is not up-to-date and it has been also considered assuming correlation between the Bernoulli random variables; in these situations, recursive algorithms have been obtained to estimate the signal considering that the state-space model is known as well as using covariance information. In this paper, we consider that the delay is modelled by a Markov chain capturing the dependence between delays. Under this assumption the least-squares linear estimation problem of the signal based on delayed observations is addressed using the information provided by the covariance functions of the process involved in observation equation. The recursive estimation algorithms proposed are obtained by using an innovative approach.

E327: **Covariance-based quadratic filtering using correlated uncertain observations from different sensors**

Presenter: **Josefa Linares-Perez**, Granada University, Spain
Co-authors: Aurora Hermoso-Carazo, Raquel Caballero-Aguila

The quadratic least-squares estimation problem of signals from observations coming from multiple sensors is addressed when there is a non-zero probability that each observation does not contain the signal to be estimated. We assume that, at each sensor, the uncertainty about the signal being present or missing in the observation is modelled by correlated Bernoulli random variables, whose probabilities are not necessarily the same for all the sensors. A recursive algorithm is derived without requiring the knowledge of the signal state-space model but only the moments (up to the fourth-order ones) of the signal and observation noise, the uncertainty probabilities and the correlation between the variables modelling the uncertainty. The estimators require the autocovariance and crosscovariance functions of the signal and their second-order powers in a semidegenerate kernel form. The recursive quadratic filtering algorithm is derived from a linear estimation algorithm for a suitably defined augmented system. The theoretical results are applied to a numerical simulation example wherein a signal is estimated from uncertain observations with correlated uncertainty coming from two sensors; the linear and quadratic estimation error covariance matrices are compared, showing the superiority of the quadratic estimators over the linear ones.

E589: **Comparative study of flexible quantile regression techniques: constructing reference curves in Pediatrics**

Presenter: **Isabel Martinez-Silva**, University of Santiago de Compostela, Spain
Co-authors: Javier Roca Pardinas, Carmen Cadarso-Suarez, Rosaura Leis Trabazo, Rafael Tojo Sierra

In many applications, it is often of interest to assess the possible relationships between covariates and quantiles of a response variable through a regression model. In some instances, the effects of continuous covariates on the outcome are highly nonlinear. Consequently, appropriate modeling has to take such flexible smooth effects into account. To this aim, several additive quantile regression methodologies were recently proposed in the statistical literature, allowing for semiparametric predictors including linear and nonlinear effects. To date, only partial comparative simulation studies of the existing methodologies have been published. Accordingly, one of the main goals of this work is to perform an in-depth simulation study to compare statistically different additive quantile regression approaches. Also, all the reviewed techniques were used to construct the overall sex- and age-specific reference curves of anthropometric measures in a population of children and adolescents aged 6 to 18 years from randomly selected schools in 14 municipalities in Galicia (Northwest Spain).

E631: **Quantifying dispersion in nanocomposites**

Presenter: **David Bray**, Imperial College London / Queen Mary University of London, UK
Co-authors: Steven Gilmour, Ambrose Taylor, Felicity Guild

Characterising dispersion quality of nanocomposites presents a challenging statistical problem. Engineers need to know the spread and arrangement of nanoparticles across the system in order to determine their affects on the physical properties. Currently there are very few direct methods for measuring dispersion and none have been fully adopted. A high precision measure is required based on data from a single sample of a material. Our approach is based on the Delaunay Network of particles. The Area Disorder, AD, (previously used for Voronoi tessellation) measures the variance in areas of Delaunay triangles scaled by expected value. Any material sample can be categorised into well or poorly dispersed using AD. AD analysis is applied to several micrographs of nanocomposites and found to reliably classify the type of dispersion.

E751: **Structure functions in statistically stationary isotropic turbulence**

Presenter: **Marios Fyrillas**, Cyprus University of Technology, Cyprus
Co-authors: Elias Gravanis, Evangelos Akylas

The physics of the linear forcing of isotropic turbulence, allows for some useful estimates of the characteristic length scales of the turbulence produced during the statistically stationary phase. With such estimates we could practically define uniquely the stationary statistics by means of the box-size of the simulation, the linear forcing parameter and the viscosity of each case. We use such estimations in the Karman Howarth equation and we solve it in terms of the second and third order structure functions using a generalized Oberlack Peters closure scheme. The resulting forms and the respective spectra are in very good agreement with experimental and DNS data. They also provide strong insight into the physical process that produces the statistical stationarity, reflecting the role of the periodic boundary condition in the resulting spectral forms.

E876: **Mixture of bivariate Poisson regressions with an application to insurance**

Presenter: **Lluis Bermudez**, University of Barcelona, Greece
Co-authors: Dimitris Karlis

With the problem of pricing an insurance contract that contains different types of coverages, several works have shown that there is some correlation between types of coverages. Hence multivariate models are needed. Among them the bivariate (multivariate) Poisson regression model is perhaps the most widely used. Since overdispersion may be present, we need to extend this model to this direction. Assuming a continuous mixing distribution leading to a bivariate mixed Poisson distribution can be a solution but typically it leads to complicated likelihoods and hence estimation is not simple. In the present paper we assume a finite mixture of bivariate (multivariate) Poisson regression. The model is easy to be fitted to real data since being a finite mixture standard EM is applicable. The model also is very flexible allowing for both positive and negative correlation and at the same time overdispersion of a different level in each variable. We discuss properties and estimation of the model and then we apply the model to insurance data discussing the added value of the proposed model.

Friday 10.12.2010 15:50 - 17:05 Parallel Session D – CFE

CS21 Room MAL B20 BAYESIAN ECONOMETRICS **Chair: Cathy Chen**

C323: **Forecasting tail risk in financial time series using a mixture of distribution approach**
Presenter: **Mike So**, The Hong Kong University of Science and Technology, Hong Kong
Co-authors: Ka Shing Chan

Financial return distribution is well recognized for its fat-tail behavior and its tail asymmetry. One natural question is whether the tail asymmetry is still significant after accounting for the conditional heteroskedasticity in returns. In this paper, we propose a mixture of distribution approach under a GARCH framework to capture the tail asymmetry. A key idea is to use the Peak-over-Threshold method of extreme value theories to construct a mixture distribution for the innovation in GARCH models. This mixture distribution combines a common distribution, like normal, and two generalized Pareto distributions for the two tail parts to capture both the leptokurtosis and the tail asymmetry. A Bayesian approach is adopted to estimate unknown parameters by using Markov Chain Monte Carlo (MCMC) methods. A Bayesian test for tail asymmetry is also established. We also study the performance of our approach in forecasting volatility and the tail risk, like Value at Risk and expected shortfall, using real data.

C259: **Forecasting risk via nonlinear models and the two-sided Weibull 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 range of symmetric, threshold nonlinear and smooth transition GARCH models are considered for the volatility process. To capture potential skewness and heavy tails, the model proposes an asymmetric two-sided Weibull as the conditional distribution of the series. Estimation is via an adaptive Markov Chain Monte Carlo (MCMC) sampling scheme, employing the Metropolis-Hastings (MH) algorithm with a mixture of Gaussian proposal distributions. The model is illustrated by applying it to multiple international stock market indices, exchange rates and a single asset return series, generating one step-ahead forecasts of VaR and ES. Standard and non-standard tests are applied to these forecasts, finding that the proposed model performs favourably compared to some popular competitors both before, during and after the global financial crisis: in particular it is the only consistently accurate model of risk over the period studied.

C421: **Bayesian hierarchical non-parametric mixture of dynamic GARCH models**
Presenter: **John W. Lau**, University of Western Australia, Australia
Co-authors: Ed Cripps

We will describe a hierarchical mixture of dynamic GARCH models, where by dynamic we mean the GARCH parameters are permitted to vary over time. The model parameters are mixed over the Poisson-Kingman process based on the stable subordinator. This process is a generalisation of the Dirichlet process typically used in hierarchical models, provides a richer clustering structure and its application to time series data is novel. Inference is Bayesian and a Markov chain Monte Carlo algorithm is used to explore the posterior distribution. The method is applied to recent Dow Jones Industrial Average log-returns. A comparison of the performance of the Dirichlet process, two special cases of the Kingman Poisson process and a standard GARCH model is presented.

CS34 Room MAL 538 PRICING AND HEDGING IN INCOMPLETE MARKETS **Chair: Dominique Guegan**

C223: **Option pricing GARCH-type models with generalized hyperbolic innovations**
Presenter: **Dominique Guegan**, University of Paris 1 Pantheon - Sorbonne, France
Co-authors: Christophe Chorro, Florian Ielpo

We provide a new dynamic asset pricing model for plain vanilla options and we discuss its ability to produce minimum mispricing errors on equity option books. Given the historical measure, the dynamics of assets are modeled by GARCH-type models with generalized hyperbolic innovations and the pricing kernel is an exponential affine function of the state variables, we show that the risk neutral distribution is unique and implies again a generalized hyperbolic dynamics with changed parameters. We provide an empirical test for our pricing methodology on two data sets of options respectively written on the French CAC 40 and the American SP 500. Then, using our theoretical result associated with Monte Carlo simulations, we compare this approach to natural competitors in order to test its efficiency. More generally, our empirical investigations analyze the ability of specific parametric innovations to reproduce market prices in the context of an exponential affine specification of the stochastic discount factor.

C447: **Multivariate option pricing with time varying volatility and correlations**
Presenter: **Lars Stentoft**, HEC Montreal, Canada
Co-authors: Jeroen Rombouts

We consider option pricing using multivariate models for asset returns. Specifically, we demonstrate the existence of an equivalent martingale measure, we characterize the risk neutral dynamics, and we provide a feasible way for pricing options in this framework. Our application confirms the importance of allowing for dynamic correlation, and it shows that accommodating correlation risk and modeling non-Gaussian features with multivariate mixtures of normals substantially changes the estimated option prices.

C497: **Semi-static hedging strategies in incomplete markets. An application for carbon allowances**
Presenter: **Marius Frunza**, Sagacarbon, France
Co-authors: Dominique Guegan

We focus on the ability of discrete GARCH-type and Generalized Hyperbolic models to provide with efficient hedging alternatives for the CO2 emission derivatives, taking into the account the efficiency and compete information issues of the market as well as the lack of liquidity on options exchange. First we perform backtesting for delta-hedging on standard options portfolios using various models and we underline the lack of efficiency of this alternative. Second we benchmark two approaches proposed in the financial literature applied to the discrete case - the variance optimal hedging and the static hedging. Next we compare these methods with a new approach based on semi-static hedging, a compromise between the static and the delta-dynamic in order to optimize the rally cost generated by the asymmetry and jumps structure of the CO2 prices. We test on a historical basis the different approaches and we show that static and semi-static approaches built on Normal Inverse Gaussian distribution provide us with the best performance.

CS44 Room MAL 532 FORECASTING THE EQUITY PREMIUM: METHODS AND NEW RESULTS Chair: Marcelo Medeiros

C243: **Mean reversion expectations and the 1987 stock market crash: An empirical investigation**
Presenter: **Eric Hillebrand**, Louisiana State University, USA

After the stock market crash of 1987, Fischer Black proposed an equilibrium model in which he explained the crash by inconsistencies in the formation of expectations of mean reversion in stock returns. This study derives testable hypotheses implied by this explanation. An Ornstein–Uhlenbeck process that is consistent with Black's model is estimated on daily stock index data before and after the crash of 1987. The results strongly support the hypothesis. Simulations show that on Friday October 16, 1987, a crash of 20% or more had a probability of more than 7% .

C311: **Evaluating the predictability of stock market returns via STARX-Tree models**
Presenter: **Alvaro Veiga**, PUC-Rio, Brazil
Co-authors: Camila Epprecht

STARX-Tree models (Tree-structured Smooth Transition autoregression with external variables) are switching regime formulations where a different ARX model is activated at each instant according to a decision tree on a vector of observable variables, called the transition variables. In this paper we propose an application of the STAR-Tree formulation to automatic trading based on the simultaneous modelling of daily returns and realized volatilities. Variable selection for both parts of the model is completely automatic, implemented through a series of statistical tests. The one-day ahead forecasts of return and volatility are then used to produce buy-sell signals and estimate the best leverage level under a given risk constraint. The methodology was tested on 23 stocks negotiated at the NYSE and compared to other formulations including the simple historical mean (the most difficult to beat), the naive forecasting, neural networks and linear ARX models. The explanatory variables include past returns, past volatilities, trading volume and some technical analysis indexes. The results show that the STAR-Tree formulation outperforms traditional models and is a competitive alternative to neural networks. Also, the analysis of the set of variables gives some interesting clues about the effect of volatility changes and technical indexes on returns.

C414: **Multi-factor model selection for predicting cross-sectional variation in stock returns**
Presenter: **Artur Passos**, Pontifical Catholic University of Rio de Janeiro, Brazil

Multi-factor models for excess returns of individual stocks or portfolios are evaluated. The main factor is the market excess return, while others are zero-investment stock portfolios with weights defined according to stock characteristics. The characteristics cover trend, leverage and value. The evaluation of factors considers explanatory and predictive performance. Specifically, predictive performance evaluation takes into account the impact of timing in portfolio-factor expected returns. We define a large set of possible factors and evaluate each one according to explanatory and predictive performance. Besides, we compare and optimize multi-factor models found in literature. Finally, we propose a technique to obtain quasi-optimal multi-factor models based on predictive performance.

CS45 Room MAL 151 MULTIVARIATE UNOBSERVED COMPONENTS MODELS Chair: Gian Luigi Mazzi

C610: **Extracting common and specific components from the vector of prices in several European power markets**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Julio Rodriguez, Maria Jesus Sanchez

Dynamic Factor Analysis provides a decomposition of a m-dimensional vector of time series y_t into common and specific parts, Ωf_t and ε_t respectively, where Ω is the loading matrix that relates the m observed series and the r unobserved common factors f_t that evolve over time according to a VARIMA model. It is usually assumed that the specific factors ε_t are both independent across elements of a single ε_t and between all pairs and ε_t and ε_s, $s \neq t$. Although mathematically convenient, the second assumption may be suspect in practice when dealing with real data. A two-step estimation procedure that produces, under certain conditions, biased estimators of the parameters is often used. To overcome this problem, a joint estimation procedure is developed. The complete model including not only the decomposition of the original series into common and specific components and the common dynamics, but also the univariate modeling of specific factors is written under its state-space formulation, and a joint estimation is carried out. A comparison between these two procedures is carried out in a Monte Carlo study. The new estimation procedure is applied to decomposing the vector of daily prices in several European electricity markets (EXAA, EEX, Powernext and MIBEL).

C681: **Deriving a euro area monthly indicator of employment: a real time comparison of alternative model-based approaches**
Presenter: **Filippo Moauro**, Eurostat, Luxembourg

The results of an extensive real time analysis of alternative model-based approaches to derive a monthly indicator of employment for the euro area are presented. In the experiment the Eurostat quarterly national accounts series of employment is temporally disaggregated using monthly unemployment. The strategy benefits of the contribution of the information set of the euro area and its 6 larger member states, as well as the split into the 6 sections of economic activity. The models under comparison include univariate regressions of the Chow and Lin' type where the euro area aggregate is directly and indirectly derived, as well as multivariate structural time series models of small and medium size. The specification in logarithms is also systematically assessed. The largest multivariate setups, up to 58 series, are estimated through the EM algorithm. Main conclusions are the following: mean revision errors of disaggregated estimates of employment are overall small; a gain is obtained when the model strategy takes into account the information by both sector and member state; the largest multivariate setups outperform those of small size and the strategies based on classical disaggregation methods.

C685: **Euromind: a Euro area monthly indicator of economic activity**
Presenter: **Gian Luigi Mazzi**, European Commission, Luxembourg
Co-authors: Cecilia Frale, Stefano Grassi, Massimiliano Marcellino, Tommaso Proietti

EuroMind-C is a monthly indicator of the Euro area gross domestic product (GDP), and of its largest member states, based on the temporal disaggregation of the quarterly national accounts estimates, within a medium-size dynamic factor model of a set of coincident indicators. It provides a timely assessment of the economic situation on a well understood quantitative scale. The methodology is based on a bottom up strategy which hinges on the separate disaggregation of the components of GDP by expenditure type and by sector. For this purpose, we formulate a monthly model of coincident indicators featuring a common euro area cycle and country specific cycle, that is estimated under temporal and contemporary aggregation constraints. Estimation is carried out by the EM algorithm. Since the model is formulated in logarithms, a nonlinear smoothing algorithm has to be applied for the disaggregation of the quarterly national accounts aggregates.

CS48 Room MAL B29 BOOTSTRAP METHODS IN FINANCE Chair: John Nankervis

C263: **Evaluating performance measures stability**

Presenter: **Giovanna Menardi**, University of Padova, Italy

Co-authors: Francesco Lisi

Managed portfolio performance evaluation is an important issue both from the academic and the investor's viewpoint. A relevant aspect concerns the choice of the measure used for assessing the performance, because each index offers a different perspective about the trade-off between return level and risk exposure. A related crucial issue, then, is the stability of such measures throughout the time and, hence, their predictive power about future results. In this work we address the problem of the evaluation of the stability of a performance measure. Firstly, we discuss the use of alternative criteria to measure stability and we propose a stability index (SI) based on the changes of ranks over several periods of time. Also, we propose a bootstrap based statistical test to evaluate the homogeneity of stabilities of alternative performance measures. In a second step, we suggest a composite performance measure, built as a linear combination of different indexes, specifically conceived to maximize the stability over the time while preserving information about the risk-adjusted return behaviour. To this aim we use an approach based on methods for dimensionality reduction. An application to a large set of US mutual funds belonging to the Large Blend category shows how the proposed index and test work.

C653: **Boosting, bagging and bragging applied to nonparametric regression - an empirical approach**

Presenter: **Cornelia Swanepoel**, North-West University, Potchefstroom Campus, South Africa

Co-authors: Lusilda Boshoff

Combinations of boosting, bagging and bragging methods are applied to nonparametric kernel regression estimation. A variety of bandwidths and other parameters are estimated by minimizing a cross-validation function. Different regression estimates are evaluated by minimizing the global mean integrated squared error discrepancy measure. Boosting is a general method for improving the accuracy of any given learning algorithm and has its roots in the machine learning context. For example, boosting has been shown in the literature to improve the Nadaraya-Watson learning algorithm. Bagging, an acronym for 'bootstrap aggregating', is a method involving the generation of multiple versions of a predictor. These replicates are used to get an aggregated estimator. In the regression setting, the aggregation calculates an average over multiple versions which are obtained by drawing bootstrap samples from the original training set. We apply some modifications of the method such as bragging where, instead of the average, a robust estimator is calculated. Boosting, bagging and bragging methods can be seen as ensemble methods. Ensemble methods train multiple component learners and then combine their predictions. Results and conclusions verifying existing literature are provided as well as new results for the new methods.

C579: **Calendar anomalies and data snooping in European stock market indices**

Presenter: **Michele Marzano**, University of Essex, UK

Co-authors: Jerry Coakley, John Nankervis

An investigation into the economic significance of three of calendar anomalies that are robust to data snooping in fourteen European stock markets from the beginning of 1990 to the end of 2009 is proposed. The anomalies studied are the day of the week effect, the month of the year effect and the Summer effect. For each dataset we tested a universe of $12,526$ trading rules. Ignoring data-snooping, we find that the number of trading rules found to be profitable is extremely high for all sample countries. To account for data-snooping, we employ a novel and powerful bootstrap snooper procedure, the SSPA (Stepwise Superior Predictive Ability) test which not only removes data-snooping but allows us to identify the rules which are significant. Applying this methodology, we find that the number of significant trading rules decreases dramatically. We establish significance across time and countries only for the month of the year trading rules. The results suggest that data-snooping plays an important role in determining the apparent significance of calendar effects.

CS54 Room MAL B18 MODELLING AND CAUSALITY Chair: Marco Reale

C649: **Commodity price–exchange rate causality in daily and intra-day data**

Presenter: **Hui Jun Zhang**, McGill University, Canada

Co-authors: Jean-Marie Dufour, John Galbraith

We investigate the causal relationship between 'commodity currency' exchange rates (CAD/USD, CLP/USD, and SAR/GBP) and commodity prices (crude oil, copper, and gold) in daily and intra-day data. Particularly, we use the concept of multi-horizon causality measure to compare the strength of different causal relationships, which cannot be assessed by non-causality tests. In contrast with previous studies, in which exchange rates have been suggested to have power in predicting future commodity prices in low-frequency data, typically quarterly, our results demonstrate the stronger causality running from commodity prices to exchange rates than in the opposite direction at both daily and 5-minute frequencies. To examine the possible indirect causality in the presence of the auxiliary variable, we also include S&P500 index price data. The causality from S&P500 index price to exchange rates has been shown to be evident. We also find that price of crude oil and gold can help to predict the movements of S&P500 index price at the daily base, while copper price tends to follow the changes of S&P500 index price.

C719: **Relating confidence to information uncertainty in qualitative reasoning**

Presenter: **Gregory Chavez**, Los Alamos National Laboratory, USA

Co-authors: David Zerkle, Brian Key, Daniel Shevitz

Qualitative reasoning models, i.e. approximate or evidential reasoning model, makes use of qualitative assessments provided by subject matter experts to model factors such as security risk. Confidence in a result is important and useful when comparing competing security risk results but a method for the determination in qualitative reasoning models is currently not available. Quantifying the confidence in a qualitative reasoning model result must be consistent and based on the available information. A novel method is proposed to determine a linguistic measure of confidence in a reasoning result from the available information uncertainty in the result using membership values in the fuzzy sets of confidence. In this study information uncertainty is quantified through measures of non-specificity and conflict. Fuzzy values for confidence are established from information uncertainty values that lie between the measured minimum and maximum information uncertainty values. Measured values of information uncertainty in each result is used to obtain the confidence. The determined confidence values are used to compare competing scenarios and understand the influences on the desired result.

C727: **Do long memory time series suffer amnesia?**

Presenter: **Marco Reale**, University of Canterbury, New Zealand

Co-authors: William Rea, Les Oxley, Chris Price

It is now widely recognized that the statistical property of long memory may be due to reasons other than the data generating process being fractionally integrated. We propose a new procedure aimed at distinguishing between a null hypothesis of unifractal fractionally integrated processes

and an alternative hypothesis of other processes which display the long memory property. The procedure is based on a pair of empirical, but consistently defined, statistics namely the number of breaks reported by a theoretical Regression Trees (ART) and the range of the Empirical Fluctuation Process (EFP) in the CUSUM test. The new procedure establishes through simulation the bivariate distribution of the number of breaks reported by ART with the CUSUM range for simulated fractionally integrated series. This bivariate distribution is then used to empirically construct a test which rejects the null hypothesis for a candidate series if its pair of statistics lies on the periphery of the bivariate distribution determined from simulation under the null. We apply these methods to the realized volatility series of 16 stocks in the Dow Jones Industrial Average and show that the rejection rate of the null is higher than if either statistic was used as a univariate test.

CS62 Room MAL G16 FINANCIAL DATA MINING Chair: Philip Yu

C244: **Conditional jump patterns mining using bidirectional Hebbian clustering and sorting**

Presenter: **Kai Lam**, Chinese University of Hong Kong, China

Co-authors: Hao Ding, Terrence Mak

Financial jump patterns are often related with macroeconomics news announcements, but they are difficult to have meaningful interpretations. In data mining, the statistical techniques of k-means clustering and correlation-based sorting are used for pattern discovery and analysis. We synthesize a financial jump process using a conditional jump representation with a MEM (Multiplicative Error Model) for non-negative input from the difference between realized volatility and bipower or multipower quadratic variation. A probability density function of Weibull-type with or without the addition of a Poisson process is used for the MEM innovations. A row-based and column-based bidirectional Hebbian learning is constructed to estimate robustly the major principal components of the jump process using the interpretation of singular value decomposition. The learned principal components allow the simple application of statistical clustering using the orthogonal projections, which can also be used to define a generalized Manhattan distance measure for sorting jump patterns which have a close proximity in the eigenspace. The synthesized MEM models are useful for studying the Hebbian clustering and sorting results under a controlled simulation environment with specified probability distributions. We also applied the algorithm for handling 247 days of high-frequency S&P 500 data with an intraday sampling at 10 minutes interval.

C627: **Measuring reputational risk**

Presenter: **Paola Cerchiello**, University of Pavia, Italy

Co-authors: Paolo Giudici

Reputation can be defined as how an entity (private or public) is perceived by each of its stakeholder groups and reputation risk as the risk that an event will negatively influence stakeholder perceptions. Since reputation involves intangible assets (public opinion, perception, reliability, merit), it is not simple to define and consequently to measure and to monitor the correlated risk. In the first part of our contribution we propose non parametric indexes aimed at producing a reputation based ranking of public companies. In the second part we propose a parametric approach based on a mixture of random variables to measure and predict reputation. The proposed models are applied to real data on Italian public companies taken from financial media corpora.

C277: **Mining optimal technical chart patterns with genetic algorithms**

Presenter: **Philip Yu**, The University of Hong Kong, Hong Kong

Co-authors: Laraine Shen

Technical analysis is believed and applied by many professional traders that profitable trading signals can be generated by analyzing historical security prices. In the literature, genetic algorithms have been successfully applied to assess the profitability of technical trading strategies based on various technical indicators. We will use genetic algorithms to find the optimal technical chart patterns and investigate their profitability in emerging financial markets.

CS64 Room MAL B34 HEAVY-TAILED FINANCIAL ECONOMETRICS Chair: David Veredas

C897: **Tailing tail risk in the hedge fund industry**

Presenter: **Marcelo Fernandes**, Queen Mary, University of London, UK

Co-authors: Walter Distaso, Filip Zikes

The aim is to assess dynamic tail risk exposure in the hedge fund sector using daily data. We use a copula function to model both lower and upper tail dependence between hedge funds, bond, commodity, foreign exchange, and equity markets as a function of market uncertainty, and proxy the latter by means of a single index that combines the options-implied market volatility, the volatility risk premium, and the swap and term spreads. We find substantial time-variation in lower-tail dependence even for hedge-fund styles that exhibit little unconditional tail dependence. In addition, lower-tail dependence between hedge fund and equity market returns decreases significantly with market uncertainty. The only styles that feature neither unconditional nor conditional tail dependence are convertible arbitrage and equity market neutral. We also fail to observe any tail dependence with bond and currency markets, though we find strong evidence that the lower-tail risk exposure of macro hedge funds to commodity markets increases with liquidity risk. Finally, further analysis shows mixed evidence on how much hedge funds contribute to systemic risk.

C385: **Estimating GARCH volatility in the presence of outliers**

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

Co-authors: Angeles Carnero, Daniel Pena

When GARCH models are fitted to real data, the residuals often have excess kurtosis which could be explained, among other reasons, by the presence of outliers. Previous research has focused on the effects of outliers on the Gaussian Maximum Likelihood (ML) estimator of GARCH parameters. However, the main interest of practitioners is the estimation of the underlying volatilities. Maximizing the likelihood based on a heavy tailed distribution protects against outliers when estimating the GARCH parameters. Consequently, we consider a Quasi Maximum Likelihood estimator based on maximizing the Student-t log-likelihood function and, alternatively, a Bounded-M estimator based on the maximization of a conveniently modified likelihood. Our objective is to study the effects of outliers on the estimated GARCH volatilities computed by using the ML estimates of the parameters and compare the performance of robust volatility estimators. We show that, in the presence of outliers, GARCH volatilities estimated by Maximum Likelihood can have very large biases even if those of the estimated parameters are small. The reason is that volatilities depend on the unconditional variance, which is a non-linear function of the parameters. We also show that robust volatility estimates have clearly smaller biases. Results are illustrated using S&P500 returns.

C389: **Spline-based tail index estimation**

Presenter: **Dieter Schell**, University of Konstanz, Germany

Co-authors: Jan Beran

Heavy-tailed densities are approximated using a combination of the Pareto density and varying number of B-Splines. The tail index is estimated simultaneously with the B-Spline coefficients. Consistency and asymptotic normality are derived. A simulation illustrates the performance of the new estimator in a finite sample setting.

C512: **Rank based testing and estimation in the general linear model with stable errors**

Presenter: **Yvik Swan**, ULB, Belgium

Co-authors: Marc Hallin, Thomas Verdebout, David Veredas

Linear models with stable error densities are considered. The local asymptotic normality of the resulting model is established. We use this result, combined with Le Cam's third lemma, to obtain local powers of various classical rank tests (Wilcoxon's and van derWaerden' s test, the median test, and their counterparts for regression and analysis of variance) under α-stable laws. The same results are used to construct new rank tests achieving parametric optimality at specified stable densities. We also propose rank based estimators of the regression parameter.

CS65 Room MAL B30 BAYESIAN METHODS IN ECONOMETRIC AND FINANCIAL APPLICATIONS Chair: Ioannis Vrontos

C209: **Estimation of time-varying GARCH-M models**

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 the recursive nature of the conditional variance makes exact likelihood analysis of these models computationally infeasible. We outline the issues and suggest to use the Markov chain Monte Carlo 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, where T is the sample size. In order to investigate the practical performance of the proposed procedure, we estimate within a Bayesian context, various GARCH-M models for weekly excess stock returns from the Dow-Jones, Nikkei and FTSE index.

C366: **The dynamics of US and UK inflation expectations**

Presenter: **Simon Potter**, Federal Reserve Bank of New York, USA

Co-authors: Deborah Gefang, Gary Koop

This paper investigates the relationship between short term and long term inflation expectations in the US and the UK with a focus on inflation pass through (i.e. how changes in short term expectations affect long term expectations). A Bayesian econometric methodology is used based on the smoothly mixing regression approach is applied to inflation expectations measured as the difference between nominal and real yield curves. This allows us to uncover the relationship between inflation pass through and various explanatory variables. We relate our empirical results to theoretical models of anchored, contained and unmoored inflation expectations. For neither country do we find anchored or unmoored inflation expectations. For the US, contained inflation expectations are found. For the UK, our findings are not consistent with the specific model of contained inflation expectations presented here, but are consistent with a more broad view of expectations being constrained by the existence of an inflation target.

C541: **Understanding liquidity and credit risks in the financial crisis**

Presenter: **Deborah Gefang**, Lancaster University, UK

Co-authors: Gary Koop, Simon Potter

This paper develops a structured dynamic factor model for the spreads between London Interbank Offered Rate (LIBOR) and overnight index swap (OIS) rates for a panel of banks. Our model involves latent factors which reflect liquidity and credit risk. Our empirical results show that surges in the short term LIBOR-OIS spreads during the 2007-2009 financial crisis were largely driven by liquidity risk. However, credit risk played a more significant role in the longer term (twelve-month) LIBOR-OIS spread. The liquidity risk factors are more volatile than the credit risk factor. Most of the familiar events in the financial crisis are linked more to movements in liquidity risk than credit risk.

CS23 Room MAL 541 BAYESIAN METHODS IN MACROECONOMICS AND FINANCE I Chair: Alessia Paccagnini

C420: **Unconventional monetary policy and the great recession**

Presenter: **Christiane Baumeister**, Bank of Canada, Canada

Co-authors: Luca Benati

We explore the macroeconomic impact of a compression in the long-term bond yield spread within the context of the Great Recession of 2007-2009 via a Bayesian time-varying parameter structural VAR. We identify a *pure* spread shock which, leaving the short-term rate unchanged by construction, allows us to characterise the macroeconomic impact of a compression in the yield spread induced by central banks' asset purchases within an environment in which the short rate cannot move because it is constrained by the zero lower bound. Two main findings stand out. First, in all the countries we analyse (U.S., Euro area, Japan, and U.K.) a compression in the long-term yield spread exerts a powerful effect on both output growth and inflation. Second, conditional on available estimates of the impact of the FED's and the Bank of England's asset purchase programmes on long-term government bond yield spreads, our counterfactual simulations indicate that U.S. and U.K. unconventional monetary policy actions have averted significant risks both of deflation and of output collapses comparable to those that took place during the Great Depression.

C615: **Model validation in the DSGE approach: A comparison**

Presenter: **Alessia Paccagnini**, Bicocca University, Italy

How model validation has been applied to Dynamic Stochastic General Equilibrium (DSGE) models is illustrated. Model validation consists in first giving a statistical representation of the model of interest, then in using this representation to explain the actual data. The theoretical form of a macroeconomic model is the starting point for the model validation. This model can be represented by using a model of the family of Vector Autoregressive Moving Average (VARMA). Actual data are best represented through an Unrestricted Vector Autoregressive (UVAR) which is the reduced form. The combination of a VARMA representation derived from the theory with the UVAR model for the real data has a hybrid or mixture model as outcome. This kind of model is a restricted VAR which is the empirical model. The crucial point in the concept of model validation in the DSGE approach is to understand if this mixture model, which is the statistical representation of the theory, is valid to explain the real data. Two examples of DSGE models are illustrated in a comparison exercise in order to introduce two problems: omitted variables within the statistical identification problem and the finite-order representation by a Vector Autoregressive (VAR) of a DSGE model.

C676: **Modeling and estimation of the synchronization in multivariate regime-switching models**

Presenter: **Cem Cakmakli**, Erasmus University Rotterdam, Netherlands

Co-authors: Dick van Dijk, Richard Paap

Often neither independence nor perfect synchronization (PS), single common cyclical dynamics, is adequate for representing synchronization of cyclical behaviour of variables. When multivariate regime-switching models are considered, PS is often imposed. We develop a general framework using Bayesian inference where we can simultaneously estimate degree of synchronization together with regimes. We extend related models in literature in two important directions. First, we allow for multiple regimes and consider possibility of different phase shifts for different regimes. We allow for both symmetric and asymmetric synchronization, namely, one cycle leads/lags the other symmetrically where turning points differ by identical intervals or asymmetrically where turning points can differ by distinct intervals depending on regime. Second, we allow for a regime-dependent covariance matrix. In empirical application, we consider cyclical behaviour of US Industrial Production and the Leading Economic Index (LEI). We find strong evidence that an MS-VAR model with three regimes is more appropriate than two-regime models. We find that third regime is necessary to capture severe recessions in 1975, 1980 and 2008. Additionally, a three-regime model where additional regime is forced to capture recovery phase is found to perform considerably worse. Lead time of LEI varies considerably across regimes, being two months for expansions, six for severe and twelve months for normal recessions.

CP03 Room IoE Crush Hall POSTERS III Chair: Christos Savva

C499: **Comovement of energy commodities revisited: Evidence from wavelet coherence analysis**

Presenter: **Lukas Vacha**, Academy of Sciences of the CR, Czech Republic

Co-authors: Jozef Barunik

The energy market comovement and contagion is considered. The novelty of our approach lies in the application of wavelet tools to commodity market data. A major part of economic time series analysis is performed in the time or frequency domain separately. Wavelet analysis can combine these two fundamental approaches, so we can work in the time-frequency domain. Using this framework, we propose a new, model- free way of estimating time-varying correlations. In the empirical analysis, we compare our approach with the dynamic conditional correlation on the main components of the energy sector. Namely, we use crude oil, gasoline, heating oil, and natural oil on a nearest-future basis over a period of approximately 16 and 1/2 years beginning on November 1, 1993 and ending on July 21, 2010. Using wavelet coherence, we uncover very interesting dynamics of correlations between energy commodities.

C363: **Measuring information flows between financial markets using transfer entropy**

Presenter: **Franziska Julia Peter**, University of Tuebingen, Germany

The concept of transfer entropy to examine the information flow between financial markets is employed. Transfer entropy is designed as a Kullback-Leibler distance of transition probabilities. This non-parametric method does not rely on restrictive modelling of the underlying processes and is also able to capture nonlinear statistical dependencies. The information flow between the CDS and bond market is measured using data on 36 iTraxx Europe companies in order to assess the question of causality with respect to the pricing of credit risk. It is found that overall information flows almost to an equal amount into both directions, while the information transfer from the CDS to the bond market is slightly larger. Furthermore, the causality between market risk and credit risk is examined by measuring the information flow between the iTraxx and the VIX. Unidirectional information flow from the VIX into the iTraxx is found.

C488: **Estimation of time varying adjusted PIN and PSOS using high-frequency transaction data**

Presenter: **Daniel Preve**, Uppsala University, Sweden

Co-authors: Yiu Kuen Tse

Informed trading (PIN) and probability of symmetric order-flow shock (PSOS) have been proposed in the literature as measures of asymmetric information and illiquidity, respectively. Our method is an extension of an existing method to estimate time varying PIN using high-frequency transaction data, which is based on an asymmetric autoregressive conditional duration model of expected durations of buy and sell orders. Our empirical results indicate that daily adjusted PIN is much more stable than daily PIN and that, in contrast to PIN, adjusted PIN is negatively correlated with integrated volatility (IV). Instead, daily PSOS is the component that is positively correlated with IV. Moreover, by comparison with the daily adjusted PIN, the daily PSOS seems to be quite volatile over time.

C755: **The model of the best quotes (bid and ask) with endogenous limit order books: a verification by HF data**

Presenter: **Martin Smid**, UTIA AS CR, Czech Republic

Recently, several models of the limit order markets have been constructed: Poisson orders' arrivals and the bounded uniform relative limit prices, the static model with non-uniform continuous absolute limit prices, a discrete-price uniform model and, finally, the recent generalization of the last model incorporating several properties of real-life limit order markets. A general setting covering all the mentioned models (except of the cancellations in the last one) was previously developed and its distribution was (recursively) described. Here, we estimate the parameters of a generalization of the latter model using several years' data of five stocks' prices from four different (limit order) markets (counting several hundreds of millions of records). We show that (a) the times between the jumps of the quotes are (close to) log normal, (b) the degree of the oscillation of the spread decreases with the increasing spread, (c) the moves of the mid-price are uncorrelated with the past (which is a possible argument supporting the EMH), (d) the dependence caused by the order books is quite strong (lasting several hours), and (e)the limit order flow (near the quotes) can be satisfactorily described by an estimable sub-model.

C269: **A single index model procedure for interpolation intervals**

Presenter: **Ana Elizabeth Garcia Sipols**, Universidad Rey Juan Carlos, Spain

Co-authors: Andres Alonso, Silvia Quintas, Clara Simon de Blas

A method for constructing interpolation intervals in time series is presented. It is an extension of an existing method for the MA(1) model. In this method one of the points that can be improved is the estimation of a conditional distribution, that is subject to the curse of dimensionality. The idea is to improve this step, by assuming a single-index like model: it is well known that curse of dimensionality can be avoided in these models, provided one can find good estimations of the directions of the model. The directions are estimated using a method where the conditional distribution is estimated using a classical kernel estimate. A numerical study based on simulated data and real data is presented.

C517: **Monte Carlo-based tail exponent estimator**

Presenter: **Miloslav Vosvrda**, Academy of Sciences of the CR, Czech Republic

Co-authors: Jozef Barunik, Lukas Vacha

We propose a new approach to estimation of the tail exponent in financial stock markets. We begin the study with the finite sample behavior of the Hill estimator under α-stable distributions. Using large Monte Carlo simulations, we show that the Hill estimator overestimates the true tail exponent and can hardly be used on samples with small length. Utilizing our results, we introduce a Monte Carlo-based method of estimation for the tail exponent. Our proposed method is not sensitive to the choice of tail size and works well also on small data samples. The new estimator also

gives unbiased results with symmetrical confidence intervals. Finally, we demonstrate the power of our estimator on the international world stock market indices. On the two separate periods of 2002-2005 and 2006-2009, we estimate the tail exponent.

C725: **Stochastic volatility option pricing with use of copula functions**

Presenter: **Ivan Saroka**, State University - Higher School of Economics, Russia

Co-authors: Dean Fantazzini

We present an option pricing model with probabilistically linked stochastic underlying price and stochastic volatility. The joint behaviour of the processes is described using a copula function approach. It is proposed to link the two cumulative distribution functions induced by these processes. It is shown that a fair option price could be recovered analytically under such model setting. The proposed approach involves derivation of a conditional distribution induced by the price process from the copula function. The relevant SDE is constructed solving the inverse problem employing a Kolmogorov Forward Equation. The copula-defined dependency becomes built into the coefficients of the SDE of the price process and the Wiener kernels of the price and volatility processes become independent. This leads to a straightforward recovery of the price of an option contract. We discuss how such an approach could lead to the elimination of the volatility smile bias. Evidence is presented of a direct link between the choice of the copula function and the shape of the volatility smile.

C236: **An approach to optimize credit portfolio**

Presenter: **Zhi Wang**, University of Essex, UK

Co-authors: John Nankervis, Xiaoquan Liu

A new approach to optimize credit portfolio by minimizing default risk is presented. First, we propose a binomial intensity system to model the joint default intensity. In this binomial model, the intensity of one name is linear affected by the other. Also a default risk measure of credit portfolio is defined and a Monte Carlo simulation is implemented to discuss the effectiveness of the measure in representing the portfolio default risk. In term of optimization, we apply Genetic Algorithm method with purpose of solving the difficulties caused by large size and various constraints. Empirically, we apply the approach to optimize CDS portfolio. We examine all memberships in iTraxx and CDX index during the crisis. The sample size is round $56,000$. The estimates of binominal model are studied in detail. Also we mimic small, medium and large portfolios, optimize the portfolios separately and discuss the results. Empirical evidence shows that CDX market has averagely higher spread and default intensities than iTraxx market. Moreover, CDX market is more volatile in both spreads and intensities. However, the pattern of default correlation does not have significant difference in two markets. Optimization results show that diversification would reduce the default risk.

C177: **The Value-at-Risk (VaR) of South East Asian countries: forecasting long memory in extreme value estimators.**

Presenter: **Prasert Chaitip**, Faculty of Economics, Chiang Mai University, Thailand

Co-authors: Arreyah Chaitip, Chukiat Chaiboonsri

Accurate modeling of Value-at-Risk (VaR) is important in finance econometrics, particularly as it relates to the modeling and forecasting of Value-at-Risk (VaR). Recently, the new ideas of forecasting using both Long Memories in VaR and Extreme Value Estimators in VaR were employed to estimate the loss at a predefined confidence level. It proposes a construction to put together the analysis of long-memory (non-stationary) time series fitted by an extreme-value distribution. And the study focuses on these two extraordinary methods to forecast the Value-At-Risk (VaR) of selected South East Asian stock markets consisting of SET index (Thailand), KLSE index (Malaysia), FTSI index (Singapore), and JKSE index (Indonesia)). The results indicated that selected South East Asian stock markets have a higher VaR overtime. The findings suggest that those countries identified as high-risk will encounter financial difficulties in their stock markets over the future year to come. Stock market risk supervision is one of the most indispensable skills for any investor to master.

Friday 10.12.2010 16:55 - 19:00 Parallel Session E – ERCIM

ES22 Room Senate Gordon Room FLEXIBLE PARAMETRIC FAMILIES OF DISTRIBUTIONS Chair: Adelchi Azzalini

E219: **Some alternative parametric families and their interconnections**
Presenter: **Chris Jones**, The Open University, UK

Let $g(x)$ be the density of a symmetric univariate continuous distribution. I will first introduce a novel class of "transformation of scale" densities of the form $2g(t(x))$ by identifying appropriate functions $t(x) = W^{-1}(x)$, and show some of the nice properties of members of the class. t, of course, contains any shape parameters one might wish to introduce. Other, existing, families of distributions will then be introduced via their connections to the transformation of scale densities: two-piece densities, Azzalini-type skew distributions, distributions of transformations of random variables and perhaps more. W, and its essential properties, will be key to the interconnections. The Sinh-Arcsinh function, $\sinh(a + b\sinh^{-1}(x))$, and the closely related t_2 distribution function will play a prominent role in the talk. Despite its apparent arcaneness, this talk will have an element of overview of a number of flexible parametric families of distributions.

E326: **A latent-state generalized additive model for location, scale and shape for multivariate pollutant concentrations**
Presenter: **Antonello Maruotti**, Universita di Roma Tre, Italy
Co-authors: Walter Zucchini

The model developed in this paper was motivated by the analysis of a daily multivariate time series of concentrations of three pollutants at a monitoring station in Rome, Italy. The objective is to integrate pollution and meteorological data in order to identify the important determinants of pollution levels. To represent such series appropriately, it is necessary to characterize the complex relationships between the pollutant series, and between these and the covariates. Some of the distributions are long-tailed; the location, scale and shape of some outcomes can depend on various covariates. We propose a generalized additive model for location, scale and shape (GAMLSS) by introducing time-dependent and outcome-specific random effects in the link functions that relate the distribution parameters to the explanatory variables. The random effects are assumed to be discrete and to follow a first-order finite-state Markov chain. The resulting model is a multivariate latent-state GAMLSS that not only controls for unobserved heterogeneity due to omitted covariates and time dependence, but also, as a by-product, provides a measure for the association structure of the outcomes. Parameter estimation by maximum likelihood can be carried out using an EM-based algorithm in latent-state framework.

E080: **Modelling of multivariate data via mixtures of skew normal and skew-t distributions**
Presenter: **Geoff McLachlan**, University of Queensland, Australia

Mixture distributions are applied to data with two main purposes in mind: (1) to provide an appealing semiparametric framework in which to model unknown distributional shapes, as an alternative to, say, the kernel density method (for example, the set of all normal mixture densities is dense in the set of all density functions under the L1 metric.); (2) to use the mixture model to provide a probabilistic clustering of the data into g clusters. In the latter case, an outright clustering is obtained by assigning a data point to the component to which it has the greatest posterior probability of belonging. In both situations, there is the question of how many components to include in the mixture model. In the typical application of normal mixture models to clustering, the number of clusters corresponds to the number of components in the mixture model. But in cases where the clusters are not elliptically symmetric, this correspondence will not hold if additional components are needed to cover the asymmetry in the data. One way to enable the number of components to correspond to the number of clusters in such situations is to fit mixture models with skew normal components or skew t-components where there might be outliers as well as asymmetry in the data. Examples of this approach are given in flow cytometry where mixture of skew t-densities are adopted to handle asymmetry and outliers in the data.

E536: **Modeling skew-symmetric distributions using B-spline and penalties**
Presenter: **Patrizio Frederic**, University of Modena, Italy

We propose a new flexible procedure for modeling Skewed-Symmetric (SS) distributions via B-splines. To avoid over-fitting we follow a penalized likelihood estimation method. The structure of B-splines SS with penalties provides a flexible and smooth semi-parametric setting allowing estimates that capture many features of the target function such as asymmetry and multimodality. After outlining some theoretical results, we propose an effective computational strategy. Finally, we present some empirical results on both simulated and real data. These exemplify the advantages of the proposed method.

E278: **Some properties of skew-symmetric distributions**
Presenter: **Adelchi Azzalini**, Univ. Padova, Italy
Co-authors: Giuliana Regoli

The family of skew-symmetric distributions is a wide set of probability density functions obtained by combining in a suitable form a few components which are selectable quite freely provided some simple requirements are satisfied. Intense recent work has produced several results for specific sub-families of this construction, but much less is known in general terms. The present paper explores some questions within this framework, and provides conditions on the above-mentioned components to ensure that the final distribution enjoys specific properties.

ES40 Room Senate Bedford Room COMPUTATIONAL ISSUES ON TIME SERIES MODELING (ANSET) Chair: Cira Perna

E618: **Inference on a stochastic two-compartment model in tumor growth**
Presenter: **Giuseppina Albano**, University of Salerno, Italy
Co-authors: Virginia Giorno, Patricia Roman-Roman, Francisco Torres-Ruiz

Inference on the parameters of a previous continuous-time model proposed by some of the authors is discussed. Such a model incorporates several key elements in tumor dynamics. More precisely, the form of proliferating and quiescent cell lines comes out from their relations with the whole tumor population, giving rise to a two-dimensional diffusion process, generally time non-homogeneous. The proposed model is able to incorporate the effects of the mutual interactions between the two subpopulations. Estimation of the rates of the two populations based on some characteristics of the involved diffusion processes is discussed when longitudinal data are available. Some simulation results are finally presented.

E636: **On discriminating between structural breaks and long memory process**
Presenter: **Bianca Di Lorenzo**, Sapienza University, Italy

The problem of discriminating between long-range dependent time series and weakly dependent time series with break points in the mean is addressed. A long memory time series X_t with fractional parameter $d \in (0, 1/2)$, shows an autocovariance function that is not absolutely summable.

This means that, in frequency domain, the spectral density diverges at low frequencies and has an hyperbolic decay. We also often observe that autocorrelation estimate of time series with structural breaks decay slowly. For this reason to distinguish these phenomena is difficult and standard methodology fails. In order to solve this problem, the idea is to apply a linear filter $F(\cdot)$ which reduces the jump at the break point, leaving the memory type. We focus our attention on the spectral density near the origin. This is equivalent to studying the behaviour of the gain function in the neighbourhood of zero after having filtered the process. In particular, the filtered time series behaves differently depending on the structure of the gain function of the filter $F(\cdot)$ near the origin. The purpose of this work is to construct a gain function which permits us to obtain a time series that still exhibits the long memory property.

E635: **Meta-heuristic methods for outliers detection in multivariate time series**

Presenter: **Antonietta di Salvatore**, Sapienza University, Rome, Italy

Co-authors: Domenico Cucina, Mattheos K. Protopapas

Meta-heuristic procedures are used to detect additive outliers in multivariate time series. The implemented algorithms are: simulated annealing, threshold accepting and two different versions of genetic algorithm. All of them use the same objective function, the generalised AIC-like criterion, and in contrast with many of the existing methods, they do not require to specify a vector ARMA model for the data and are able to detect any number of potential outliers simultaneously. A great amount of work has been done to choose suitable values for the AIC penalty constant according to time series length and number of components. The suggested method is based on the extreme value distribution theory. We used simulated time series and real data to evaluate and compare the performance of the proposed methods. We compare also the performance of our meta-heuristic algorithms with the corresponding results coming from detection procedure based on vector ARMA model.

E742: **Bias estimation in local polynomial regression**

Presenter: **Maria Lucia Parrella**, University of Salerno, Italy

The analysis of the mean squared error (MSE) represents a fundamental step in kernel based regression, for many reasons. First of all, it is useful to derive the assumptions for the consistency of this nonparametric estimator. Moreover, its decomposition into bias and variance, together with the stated results on the asymptotic normality of the kernel estimator, allows us to derive the pointwise confidence bands useful for inferential decisions. Last but not least, the evaluation of the MSE is useful in order to select the best value for the bandwidth parameter. The exact expressions of the bias and the variance of kernel-based estimators are well known, both from a finite than from an asymptotic point of view. Anyway, since they depend on some unknown functionals of the underlying process, it is very difficult to get a reliable estimate of the MSE. The main difficulties are encountered when estimating the bias, given its dependence on some derivative function. Our goal here is to present some recent proposals on bias estimation in local polynomial regression. We compare the finite sample behaviour and the computational burden of the alternative estimators through a simulation study.

E677: **Identification of nonlinear time series based on kernel predictors**

Presenter: **Cira Perna**, University of Salerno, Italy

Co-authors: Francesco Giordano, Marcella Niglio

The identification of time series models is not an easy task that becomes more difficult when we go from the linear to the nonlinear domain. In this contribution we investigate the possibility of identifying nonlinear time series models, belonging to the Markov class, using nonparametric kernel predictors. To evaluate the performance of the proposed approach we have implemented a Monte Carlo simulation study where the identification of some widely known nonlinear time series models are investigated and compared. An application to economic time series is given to discuss the proposed identification procedure in the empirical domain.

ES72 Room Senate Beveridge Hall ROBUSTNESS FOR COMPLEX MODELS Chair: Stefan Van Aelst

E761: **Outlier detection in Geostatistics**

Presenter: **Sonja Kuhnt**, TU Dortmund University, Germany

Co-authors: Alexander Ullmann, Thomas Muehlenstaedt, Juergen Tiedge

A standard technique to observe geostatistical data is Airborne Laser Scanning. By airplane the ground surface is scanned based on high frequency laser impulses. A basic problem of this method is that the laser beam often cannot reach the ground due to trees or bushes. Reconstructing the ground surface ignoring outlier contamination yields unacceptable modeling errors. Therefore, we construct outlier detection rules based on robust regression techniques like repeated median and trimmed least squares. Furthermore, outliers mainly occur in clusters and it happens that in a sub region more than half of the observations are outliers. In order to deal with this kind of problem, special window techniques are developed. Then a classical modeling technique for geostatistical data like Kriging can be used. An application of the methodologies to a real world data set is presented, comparing Kriging models for contaminated and outlier-free data sets.

E701: **S-estimation of mixtures**

Presenter: **Luca Greco**, University of Sannio - Benevento, Italy

Co-authors: Alessio Farcomeni

A modification of the EM algorithm is proposed, which is aimed at performing robust estimation of mixtures of multivariate distributions in the elliptically symmetric family. The M step is replaced by S-estimation of location and scatter. In particular, the bisquare multivariate S-estimator is considered. The estimates are computed by solving a system of estimating equations that are characterized by component-specific sets of weights, based on squared Mahalanobis distances. The main properties and the good behavior of the method are investigated and illustrated by some numerical studies and applications.

E357: **Depth-based runs tests for bivariate central symmetry**

Presenter: **Christophe Ley**, Universite Libre de Bruxelles, Belgium

Co-authors: Davy Paindaveine

McWilliams introduced a simple nonparametric procedure based on runs for the problem of testing univariate symmetry about a specified center m. His procedure first orders the observations according to their distances from m, then replaces each observation with its sign (with respect to m), and finally rejects the null for small values of the number of runs in the resulting series. This test has been shown to be universally consistent and enjoys nice robustness properties, but is unfortunately limited to the univariate setup. In this paper, we extend McWilliams' test into a bivariate test of central symmetry, still about a specified center. We first order the observations according to a certain concept of depth (e.g. the halfspace depth or the simplicial depth), then replace each observation with its spatial sign, and finally reject the null when an original concept of runs related to the simplicial depth is too small. The resulting runs test is affine-invariant, distribution-free and robust. We derive its exact and limiting distribution under the null hypothesis and study its performances through Monte Carlo simulations. We conclude by a short generalization of these bivariate

runs to any dimension.

E571: **On the robustness of change detection algorithms for data streams**
Presenter: **Tamraparni Dasu**, AT&T Labs Research, USA
Co-authors: Shankar Krishnan, Gina-Maria Pomann

Distributional change detection in data streams is an important data mining problem with applications to sensor networks, internet data traffic and financial ticker data. Known change detection (CD) algorithms produce a binary output of "Change" or "No Change" but shed no light on the properties of an algorithm such as robustness, e.g. in terms of stability with respect to small perturbations in the distribution ("holes" in the data, outliers, location and scale shifts). Current evaluative frameworks benchmark through experiments on data sets. In this paper, we propose a statistically rigorous framework based on an objective performance measure, *streaming power*. Since streaming power is a probability, it represents a measure that is comparable for disparate algorithms. We demonstrate the framework using major statistical change detection algorithms known in literature and real world applications. We also provide an extensive suite of experiments. The multidimensional test streams consist of real and synthetic data stream segments that can be injected with different types of changes and combined in a modular fashion to form streams of desired length, dimensionality and variability. As a service to the community, we will make our repository of test data streams publicly available.

E369: **Instrumental weighted variables under heteroscedasticity and with constraints**
Presenter: **Jan Amos Visek**, Charles University in Prague, Czech Republic

Situations where the explanatory variables are correlated with the disturbances in linear regression models are not rare, especially in social sciences. In this case, the ordinary least squares may be biased and inconsistent. The instrumental variables are one of the ways to cope with this situation. The dependence of regressors and disturbances may (likely) imply heteroscedasticity. Moreover if data are contaminated, it should be taken into account. That is why the instrumental weighted variables were defined. The weights should reduce the influence of the points with contamination. Of course, these points are unknown. Hence, it seems quite natural to let the method to decide itself which residuals are to be weighted down. Thus, the weights are assigned to the order statistics of the squared residuals rather than directly to the squared residuals. This contribution will show why it may be sometimes profitable to impose some constraints on the estimates of the regression coefficients (e.g. as a remedy for collinearity, improvement of combining the forecasts). Then the $\sqrt{n}$-consistency of the estimator is proved under heteroscedasticity imposing general constraints on the estimates.

ES63 Room Senate Bloomsbury Room BAYESIAN NONPARAMETRICS: MODELS AND APPLICATIONS Chair: Antonio Lijoi

E439: **A Bayesian model of NMR spectra for the deconvolution and quantification of metabolites in complex biological mixtures**
Presenter: **Maria De Iorio**, Imperial College London, UK
Co-authors: William Astle, Sylvia Richardson, Timothy Ebbels, David A. Stephens

Nuclear Magnetic Resonance (NMR) spectroscopy is a technique for obtaining structural and quantitative information about molecules. We present a model for proton NMR spectra, which allows us to identify and quantify the metabolic compounds present in complex mixtures such as biofluids. Resonance signals combine additively so our modelling is based on a Bayesian linear regression. We assume the spectral data are generated with Gaussian error from $\sum_j w_j S_j(\delta) + \eta(\delta)$, sampled at regular intervals of the chemical shift δ (a parameter proportional to the frequency of exposure radiation). Here, S_j is a template function determined by the chemistry of the j^{th} metabolite. We aim to deconvolve spectra into "metabolite" components, which allows us to make inference about the concentration parameters w_j. We take advantage of prior information about the shape of these templates for selected metabolites, which is available experimentally or from online databases. To complement our model for the spectral signal generated by compounds with well characterised NMR signatures we model the residual signal by a flexible nonparametric component η. We aim to use the w_j and η for supervised and unsupervised classification of individuals by their biofluid metabolite profile.

E538: **False discovery rates in somatic mutation studies of cancer**
Presenter: **Lorenzo Trippa**, Dana–Farber Cancer Institute and Harvard University, USA
Co-authors: Giovanni Parmigiani

The purpose of cancer genome sequencing studies is to determine the nature and types of alterations present in a typical cancer and to discover genes mutated at high frequencies. We discuss statistical methods for the analysis of somatic mutation frequency data generated in these studies. In this context, we describe and compare statistical methods for constructing scores that can be used to prioritize candidate genes for further investigation and to assess the statistical significance of the candidates thus identified. Controversy has surrounded the reliability of the false discovery rates estimates provided by the approximations used in early cancer genome studies. To address these we develop a semi-parametric Bayesian model that provides an accurate fit to the data. We use this model to generate a large collection of realistic scenarios, and evaluate alternative approaches on this collection. Our assessment is impartial in that the model used for generating data is not used by any of the approaches compared. And is objective, in that the scenarios are generated by a model that fits data. Our results quantify the conservative control of the false discovery rate with the Benjamini and Hochberg method compared to the Empirical Bayes approach and shows a negligible departure from the target false discovery rate of the methodology used in recent cancer studies.

E089: **Simplex factor models for multivariate unordered categorical data**
Presenter: **Anirban Bhattacharya**, Duke University, USA
Co-authors: David Dunson

Gaussian latent factor models are routinely used for modeling of dependence in continuous, binary and ordered categorical data. For unordered categorical variables, Gaussian latent factor models lead to challenging computation and overly complex modeling structures. As an alternative that is more natural for unordered categorical data, we propose a novel class of simplex factor models. In the single factor case, the model treats the different categorical outcomes as independent with unknown marginals. The model can characterize highly flexible dependence structures parsimoniously with few factors, and as factors are added, any multivariate categorical data distribution can be accurately approximated. Using a Bayesian approach for computation and inferences, a highly efficient Gibbs sampler is proposed that scales well with increasing dimension, with an adaptive Gibbs step enabling selection of the number of factors. Theoretical properties are described and we evaluate the approach through simulation examples. Applications are described for modeling dependence in nucleotide sequences and prediction from high-dimensional categorical features.

E392: **Long-range dependent Dirichlet processes with canonical correlation kernels**
Presenter: **Dario Spano**, University of Warwick, UK

We will show a simple construction of a family of long-range dependent measure-valued processes with Ferguson-Dirichlet stationary distribution. The evolution of such processes is governed by a kernel with orthogonal polynomial eigenfunctions. The eigenvalues are Mittag-Leffler functions

and, as such, they give rise to subdiffusive sample paths. Dual processes and sampling distributions are easily derived. This class is a generalization of a well-known family of Fleming-Viot processes from Population Genetics.

E076: **Bayesian nonparametric modeling for dynamic interacting systems**
Presenter: **Matteo Ruggiero**, University of Pavia, Italy
Co-authors: Igor Pruenster

We review some recent contributions to Bayesian nonparametric modeling for the representation of dynamical stochastic interacting systems, with applications to population genetics and economics. These are based on nonparametric hierarchical models which produce generalised Blackwell-MacQueen Polya urns and allow to induce the dependence in both the static and the dynamic framework. The joint use of Markov chain Monte Carlo principles then leads to the construction of a system of particles which interact over time, providing meaningful representation of phenomena such as competition among idiosyncratic economic agents or species coexistence within and between different sites.

EP02 Room Senate Crush Hall POSTERS II Chair: Klea Panayidou

E643: **A data-based power transformation for compositional data**
Presenter: **Michail Tsagris**, University of Nottingham, UK
Co-authors: Simon Preston, Andrew T.A. Wood

Compositional data is a special type of multivariate data in which the elements of each observation vector are non-negative and sum to one. Data of this type arise in many areas, such as geology, archaeology, biology amongst others. Compositional data analysis is carried out either by neglecting the compositional constraint and applying standard multivariate data analysis, or by transforming the data using the logs of the ratios of the components. In this work we introduce a more general transformation which includes both approaches as special cases. It is a Box-Cox type transformation and involves a single parameter, α. We discuss various methods to estimate the value of the parameter, such as profile likelihood, linear regression and discriminant analysis.

E667: **Probabilistic record linkage with EM algorithm and the R system**
Presenter: **Rafael Pino**, University of Seville, Spain
Co-authors: Maria-Dolores Cubiles de la Vega, Elisa Caballero-Ruiz, Fancisco Garcia-Gonzalez

EM algorithm is a traditional tool for probabilistic record linkage. We present a collection of R programs for performing this task. For each pair of records, a vector of binary (0-1) comparison values is supposed available, where 1 means that both values agree, and 0 that they are completely different. Our programs implement the steps of the EM algorithm, and display numerical and graphical information to help the user in the decision about the threshold values to be used in the final decision. Some real examples are provided, according to our experience in a project of the Andalousian Statistical Institute. Other comparison criteria and their computational treatment are being considered.

E779: **Global approach for evaluating the quality of clustering results**
Presenter: **Osvaldo Silva**, Azores University, Portugal
Co-authors: Helena Bacelar-Nicolau, Fernando Nicolau

The discovery of knowledge is a complex and subjective process and in the particular case of Hierarchical Cluster Analysis (HCA) depends on many factors, such as the clustering algorithms applied and the strategies developed in the initial stage in Cluster Analysis, which can lead to the attainment of different clustering results. In the present work, we suggest a global approach for evaluating the quality of clustering results and a performance comparison among different clustering algorithms using all information available, namely based on stability, isolation and homogeneity indices of the clusters, among others. In addition, we present a visual method to facilitate the evaluation of the quality of the partitions, being allowed to enhance the similarities and the differences between the partitions, as well as the behaviour of the elements in the partitions. We illustrate our approach using real horse data referring to complex and heterogeneous data units (symbolic data). We applied HCA based on the generalized affinity coefficient (similarity coefficient) combined with 26 clustering algorithms (classic and probabilistic). Finally, we discuss the obtained results and the contribution of this approach how to know better the cluster structure of a data set.

E756: **Fitting Weibull distribution on wind speed data series**
Presenter: **Evangelos Akylas**, Cyprus University of Technology, Cyprus
Co-authors: Panagiotis Zavros, Dimitrios Skarlatos, Marios Fyrillas

This study is concerned with the two-parameter Weibull distribution which is widely employed as a model for wind speed data. The the maximum likelihood method for the estimation of the Weibull parameters was proved by numerous studies to be a more accurate technique than the commonly used log-linear regression. However it depends on detailed original information and demands more computational time. We present a modified technique which is based on the estimation of higher moments of the Weibull distribution. The technique demands only basic averaged statistical information, it is computationally cheap and renders practically equivalent results with the maximum likelihood method.

E858: **Evaluation of the conflict between sensors in fusion data**
Presenter: **Arnaud Roquel**, Paris sud, France
Co-authors: Sylvie Le-Hegarat-Mascle, Isabelle Bloch

Multi-sensor systems are used in various domains such as assistance to the person or shape recognition. The aim is to collect relevant and reliable information on the observed phenomenon, using physical or logical sources. Incompleteness, imprecision and uncertainty of this information may lead to conflicting situations in fusion procedures. In this paper, we suggest to use a set of measurements to detect the conflicting states, modeled in the belief function theory. We have considered several distances between bbas (basic belief assignments) proposed in the literature, including the classical measure of conflict, equal to the empty set mass. We show that the conflict can be better analyzed and understood by not considering the sources globally but distinguishing between the different hypotheses (more precisely each hypothesis is labeled according to its mass value after fusion). This has been assessed using numerical simulations with randomly chosen bba values, and by studying the distribution of the samples in 3D space (conflict evaluated by $m_{[1+2]}(\emptyset)$, distance between bbas evaluated for H hypothesis by $\mid q_{[1]}(H) - q_{[2]}(H) \mid$, fusion bba evaluated by $q_{[1+2]}(H)$). We apply such a conflict detection method to real data acquired by a system of prevention of motorcycle fall off.

E559: **Modelling environmental data incorporating a hydrometeorological factor and latent variables in dynamic models**
Presenter: **Marco Costa**, Universidade de Aveiro, Portugal
Co-authors: A. Manuela Goncalves

Water quality monitoring is an important tool in the management and assessment of surface water quality. This study focuses on a rather extended

data set relative to the River Ave basin (Portugal) and consists mainly of monthly measurements of biochemical variables in a network of monitoring water quality stations. A hydrometeorological factor is constructed for each monitoring station based on monthly estimates of precipitation obtained by means of a rain gauge network. For each water quality-monitoring site is identified a region of influence that takes into account the region's topography and the land's drainage dynamics. Through stochastic and deterministic methods of interpolation (Kriging / Polygons of Thiessen) it is estimated the mean area rainfall during each month in the area of influence of each water quality monitoring site. These estimates are based on rain gauges located in the respective area of influence. The hydrometeorological factor is incorporated as a covariate in multivariate state space models fitted to homogeneous groups of monitoring sites. Additionally, in the modelling process it is considered a latent variable that allows incorporating a structural component, such as seasonality, in a dynamic way.

E581: **Small area estimation: Application to the estimation of the proportion of people affected by dengue**

Presenter: **Silvia Gonzalez**, Universidad de Granada, Spain

Co-authors: Agustin Santiago, Maria del Mar Rueda, Antonio Arcos, Encarnacion Alvarez

The estimation of a proportion in the context of small areas is discussed assuming that the available auxiliary information corresponds to qualitative variables. We propose design-based estimators of a proportion based upon the ratio, difference and calibration methods, which assume that the proportions of the auxiliary variables are known at the population level or the domain level. The proposed estimators are used to estimate the proportion of being affected by dengue (classic) and dengue hemorrhagic fever by considering data extracted from a study in the state of Guerrero (Mexico). The behavior of proposed estimators is compared to alternative model-based estimators in this context.

E774: **Generation of multivariate discrete data**

Presenter: **Alessandro Barbiero**, Universita degli Studi di Milano, Italy

Co-authors: Pier Alda Ferrari

Over the recent years, great interest has been addressed to ordinal variables and to the development of multivariate statistical techniques for their analysis. The empirical comparison of such techniques, either exploratory or inferential, often requires simulating ordinal data under different experimental conditions. Several methods for generating multidimensional data from continuous or discrete variables have been proposed. This paper focuses on an algorithm for generating ordinal data with assigned marginal distributions and correlation matrix. The procedure consists of two steps: the first one aims at setting up the desired experimental conditions, employing a straightforward discretization procedure from a standard multinormal variable, whose correlation matrix is computed through an iterative algorithm in order to achieve the target correlation matrix for ordinal data. The second step actually implements the sampling under the experimental conditions and allows performing a Monte Carlo simulation study. The algorithm does not suffer from some drawbacks encountered by other existing techniques and has a large application. It is implemented in R through a function which allows the user to choose the sample of size, the number of variables, their distribution and the target correlation matrix which is also checked for its feasibility. Two examples of application are provided.

E783: **Distribution of the affinity coefficient between variables based on the Monte Carlo simulation method**

Presenter: **Aurea Sousa**, Azores University, Portugal

Co-authors: Osvaldo Silva, Helena Bacelar-Nicolau, Fernando Nicolau

The affinity coefficient and its extensions have been used both in hierarchical and non-hierarchical Cluster Analysis. The asymptotic distribution of the affinity coefficient between variables under different assumptions of reference has already been studied and a probabilistic approach to classification based on this coefficient has been introduced. We present some results concerning an empirical study about the asymptotic distribution of the basic/generalized affinity coefficient, with or without standardization by the W method of Wald and Wolfowitz, and with or without global standardization. This study is based on the Monte Carlo simulation method and includes the comparison between those coefficients and the Pearson's correlation coefficient. We continue an approach already developed on the asymptotic behaviour of the coefficient of affinity. Even then it was crucial to invest in knowledge of the exact or asymptotic distributions of these coefficients, in order to approach the decision-process in classification and the classical methods of statistical decision. We present the main conclusions that summarize the asymptotic result in the case where the variables follow distribution models set a priori.

E807: **Optimal sampling strategies for populations with a general type of correlation**

Presenter: **Ioulia Papageorgiou**, Athens University of Economics and Business, Greece

The presence of correlation amongst population units is very common in practice and consequently sampling from such populations is an important problem. In this talk we consider the problem of deriving the best sampling strategy for populations with correlated units. The strategy consists of a sampling design and an estimator with optimal properties. Deriving the best sampling strategy requires to simultaneously considering the problems of producing estimates and proposing the best sampling procedure. For a general form of correlation structure, we formulate the problem and derive an objective function that we need to optimize in order to obtain the design or family of designs that lead to the best unbiased estimator. The complexity and high dimensionality of the optimization problem generate a number of practical difficulties. We attempt to overcome these by proposing two different approaches. One is numerical and is based on a continuous extension of the autocorrelation function of the population and the second is making use of splines for the same purpose. The work is illustrated with real and simulated data that demonstrate the optimality of the approach compared to some classical sampling designs.

Saturday 11.12.2010 08:40 - 10:45 Parallel Session F – CFE

CS11 Room MAL B20 FINANCIAL TIME SERIES MODELLING AND FORECASTING I **Chair: Alessandra Amendola**

C255: **A non-parametric procedure for combining high dimensional multivariate volatility forecasts**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Alessandra Amendola

A computationally efficient procedure for the combination of high dimensional multivariate volatility forecasts is proposed. The problem of combining multivariate volatility forecasts had already been addressed. It considered the estimation of a high number of bivariate systems in order to avoid the curse of dimensionality. However, for large dimensional portfolios, which in risk management represent the rule rather than exception, this approach is still highly demanding in terms of computing times. In order to specifically deal with these situations, we propose a novel approach to the estimation of combination weights based on the Composite GMM. This is a variant of the GMM method specifically tailored for very large dimensional estimation problems. Finally, we empirically compare the performances of different combination techniques by means of a Monte Carlo simulation study and an application to real data.

C543: **Extreme events in financial time series: a heuristic procedure for multivariate asset selection**
Presenter: **Paola Zuccolotto**, University of Brescia, Italy
Co-authors: Giovanni De Luca

The occurrence of abnormal market conditions can be particularly bleeding for investors. When a financial crisis period is expected, investments can be protected from extreme events by choosing assets with low association in the lower tails of their distributions. From a statistical point of view, this requires to deal with multivariate probability laws. However, highly accurate models could involve the estimation of a high number of parameters, which in turn leads to less robust and less reliable results. Given a set of p financial assets, this paper proposes a heuristic strategy for selecting K subsets of k assets with a low risk of the joint occurrence of extreme events. The selection is carried out through an algorithmic procedure based on Random Forest (RF), which generalizes a formerly proposed algorithm which was sensitive to the initial step. Among the K possible subsets of k assets, we propose to choose the one minimizing a measure of the relationship among extremely low values given by a multivariate tail dependence coefficient, which can be computed after the estimation of an appropriate multivariate copula function.

C544: **On the relation between firm characteristics and volatility dynamics with an application to the 2007-2009 financial crisis**
Presenter: **Christian Brownlees**, NYU, USA

Despite the extensive literature on the analysis of firm equity volatility, relatively little is known about the relation between firm characteristics and volatility dynamics. This is partly due to the lack of an appropriate modelling framework in which these questions can be addressed adequately. This work proposes a Hierarchical Factor GARCH model for multivariate volatility analysis in large panels of assets. The novelty consists of augmenting the specification with equations linking the volatility dynamics parameters of each firm to observed and unobserved characteristics. The hierarchical approach has features that are useful in economic and forecasting applications. It permits one to investigate how variation of firm variables explains variation in volatility dynamics. Moreover, it allows for a parsimonious parameterization of a multivariate system that is independent of its dimension, yet capable of retaining flexibility in the individual series dynamics thanks to the random effect structure. The model is estimated via Maximum Likelihood. The methodology is used to analyse the volatility dynamics of top U.S. financial institutions during the crisis. Dynamics are a function of firm size, leverage, distance to default and liquidity. Results show that leverage is the most influential variable, and firms with high leverage have high exposure to systematic shocks.

C293: **Disentangling memory from cycles**
Presenter: **Walter Distaso**, Imperial College Business School, UK

When estimating the extent of persistence, two other important types of time-influences should be taken into account: time trends (including possible occasional breaks in the trends) and the frequency of the cycles. Work on tackling these issues solve the technically difficult problem of estimating ω and d, when $d \in (0, \frac{1}{2})$. We provide a method to jointly estimate and test hypotheses about: (a) the location of the spectral singularity (frequency of the cycle); (b) the persistence parameter d that will be allowed to have a wider domain than $(0, \frac{1}{2})$; and (c) deterministic time trends that may be present in the series. We do this in a semiparametric framework by extending local Whittle estimators and allowing for nonlinear and nonnormal processes. A stepwise correction factor to the Whittle objective function will be needed for $d > 1/2$. We stress that omitting time trends or cycles can lead to erroneous inferences on the degree of persistence, hence our desire to add both of these components to the modelling exercise. The estimate of d in I(d) is inconsistent if the location ω of the spectral singularity is incorrectly forced to be zero, as is the case with I(d) models.

C686: **Beatlestrap**
Presenter: **Alessandro Palandri**, University of Warwick, UK

The Bootstrap of test statistics requires the re-estimation of the model's parameters for each bootstrap sample. When parameter estimates are not available in closed form, this procedure becomes computationally demanding as each replication requires the numerical optimization of an objective function. This paper investigates the feasibility of the Beatlestrap, an optimization-free approach to bootstrap. It is shown that, ex-post, M-estimators may be expressed in terms of simple arithmetic averages therefore reducing the bootstrap of Wald statistics to the bootstrap of averages. Similarly, it is shown how the Lagrange Multiplier and the Likelihood Ratio statistics may be bootstrapped bypassing the objective function's multiple optimizations. The proposed approach is extended to simulation based Indirect Estimators. The finite sample properties of Beatlestrap are investigated via Monte Carlo simulations.

CS49 Room MAL B30 BAYESIAN ECONOMETRICS AND APPLICATIONS I: APPLICATION IN FINANCE **Chair: Teruo Nakatsuma**

C155: **Contingent claims valuation for low frequency data with Knightian uncertainty**
Presenter: **Radu Tunaru**, University of Kent at Canterbury, UK

Pricing contingent claims for low frequency data such as real-estate prices presents some difficult challenges, mainly related to slow trading and to Knightian uncertainty associated with parameters inference, the latter being very important for model risk. A viable framework to solve these important problems is described in this paper, focusing on using the proper bridge process to augment the unobserved data. Closed formulae are derived when the transition densities of the underlying diffusion process are known analytically. The calibration process is supported by MCMC techniques while the contingent claims pricing procedure is also performed under a Bayesian paradigm. In the paper we also present some applications involving the forwards, swaps and European options contingent on IPD commercial real-estate index. For this application, a mean-reverting

model with predictable trend in the drift is considered and it is shown that the posterior distribution of main parameters of interest is a normal inverse gamma distribution.

C343: **Extensions of the Lee-Carter model for mortality projections**
Presenter: **Udi Makov**, University of Haifa, Israel
Co-authors: Shaul Bar-Lev, Yaser Awad

The literature on mortality projections was 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. The paper will discuss potential extensions to the Lee-Carter model along the following lines: a) Presentation of the L-C model as a state-space model. b) Bayesian Implementation of the L-C model with a broad family of embedded ARIMA models. c) Adaptation of the L-C model for simultaneous projections of several populations.

C527: **A model of credit retail premia**
Presenter: **Jan Bruha**, Czech National Bank, Czech Republic

A model of credit retail premium is proposed. A theoretical model of pass through from the money market to retail rates is formulated. The theoretical model is then converted to a state space form and estimated on data of the Czech Republic. The state space form can be used to filter unobserved states (the price of objective risk of loans), to forecast the premium, to do shock decomposition, and to run conditional forecasts. It is shown that the observed credit retail premia can be explained by the cyclical position of the economy. The model has a success in explaining the mortgages rates, and its performance is relatively worst for large firm loans. Possible reasons behind this pattern are discussed.

C403: **Screening massive numbers of funds: Parallel computing and Bayesian methods in finance**
Presenter: **Kenichiro McAlinn**, Keio University Graduate School of Economics, Japan
Co-authors: Teruo Nakatsuma

The Achilles' heel of Bayesian statistics in terms of its general acceptance among professionals in finance is the time it takes for computation. Ever since the introduction of Bayesian statistics into finance, numerous Bayesian methods that utilize subjective views of investors have been proposed and not a few of them were proved to be successful in solving portfolio selection or other problems. Despite mounting affirmative evidence for Bayesian methods, investors are still reluctant to use them in practice because of their time consuming nature. With the advent of fast and inexpensive devices for massively parallel computing, however, formerly impractical computations that take hours or days can be done in minutes or even seconds. In this paper, we compare several computationally intensive Bayesian methods for screening thousands of funds for portfolio selection, and investigate which method is superior in terms of computational time and portfolio performance.

C621: **Bayesian estimation of probability of informed trading**
Presenter: **Kosuke Oya**, Osaka University, Japan

We consider a Bayesian estimation of probability of informed trading and the related model selection problem. The original model for estimation of informed trading probability is based on sequential trading model. If a private information event occurs on particular day and informed trader receives a private signal which is positive, buy order flow for that day arrives according to a Poisson distribution with the intensity which is greater than the intensity when there is no positive private information. A similar argument is established for the sell order flows. The model is represented as the mixtures of Poisson distributions with different intensities. However the model cannot match the positive correlation between buyer and seller initiated order flow that is observed in actual market. Although the extended model has been proposed to remedy the problem, it is not easy to estimate the proposed model using the likelihood method. Further, the regularity condition for the asymptotic theory of the test statistic to select the model with different parameter restrictions is not satisfied for the extended model in some situations. We apply MCMC method to estimate the model and propose the model selection procedure with marginal likelihood to elude the testing problem.

CS82 Room MAL 355 MICROECONOMETRICS Chair: Ana-Maria Fuertes

C117: **A new class of conditional mean tests for binary regression models**
Presenter: **Jose Murteira**, Faculdade de Economia, Universidade de Coimbra, Portugal
Co-authors: Esmeralda Ramalho, Joaquim Ramalho

Typically, the regression analysis of binary and fractional responses requires specification of the conditional expectation of the dependent variable. The recent econometric literature proposes several Lagrange multiplier (LM) tests of conditional mean assumptions in binary and fractional models, that include traditional approaches (RESET, goodness-of-link and non-nested tests) as well as "goodness-of-functional form" (GOFF) tests. GOFF tests have been found to possess very good size and power properties in various situations. However, in some cases, the GOFF tests exhibit very low power when compared with other tests. We propose a new class of LM tests that include GOFF tests as special cases. These new tests are based on a model that generalizes common specifications for binary (and fractional) regression models to accommodate several symmetric and asymmetric shapes and, hence, are potentially powerful against multiple departures from the postulated model. As a nuisance parameter in the generalized model is not identified under the null, and as the power of the tests is sensitive to the value of this parameter, use of the supreme or the average of the corresponding LM statistics is investigated. The asymptotic null distribution of all test versions is derived, and use of the bootstrap to approximate their null distributions is discussed. A very promising finite sample performance is found for the bootstrap-based supremum LM statistic in a Monte Carlo study.

C492: **Measuring returns to education: Bayesian analysis using weak or invalid instrumental variables**
Presenter: **Nalan Basturk**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide, Herman K. van Dijk

A simple regression of earned income on years of education in order to measure the education-income effect may suffer from endogeneity problems, for example as a result of unobserved individual capabilities. The typical treatment of such issues is the use of Instrumental Variable (IV) models, where instruments are used to make inference about the effect of the endogenous explanatory variable. Several issues in IV models, such as weak instruments, endogenous instruments and heterogenous effects of regressors are addressed. We define alternative models accounting for these issues, and assess the extent to which these models are suitable to the data. For assessing model performance, we rely on Bayesian methods, as they provide general probabilistic tools to explicitly account for parameter and model uncertainty. We propose a predictive likelihood approach instead of the conventional marginal likelihood approach, where the former allows us to refrain from imposing strong prior information. We apply the proposed method to simulated data with differing degrees of endogeneity and instrument strength, and two datasets on the income-education relationship. We show that this method can be used to weight the evidence of different models and to assess some of the highly debated issues in IV models.

C154: **Fractional regression models for second stage DEA efficiency analyses**

Presenter: **Esmeralda Ramalho**, Universidade de Evora, Portugal

Co-authors: Joaquim Ramalho, Pedro Henriques

Data envelopment analysis (DEA) is commonly used to measure the relative efficiency of decision making units. Often, in a second stage, a regression model is estimated to relate DEA efficiency scores and exogenous factors. We argue that the traditional tobit approach for second stage DEA analysis does not define a reasonable data generating process for DEA scores and discuss various alternative models that may be more useful to deal with the fractional nature of DEA scores. Under the assumption that DEA scores may be treated as descriptive measures of the relative performance of units in the sample, we consider various alternative regression models appropriate to deal with fractional dependent variables, propose some generalizations to these models, and, given that DEA scores take the value 1 with a non-zero probability, examine the use of two-part models in this framework. Several tests suitable for assessing the specification of each alternative model are also discussed.

C810: **Algorithmic trading with human agents and computer agents in a virtual stock market & virtual futures market**

Presenter: **Daniel Paraschiv**, National University of Ireland, Galway, Ireland

Co-authors: Srinivas Raghavendra, Laurentiu Vasiliu

Algorithmic trading in financial markets has grown rapidly over the past few years. It has been estimated that algorithmic trading constitute about a third to forty percent of all US and EU stock trades. More remarkably during the depths of the current recession high-frequency trading firms account for nearly 73 percent of all US equity trading volume. With both futures and options markets entering into algorithmic trading, it is pertinent to study the impact of such a trading strategy on the overall stability of the market. With this objective, we present a virtual market where both human and computer agents trade Stocks, one Exchange Traded Fund (ETF) and Futures on Stocks and ETF. Computer agents (algo traders) trade using rules based on technical and fundamental indicators. Whereas human agents trade using both fundamental and technical indicators like moving average convergence divergence (MACD) or relative strength index, and they were also allowed to build their own technical indicators using the historical prices and volumes. We present the evidence of some of the stylized facts of the real financial markets using econometric and statistical analysis of both the daily and intra-day data of the virtual stock market.

C318: **Value at Risk from probability forecasts**

Presenter: **Robert Sollis**, Newcastle University Business School, UK

An alternative approach to computing the popular financial risk measure Value at Risk (VaR) is proposed. It is based on indirectly forecasting the relevant portfolio return quantile using probability forecasts computed from binary response models. Compared to directly forecasting the relevant quantile using quantile regression this approach has the practical advantage that maximum likelihood techniques can be used for parameter estimation and inference. An empirical application to US stock market data finds that this alternative approach offers similar performance to quantile regression and is preferred to the orthodox RiskMetrics and historical simulation approaches to computing VaR.

CS52 Room MAL G16 QUANTITATIVE RISK MANAGEMENT I Chair: Marc Paolella

C093: **Multivariate asset return prediction with mixture models**

Presenter: **Marc Paolella**, University of Zurich, Switzerland

The use of mixture distributions for modeling asset returns has a long history in finance. New methods of demonstrating evidence for their necessity in the multivariate case is provided. The use of a two-component multivariate normal mixture distribution, coupled with shrinkage via a quasi-Bayesian prior, is motivated, and shown to be numerically trivial and reliable to estimate, unlike the majority of multivariate GARCH models in existence. Equally important, it provides a clear improvement over use of GARCH models feasible for use with a large number of assets, such as CCC, DCC, and their extensions, with respect to out-of-sample density forecasting. A generalization to a mixture of multivariate Laplace distributions is motivated via univariate and multivariate analysis of the data, and an EM-algorithm is developed for its estimation in conjunction with a quasi-Bayesian prior. It is shown to deliver significantly better forecasts than the mixed normal, with fast and numerically reliable estimation. Crucially, the distribution theory required for portfolio theory and risk assessment is developed.

C095: **Mixture dynamic conditional correlation model**

Presenter: **Maria Putintseva**, University of Zurich, Switzerland

Prediction of the mean and covariance structure of a set of assets is the key aspect of portfolio optimization, hedging, risk management and derivative pricing. Multivariate dynamic conditional correlation model of Engle extrapolates the ideas of GARCH-type univariate modeling to the study of the dynamics of correlation matrices. In working with this model, we found out that as the size of the sample decreases, the quality of the quasi-maximum likelihood estimator worsens significantly. The potential explanation for this observation is the fact that, in relatively small samples, discrepancies between actual and simplified quasi likelihood functions may not always be neglected. Therefore, we suggest replacing the assumption of multivariate normality by a finite mixture of multivariate normal distributions. This kind of distributions is able to capture many important characteristics of the financial data, including long memory and fatter tails of the returns. The number of model parameters remains small in comparison with many other multivariate models used for the same purposes, and the combination of EM-algorithm and two-step likelihood maximization procedure allows to estimate the parameters reasonably fast. Out-of-sample testing of predictability shows that the new model outperforms Engle's dynamic conditional correlation model on both large and small samples.

C100: **Stable mixture GARCH models**

Presenter: **Jochen Krause**, University of Zurich, Switzerland

Co-authors: Simon Broda, Markus Haas, Marc Paolella, Sven Steude

A new model class for univariate asset returns is proposed. It nests the unconditional stable Paretian, unconditional mixtures of normals, normal-GARCH, stable-GARCH, and mixed-normal GARCH models. To resolve the well-known difficulties associated with maximum likelihood estimation of mixture models, a new estimator is devised which augments the likelihood and circumvents the impedimental singularities in the likelihood function. An extensive out-of-sample risk forecasting exercise for six major FX and equity indices confirms the superiority of the general model compared to its special cases and other competitors. The results provide evidence that the stable Paretian assumption, in a multi-component setting coupled with a rich GARCH-type structure, is a tenable assumption for modeling asset returns. The model is extended to the multivariate case by using an independent component analysis framework, which is operational for small and medium-sized (in the hundreds) numbers of assets. In the context of this multivariate extension, the portfolio selection problem using expected shortfall as the downside risk measure can be solved efficiently.

C291: **Modelling asset conditional correlations during the recent financial crisis**
Presenter: **Isabel Casas**, Aarhus University, Denmark
Co-authors: Nektarios Aslanidis

An accurate assessment of the degree of co-movement between asset returns is of interest for a number of reasons. From the point of view of investors, the optimal design of a well-diversified trading portfolio depends on a proper understanding of asset correlations. Changes in co-movement patterns call for an adjustment of portfolios. Policy makers are also interested in these links because of their implications for systemic risk. This article aims at estimating the conditional correlations of two portfolios: a) a portfolio consisting of nine equity sectors SPDRs and the S&P 500 composite during 2004-2010; and b) a portfolio consisting of seven exchange rates. We compare the DCC estimator with a semi-parametric and a non-parametric estimator.

C737: **Phase classification by support vector machine**
Presenter: **Teruko Takada**, Osaka City University, Japan
Co-authors: Takahiro Kitajima

Forecasting phase transitions is difficult for conventional time series approaches, and the basis of the forecast is criticized as black box. The aim of this article is to solve these problems by using Support Vector Machine (SVM), which has contributed to the progress of various research fields as an excellent classifier. Based on SVM, we propose a new approach for data adaptive nonparametric identification of different phases given training data. By analyzing the dividing hyper plane, support vector, capturing statistical properties of transitional periods is enabled. Extracting information from support vector is hardly seen in the literature. Moreover, application of SVM to finance and economics is hardly seen. By learning monthly moment behavior of New York Stock Exchange index and Tokyo Stock Exchange index daily returns from 1967 to 2010, bull/bear market phases are classified by SVM with success probability about 70% . Given the longest available training data up to the base date, the investment performance during the next 1 year achieved on average 9-10% at annual rate, significantly outperforming annual index returns. The provided hazard map describes the probability that the market is in the transitional period as a function of the key explanatory variables.

CS60 Room MAL B33 NONLINEAR FINANCIAL ECONOMETRIC MODELS Chair: Elias Tzavalis

C432: **Regime specific predictability in predictive regressions**
Presenter: **Jean-Yves Pitarakis**, University of Southampton, UK
Co-authors: Jesus Gonzalo

Predictive regressions are linear specifications linking a noisy variable such as stock returns to past values of a more persistent regressor such as valuation ratios, interest rates etc with the aim of assessing the presence or absence of predictability. Key complications that arise when conducting such inferences are the potential presence of endogeneity, the poor adequacy of the asymptotic approximations amongst numerous others. We develop an inference theory for uncovering the presence of predictability in such models when the strength or direction of predictability, if present, may alternate across different economically meaningful episodes. This allows us to uncover economically interesting scenarios whereby the predictive power of some variable may kick in solely during particular regimes or alternate in strength and direction (e.g. recessions versus expansions, periods of high versus low stock market valuation, periods of high versus low term spreads etc). The limiting distributions of our test statistics are free of nuisance parameters and some are readily tabulated in the literature. Finally our empirical application reconsiders the literature on Dividend Yield based stock return predictability and contrary to the existing literature documents a strong presence of predictability that is countercyclical, occurring solely during bad economic times.

C795: **Stochastic volatility driven by large shocks**
Presenter: **Yiannis Dendramis**, Athens University of Economics and Business, Greece
Co-authors: George Kapetanios, Elias Tzavalis

This paper presents a new model of stochastic volatility which allows for infrequent shifts in the mean of volatility, known as structural breaks. These are endogenously driven from large innovations in stock returns arriving in the market. The model has a number of interesting properties. Among them, it can allow for the shifts in volatility which are of stochastic timing and magnitude. This model can be used to distinguish permanent shifts in volatility coming from large pieces of the news arriving in the market, from ordinary volatility shocks.

C523: **High-frequency jump filtering in a microstructure model**
Presenter: **Michael Rockinger**, University Lausanne/ Swiss Finance Institute and HEC, Switzerland
Co-authors: Jerome Lahaye, Eric Jondeau

We present a general microstructure model written as a state-space model which allows for jumps. Our microstructure model also incorporates short-term GARCH effects, daily, and intradaily seasonality. Guided by outlier removal techniques, we adapt the Kalman filter to iteratively detect jumps. Simulations demonstrate good power properties for jump identification. We apply our methodology to an illiquid and a liquid stock traded at Euronext Paris over one month. We find that jumps and market microstructure noise cannot be considered in isolation. On average we detect one jump per day. Our methodology could be used to detect jumps in real time.

C272: **Semi-parametric estimation of American option prices**
Presenter: **Diego Ronchetti**, Swiss Finance Institute at the University of Lugano, Switzerland
Co-authors: Patrick Gagliardini

We introduce a new semi-parametric estimator of the price of American options. The estimator is based on a parametric specification of the stochastic discount factor and is non-parametric w.r.t. the historical dynamics of the Markov state variables. The estimation method exploits the no-arbitrage conditions for the short-term risk-free bond, the underlying asset and a cross-section of observed prices of American options written on it. We obtain an estimator of the transition density of the state variables process by minimizing a statistical measure based on the Kullback-Leibler divergence from a kernel-based transition density. We use the estimator to compute the price of American options not traded in the market by recursive valuation. Other functionals of the transition density interesting for financial applications can be estimated in a similar way.

C148: **A volatility smirk that defaults: The case of the S&P 500 index options**
Presenter: **Panayiotis Andreou**, Cyprus University of Technology, Cyprus

There is limited empirical evidence to examine the economic determinants of the volatility smirk anomaly as implied by market option data. The time-series economic determinants that affect the shape of the S&P 500 index implied volatility functions is investigated. One of the most important contributions is to assess the role of market default risk on the pricing of the S&P 500 index options. The analysis shows that market default risk has a dual role to play, since it can potentially capture both, the market leverage effect, as well as, the market's perceptions about the future growth/state

of the economy. More importantly, in a regression analysis where market leverage is disentangled from asset returns, it is also shown that the contemporaneous S&P 500 index return is still important in explaining the shape of the S&P 500 index implied volatility functions. The overall results suggest that, besides options pricing models that admit stochastic volatility and random jumps, it is also worthwhile to exploit models that take into account market leverage as an additional factor.

CS70 Room MAL B34 COMPUTATIONAL ECONOMETRICS WITH R Chair: Achim Zeileis

C108: **Improved interval estimation of long run response from a dynamic linear model**

Presenter: **Jae Kim**, La Trobe University, Australia

Co-authors: Iain Fraser, Rob Hyndman

A new method of interval estimation for the long run response (or elasticity) parameter from a general linear dynamic model is proposed. We employ the bias-corrected bootstrap, in which small sample biases associated with the parameter estimators are adjusted in two stages of the bootstrap. As a means of bias-correction, we use alternative analytic and bootstrap methods. To take atypical properties of the long run elasticity estimator into account, the highest density region (HDR) method is adopted for the construction of confidence intervals. From an extensive Monte Carlo experiment, we found that the HDR confidence interval based on indirect analytic bias-correction performs better than other alternatives, providing tighter intervals with excellent coverage properties. Two case studies (demand for oil and demand for beef) illustrate the results of the Monte Carlo experiment with respect to the superior performance of the confidence interval based on indirect analytic bias-correction. All computations are conducted using R and its hdrcde package.

C430: **The convergence properties of the BLP (1995) contraction mapping and alternative nonlinear algorithms in R**

Presenter: **Jo Reynaerts**, Katholieke Universiteit Leuven, Belgium

Co-authors: Ravi Varadhan, John C. Nash

Given its linear rate of convergence and the typical sample size encountered, the BLP Contraction Mapping used in estimating random coefficients logit models of demand is a time-consuming procedure. By reformulating the fixed-point problem as a nonlinear rootfinding problem, this paper introduces alternative methods to accelerate convergence of the Contraction Mapping, specifically (1) the classical Newton-Raphson (N-R) method, (2) the derivative-free spectral algorithm for nonlinear equations (DF-SANE), and (3) the squared polynomial extrapolation method (SQUAREM). Using a Monte Carlo study in R, we find that under the worst of circumstances SQUAREM has better convergence properties in terms of speed (up to more than five times faster) and stability (up to 14 percentage points better convergence rate) than BLP while attaining a quality of approximation that only differs at the 7th or 8th decimal. We also find that SQUAREM outperforms DF-SANE in nearly all scenarios. Newton-Raphson with an analytical Jacobian, used as a benchmark (and never better than SQUAREM and DF-SANE), is faster than BLP for small sample sizes and situations close to the true solution of the outer loop GMM minimization problem, but its performance deteriorates rapidly for larger sample sizes and more demanding circumstances.

C508: **A comparison of semiparametric estimators for the binary choice model**

Presenter: **Alicia Perez-Alonso**, University of Vigo, Spain

Co-authors: Maja Rynko, Christoph Weiss

This paper aims at being a practitioner's guide to computation and comparison of six popular semi-parametric estimators for binary response models. The comparison is restricted to the latent variable framework with single index form, and requires to impose some normalization on the parameter β of the index. The inferential problem is to estimate β, and to the extent possible, standard errors and probabilities. In particular, the paper focuses on the Maximum Score Estimator of Manski, its smooth version by Horowitz, Ichimura's Semiparametric Least Squares Estimator, the estimator of Klein and Spady, the Semi-Nonparametric (SNP) approach of Gabler et al., and Lewbel's estimator. The paper provides a unified R code for these estimators that contains different numerical and estimation approaches. Among them, the Great Circle Search and the Mixed Integer Programming for Manski's estimator, different choices of basis function for the SNP, and bandwidth selection through cross-validation and penalized functions for the estimators that require kernel smoothing. Several Monte Carlo experiments compare the performance of the estimators across different settings that accommodate many different kinds of heteroskedasticity. As empirical illustration, the paper models the probability of smoking and discusses the comparability of the results across different estimation strategies.

C570: **Some computational aspects of count data regression**

Presenter: **Christian Kleiber**, Universitaet Basel, Switzerland

Modern count data regression goes beyond the classical Poisson and negative binomial regression models that are conveniently discussed in the framework of generalized linear models (GLMs). For historical reasons, fields such as econometrics have never adopted GLMs, and so many tools from the GLM world are underused there. The talk will argue that recent *generalized* or *modified* econometric regression models for count data, although not GLMs themselves, benefit from a discussion that relates their building blocks to the unifying framework of GLMs. This framework is also a natural setting from a computational point of view, and hence for implementations in statistical software. The talk will address, among further topics, GLM aspects of truncated count data regression and non-classical link functions in hurdle models. It will also be argued that certain *pathologies* such as complete separation, which are inherited by the new models from their GLM building blocks, are better understood and handled by exploiting the GLM framework. Where appropriate, illustrations will make use of the R language.

CS87 Room MAL 151 COMPUTATIONAL ECONOMETRICS AND DATA ANALYSIS Chair: Paolo Foschi

C267: **Measuring inflation expectations using interval-coded data**

Presenter: **Yasutomo Murasawa**, Osaka Prefecture University, Japan

To quantify qualitative survey data, the Carlson–Parkin method assumes normality, a time-invariant symmetric indifference interval, and long-run unbiased expectations. Interval-coded data do not require these assumptions. Since April 2004, the Monthly Consumer Confidence Survey in Japan asks households their price expectations a year ahead in seven categories with partially known boundaries. Thus, one can identify up to six parameters including an indifference interval each month. This paper compares normal, skew normal, and skew t distributions, and finds that the skew t distribution fits the best throughout the period studied. The results help to understand the dynamics of heterogeneous expectations.

C325: **A K-sample homogeneity test based on the quantification of the p-p plot: the harmonic weighted mass index**

Presenter: **Rien Wagenvoort**, European Investment Bank, Luxembourg

Co-authors: Jeroen Hinloopen, Charles van Marrewijk

We first introduce the K-dimensional p-p plot. Next, we propose a quantification of this K-dimensional p-p plot that assigns equal weight to all distances between the respective distributions: The surface between the p-p plot and the diagonal. This surface is labelled the Harmonic Weighted

Mass (HWM) index. The HWM index for two samples differs from the L1-FCvM statistic when there are ties in that it is invariant to the position of the tie in the sequence of order statistics. This makes it a more robust statistic. We introduce the diagonal-deviation (d-d) plot that allows the index to be computed exactly under all circumstances. An example involving economic growth rates of the G7 countries illustrates that the HWM test can have better power than alternative Empirical Distribution Function tests.

C805: **Causality and club convergence: A nonparametric framework for cross-country analysis**
Presenter: **Arnab Bhattacharjee**, University of Dundee, UK
Co-authors: Subhadip Bandyopadhyay, Snigdhansu Chatterjee

We develop new framework and methods for nonparametric analysis of economic performance across countries and regions. The proposed method proceeds sequentially in three steps. First, we use a time series alignment method to estimate leads and lags and thereby synchronise multivariate time series data over a cross section of countries. At the second step, we use clustering methods on the aligned time series data and classify countries according to their levels of income. Finally, we use notions of multivariate depth to test for universal convergence and club convergence based on the a priori clustered classes of countries. We also develop a useful graphical tool to represent patterns of causation, classification and convergence across countries. The proposed methodology is computation intensive, but is robust against departures from traditional time series and panel data models. Application to data on output and unemployment at the business cycle frequency across a wide range of countries offers exciting new evidences on the causal lags and leads across countries as well as club convergence.

C258: **Tree-structured vector autoregressive model with smooth transition - STVAR-tree**
Presenter: **Alexandre Santos**, PUC-Rio, Brazil
Co-authors: Alvaro Veiga

The main goal is to introduce a nonlinear multivariate model, which combines the STVAR (Smooth Transition Vector Autoregressive) model with the CART (Classification and Regression Tree) method and use it for generating scenarios and forecasting. The resulting model is a Tree-Structured Vector Autoregressive model with Smooth Transition, called STVAR-Tree, which is based on the concept of multiple regimes, defined by a binary tree. The model specification is based on Lagrange Multiplier tests. Thus, the growth of the tree is conditioned on the existence of nonlinearity in the time series, which indicates the node to be split and the corresponding transition variable. In each division, linear parameters are estimated by Multivariate Least Squares, and nonlinear parameters by Nonlinear Least Squares. As a way of checking the STVAR-Tree model, several Monte Carlo experiments were performed in order to see the functionality of both the LM test and the model estimation. Best results were obtained with medium and large samples. Besides, the STVAR-Tree model was applied to Brazilian time series of rivers flow and electricity spot price. Adding both the experiments and the two applications results we conclude that the STVAR-Tree model may be applied to solve real problems, having good results.

C882: **Forecasting medium and large datasets with vector autoregressive moving average (VARMA) models**
Presenter: **Gustavo Fruet Dias**, Queen Mary University of London, UK
Co-authors: George Kapetanios

We address the issue of forecasting key macroeconomics variables using medium and large datasets (from 10 to 40 variables). As an alternative to the standard VAR and AR models, we propose the use of Vector Autoregressive Moving Average (VARMA) models. We overcome the estimation issue that usually arises in high dimensional VARMA models by adopting the Iterative Least Squares (IOLS) methodology as a feasible estimation procedure. We derive theoretical results which show that IOLS estimator is consistent even for large datasets. We report results from Monte Carlo simulations considering estimator performance and forecast accuracy. In the former one, we report results that show IOLS estimator is consistent and feasible for large systems and it performs better than Maximum Likelihood estimator (MLE), whereas in the latter one, we show that VARMA models outperform both VAR and AR(1) models under different sample and system size. On empirical application, we show that different specifications of VARMA models estimated using IOLS framework provide more accurate forecasts than VAR and AR(1) models considering different model dimensions. We conclude that the IOLS framework allows VARMA models to become a valid and competitive alternative for forecasting key macroeconomic variables using medium and large datasets.

CS73 Room MAL B29 TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

C587: **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 coefficient estimator to first order. Illustrative numerical and simulation results on the magnitude of 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.

C222: **Panel autoregressive models with cross-sectional dependence**
Presenter: **Jan Mutl**, Institute for Advanced Studies, Austria
Co-authors: Jaroslava Hlouskova

We consider dynamic panel data models with several types of cross-sectional dependence of the disturbances that include spatial autoregressive models, factor structure as well as various types of heteroscedasticity and error components. For each model we describe and discuss appropriate estimation methods. We provide an overview of the large sample results available for these estimation procedures under the various model specifications. Finally, we compare the small-sample performance of the estimation strategies in a Monte Carlo study.

C261: **On the usefulness of the Diebold-Mariano test in the selection of prediction models: Some Monte Carlo evidence**
Presenter: **Robert Kunst**, University of Vienna, Austria
Co-authors: Mauro Costantini

In evaluating prediction models, many researchers flank comparative ex-ante prediction experiments by significance tests on accuracy improvement, such as the Diebold-Mariano test. We argue that basing the choice of prediction models on such significance tests is problematic, as this practice may favor the null model, usually a simple benchmark. We explore the validity of this argument by extensive Monte Carlo simulations with

ARMA generating processes. For most parameter constellations, we find that utilization of additional significance tests in selecting the forecasting model fails to improve predictive accuracy.

C505: **Bounds, breaks and unit root tests**

Presenter: **Lola Gadea**, University of Zaragoza, Spain

Co-authors: Josep Lluis Carrion i Silvestre

The unit root testing when the range of the time series is limited considering the presence of multiple structural breaks is addressed. Contrary to the standard result, level shifts with large magnitude effects change the limit distribution of the unit root tests when the time series is bounded. The paper designs a new procedure to accommodate the presence of multiple level shifts when testing the unit root hypothesis using the M tests. The approach is illustrated through the analysis of the U.S. unemployment rate in order to show the relevance of our proposal.

C165: **Estimating cointegrating relationships: A tuning parameter free approach**

Presenter: **Martin Wagner**, Institute for Advanced Studies Vienna, Austria

Co-authors: Tim Vogelsang

We propose a new tuning parameter free estimator for cointegrating relationships. The estimator is based on an alternative method of (endogeneity and serial correlation) bias removal that is based on a simple partial sum transformation of the regression model. The analytical properties of the estimator as well as subsequent inference based on this estimator are investigated using asymptotic theory as well as extensive simulations. Fixed-b asymptotic theory and associated critical values for hypothesis tests are provided. In the simulations the estimator and tests are compared with standard approaches including FM-OLS, DOLS and VAR models.

Saturday 11.12.2010 08:40 - 10:45 Parallel Session F – ERCIM

EI84 Room Senate Beveridge Hall INVITED SESSION: HEURISTIC METHODS IN STATISTICS Chair: Sandra Paterlini

E129: **From heuristic optimization by differential evolution to adaptive MCMC in real and discrete parameter spaces**

Presenter: **Cajo J.F. ter Braak**, Wageningen University and Research Centre, Netherlands

Co-authors: Jasper Vrugt

Differential evolution (DE) is a heuristic evolutionary algorithm for numerical optimization in real parameter spaces. In a statistical context one would not just want the optimum but also its uncertainty. The uncertainty distribution can be obtained by Markov Chain Monte Carlo (MCMC) simulation using the Metropolis-Hasting algorithm. DE and MCMC have recently been integrated, resulting in population MCMC methods (DE-MC and DREAM) that adaptively optimize the offspring distribution. In this talk I explore whether other heuristic methods can usefully be turned into MCMC methods. I also extend DE-MC to binary random variables, add a non-parametric snooker update and compare DE-MC with the Cross Entropy method and estimation of distribution algorithms (EDA). The efficiency is evaluated on examples by simulation.

E356: **Replicating hedge fund indices with optimization heuristics**

Presenter: **Manfred Gilli**, University of Geneva, Switzerland

Co-authors: Enrico Schumann, Gerda Cabej, Jonela Lula

Hedge funds offer desirable risk-return profiles; but we also find high management fees, lack of transparency and worse, very limited liquidity (they are often closed to new investors and disinvestment fees can be prohibitive). This creates an incentive to replicate the attractive features of hedge funds using liquid assets. We investigate this replication problem using monthly data of CS Tremont for the period of 1999 to 2009. Our model uses historical observations and combines tracking accuracy, excess return, and portfolio correlation with the index and the market. Performance is evaluated considering empirical distributions of excess return, final wealth and correlations of the portfolio with the index and the market. The distributions are compiled from a set of portfolio trajectories computed by a resampling procedure. The nonconvex optimization problem arising from our model specification is solved with a heuristic optimization technique. Our preliminary results are encouraging as we can track the indices accurately and enhance performance (e.g. have lower correlation with equity markets).

E678: **Promoting the generalisation of genetically induced trading rules**

Presenter: **Alexandros Agapitos**, University College Dublin, Ireland

Co-authors: Michael O'Neill, Anthony Brabazon

The goal of Machine Learning is not to induce an exact representation of the training patterns themselves, but rather to build a model of the underlying pattern-generation process. One of the most important aspects of this computational process is how to obtain general models that are representative of the true concept, and as a result, perform efficiently when presented with novel patterns from that concept. A particular form of evolutionary machine learning, Genetic Programming, tackles learning problems by means of an evolutionary process of program discovery. In this paper we investigate the profitability of evolved technical trading rules when accounting for the problem of over-fitting. Out-of-sample rule performance deterioration is a well-known problem, and has been mainly attributed to the tendency of the evolved models to find meaningless regularities in the training dataset due to the high dimensionality of features and the rich hypothesis space. We present a review of the major established methods for promoting generalization in conventional machine learning paradigms. Then, we report empirical results of adapting such techniques to the Genetic Programming methodology, and applying it to discover trading rules for various financial datasets.

ES37 Room MAL B36 STATISTICS FOR INTERVAL DATA Chair: Gil Gonzalez-Rodriguez

E596: **Predicting financial asset returns and volatility - interval data regressions**

Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain

Co-authors: Henning Fischer, Peter Winker

Several flexible interval-arithmetic regression models for interval data have been recently presented, and a coherent estimation for the regression parameters has been proposed. The statistical processing of random intervals is a convenient tool when the available data have got an interval nature; this is the case of fluctuations, ranges of values, censored observations, to name but a few. This type of data appears in multiple areas: Economics and Finance, Medicine, Environment studies, Agriculture, among others. The aim in this work is to investigate the performance of some interval linear models for explaining and modelling financial asset returns and volatility. Classical CAPM-type regression models use single inter-daily stock market returns, so it is necessary to consider also a second regression process for modelling the volatility. By considering interval stock market returns, and applying interval regression techniques, both quantities can be jointly modelled. An empirical study is shown, to compare forecasting for financial asset returns when interval or classical regression models are applied. Employing data correspond to inter-daily returns for DAX stock index (Germany) and its 30 constituents, in the period 2005-2009.

E792: **Predicting financial asset returns and volatility - Point data vis-a-vis interval data models**

Presenter: **Henning Fischer**, University of Giessen, Germany

Co-authors: Angela Blanco-Fernandez, Peter Winker

Modeling and forecasting financial asset returns and volatility, in particular, is a key issue in applications such as portfolio management, derivative securities pricing, and risk management. Recent research often proposes measures derived from high-frequency data as the most accurate proxies of the unobservable volatility process. However, data sampled at high frequency are costly, not readily available, and prone to all sorts of microstructure noise which may bias volatility estimates. Thus, models dealing with inter-daily data do not necessarily yield inferior forecasts. By also considering the range, i.e. the interval spanned by the highest and the lowest recorded daily price, respectively, one might effectively implement all intra-daily trade information into these lower-frequency models and hence improve forecasts. We investigate that thought by comparing CAPM-type regression models using (i) point data and (ii) interval data, both of daily frequency. While the former use single inter-daily returns based on closing prices only, possibly modeling volatility as a GARCH process, the latter directly incorporate return intervals as their variables when being estimated by linear regression techniques. Employing data on German stocks, we analyze if interval data regressions lead to improved forecasts in comparison to those of models based solely on single closing prices.

E760: **Least-squares estimation of a multiple regression model for interval data**

Presenter: **Marta Garcia-Barzana**, University of Oviedo, Spain

Co-authors: Ana Colubi, Erricos J. Kontoghiorghes

A multiple regression model for interval-valued explanatory and response random elements is firstly presented. The model is a generalization of an existing simple model for interval data based on the natural set-arithmetic. The non-negativity condition on the spreads of the predicted interval-valued entails the estimation to be a computationally infeasible combinatorial problem. This is offset by obtaining a new expression of the parameters in terms of the moments of the mid-points and spreads of the involved random intervals. It is shown that the parameters can be estimated by separating the estimation of the absolute values and the sign, which depends only on the mid points. In turn this implies that the least-squares problem can be written in terms of a quadratic minimization problems with linear constraints which can be solved with standard procedures. The empirical behaviour of the estimators is illustrated by means of simulations. The results are applied to a real-life case study.

E633: **Inclusion test of the expected value of a random interval in a given interval**

Presenter: **Ana Belen Ramos-Guajardo**, European Centre for Soft Computing, Spain

Co-authors: Gil Gonzalez-Rodriguez, Maria Angeles Gil

A bootstrap procedure for testing the inclusion of the expected value of a random compact interval in a given closed interval is developed. The expected value in this case is understood in the sense of the Aumann integral, which leads to a compact interval. In this way, a measure for the inclusion of the interval-valued mean value in another interval is analyzed. The asymptotic distribution of such a measure is not suitable to handle in practice. Thus, a bootstrap procedure is developed and empirically analyzed by means of simulation studies. This test generalizes the classical inclusion test of the expected value of a random variable in an interval, and can be applied when the nature of the experimental data can be better captured through intervals. For instance, when data represent subjective valuations of the quality of a given product, or the variation of the tides or the blood pressure during a day. An illustrative example in this line is provided.

ES43 Room MAL 152 STATISTICS FOR FUNCTIONAL DATA Chair: Pascal Sarda

E180: **Prediction in functional linear regression with functional response**

Presenter: **Christophe Crambes**, Universite Montpellier 2, France

Co-authors: Andre Mas

We study the problem of prediction in the functional linear model when the output is also functional. In other words, we are interested in the model writing $Y(t) = S(X)(t) + \varepsilon(t)$, where the operator S is defined by $S(f)(t) = \int_0^1 \mathcal{S}(s,t)f(s)ds$ for all square integrable function f. Given a sample $(X_i, Y_i)_{i=1,\ldots,n}$, the objective is to study a predictor of Y_{n+1} given a new curve X_{n+1}. Our predictor is based on a Karhunen-Loeve decomposition of the curves X (covariate) and Y (output). Our results give an asymptotic development of the mean square error prediction. We also get an optimality result for these rates of convergence in a minimax sense, as well as a central limit theorem for the predictor.

E458: **Applications of bootstrap calibration to functional linear regression**

Presenter: **Adela Martinez-Calvo**, University of Santiago de Compostela, Spain

Co-authors: Wenceslao Gonzalez-Manteiga

Numerous papers have focused on the Functional Data Analysis (FDA) in the recent literature. One of the most interesting research lines is the study of the functional linear regression model with scalar response. In this work we propose a bootstrap resampling procedure (naive and wild bootstrap) which allows us to calibrate different distributions and test hypotheses related with the functional linear regression model. Simulations and applications to a real environmental dataset show the behaviour of the bootstrap calibration from a practical point of view.

E396: **Teaching an Englishman to speak Chinese using functional data analysis**

Presenter: **John Aston**, University of Warwick, UK

Co-authors: Jeng-Min Chiou, Jonathan Evans

Fundamental frequency (F0, broadly "pitch") is an integral part of spoken human language and in many languages such as Mandarin Chinese, relative pitch information is part of each word's dictionary entry. However, a comprehensive quantitative model for F0 can be a challenge to formulate due to the large number of effects and interactions between effects that lie behind the human voice's production of F0, and the very nature of the data being a contour rather than a point. A semi-parametric functional response model for F0 will be formulated by incorporating linear mixed effects models through the functional principal component scores. This model is applied to the problem of modelling F0 in the tone languages such as Mandarin and Qiang (a dialect from China).

E613: **Nonparametric regression when both response and predictor are random curves**

Presenter: **Frederic Ferraty**, Institut de Mathematiques de Toulouse, France

Co-authors: Philippe Vieu

Investigations involving functional data focused on linear modelling and successful functional statistical methods have been achieved from the 1990's. More recently, nonparametric approaches have been widely investigated, especially in regression model in order to take into account non-linear relationship between scalar response and functional predictor (i.e. random variable valued into infinite-dimensional space: random curves, random surfaces, etc). In such a nonparametric functional setting, theoretical and practical developments have been done for estimation and/or prediction. The case of random curve response is not so developed. This talk aims to present recent advances on nonparametric regression when both response and predictor are random curves. We start by giving theoretical properties about a kernel estimator of the regression operator. It appears that getting confidence areas from asymptotic distribution seems to be unrealistic in this functional data situation (the quantities involved in the asymptotic distribution are not computable). An alternative standard tool is the so-called bootstrap method. In a second part, we investigate on a bootstrap methodology. Theoretical properties will be stated whereas simulations and real datasets will illustrate the good practical behaviour of the bootstrap in this functional situation. At last, we will show how functional pseudo-confidence areas can be built.

E804: **Robust analysis of functional data**

Presenter: **Sara Lopez-Pintado**, Columbia University, USA

Co-authors: Juan Romo

In many research areas such as medicine, biology and economics, there is a growing need for the statistical analysis of functional data where the basic observation unit is a curve. Many multivariate methods such as analysis of variance and classification have been extended to functional data. The statistical analysis of curves can be significantly improved using robust estimators. In this work we discuss new graph region-based definitions of depth for functional data which provide a center-outward ordering of curves. Robust estimators such as the median and the trimmed mean curves are defined for functional data. A simulation study of contaminated models is described where location estimates based on our new notions of depth are compared with more classical estimators such as the mean and coordinate-wise median. We propose robust classification methods for curves based on the new notions of depth. Applications of these tools to real data sets such as growth curves and gene expression microarray data are presented.

ES62 Room MAL 421 CHANGE POINTS IN TIME SERIES Chair: Konstantinos Fokianos

E741: **Breaks in long memory time series**

Presenter: **Heiko Rachinger**, Universidad Carlos III de Madrid, Spain

There has been a long discussion about whether time series that appear persistent have a break in the level or are fractionally integrated. In this paper, we analyze least squares (LS) estimation of breaks in a long memory time series with an intercept. Taking into account the estimation of the memory parameter, the problem becomes a nonlinear one and we have to consider specific arguments to derive the asymptotic properties, in particular we propose a filter that renders the breaking fractionally integrated series asymptotically I(0) series. We show that the estimator of the break fraction is consistent with a rate T convergence equally if the break occurs in the mean, in the memory or in both and determine its asymptotic distribution for fixed and shrinking magnitude of breaks. Finally, we analyze tests for the number of breaks in this setting. For a break in the memory, the asymptotics results correspond to well known results in the literature. For a break in the intercept, the results differ in terms of the asymptotic distribution of the break fraction estimate and of the test statistic under the null distribution of no break and the LS-procedure loses some of its nice properties, such as nuisance parameter free test distributions. The rate of divergence under the alternative depends on which parameters are breaking and on the actual memory levels. In a simulation exercise, we find that in finite samples, testing for breaks in the intercept as well as the memory leads to an oversized test. In consequence, we suggest testing in practice only for a break in the memory. We illustrate the procedure with an empirical example.

E486: **Robustifications of the EM algorithm for state space models**

Presenter: **Peter Ruckdeschel**, Fraunhofer ITWM, Germany

Co-authors: Bernhard Spangl, Irina Ursachi

We report on recent progress made in robustness for Kalman filtering for linear, time discrete, and time-invariant state space models with Euclidean state space. Algorithms for robustifications of the Kalman filter in this context have been around for some time. All these approaches though assume knowledge of the state space model as to its hyper parameters, i.e.; innovation and error covariances Q and V, as well as transition and observations matrices F and Z. In many applications these have to be estimated from the data. For this purpose Shumway and Stoffer have proposed an EM algorithm, which has been improved by several authors since. These approaches are non-robust, though. We discuss an approach where we propose robustifications for each stage of the EM-algorithm, i.e., initialization, expectation and maximization step. This approach is backed by a corresponding implementation to R, package robKalman, developed in r-forge.

E599: **Clustering multivariate time series for robust change point detection**

Presenter: **Uwe Ligges**, TU Dortmund, Germany

Co-authors: Sebastian Krey

Clustering multivariate time series for robust change point detection is a common problem and several fields. The method presented in this talk has been developed for applications in music analysis, particularly timbre recognition (which instrument / voice do we hear), where the aim is to segment tones at change points into different phases. This can lead to an enormous reduction of complexity since only one phase of each tone rather than hundreds of windows of a multivariate time series can be used for further post processing. For monophonic recordings clustering methods based on classic sound features like static Mel-frequency cepstral coefficients (MFCC) often result in good and interpretable results. For more difficult situations different segmentation approaches for multivariate time series can be used and their results are presented and compared to the clustering methods. Additionally the suitability of more modern sound features like cepstral modulation ratio regression (CMRARE) are investigated. Other applications are hearing aid development, sheet music transcription as well as speech recognition.

E397: **Quantifying the uncertainty of change point estimates in time series**

Presenter: **Christopher Nam**, University of Warwick, UK

Co-authors: John Aston, Adam Johansen

Quantifying the uncertainty in the number and locations of change points in time series has become an important topic of interest, particularly in a variety of applications including econometrics and finance. Current existing methods however generally fail to capture fully or explicitly the uncertainty regarding estimates of these quantities. A new methodology is proposed to compute the posterior distributions for these change point characteristics in light of parameter uncertainty. This provides full and explicit quantification of the uncertainty in these quantities. This methodology combines recent work of exact change point distributions conditional on model parameters via Finite Markov Chain Imbedding in a Hidden Markov Model context, and accounts for parameter uncertainty via Sequential Monte Carlo. The combination of the two leads to a flexible and computationally efficient procedure which does not require estimates of the underlying state sequence. Good estimation of the posterior distributions of change point characteristics is illustrated for a variety of data, including econometric and medical time series.

E176: **On hierarchical epidemic changes in dependent functional data**

Presenter: **Claudia Kirch**, Karlsruhe Institute of Technology, Germany

Co-authors: John Aston

Recently, change-point analysis has been highlighted as a useful technique in psychological experiments performed with functional Magnetic Resonance Imaging (fMRI) where different subjects react differently to stimuli such as stress or anxiety. While current methodology is applied pointwise across spatial locations, our approach is based on recently developed change-point procedures for functional data. Of specific interest are procedures for dependent data and epidemic changes. Because of the very high-dimensionality of the data an approach based on a general covariance structure is computationally not feasible. Therefore, a special case, that of multidimensional separable functional covariance structures will be considered. In the above application multiple subjects are usually scanned, indicating a hierarchical nature of the change-points within the experiments, with the distribution of the change-points over all subjects an item of interest. In this case it is possible to estimate the distribution and density of a change-point based on estimators derived from the above procedures.

ES64 Room MAL 532 BAYESIAN NONPARAMETRICS: METHODS AND COMPUTATION Chair: Igor Pruenster

E427: **Bayesian computation with unknown normalising constant**

Presenter: **Stephen Walker**, University of Kent, UK

An approach by which to undertake exact posterior inference using MCMC methods when the normalising constant for the model is unknown and not computable are presented. Strategic latent variables are introduced which allow the removal of the normalising constant and subsequently permit a Gibbs sampler to be completely specified.

E078: **Vectors of Poisson-Dirichlet processes**
Presenter: **Fabrizio Leisen**, Carlos III University, Spain
Co-authors: Antonio Lijoi

The definition of vectors of dependent random probability measures is a topic of interest in applications to Bayesian statistics. They, indeed, define dependent nonparametric prior distributions that are useful for modeling observables whose values depend on covariates. In this paper we propose a vector of two-parameter Poisson-Dirichlet processes. It is well- known that each component can be obtained by resorting to a change of measure of a σ-stable process. Thus dependence is achieved by applying a Levy copula to the marginal intensities. In a two-sample problem, we evaluate the corresponding partition probability function which turns out to be partially exchangeable.

E088: **Nonparametric Bayes conditional distribution estimation and mean regression: posterior consistency and applications**
Presenter: **Debdeep Pati**, Duke University, USA
Co-authors: David Dunson

A wide variety of priors have been proposed for nonparametric Bayesian estimation of conditional distributions, and there is a clear need for theorems providing conditions on the prior for large support, as well as weak and strong posterior consistency. Focusing on a broad class of priors formulated as predictor-dependent mixtures of Gaussian kernels, we provide sufficient conditions under which weak and strong posterior consistency hold. This theory is illustrated by showing that the conditions are satisfied for a class of generalized stick-breaking process mixtures in which the stick-breaking lengths are constructed through mapping continuous stochastic processes to the unit interval using a monotone differentiable link function. Probit stick-breaking processes provide a computationally convenient special case. We develop a robust Bayesian regression procedure based on a Gaussian process prior for the mean regression function and probit stick breaking mixtures of Gaussians for the collection of residual densities indexed by predictors. Based on our theory, we prove strong posterior consistency in estimating the mean regression function and the residual density. The methods are illustrated using simulated and real data applications.

E085: **A construction of dependent Dirichlet processes**
Presenter: **Bernardo Nipoti**, University of Pavia, Italy

We present the construction of two dependent Dirichlet processes (DPs) that we obtain by normalizing the increments of two dependent gamma completely random measures (CRMs). These are constructed as linear functionals of dependent Poisson random measures. If p_0, p_1 and p_2 are independent Dirichlet processes with suitable parameters, such a construction gives rise to dependent DPs, $\tilde{p}_1$ and $\tilde{p}_2$, with the property that $\tilde{p}_1$ is a mixture of p_0 and p_1 while $\tilde{p}_2$ is a mixture of p_0 and p_2. The vector of CRMs that we obtain is analytically tractable and the dependence structure between $\tilde{p}_1$ and $\tilde{p}_2$ can be described in terms of a parameter in $[0,1]$. Given two exchangeable samples, respectively from $\tilde{p}_1$ and $\tilde{p}_2$, such a parameter suggests how distant we are from the two extreme situations of full exchangeability ($\tilde{p}_1 = \tilde{p}_2 = p_0$) and partial exchangeability ($\tilde{p}_1 = p_1$ and $\tilde{p}_2 = p_2$). The vector $(\tilde{p}_1, \tilde{p}_2)$ may be also used to construct dependent mixture models. An extension of the well-known Blackwell-MacQueen sampling scheme allows one to implement a Gibbs sampler and achieve a full Bayesian analysis for clustering and density estimation. This approach also leads one to construct $k \geq 2$ dependent DPs and more general dependent random probability measures for Bayesian nonparametric inference.

ES56 Room MAL 509 SUPERSATURATED AND FRACTIONAL FACTORIAL DESIGNS Chair: Steven Gilmour

E319: **The statistical analysis of optimal mixed-level supersaturated designs using group screening**
Presenter: **Kalliopi Mylona**, University of Antwerp, Belgium
Co-authors: Christos Koukouvinos

Supersaturated designs (SSDs) are used for screening out the important factors from a large set of potentially active variables. The huge advantage of these designs is that they reduce the experimental cost drastically, but their critical disadvantage is the high degree of confounding among factorial effects. In this contribution, we focus on mixed-level factorial designs which have different numbers of levels for the factors. Such designs are often useful for experiments involving both qualitative and quantitative factors. When analyzing data from SSDs, as in any decision problem, errors of various types must be balanced against cost. In SSDs, there is a cost of declaring an inactive factor to be active (i.e. making a Type I error), and a cost of declaring an active effect to be inactive (i.e. making a Type II error). Type II errors are usually considered much more serious than Type I errors. We present a group screening method for analyzing data from $E(f_{NOD})$-optimal mixed-level supersaturated designs possessing the equal occurrence property. Based on the idea of the group screening methods, the f factors are sub-divided into g "group-factors". The "group-factors" are then studied using the penalized likelihood methods involving a factorial design with orthogonal or near-orthogonal columns. The penalized likelihood methods indicate which "group factors" have a large effect and need to be studied in a follow-up experiment. We will compare various methods in terms of Type I and Type II error rates using a simulation study.

E394: **The effectiveness of supersaturated screening experiments**
Presenter: **Chris Marley**, University of Southampton, UK
Co-authors: Dave Woods

Although supersaturated experiments have received much attention, there is still some doubt as to how useful they can be for screening in practical experiments. We demonstrate what can be achieved when using a supersaturated experiment that is both designed and analysed appropriately. Simulation studies show that a high power to detect active effects and a low type I error rate can be attained provided that the concept of effect sparsity is loosely obeyed. We also illustrate that for a given analysis method, there is typically little difference between designs from different criteria. Further, it is shown that the columns of the design to which the active factors are assigned can have a considerable impact on how often they are correctly detected in the analysis. Along the way, we compare several different analysis methods which have been previously proposed.

E476: **Methods for analysing supersaturated designs: A computational comparison and some new proposals**
Presenter: **Stelios Georgiou**, University of the Aegean, Greece
Co-authors: Stella Stylianou

Supersaturated designs are fractional factorial designs in which the run size is too small to estimate all the main effects. Under the effect sparsity assumption, the use of supersaturated design can provide a low-cost identification of the few possibly dominate factors (screening). Methods for analysing supersaturated designs have been proposed. Here, we compare the performance and efficiency of some known methods. The ability of each method to identify the true underling model is evaluated, in terms of Type I and Type II error rates, by using simulated experiments. Some new proposals and suggestions on the use of supersaturated designs, in screening experiments, are also discussed.

E440: **Linear programming approach for blocking strength-three designs**
Presenter: **Bagus Sartono**, Universiteit Antwerpen, Belgium
Co-authors: Peter Goos, Eric Schoen

Linear Programming (LP) can handle a broad range of optimization problems, because both the objective function and the constraints are easy to adjust. Here, we apply LP to the problem of blocking strength-three orthogonal arrays. Such arrays are useful as statistical designs when two-interaction effects are expected to be active. In that case, the blocking should interfere to the smallest extent possible with the estimation of the two-factor interaction components. We propose an integer LP model to deal with the problem of finding the best blocking, and apply this technique to a comprehensive catalog of non-isomorphic designs for up to 81 runs. For most of these designs, we obtained orthogonal blocking of main effects with minimum confounding between blocks and interactions for all feasible block sizes.

E567: **On optimal design of experiments for model discrimination using Akaike weights design criterion**

Presenter: **Danan S. Wicaksono**, RWTH Aachen University, Germany

Co-authors: Wolfgang Marquardt

Akaike Weights Design Criterion (AWDC) has been proposed as a measure of distinguishability between competing model candidates in the context of optimal design of experiments for model discrimination. AWDC aims at a better discrimination in cases with more than two competing model candidates and takes into account the relative difference of information loss between model candidates via Akaike's Information Criterion (AIC). The efficacy of optimal design of experiments for model discrimination using AWDC is illustrated by means of two cases involving numerous competing model candidates. The first case includes the effect of measurement error while the second case considers a situation in which a structurally correct model candidate is not available. In the presence of multiple locally optimal designs due to the nonconvexity of the arising optimization problem, this study demonstrates the benefit of employing global optimization approach rather than local optimization approach to design smaller number of experiments required to discriminate the competing model candidates. The local and global optimization approaches are enabled through a modeling system interfacing a collection of optimization solvers. The local optimization approach uses an interior point method while the global optimization approach utilizes a branch and bound algorithm applying deterministic bounds on global optimality.

ES73 Room MAL B35 ROBUST MULTIVARIATE METHODS Chair: Ruben Zamar

E132: **Test for outlier detection and optimal choice of efficiency for MM-regression estimators**

Presenter: **Catherine Dehon**, Universite libre de Bruxelles, Belgium

Co-authors: Marjorie Gassner, Vincenzo Verardi

In regression, it is well known that the least-squares estimation is not robust and may be excessively influenced even by a very limited number of outliers. In a first step we proposed a Hausman-type test for practitioners based on the trade-off between robustness and efficiency to find out if a linear least-squares estimation is appropriate or if a robust method (in this case an S-estimator) should be preferred. Unfortunately this test relies on classical symmetry, homoskedasticity and normality assumptions of the error term. No such assumptions are necessary for the test we propose in the second step based upon the fact that MM-estimators are exactly identified generalized method of moments estimators. Using the same methodology, we propose a data-based method allowing to find the maximal efficiency an MM-estimator can withstand before the bias become too large with respect to efficiency. At each step of the procedure, a highly robust estimator is compared with an MM-estimator with increasing efficiency. If the gain in efficiency is not balanced by excessive bias, the two vectors of slope estimators will be comparable and then the procedure continue to the highest possible efficiency. Some simulations are shown to assess the relevance of this new procedure.

E098: **Covariance regularization with robust estimators**

Presenter: **Gentiane Haesbroeck**, University of Liege, Belgium

Co-authors: Christophe Croux

In high dimensional settings, covariance regularization techniques are quite popular for estimating a sparse covariance or concentration matrix. Usually, the regularized scatter estimators are obtained by maximizing the log-likelihood of the data, subject to a penalty. In this talk, robust alternatives to this classical approach will be presented. More specifically, an M and an MCD -type regularized estimators will be defined and shown to achieve the highest possible breakdown value. Algorithms to compute these estimators will also be discussed and their performance compared by means of simulations. Some applications will also be illustrated. First, a robust covariance-regularized regression will be performed using the robust proposals. Then, as regularization and robustness are also necessary in graphical modeling, it will be shown that robust graphical models can be easily constructed from the regularized robust scatter estimators.

E193: **Some asymptotics for elemental subsets in regression with applications**

Presenter: **Keith Knight**, University of Toronto, Canada

In a linear regression model, elemental subsets consist of the minimum number of observations needed to estimate the regression parameter where the resulting estimators are called elemental estimators. Elemental estimators have a long history in statistics; many estimators are functions of elemental estimators and elemental estimators are used for outlier detection as well as for computation of highly robust estimators. In this paper, we consider some asymptotic theory for elemental subsets in linear regression. In particular, we derive the limiting distribution of the elemental subsets that produce good elemental estimators. A number of applications of this theory are also given, including a diagnostic for homoscedasticity and a heuristic for sampling elemental subsets in the computation of least median of squares estimates.

E144: **Robustness and efficiency of the Spearman and Kendall correlation measures**

Presenter: **Christophe Croux**, K.U.Leuven, Belgium

Co-authors: Catherine Dehon

Nonparametric correlation estimators as the Kendall and Spearman correlation are widely used in the applied sciences. They are often said to be robust, in the sense of being resistant to outlying observations. In this paper we formally study their robustness by means of their influence functions and gross-error sensitivities. Since robustness of an estimator often comes at the price of an increased variance, we also compute statistical efficiencies at the normal model. We conclude that both the Spearman and Kendall correlation estimators combine a bounded and smooth influence function with a high efficiency. In a simulation experiment we compare these nonparametric estimators with correlations based on a robust covariance matrix estimator, and with the Quadrant correlation.

E708: **Applications of fast robust bootstrap inference**

Presenter: **Stefan Van Aelst**, Ghent University, Belgium

Co-authors: Gert Willems

The fast and robust bootstrap (FRB) is a powerful tool for robust inference related to high breakdown estimators based on a smooth loss function. Initially the FRB has been used to obtain robust standard errors and confidence intervals for the parameters of interest. In later developments it has been shown that the FRB has a wider range of applications and can be used for robust testing, robust model selection and studying the stability of a model, etc. In this talk we illustrate some of the recent applications of FRB that can be performed with the R package FRB available on CRAN.

Saturday 11.12.2010 11:10 - 12:50 Parallel Session G – CFE

CI98 Room Senate Beveridge Hall INVITED SESSION: RECENT DEVELOPMENTS IN ECONOMETRICS Chair: Elias Tzavalis

C069: **Bayes procedures for optimal measurement of policy effects and risk**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide

New simulation based Bayesian methods are introduced for two issues in economics: (i) rare events like financial risk measurement of which the importance is stressed by the economic crisis; (ii) treatment effects of education on income, entrepreneurial performance and occupational choice, which is gaining importance due to the increasing knowledge economy.

C081: **Testing earnings management**
Presenter: **Philip Hans Franses**, Erasmus School of Economics, Netherlands
Co-authors: Dennis Fok

Earnings management to avoid earnings decreases and losses implies that the time series properties of the last quarter in the fiscal year differ from those of the other three quarters. We propose a simple parametric methodology to diagnose such differences. Application to a random sample of 390 firms in the Compustat database gives strong evidence of earnings management.

C706: **Constructing general smooth tests based on the Karhunen-Loeve expansion**
Presenter: **Chung-Ming Kuan**, National Taiwan University, Taiwan
Co-authors: Tzu-Chi Lin

Many specification tests are the Cramer-von Mises (CvM) test whose limit can be expressed as a weighted sum of independent χ^2 random variables, with the weights being the eigenvalues in the Karhunen-Loeve (KL) expansion of the limiting process and diminishing to zero. Such tests thus lack power against *high frequency* alternatives. To circumvent this power deficiency problem, it is natural to construct smooth tests whose limits are non-weighted sum of χ^2 random variables. In this paper we construct a general smooth test that is the sum of squared principal components of the KL expansion of an associated Gaussian process. We propose an estimation method for the eigenvalues, eigenfunctions, and the principal components of the KL expansion and determine the number of components in the statistic by a combination of the AIC and SIC. As examples, we extend several existing tests to smooth tests and evaluate their finite-sample performance. It is shown that the smooth tests have power advantages relative to the original CvM tests.

CS18 Room MAL B34 RISK MODELLING FOR FINANCIAL DERIVATIVES Chair: Felix Chan

C257: **Numerical Weather Forecasting For Weather Derivatives**
Presenter: **Michael Friedlander**, University of Hong Kong, Hong Kong

With the launch of the exchange traded weather derivatives market in 1999, investors and business leaders expressed hope that the tools necessary to manage weather risk were at hand. While the market has experienced rapid growth, the widespread implementation of weather risk management principles remains but a small fraction of its potential, not only in the developed markets, but in emerging markets globally. It is postulated that in part, the market's slow adoption of what would appear to be sound risk management principles can be traced to a suite of risk management tools that has been less than satisfactory. The inconsistent performance of existing weather derivative valuation methodologies is a consequence of the mistaken premise that weather is governed by stochastic processes. This research demonstrates that by considering weather (and hence weather indices) by their mathematical description as expressed by the science of meteorology, market participants can enjoy the full benefits of weather risk management strategies. Further, these concepts are immediately deployable anywhere on Earth, irrespective of a historical weather record thereby providing emerging market businesses with the same benefits as the more developed markets.

C617: **Valuation and Risk Analysis of Accumulators with Mean Reversion**
Presenter: **Fo Chun Ng**, University of Hong Kong, Hong Kong
Co-authors: Philip Leung Ho Yu

Mean reversion is a phenomenon in which asset prices tend to return to a long-run average value over time. It has material impact on derivatives value, especially those which are highly path dependent, such as accumulators. Despite the popularity of accumulators in recent years, the impact of mean reversion in the price of the underlying asset on its value has not been well explored. This paper aims at studying the price and risk of accumulators under a mean reverting process. A number of common underlying assets of accumulators, including different asset classes such as commodities, foreign currencies and stocks, are investigated and the analysis reveals significant mean reversion in some of their prices. An example is also given to illustrate the effect of mean reversion on an accumulator. It is found that assuming mean reversion in the price of the underlying asset yields much different result in the fair value, the zero-cost strike price and the value-at-risk of the accumulator.

C816: **Econometric analysis of volatile art markets**
Presenter: **Fabian Bocart**, Universite Catholique de Louvain, Belgium
Co-authors: Christian Hafner

It is well documented that volatility of many commodities and stocks display a certain degree of time variation. While considerable efforts have been devoted to assess returns in the art market, few studies attempt to investigate volatility structure of art as a function of time. Yet, volatility of fine art is worth investigating, and a better understanding of its structure may be of practical use for market participants, more particularly for participants exposed to derivatives on art. Such derivatives include price guarantees underwritten by auction houses. As far as returns are concerned, two main methodologies have been developed: the repeat sale methodology and the hedonic regression. Methodology-wise, ordinary least squares are usually employed to estimate parameters. In this paper, we recommend using Nadaraya-Watson estimator to obtain time dependent estimation of variance of residuals. This function is used to compute robust estimation of standard errors of parameters and to derive a consistent estimator of art market's volatility through time. We illustrate our methodology by building a blue chips art index from 2005 to 2010 and a corresponding estimator of time-varying volatility.

C070: **Forecast combinations of risk under different forecast criteria**
Presenter: **Felix Chan**, Curtin University of Technology, Australia
Co-authors: Adam James

It is well known in the empirical literature that combining forecasts from different models can often lead to superior forecast performance than forecasts from individual models, at least in the Mean Squares Error (MSE) sense. It has also been noted that using simple combination of forecasts often performs better than more sophisticated weighted average schemes, although simple combinations tend to ignore correlations between forecast errors. However, it is unclear whether these stylized facts hold under different forecast criteria. Given forecasts for Value-at-Risk threshold cannot be evaluated using MSE, it provides an interesting platform for evaluating the empirical usefulness of forecast combinations beyond the conventional forecast criteria. This paper proposes simple combinations for Value-at-Risk forecasts and evaluates their performances following the Back-testing procedure proposed by Basel Accords.

CS25 Room MAL B20 NON-STATIONARY TIME SERIES Chair: H. Peter Boswijk

C238: **Bootstrap co-integration rank testing: deterministic variables and initial values**

Presenter: **Robert Taylor**, University of Nottingham, UK

Co-authors: Giuseppe Cavaliere, Carsten Trenkler

The role of deterministic components and initial values in bootstrap likelihood ratio type tests of co-integration rank is investigated. A number of bootstrap procedures have been proposed in the recent literature some of which include estimated deterministic components and non-zero initial values in the bootstrap recursion while others do the opposite. To date, however, there has not been a study into the relative performance of these two alternative approaches. In this paper we fill this gap in the literature and consider the impact of these choices on both OLS and GLS de-trended tests, in the case of the latter proposing a new bootstrap algorithm as part of our analysis. Overall, for OLS de-trended tests our findings suggest that it is preferable to take the computationally simpler approach of not including estimated deterministic components in the bootstrap recursion and setting the initial values of the bootstrap recursion to zero. For GLS de-trended tests, we find that the approach which includes a restricted estimate of the deterministic component in the bootstrap recursion, can improve finite sample behaviour further.

C402: **Summability of stochastic processes: A generalization of integration and co-integration valid for non-linear processes**

Presenter: **Jesus Gonzalo**, U. Carlos III de Madrid, Spain

Co-authors: Vanessa Berenguer-Rico

The order of integration is valid to characterize linear processes; but it is not appropriated for non-linear worlds. We propose the concept of summability (a re-scaled partial sum of the process being Op(1)) to handle non-linearities. The paper shows that this new concept, S(delta): (i) generalizes I(delta); (ii) measures the degree of persistence as well as of the evolution of the variance; (iii) controls the balancedness of non-linear regressions; (iv) co-summability represents a generalization of co-integration for non-linear processes. To make this concept empirically applicable asymptotic properties of estimation and inference methods for the degree of summability, delta, are provided. The finite sample performance of these methods is analyzed via a Monte Carlo experiment. The paper finishes with the estimation of the degree of summability for the Nelson-Plosser extended database.

C866: **Bootstrap sequential determination of the co-integration rank in VAR models**

Presenter: **Anders Rahbek**, University of Copenhagen, Denmark

Co-authors: Giuseppe Cavaliere, Robert M. Taylor

Determining the co-integrating rank of a system of variables has become a fundamental aspect of applied research in macroeconomics and finance. It is well known that standard asymptotic likelihood ratio tests for co-integration rank can be unreliable in small samples with empirical rejection frequencies often very much in excess of the nominal level. As a consequence, bootstrap versions of these tests have been developed. To be useful, however, sequential procedures for determining the co-integrating rank based on these bootstrap tests need to be consistent, in the sense that the probability of selecting a rank smaller than (equal to) the true co-integrating rank will converge to zero (one minus the marginal significance level), as the sample size diverges, for general I(1) processes. No such likelihood based procedure is currently known to be available. We fill this gap in the literature by proposing a bootstrap sequential algorithm which we demonstrate delivers consistent cointegration rank estimation for general I(1) processes. Finite sample Monte Carlo simulations show the proposed procedure performs well in practice.

C405: **Inference on parameters in cointegrated vector autoregressive models with non-stationary volatility**

Presenter: **Peter Boswijk**, University of Amsterdam, Netherlands

Co-authors: Giuseppe Cavaliere, Anders Rahbek, Robert Taylor

Given the well established fact that many key macro-economic and financial variables are subject to permanent changes in unconditional volatility, in this paper we consider estimation and hypothesis testing on the cointegrating relations in vector autoregressions with non-stationary (unconditional) volatility of a very general form, which includes single and multiple volatility breaks as special cases. We show that the conventional results that the Gaussian maximum likelihood estimator of the cointegrating vector are mixed normal, with the associated likelihood ratio tests for linear restrictions being asymptotically chi-squared, break down under permanent volatility changes. As a consequence, standard confidence intervals and tests of hypothesis on the cointegrating vectors are potentially unreliable. As a solution, we propose wild bootstrap inference methods which do not require the practitioner to specify a parametric model for volatility, nor to assume that the pattern of volatility is common to, or independent across, the vector of series under analysis. We formally establish that the wild bootstrap allows to replicate the relevant asymptotic distributions and that it has very good finite sample properties under a wide range of volatility models.

CS39 Room MAL B29 INSTRUMENT SELECTION Chair: Jan F. Kiviet

C424: **Exogeneity tests, weak identification and IV estimation**

Presenter: **Jean-Marie Dufour**, McGill University, Canada

Co-authors: Firmin Doko Tchatoka

We study the effects of weak identification on Durbin-Wu-Hausman (DWH) specification tests. We provide finite-sample and asymptotic theoretical results showing that DWH tests are robust to weak identification, but may lack power. However, contrary to previous findings, power can be substantial as soon as one instrument is strong. This suggests that DWH remain useful in practice. On IV estimation, we present evidence showing that: (1) OLS often performs better than IV; (2) pretest-estimators based on exogeneity tests exhibit the best performance. This suggests that IV estimation should be used only when both endogeneity and instruments are strong.

C588: **Evaluating the performance of tests of overidentifying restrictions**

Presenter: **Jerzy Niemczyk**, European Central Bank, Germany

In linear regression models with endogenous explanatory variables a researcher utilizes instruments that have known correlation with the error term and that provide necessary information for the consistent estimation of the parameters. In this paper, for a simple linear model, we investigate in

Monte Carlo setups several versions of classic GMM tests and Likelihood Ratio (LR) type tests based on GEL estimators for testing overidentifying restrictions. We investigate incremental versions of those tests, that is the difference between test statistics of the validity of a set of instruments and of a subset of those instruments. We examine several bootstrap implementations that are meant to size correct GMM type test statistics. We find that a version of the Sargan test, size corrected using bootstrap, performs very well (even in models with heteroscedastic errors) in comparison to the other tests we analyze. We find that the LR type tests have substantial size problems, which can be fixed possibly by using bootstrap techniques. Size corrected versions do not perform better than Sargan tests in the homoscedastic case, but for the heteroscedastic case they show some potential. The GMM (Sargan) tests seem to be the safest option for the model and circumstances we analyzed.

C092: **Wald-type tests for error-regressors covariances, partial exogeneity tests and partial IV estimation**

Presenter: **Firmin Doko Tchatoka**, McGill University and CIRANO, Canada

Co-authors: Jean-Marie Dufour

Testing whether a subset of explanatory variables is exogenous is an important challenge in econometrics. Standard exogeneity tests of the type proposed by Durbin-Wu-Hausman are unreliable and unusable for testing such hypotheses. The only procedure so far with correct size when testing partial exogeneity is the generalized Wald-test. However, such a test assumes that errors are Gaussian. Its properties are not established when errors are non Gaussian. This paper develops a new version of earlier test which is typically valid even when errors are non Gaussian. We present a Monte Carlo experiment which confirms our theoretical results. Moreover, we provide an analysis of the performance of different pretest-estimators based on Wald-type tests. Our analysis allows us to propose two new pretest-estimators which have a good performance over a wide range cases compared with usual IV estimators. Therefore, our procedure may be viewed as an instrument selection procedure where a Wald-type test is first undertaken to decide which variables should be instrumented and which ones are valid instruments.

C173: **The performance of tests on endogeneity of subsets of explanatory variables scanned by simulation**

Presenter: **Milan Pleus**, University of Amsterdam, Netherlands

Co-authors: Jan Kiviet

Tests for classification of regressor subsets as endogenous or predetermined are derived and examined. Correct classification is important because misclassification leads either to estimator inefficiency or inconsistency. The derivations formulate the tests as significance tests in auxiliary IV regressions. A simulation approach is developed by which relevant data generating processes are designed by setting values to its parameters indirectly by first fixing salient features of the relationship, viz.: the degree of simultaneity of individual explanatory variables, degree of multi-collinearity between explanatory variables, individual and joint strength of instruments for endogenous variables. This allows scanning the relevant parameter space of wide model classes for flaws in performance regarding type I and II errors of the tests and their bootstrapped versions. Besides synthetic models, we also simulate models for empirical data borrowed from the literature. Like bootstrapping tests requires resampling under the null, we find that testing for endogeneity by auxiliary regressions benefits from estimating variances under the null, as in Lagrange multiplier tests, rather than under the alternative, as in Wald tests. Also, one better avoids hybrid variance estimates resulting from mechanical implementation of the Hausman test principle. Testing endogeneity of subsets by standard methods is impeded for weakly identified regressors.

CS63 Room MAL B30 INDEPENDENT COMPONENT ANALYSIS Chair: David Veredas

C339: **Rank-based inference in independent component models**

Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium

Co-authors: Pauliina Ilmonen, Klaus Nordhausen, Hannu Oja

Independent component (IC) models are location-scatter models obtained by applying an invertible affine transformation to a noise vector with independent marginals. Such models constitute an interesting and quite flexible alternative to the celebrated elliptical models, which are generated in the same way from a noise vector that is spherically symmetric. We develop a methodology that can be used to address, in the special context of IC models, a broad class of classical problems from multivariate analysis—including the one-sample location problem, the one-sample scatter problem, the problem of testing independence between two subvectors, etc. The procedures we define combine efficiency, robustness, and validity under very mild distributional assumptions. Broad validity (root-n consistency for point estimation, null asymptotic size $alpha$ for hypothesis testing) is achieved by resorting to ranks—or more precisely, in the symmetric context we actually consider in this work, to signed-ranks. As for efficiency, it enters the picture through the locally and asymptotically normal structure of IC models. We describe our procedures for some particular inference problems, state their main theoretical properties, and investigate their finite-sample performances via simulations.

C496: **Estimation of multivariate stable processes with discrete spectral measure.**

Presenter: **Yves Dominicy**, Universite Libre de Bruxelles (U.L.B.) / E.C.A.R.E.S., Belgium

Co-authors: David Veredas

This paper deals with the estimation of multivariate α-stable distributions. To this end, three techniques will be put together. The proposed procedure combines the method of simulated quantiles (MSQ), the properties of multivariate α-stable distributions, and the estimation of independent component analysis (ICA). Indeed, by the properties of the multivariate α-stable distributions, a mix of vector of i.i.d. stable random variables is stable distributed as long as all the elements of the vector share the same tail index. The idea is to use scatter matrices to estimate the mixing matrix, then compute the ICA model, estimate the parameters of each independent component with MSQ and finally obtain the multivariate α-stable distribution. We report some simulation experiments and an empirical application to financial data.

C534: **Conditionally heteroskedastic dynamic factor models**

Presenter: **Matteo Barigozzi**, London School of Economics and Political Science, UK

Co-authors: Lucia Alessi

Our model combines factor and multivariate GARCH models. The information contained in large datasets is captured by few dynamic common factors, which we assume being conditionally heteroskedastic. After presenting the model, we outline a multi-step estimation technique which combines asymptotic principal components and multivariate GARCH. We also prove consistency of the estimated conditional covariances. We estimate this model on a large dataset for the euro area, including more than 200 variables. We provide a structural interpretation of the dynamic common factors and analyze their conditional volatilities and covolatilities.

C733: **A conditionally heteroskedastic independent factor model with an application to financial stock returns**

Presenter: **Ester Gonzalez-Prieto**, MPIDR, Germany

Co-authors: Antonio Garcia-Ferrer, Daniel Pena

We propose a new conditionally heteroskedastic factor model, the GICA-GARCH model, which combines independent component analysis (ICA) and multivariate GARCH (MGARCH) models. This model assumes that the data are generated by a set of underlying independent components

(ICs) that capture the co-movements among the observations (which are assumed to be conditionally heteroskedastic). The GICA-GARCH model separates the estimation of the ICs from their fitting with a univariate ARMA-GARCH model. Here we will use two ICA approaches to find the ICs: the first one estimates the components maximizing their non-Gaussianity, and the second approach exploits the temporal structure of the data. After estimating and identifying the common ICs, we fit a univariate GARCH model to each of them in order to estimate their univariate conditional variances. Then, the GICA-GARCH model provides a new framework for modelling multivariate conditional heteroskedasticity in which we can explain and forecast the conditional covariances of the observations by modelling univariate conditional variances of a few common ICs. We report some simulation experiments to show the ability of ICA to discover leading factors in a multivariate vector of financial data. Finally, we present an empirical application to the Madrid stock market where we evaluate the forecasting performance of the GICA-GARCH, and two additional factor GARCH models: the orthogonal GARCH and the conditionally uncorrelated components GARCH.

CS68 Room MAL G16 NEW DEVELOPMENTS ON GARCH MODELS I Chair: Jean-Michel Zakoian

C203: **Two-stage QML estimation of GARCH models and testing the efficiency**

Presenter: **Guillaume Lepage**, CREST, France

Co-authors: Christian Francq, Jean-Michel Zakoian

In generalized autoregressive conditional heteroskedastic (GARCH) models, the standard identifiability assumption that the variance of the iid process is equal to 1 can be replaced by an alternative moment assumption. We show that, for estimating the original specification based on the standard identifiability assumption, efficiency gains can be expected by using a quasi-maximum likelihood (QML) estimator based on a non Gaussian density and a reparameterization based on an alternative identifiability assumption. A test allowing to determine whether a reparameterization is needed, that is, whether the more efficient QMLE is obtained with a non Gaussian density, is proposed.

C638: **The power log-GARCH model**

Presenter: **Genaro Sucarrat**, BI Norwegian School of Management, Norway

Co-authors: Alvaro Escribano

Exponential models of autoregressive conditional heteroscedasticity (ARCH) are attractive in empirical finance because they guarantee the non-negativity of volatility, and because they enable richer autoregressive dynamics. However, the currently available models exhibit stability only for a limited number of conditional densities, and the available estimation and inference methods in the case where the conditional density is unknown hold only under very specific and restrictive assumptions. Here, we provide results and simple methods that readily enables consistent estimation and inference of univariate and multivariate power log-GARCH models under very general and non-restrictive assumptions when the power is fixed, via vector ARMA representations. Additionally, stability conditions are obtained under weak assumptions, and the power log-GARCH model can be viewed as nesting certain classes of stochastic volatility models, including the common ASV(1) specification. Finally, our simulations and empirical applications suggest the model class is very useful in practice.

C192: **The diffusion limit of dynamic conditional correlation models**

Presenter: **Francesco Violante**, FUNDP Namur, Belgium

Co-authors: Christian Hafner, Sebastien Laurent

We derive a class of diffusion approximations based on dynamic conditional correlation models. We consider a modified version of the standard DCC model, the consistent DCC model. We show that this model admits a degenerate diffusion limit characterized by a diffusion matrix of reduced rank. This result is due to collinearity of the innovations in the volatility and correlation dynamics. The limit diffusion of the cDCC model is therefore driven by a number of independent Brownian motions that is smaller than the number of stochastic differential equations in the system. We investigate different sets of conditions on the rate of convergence of the parameters which allow to recover other degenerate diffusion approximations characterized by time varying but deterministic marginal variances and/or correlations. We also show what type of models can be recovered as Euler approximation of the different diffusions. Our results are validated through a comprehensive Monte Carlo simulation exercise.

C705: **Detecting multiple change-points in GARCH models using penalized quasi-likelihood method**

Presenter: **Olivier Wintenberger**, University Paris Dauphine, France

Co-authors: Jean-Marc Bardet, William Knenge

This presentation is devoted to the off-line multiple change point detection in a semiparametric framework. The time series is supposed to belong to a large class of models including GARCH and TARCH models where the coefficients change at each instant of breaks. The different unknown parameters (number of changes, change dates and parameters of successive models) are estimated using a penalized contrast built on conditional quasi-likelihood. Under Lipshitzian conditions on the model, the consistency of the estimator is proved when the moment order r of the process satisfies $r \geq 2$. If $r \geq 4$, the same convergence rates for the estimators than in the case of independent random variables are obtained.

CS42 Room MAL B33 ECONOMIC AND FINANCIAL FORECASTING II Chair: Ana-Maria Fuertes

C557: **Modelling government bonds in the Australian fixed-income market**

Presenter: **Rui Chen**, the University of Sydney, Australia

Co-authors: Jiri Svec, Maurice Peat

Accurately forecasting interest rates is of fundamental interest to both academics and market practitioners. In this paper, we study the predictive ability of several models in forecasting the government bond yields in Australia. We compare three different term structure models: the Diebold and Li (DL) model, the Functional Signal plus Noise (FSN) model, and the classic affine model. To provide a more powerful and robust method of term structure approximation, we propose a state-space specification with maximum likelihood (ML) parameter estimation. Comparing to standard estimation procedures, our models provide a more stable and robust result. Moreover, using the Root Mean Square Error (RMSE) metric, we demonstrate that all three models outperform a random walk. which is an improvement over existing literature. We compare the DL, FSN and affine model using the RMSE and loss functions from pairs trading algorithms criteria to identify the best forecasting model in Australia. These results are important for monetary policy makers, portfolio managers and individual financial planners.

C503: **An artificial neural network based heterogeneous panel unit root test: Application to exchange rates**

Presenter: **Christian de Peretti**, University Claude Bernard Lyon 1, France

Co-authors: Mario Cerrato, Carole Siani

We propose an artificial neural network (ANN) based panel unit root test, extending the neural test to a dynamic heterogeneous panel context, and following the panel methodology. This extension is not straightforward, since results are not provided on the distribution as well as on the moments

of their individual neural test statistic. In addition, in the context of panel data, cross sectional dependence has to be accounted for. The cases of uncorrelated as well as autocorrelated error terms are considered. We also propose to use the stochastic simulation based numerical method named *bootstrap* to compute the small sample distribution of the test statistics. Small sample properties of the proposed tests are investigated and compared to a reference test via Monte Carlo experiments. Bootstrapping significantly improves the performance of the neural test. The neural model can then be used to forecast data. An application to a panel of bilateral real exchange rate series with the US Dollar from the 20 major OECD countries is provided to examine the Purchase Power Parity (PPP) property.

C495: **Modelling and forecasting Italian state budget expenditures**

Presenter: **Tatiana Cesaroni**, Ministry of Economy and Finance, Italy
Co-authors: Ottavio Ricchi, Giuseppe Bianchi

We evaluate the usefulness of using high frequency data coming from budget state financial accounting in order to improve the annual forecasts accuracy of the Italian budget expenditures. In search for leading indicators we use a newly available data set of Italian State budget monthly financial microdata. Early work on the issue is encompassed with the provision of a macroeconometric model for the budget cycle linking various budget phases (i.e. appropriations, expenditures commitments and payments). Relying on start of year buget law appropriation, used as a target reference value, we provide forecasts for budget payments and we compare the results with those given by benchmark monthly ARIMA models. Secondly, we combine disaggregated forecasts results for current and capital expenditures in order to provide an aggregate forecast of total budget expenditure. To evaluate the forecast ability of our models we perform an out of sample forecast exercise based on recursive and rolling estimates. The forecasting performance evaluation shows that the models augmented with financial accounting data outperform the pure autoregressive models in terms of RM(F)SE. Furthermore we find that the combination of disaggregated forecasts improves the individual prediction.

C232: **Exchange rate pass-through revisited: what drives it.**

Presenter: **Ana Maria Fuertes**, Birkbeck College, UK
Co-authors: Kate Phylaktis, Raphael Brun-Aguerre

A comprehensive investigation of short-run pass-through from exchange rates into import prices is provided. The analysis is based on a relatively large cross section of both developed and emerging economies over the period 1980-2006, which enables us to examine the differences in pass-through between these two groups of countries. By relying on effective exchange rates this study avoids the biases related to the use of bilateral exchange rates as in much of earlier work. Furthermore, our study distinguishes itself from previous ones in exploiting effective export price indices constructed using the relative bilateral trade as weights instead of relying on domestic price indices, as in most other studies, which do not describe the behaviour of export prices. On the basis not only of in-sample goodness of fit criteria but also out-of-sample forecasting RMSE and MAE criteria, an exercise not attempted before in the literature, we demonstrate that ECM formulation(s) which include the long-run cointegrating relationship between the import price, export price and exchange rate outperforms the first-differences formulation employed in much of the literature. Our in-sample and out-of-sample forecasting frameworks highlight the nonlinear (time-varying) behaviour of import pass-through as dictated by a number of drivers including the exchange rate volatility, GDP, inflation, world economic activity sentiment, and the sign and size of exchange rate changes.

Saturday 11.12.2010 11:10 - 13:15 Parallel Session G – ERCIM

ES68 Room MAL 151 DATA ANALYSIS I **Chair: Klea Panayidou**

E896: **Adaptive PORT-MVRB extreme value index estimation: A comparison of heuristic algorithms**
Presenter: **M. Ivette Gomes**, University of Lisbon, Portugal
Co-authors: Ligia Henriques-Rodrigues

A comparison of two data-driven heuristic procedures of estimation of a positive extreme value index (EVI) is considered. Thus, we are concerned with heavy right tails. The semi-parametric EVI-estimators under consideration are the so-called PORT-MVRB EVI-estimators. These are location-invariant estimators, based on the peaks over random threshold (PORT) methodology applied to second-order minimum-variance reduced-bias (MVRB) EVI-estimators. We consider the use of two heuristic algorithms for the adaptive choice of k and q due to the stability on k of the MVRB EVI-estimates and as well as of the PORT-MVRB-estimates for adequate values of q. Both algorithms are based on the bias pattern of the EVI-estimators as a function of k. In the first algorithm we work with the original PORT-MVRB estimators. In the second algorithm, we perform a simple smoothing operation of the estimates. Applications to simulated data sets and to real data sets in the field finance will be provided.

E342: **The role of latent factors in budget allocation in the development of human capital in Trinidad and Tobago over time**
Presenter: **Linda Hewitt**, Centre for Interdisciplinary Research and Development, Trinidad/Tobago

The influence of latent factors such as attitude and preference has not been as widely studied in the social domain as in the field of psychology where tools for their measurement have been devised. Undoubtedly, these latent factors are known to influence decision making. So as to determine the extent of their intensity and interaction effect with their observed counterparts, the employment of a mixture of statistical methods and structural equation models have yielded encouraging results. Preliminary examination of Trinidad and Tobago data has revealed a consistent bias in monetary allocation towards physical infrastructure well beyond human capital. Notable shortfalls of the latter are evident along with severe social problems that are now proving very difficult to resolve.This paper taken from a larger study aims to determine the strength and direction of these latent factors so as to gain knowledge that will lead to informed decision making regarding the allocation of resource. A measure of preference is initially derived by undertaking paired comparisons on the basis of data viewed longitudinally over time. Next, measures of attitude and preference are derived by employing a latent variable structural equation model.

E736: **A longest run test for heteroscedasticity in univariate regression model**
Presenter: **Samuela Leoni-Aubin**, Insa-Lyon, France
Co-authors: Jean-Baptiste Aubin

Heteroscedasticity represents a serious problem in statistics and has to be taken into consideration when performing any econometric application that could be affected by the latter. The scope of this communication is the presentation of a test that enables to detect heteroscedasticity in univariate regression model. The test is simple to compute and very general since no hypothesis is made on normality of errors. Consider a n-sample data (x_i, Y_i) following the fix-designed regression model:$Y_i = \mu(x_i) + \sigma(x_i)e_i$. We are interested in testing the null hypothesis of homoscedasticity. We assume that the errors (e_i) of the model are a sequence of unobserved independent random variables (not necessarily normal) with continuous probability density functions symmetric with respect to 0 and such that $E(e_i^2) = 1$. Suppose that μ is known. We define for all i squared residuals $\varepsilon_i^2 := (Y_i - \mu(x_i))^2$. Let's consider the sequence in which the ith term is equal to 1 if ε_i^2 is greater than is the median of the squared residuals and 0 if ε_i^2 is smaller. The test statistic is the length L_n of the longest run of 0's or 1's of the previous sequence. The exact distribution function of the test statistic under the null hypothesis is given here. Simulations show that the presented test fairs well with respect to other nonparametric tests for homoscedasticity.

E187: **Survival analysis in LGD modeling**
Presenter: **Jiri Witzany**, University of Economics, Czech Republic
Co-authors: Michal Rychnovsky, Pavel Charamza

An application of the survival time analysis methodology to estimations of the Loss Given Default (LGD) parameter is proposed. The main advantage of the survival analysis approach compared to classical regression methods is that it allows exploiting partial recovery data. The model is also modified in order to improve performance of the appropriate goodness of fit measures. The empirical testing shows that the Cox proportional model applied to LGD modeling performs better than the linear and logistic regressions. In addition a significant improvement is achieved with the modified "pseudo" Cox LGD model.

ES25 Room MAL B36 INFORMATION FUSION AND CLASSIFICATION USING BELIEF FUNCTIONS **Chair: Thierry Denoeux**

E789: **CEVCLUS : constrained-evidential clustering of proximity data**
Presenter: **Violaine Antoine**, Universite de Technologie de Compiegne, France
Co-authors: Benjamin Quost, Marie-Helene Masson, Thierry Denoeux

We study the integration of background knowledge in the search of a partition of objects starting from their dissimilarities. Recent works have demonstrated the interest of taking side information into account, such as known associations between objects. This knowledge may be formalized as pairwise constraints: a Must-Link constraint specifies that two objects are in the same cluster, and a Cannot-Link constraint that they are in different clusters. We provide a new algorithm, CEVCLUS, an extension of EVCLUS, an algorithm which was formerly developed to derive a credal partition from relational data. A credal partition extends the existing concepts of hard, fuzzy (probabilistic) and possibilistic partition by allocating, for each object, a "mass of belief", not only to single clusters, but also to any subsets of clusters. This additional flexibility was shown to bring a deeper insight into the data. To introduce must-link and cannot-link constraints in EVCLUS, we first formulate them in the belief functions framework. Then, we introduce them in the algorithm as constraints on the partition. Some experiments study the efficiency of the proposed method.

E839: **Sensor fusion on a hierarchical structure using the theory of belief functions**
Presenter: **Yaxin Bi**, University of Ulster, UK

Sensor fusion based on uncertain, incomplete and sometime inaccurate information is necessary whenever any system interacts in an intelligent way with its environment. This is exactly the case for activity recognition systems in smart home environments. More specifically if an intelligent system is to built for multi-sensor integration, where the information available from a variety of sensors such as video camera, infrared badges, fingerprint receiver, etc. is inherently uncertain or incomplete, the system has to carry out reasoning or decision making according to the degrees

of partial belief. In this paper we describe an evidential framework developed for multi-sensor integration within smart home environments. We first formulate multi-sensors deployed in the smart home onto a hierarchical structure and then present a set of computational algorithms for belief propagation with the hierarchical structure for activity recognition.

E817: **A comparison between the transferable belief and the imprecise reliability models based on Monte Carlo simulations**
Presenter: **Mohamed Sallak**, Compiegne University of Technology, France
Co-authors: Walter Schon, Felipe Aguirre

Most methods in reliability and quantitative risk assessment assume that the precise probability distributions of the components failures are available. However, this assumption is usually wrong particularly in systems with scarce failures (railways systems, nuclear systems, chemical process, etc.). Therefore, a reliability model based on the Transferable Belief Model (TBM) is proposed in the paper. This model is compared to the probabilistic models based on Monte Carlo simulations. The TBM is an interpretation of the Dempster-Shafer theory which can be considered a generalization of classical probability and possibility theories. The present probabilistic approach consists in treating the components failure probabilities as random variables represented by a specified distributions (log-normal, normal, log-uniform...). It uses Monte-Carlo sampling simulations to repeatedly sample components failure probabilities from the appropriate distributions and calculate the system's reliability. We study the differences between the two approaches in several configurations (series, parallel, series-parallel and bridge). A case study is provided to illustrate the difference between the two approaches.

E629: **Using imprecise and uncertain information to enhance the diagnosis of a railway device**
Presenter: **Zohra Cherfi**, INRETS - UTC, France
Co-authors: Latifa Oukhellou, Thierry Denoeux, Patrice Aknin

The results of a study showing the interest of fusing partial and unreliable information to improve the diagnosis of a railway infrastructure device are presented. The general diagnosis approach is based on the noiseless Independent Factor Analysis (IFA) model. It assumes that the observed variables extracted from the inspection signal are generated by a linear mixture of independent latent variables linked to the system component states, each individual latent variable being modeled by a mixture of Gaussians. Usually, learning with this statistical model is performed within an unsupervised framework in which only unlabeled samples are used. The idea investigated here is to handle this learning process in a partially supervised learning framework involving imperfect information on the system component states. A large amount of data is indeed available but labeling all the data would be very expensive. Instead, we chose to present only a subset of data to human experts and to build labels form the combination of their different opinions within Evidential theory. These labels that can be subject to imprecision and uncertainty are introduced afterwards in the learning task. We show here that fusing partial and unreliable information on the cluster membership of some samples, can significantly improve classification results as compared to the unsupervised context.

E829: **Deriving continuous belief functions from empirical ROC spaces**
Presenter: **Pierre-Emmanuel Dore**, DRDC Valcartier, Canada
Co-authors: Anne-Laure Jousselme, Patrick Maupin, Arnaud Martin

We consider the problem of sensor placement where a set of sensors must be positioned over a terrain while maximizing the coverage, i.e. the global expected detection performance of the network. Individual sensors' performances are represented by 3-dimensional Receiver Operating Characteristic (ROC) curves including false positives (Pf), true positives (Pd) and distance to target (d). The global network coverage is then obtained by a combination of the local performances of the individual sensors. In practice, ROC curves are obtained by processing real data, estimating only a small number of operating points, and applying then an extrapolation. This decision step corresponding to the extrapolation may have a high impact in practice especially when several ROC curves will be combined, and consequently may modify the sensor placement solution. We propose here to model sensors' performance by continuous belief functions in the 3-dimensional ROC space (Pd, Pf, d). The network coverage is then obtained by combination of the individual ROC curves followed by a decision step, resulting in a 2-dimensional ROC space over the terrain to be further used in the optimization algorithm. We show the impact of such a modelization on the sensor placement solutions obtained compared to traditional approaches.

ES38 Room MAL 152 GOODNESS OF FIT Chair: M. Carmen Pardo

E067: **Assessing ARMA models for stationary time series by transforming accumulated residues processes**
Presenter: **Alejandra Cabana**, Universitat Pompeu Fabra, Spain
Co-authors: Enrique Cabana

The transformation of processes has shown to be a fruitful tool in developing goodness-of-fit tests for independent samples and regression models. In the first case, the process to be transformed is the empirical process of the sample, and in the second one, a process of accumulated residues. In both cases, the transformation provides a new process able to detect any departure from the null hypothesis of model fit, and customised to show efficiently the departures in a given direction chosen by the user. The convergence under the null hypothesis of fit of the empirical process to a Brownian bridge (or its L^2 projection on a known subspace when parameters are estimated) plays a central role in the heuristic and technical arguments applied for the transformation. When dealing with regression models, the accumulated residues process has similar asymptotic properties. We show that in the case of ARMA models, the behaviour of a marked residues process leads to construct consistent tests for ARMA(p,q) models, focused on ARMA$(p+1,q)$ and ARMA$(p,q+1)$. The resulting tests are compared in terms of power with other included in the statistical literature, including an AR tests based on the mth root of the determinant of the mth autocorrelation matrix and a test based on the discrepancy between the standardized spectral distribution and its sample estimate proposed.

E260: **Tube formula approach to testing multivariate normality and testing uniformity on the sphere**
Presenter: **Akimichi Takemura**, University of Tokyo, Japan
Co-authors: Satoshi Kuriki

Applications of tube method to null distributions of some maximum type test statistics are presented for testing multivariate normality and for testing uniformity on the sphere. For multivariate normality, we consider projection pursuit indices based on higher-order cumulants. For the uniformity on sphere we consider Anderson-Stephens statistic and its variant.

E301: **A class of goodness-of-fit tests based on transformation**
Presenter: **M. Dolores Jimenez-Gamero**, Universidad de Sevilla, Spain
Co-authors: Simos G. Meintanis, Virtudes Alba-Fernandez

A class of goodness-of-fit tests which are based on the empirical characteristic function (ecf) are considered. The test statistic of each test in this class is a weighted integral of the squared modulus of the difference between the ecf and a parametric estimator of the characteristic function of the

model in the null hypothesis, which is assumed to be composite. Under some conditions, the resulting test is consistent against fixed alternatives and its null distribution can be consistently approximated through a parametric bootstrap. A problem with these tests is the computation of the test statistic, which requires the analytical calculation of some expectations which depend on the weight function and on the null model. To overcome this problem, assuming that the distribution generating the data is absolutely continuous with respect to Lebesgue measure on R^d, in this work we consider an estimated Rosenblatt transformation of the data. The tests based on the transformed data have similar properties to those based on the original data. The advantage of the transformation is that, for not too much restrictive choices of the weight function, the expression of the resulting test statistic is readily computable.

E453: **Multinomial goodness-of-fit tests under inlier modification**
Presenter: **Ayanendranath Basu**, Indian Statistical Institute, India
Co-authors: Abhijit Mandal

The Pearson's chi-square and the log-likelihood ratio chi-square statistics are fundamental tools in multinomial goodness-of-fit testing. The power divergence family includes both statistics as special cases. This family is indexed by a single parameter, and divergences at either end of the scale are more powerful against alternatives of one type while being rather poor against the opposite type. Several other families of divergences available in the literature also show similar behaviour. We present several inlier control techniques in the context of multinomial goodness-of-fit testing which generates procedures having reasonably high powers for both kinds of alternatives. We explain the motivation behind the construction of the inlier modified test statistics, establish their asymptotic null distributions and explore their performance through simulation and real data examples to substantiate the theory developed.

E210: **Testing for the Semicircle Law**
Presenter: **Konstantin Glombek**, University of Cologne, Germany

Let X be a $d \times n$ array of centered i.i.d. random variables with variance 1 and zero excess, $S = \frac{1}{n} XX^t$ be the sample covariance matrix and denote the identity matrix as I. The talk will focus on the question of how large the sample length n has to be compared to the dimension d of the sample so that the estimator S is not biased due to d being too large compared to n. The empirical eigenvalue distribution of $\sqrt{\frac{n}{d}}(S-I)$ is investigated for this purpose. Since this distribution converges almost surely to the semicircle law as $n, d \to \infty$ and $d/n \to 0$, tests for the validity of the semicircle law are provided. The test decisions mainly depend on the ratio d/n. Simulation studies show that tests with large power require $n = O(d^3)$ to obtain a proper estimation of the true covariance matrix by the sample covariance matrix. Tests with low power require $n = O(d^2)$ instead.

ES09 Room MAL 532 OPTIMIZATION HEURISTICS IN ESTIMATION AND MODELLING Chair: Peter Winker

E515: **Rebalancing triggers for the CPPI using genetic programming**
Presenter: **Tikesh Ramtohul**, WWZ Uni Basel, Switzerland
Co-authors: Dietmar Maringer

The Constant Proportion Portfolio Insurance (CPPI) technique is a dynamic capital-protection strategy that aims at providing investors with a guaranteed minimum level of wealth at a specified time horizon. It gives an investor the ability to limit downside risk while allowing some participation in upside markets. Despite its widespread popularity in the industry, this strategy is not bereft of risk for the issuer. The inability to meet the guarantee at maturity, commonly referred to as gap risk, is a major concern. Gap risk could be limited by increasing the frequency of trading (rebalancing) but this has the adverse effect of generating more transaction costs. One way of achieving a tradeoff between gap risk and rebalancing frequency is through the use of rebalancing triggers, which constitutes the main focus of this paper. We use a genetic programming (GP) approach to obtain bounds of tolerance around the value of the multiplier implied by the portfolio composition. The experiments focus on GBM and GARCH price processes, and two different types of fitness functions, namely Tracking Error and Sortino ratio. We investigate the performance of the GP-based rebalancing strategy for different parameter settings and find that it yields better results than calendar-based rebalancing strategies in most scenarios.

E511: **New evidence on variable selection with stochastic optimization algorithms**
Presenter: **Antonio Gargano**, Bocconi/University of California San Diego, Italy

Four stochastic optimization algorithms (Brute Force, Random Search, Multiple Random Search and Genetic Algorithm) are adopted to perform model selection in a linear regression framework when the number of candidate variables is such that full enumeration of all possible models is impossible. A comparison both in absolute and relative terms is made by means of two Monte Carlo simulations. To provide a more realistic picture a final application to real data is fulfilled. Algorithms' performance in maximizing well known in-sample and out-of-sample fit functions are compared using different criteria. The main conclusion is that performance depends on the problem complexity and on the criteria used to judge it. In absolute terms Genetic Algorithm is the most powerful, robust and efficient among the algorithms tested. Nevertheless, local search algorithms represent a good alternative whether one is interested in just satisfactory solutions, or in finding a good compromise between the quality of the solution and computational effort.

E530: **Genetic algorithm for trade scheduling optimization**
Presenter: **Wei Cui**, University College Dublin, Ireland
Co-authors: Anthony Brabazon, Michael O'Neill

The optimization of large-order trade execution especially the trade scheduling problem is examined. The optimization of large-order trade execution has traditionally been considered as a difficult dynamic optimization problem, because traders need to adapt the trade execution strategies to changing market conditions and price changes over time in order to minimize trading costs. With the growth of institutional trading in financial markets, it has received increasing attention in recent years. Due to its complexity this problem has been split into two, optimal trade scheduling and optimal order execution. Traditional mathematical programming approaches for the trade scheduling problem always derive approximate solutions by simplifying this problem. In practice, this approximation error can be significant in highly volatile markets thus affecting trade execution performance. Although heuristic methods are popularly applicable to investment problems, it has not been used for the trade scheduling optimization. In this study, we apply Genetic Algorithm to solve the trade scheduling problem. Our results show that this method can evolve quality solutions to the trade scheduling problem.

E703: **Estimating stable distributions with genetic algorithms**
Presenter: **Christian Lau**, Martin Luther University Halle-Wittenberg, Germany

There already exist different ways of estimating the parameters of a stable distribution. In this paper we use a genetic algorithm. In terms of risk management, the left tail of a distribution is the most import region. The advantage of our approach is that during the estimation process more

weight can easily be put on the tail. In this way we receive a closer fit to the empirical data here. We analyse the estimated distribution in terms of goodness and calculate the Value at Risk. In comparison to the existing methodology, we find our approach to give superior results.

E743: **Data modeling using the nuclear norm heuristic**
Presenter: **Ivan Markovsky**, University of Southampton, UK

A general problem for approximate data fitting by low-complexity linear models is considered. Computationally the problem is structured low-rank approximation or equivalently rank minimization. Except for a few special cases, however, those are NP hard optimization problems. From the two major types of solution methods—local optimization methods and convex relaxations—a convex relaxation method based on the nuclear norm heuristic is used. The method can deal with an arbitrary affine structure and approximation norm, and allows regularization and inequality constraints on the approximation. Such a diverse list of features is currently not achievable by alternative methods. Literate programs, implementing the method in Matlab and results on test examples with missing data and outliers are presented.

ES11 Room MAL 509 STATISTICAL SIGNAL EXTRACTION AND FILTERING I **Chair: Manfred Deistler**

E320: **Semiparametric inference in correlated long memory signal plus noise models**
Presenter: **Josu Arteche**, University of the Basque Country, Spain

We propose an extension of the log periodogram regression in perturbed long memory series that accounts for the added noise, also allowing for correlation between signal and noise, which represents a common situation in many economic and financial series. Consistency (for $d < 1$) and asymptotic normality (for $d < 3/4$) are shown with the same bandwidth restriction as required for the original log periodogram regression in a fully observable series, with the corresponding gain in asymptotic efficiency and faster convergence over competitors. Local Wald, Lagrange Multiplier and Hausman type tests of the hypothesis of no correlation between the latent signal and noise are also proposed.

E461: **Variance profile**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Tommaso Proietti, Marco Reale

The variance profile (VP) is introduced as a tool for characterising a stationary stochastic process. The VP is defined as the power mean, or Hölder mean, of the spectral density function of the process. If p denotes the power parameter, for $p = -1$ it provides the interpolation error variance (harmonic mean), i.e. the estimation error variance when the value of the process at time t is predicted from the past and future observations. For $p = 0$ it provides the one-step-ahead prediction error variance (geometric mean, which is the usual Szëgo-Kolmogorov formula); whereas for $p = 1$ one obtains the unconditional variance of the process (arithmetic mean). Simple proofs of these results are derived and the connections with the predictability of a stochastic process are explored. Non parametric estimation of the VP is considered and the variance profiles of some processes, such as the fractional noise process, are derived.

E642: **The effect of misspecification in models for extracting trends and cycles**
Presenter: **Davide Delle Monache**, University of Cambridge and University of Rome - "Tor Vergata", UK
Co-authors: Andrew Harvey

Issues involved in specifying the trend component in an unobserved components (UC) model are firstly reviewed. The conditions under which the maximum likelihood (ML) estimators in a misspecified model converge to fixed values are obtained and these values are computed for a range of models. The robustness of various specifications is assessed by comparing the loss in terms of MSE of filtered and smoothed estimators of the trend using a recent algorithm. The paper first explores filtering and smoothing efficiencies for simple trend plus error models. Cycles are then brought into the analysis and efficiency and robustness with respect to smoothing and filtering are assessed. The comparison of trend-cycle models with correlated and uncorrelated disturbances yields some interesting insights. For example, a misspecified model with uncorrelated components can still yield a filter with high efficiency. Thus while the interpretations of the cycle obtained from smoothed estimates may differ substantially, the current estimates of the output gap may be very close.

E777: **Heart rate variability monitoring in critical illness: a time-frequency approach**
Presenter: **Ana Paula Rocha**, Universidade do Porto and CMUP, Portugal
Co-authors: Renato Soeiro

Information on time and frequency domain current indexes of Heart Rate Variability (HRV) can be a useful non-invasive tool, allowing access to the clinical evolution of patients, namely those being monitored at intensive care units. Analyzing how spectral components evolve over time requires an approach gathering joint frequency and time information in order to access non-stationary events. We applied this approach on recordings, collected from children with acute brain injury at Pediatric Intensive Care Unit of Hospital de S. Joao, Porto; the HRV series are extracted using a previously validated multilead wavelet detector. Due to the nature of HRV data in intensive care, artifacts and non-stationarities are very common, therefore filtering such as impulse rejection and advanced detrending has to be applied preceding the time-frequency processing. In this work we use TF-methods, namely the Smoothed Pseudo-Wigner-Ville transform to track and localize spectral changes. A special attention is given to the sensibility to the choice of the kernel and the other parameters involved in the data pre-processing.

E313: **Bayesian stochastic model specification search for seasonal and calendar effects**
Presenter: **Tommaso Proietti**, University of Rome "Tor Vergata", Italy
Co-authors: Stefano Grassi

We apply a Bayesian model selection technique, known as stochastic model specification search, for characterising the nature of seasonality and calendar effects in macroeconomic time series. We illustrate that the methodology can be quite successfully applied to discriminate between stochastic and deterministic trends, seasonals and trading day effects. In particular, we formulate stochastic models for the components of an economic time series and decide on whether a specific feature of the series, i.e. the underlying level and/or a seasonal cycle are fixed or evolutive.

ES52 Room MAL B35 ROBUST METHODS, COMPUTATION, AND APPLICATIONS **Chair: Peter Filzmoser**

E082: **Some thoughts on the aggregation of variables in dissimilarity design**
Presenter: **Christian Hennig**, UCL, UK

One way of analysing complex data is to define a dissimilarity measure and to use dissimilarity-based methodology such as dissimilarity-based clustering or k-nearest neighbour methods. The question then arises how to define the dissimilarities. Here we consider how to aggregate variables in order to define dissimilarity measures in high-dimensional data sets or data sets with mixed type variables. Arising questions concern the

standardisation of variables (e.g., they could be standardised by range, variance, or a robust scale statistic such as the MAD), aggregation method (e.g., Euclidean vs. Manhattan) and variable weighting (this is nontrivial at least for aggregating mixed type variables, e.g., interval, ordinal and nominal scaled ones). The general philosophy behind the presentation is that there is no objectively optimal solution to these problems and it is important to understand properly what the different approaches do and imply in order to make a well informed application-based decision about them. Furthermore, the interplay between dissimilarity design and the statistical method that is afterwards applied to the dissimilarities will be discussed.

E123: **Robust model selection in the social sciences**

Presenter: **Andreas Alfons**, Vienna University of Technology, Austria

Co-authors: Peter Filzmoser

Motivated by applications in the social sciences, a robust model selection method, combined with a strategy to reduce the number of selected predictor variables to a necessary minimum, has been developed. For better interpretability in the context of social sciences, the set of explanatory variables needs to be very small and strong dependencies among the regressor variables need to be eliminated. The fulfillment of these requirements will therefore be called context-sensitivity and the resulting strategy for variable selection can be considered a tradeoff between quality of the model and interpretability. The proposed procedure led to highly interpretable models in practical applications. A simulation study further verified that primarily only variables with potentially new information are included in the resulting model.

E135: **Computational information geometry and robustness**

Presenter: **Frank Critchley**, Open University, UK

Co-authors: Karim Anaya-Izquierdo, Paul Marriott, Paul Vos

A very short introduction to emerging geometries for statistical science is offered, followed by an overview of computational information geometry. The role of model-robustness, as well as data-robustness, is emphasised. Examples illustrate the development.

E139: **Biomarker identification in metabolomics**

Presenter: **Ron Wehrens**, Fondazione Edmund Mach, Italy

Co-authors: Pietro Franceschi, Urska Vrhovsek, Fulvio Mattivi

Knowledge on which metabolites are indicative of class differences is important for a better understanding of underlying biological processes. To achieve this, several approaches can be used. Among the most popular are the strategies based on PLSDA, either by assessing the absolute size of the model coefficients or by looking at derived statistics such as the variable-importance measure. These necessitate some form of validation, in order to determine the optimal number of latent variables. Univariate statistics based on t-tests are potentially simpler to apply but still need appropriate cut-off points, and do not use correlation information. We discuss several alternative and more robust strategies for biomarker discovery, based on the stability of the candidate biomarker set under perturbations. We evaluate the performance of these using a real LCMS data set obtained by spiking apple extracts with known metabolites at different concentrations, in order to mimic real-life experimental conditions. The scope of the data has been extended by using these apple data for simulations of much larger data sets with similar characteristics. In this way, we can assess the effect of the variable-to-sample ratio, which in metabolomics usually is quite large on the biomarker identification.

E161: **EPP-Lab: a JAVA interface for exploratory projection pursuit**

Presenter: **Anne Ruiz-Gazen**, Universite Toulouse 1, France

Co-authors: Alain Berro, Souad Larabi Marie-Sainte

Exploratory Projection Pursuit is a powerful methodology to explore multivariate data sets and detect some structures such as outliers or clusters. The theory has been studied in detail more than twenty years ago but the method is still not used in practice and we believe that an interactive implementation is necessary to make the methodology more popular. The EPP-Lab interface we propose includes several projection pursuit indices, some of them are dedicated to the detection of outliers and others to clusters. The optimization algorithms we have implemented are bio-inspired (genetic algorithms and particular swarm optimization) and do not require any regularity condition on the index function. The use of the interface is divided in two phases. First, the optimization process is launched several times. This process may be costly in terms of computation time for very large data sets but does not require the statistician to be in front of the computer. Then, many potentially interesting views can be analyzed by the statistician by means of several graphical tools available in the interface. This interface is still a work in progress and we hope that the conference participants will help us to improve the existing version by comments and remarks.

ES18 Room MAL 421 INTEGER TIME SERIES I Chair: Robert Jung

E157: **Bayesian nonparametric inference for INAR time series models**

Presenter: **Michael Wiper**, Universidad Carlos III de Madrid, Spain

Co-authors: Ana Justel

The INAR(p) time series model is of form $X_t = \sum_{i=1}^{p} Y_t(X_{t-i}, \alpha_i) + \varepsilon_t$ where $Y_t(n, p)$ represents a binomially distributed variable with parameters n and p and ε_t is an independent error term distributed on the non-negative integers. Typically, the error distribution is modeled parametrically as, e.g. a Poisson or geometric distribution, but here we prefer to use a Bayesian, nonparametric approach. We illustrate using simulated data that a naive approach based on the direct modeling of the error distribution via Dirichlet processes fails here and show that a Dirichlet process mixture model should be preferred. Applications of our approach to real real data sets are given.

E249: **A Monte Carlo study on estimation and forecasting in higher-order INAR(p) models**

Presenter: **Henriette Reinhold**, University of Erfurt, Germany

Co-authors: Robert Jung

Stationary time series models for counts with order higher than unity featuring independent binomial thinning operations are considered. Due to a number of appealing features, this so called INAR class of models has gained considerable attraction in recent years and a range of methods for the (efficient) estimation of the model parameters has been proposed. To date, however, no systematic comparison of these methods in small (medium) sized samples is available. This paper seeks to fill this gap by focusing on second-order models and by considering moment-based (MM) as well as (non-)parametric maximum likelihood (MLE) methods. In a comprehensive simulation study a series of distributional assumptions for the innovations of the INAR model are employed along with parameterisations over a wide range of the parameter space resulting in 78 data-generating processes (dgp). A multivariate mean-squared error criterion is applied to evaluate the finite sample behaviour of the estimators studied. Furthermore, probabilistic forecasts are assessed using proper scoring rules, among others. Overall, the performance of estimators is determined by an interplay of characteristics of the dgp with no uniform dominance of a single method. Parametric MLE assuming the wrong distribution does

not necessarily perform worse than MM or non-parametric MLE, even in larger samples. In empirically relevant parameter settings MLE based on flexible distributional assumptions turns out to be generally superior to non-parametric MLE.

E528: **Bayesian detection of additive outliers in INAR(1) models**

Presenter: **Maria Eduarda Silva**, Universidade do Porto, Portugal

Co-authors: Isabel Pereira

Outliers are commonly encountered in time series data analysis. When the models under study are linear ARIMA models, additive outliers, innovational outliers and other exogenous interventions like level changes and temporary changes may be represented under a unique dynamic model and there is now a voluminous amount of literature dealing with the detection and estimation of exogenous interventions in ARIMA processes and in particular with AO and IO outliers. Here, a Bayesian approach to the problem of detecting additive outliers in Poisson Integer-valued AutoRegressive models is considered. MCMC methods are used to estimate the probability of outlier occurrence at each time point. The proposed methodology is illustrated using simulated and real examples.

E436: **Detection of outliers in integer-valued AR models**

Presenter: **Isabel Silva**, FEUP and CIDMA, Portugal

Co-authors: Maria Eduarda Silva

During the last decades there has been an increasing interest in integer-valued time series models and a considerable volume of work is now available in specialized monographs. Among the most successful integer-valued time series models proposed in the literature are the first-order integer-valued autoregressive (INAR(1)) process. Although this model attempts to resemble the structure and properties of the usual linear AR models, it allows the choice of marginal distributions among a wide class of discrete distributions. Their statistical and probabilistic properties have been obtained in the literature. A problem of interest in time series modeling is to detect outliers, which can be viewed as discrepant observations, in order to assess for data quality and to study the robustness of the statistical inference in the presence of inconsistent observations. In this paper, we consider INAR models contaminated with additive and innovative outliers. We use a method based on wavelets, which are a family of basis functions used to localize a given function in both space and scaling, in order to address the problem of identifying the time point of the outlier.

E584: **Bayesian Inference in SETINAR(2;p,q) model**

Presenter: **Raquel Nicolette**, Universidade de Aveiro, Portugal

Co-authors: Isabel Pereira, Manuel Scotto

We considerer the class of self-exciting threshold integer-valued autoregressive models (SETINAR) driven by independent Poisson-distributed random variables. This paper considers the parameters estimation problem for the self-exciting threshold integer-valued autoregressive models with two regimes, denoted by SETINAR(2;p,q), where p and q are the orders of the integer-valued processes involved. The Markov chain Monte Carlo (MCMC) algorithm is used and its inference capabilities are analysed through a simulation study.

Saturday 11.12.2010 15:15 - 16:55 Parallel Session I – CFE

CS51 Room MAL B33 BAYESIAN ECONOMETRICS AND APPLICATIONS II: ECONOMETRICS Chair: Toshiaki Watanabe

C770: **Sequential detection of changes**

Presenter: **Omiros Papaspiliopoulos**, Universitat Pompeu Fabra, Spain

Co-authors: Nicolas Chopin, Pierre Jacob

We consider dynamic semi-Markov representations of change point models as a framework both for predicting non-stationary financial and macroeconomic time series but also for testing certain economic hypotheses. In particular, we investigate sequential estimation and prediction of such models under various specifications of the duration distribution and the form of the local model within each regime. In the talk we review deterministic algorithms for on-line inference, providing links to the literature, results about their computational complexity and their limitation in terms of the type of dependence between different regimes. We propose a new sequential Monte Carlo algorithm, related to the recent Particle MCMC framework which can estimate sequentially unknown parameters and change points under general model specifications.

C237: **Generalized multiple-point algorithms for approximate Bayesian computation**

Presenter: **Genya Kobayashi**, Kobe University, Japan

Co-authors: Hideo Kozumi

The well-known problems of the approximate Bayesian computation (ABC) algorithm include sensitivity to starting values, inefficiency and low acceptance rate. To overcome these problems, this paper develops the generalized multiple-point (GMP) algorithm for the ABC. The GMP extends the multiple-point Metropolis algorithm, which proposes multiple dependent candidates, and includes the generalized multiple-try algorithm as a special case. The GMP selects one point among the multiple candidates based on the weighting function which may be arbitrary chosen. Using the simulated data in a sequence of simulation studies, it is demonstrated that the GMP substantially improves the ABC algorithm on those problems. A real data example is also presented using the exchange rate data.

C308: **Characterizing economic trends by Bayesian stochastic model specification search**

Presenter: **Stefano Grassi**, University of Perugia, Italy

Co-authors: Tommaso Proietti

We apply a recently proposed Bayesian model selection technique, known as stochastic model specification search, for characterising the nature of the trend in macroeconomic time series. We illustrate that the methodology can be quite successfully applied to discriminate between stochastic and deterministic trends. In particular, we formulate autoregressive models with stochastic trends components and decide on whether a specific feature of the series, i.e. the underlying level and/or the rate of drift, are fixed or evolutive. The contribution to the stochastic model specification search is the inclusion of autoregressive terms into the model selection problem as well as the application to a large data set consisting of the original Nelson and Plosser series and a set of key economic indicators of the U.S. economy. The results shows that most annual time series in the Nelson and Plosser data set are better characterised as trend stationary. However, the posterior distribution of the sum of the autoregressive coefficients is in some cases highly concentrated on the boundary of the stationary region, leading to a quasi unit root process. Finally we conclude that the consideration of the autoregressive component is essential for the characterisation of the selected model.

C601: **A new method for the evaluation of dynamic stochastic general equilibrium models**

Presenter: **Toshiaki Watanabe**, Hitotsubashi University, Japan

Recent years have seen a surge in the econometric analysis of dynamic stochastic general equilibrium (DSGE) models using an MCMC Bayesian method. Since DSGE models are structured from micro-foundation theories, they can identify various shocks in a theoretically consistent way. It is important to evaluate the DSGE model before applying it to the policy analysis because the results may depend on the model. This article develops a new method for the evaluation of DSGE models. One of the methods used for the evaluation of DSGE models is the approach using the DSGE-VAR model. In this approach, a DSGE model is used as a prior for a VAR model and the fit is compared by relaxing the cross-equation restrictions implied by the DSGE model systematically. The problem with this approach is that a VAR model is only an approximation of a DSGE model because a DSGE model cannot be represented by a VAR model. Since DSGE model can be represented by a state-space (SS) model, we use a SS model and select its prior using a DSGE model. The resulting DSGE-SS model is compared with the DSGE-VAR model by applying them to a simple standard New Keynesian DSGE model.

CS16 Room MAL B29 BAYESIAN MODEL AVERAGING Chair: Gianni Amisano

C379: **Combining predictive densities using a Bayesian nonlinear filtering approach**

Presenter: **Roberto Casarin**, University of Brescia, Italy

Co-authors: Monica Billio, Francesco Ravazzolo, Herman van Dijk

We are concerned with the problem of combining different predictive densities in a multivariate setting. We follow a Bayesian framework and propose a general combination approach based on a distributional state space representation of the predictive densities and of the combination scheme. In the proposed approach the weights are time-varying and have random fluctuations and both parameters and model uncertainty are taken into account in the combination. We also discuss several multivariate combination schemes and assume the weight dynamics are possibly driven by the past forecasting performances of the predictive densities. We provide some applications to macroeconomics time series.

C387: **Core inflation, model averaging and structural instability**

Presenter: **Francesco Ravazzolo**, Norges Bank, Norway

Co-authors: Oyvind Eitrheim, Shaun Vahey

We construct ensemble predictive densities for core inflation based on the out of sample forecast performance of component models allowing for structural breaks for a set of five different countries, Australia, Germany, Norway, the UK and the US. Each component model is based on information from a particular disaggregate inflation series where we allow for one or more mean shifts. The median of ensemble predictive densities can be interpreted as a forecast-based measure of core inflation. We demonstrate that ensemble forecast densities for measured headline CPI inflation a) are based on information from multiple disaggregate inflation series and b) take properly account that mean shifts are well calibrated. The resulting ensemble densities often outperform aggregate autoregressive benchmarks and forecasts based on more traditional core inflation definitions as trimming and ex-measures.

C820: **Optimal prediction pools in macroeconomics**

Presenter: **Gianni Amisano**, European Central Bank, University of Brescia, Germany

Co-authors: John Geweke

We consider the properties of weighted linear combinations of prediction models, or linear pools, evaluated using the log predictive scoring rule. Although exactly one model has limiting posterior probability, an optimal linear combination typically includes several models with positive weights. We derive several interesting results: for example, a model with positive weight in a pool may have zero weight if some other models are deleted from that pool. The results are illustrated using different popular models in macroeconomics, such as a DSGE model, a dynamic factor model and a VAR. We also show how using Bayesian and frequentist estimation procedure might lead to different combination schemes.

C522: **Model selection versus model averaging for forecasting economic time series**

Presenter: **Rolf Scheufele**, Halle Institute for Economic Research (IWH), Germany

One of the biggest problems we face in econometrics is the uncertainty about the correct specification of our models. Therefore, this paper compares alternative modeling strategies for the specification of ARDL models, which are frequently used in forecasting economic outcomes. Model selection strategies are presented that are based on the sequential elimination of lags (general-to-specific), a specific-to-general strategy and sequential strategies based on information criteria (SC, AIC, AICC, HQ, PC and R^2). Furthermore, we use pooling techniques that consider various models and the different outcomes based on the respective forecasting performance. I choose simple weighting schemes such as the average as well as weighting schemes based on the proposed information criteria and Bayesian techniques. In a Monte Carlo simulation we find that sequential model selection strategies mostly fail in uncovering the true model for small and moderate sample sizes. Model averaging schemes that incorporate all possible lag combinations provide a clear improvement in terms of forecasting performance. In a real time application for German GDP we further document the superiority of model averaging compared to a model selection approach. Generally, this study recommends model averaging due to model uncertainty even when the data generating process is stable over time.

CS22 Room MAL B34 BAYESIAN METHODS IN MACROECONOMICS AND FINANCE II Chair: Andrea Carriero

C826: **Evolving macroeconomic dynamics in a small open economy**

Presenter: **Haroon Mumtaz**, Bank of England, UK

Co-authors: Philip Liu

This paper carries out a systematic investigation into the possibility of structural shifts in the UK economy using a Markov-switching dynamic stochastic general equilibrium (DSGE) model. We find strong evidence for shifts in the structural parameters of several equations of the DSGE model. In addition, our results indicate that the volatility of structural shocks has also changed over time. However, a version of the model that allows for a change in the coefficients of the Taylor rule and shock volatilities provides the best model fit. Estimates from the selected DSGE model suggest that the mid-1970s were associated with a regime characterised by a smaller reaction by the monetary authorities to inflation developments.

C294: **Forecasting with medium and large Bayesian VARs**

Presenter: **Gary Koop**, University of Strathclyde, UK

This paper is motivated by the recent interest in the use of Bayesian VARs for forecasting, even in cases where the number of dependent variables is large. In such cases, factor methods have been traditionally used, but recent work using a particular prior suggests that Bayesian VAR methods can forecast better. In this paper, we consider a range of alternative priors which have been used with small VARs, discuss the issues which arise when they are used with medium and large VARs and examine their forecast performance using a US macroeconomic data set containing 168 variables. We find that Bayesian VARs do tend to forecast better than factor methods and provide an extensive comparison of the strengths and weaknesses of various approaches. Our empirical results show the importance of using forecast metrics based on the entire predictive density, instead of relying solely on those based on point forecasts.

C720: **Methods for computing marginal data densities from the Gibbs output**

Presenter: **Leonardo Melosi**, London Business School, UK

Co-authors: Cristina Fuentes-Albero

Marginal Data Densities (MDDs) are used for Bayesian model selection, model averaging, and prior selection. This paper introduces two new methods for estimating the MDD for Vector AutoRegressive (VAR) models, Vector Equilibrium Correction Model, Markov-Switching VARs, Time-Varying Parameter VARs, Dynamic Factor Models, and Factor Augmented VARs (FAVARs). Our two estimators take advantage of the fact that a few parameter blocks can be integrated out analytically from the conditional posterior densities in those models. An empirical application on VAR models shows that our methods improve upon Chib's method in both estimation accuracy and computation time.

C885: **QUEB-VARs: QUick and Easy Bayesian VARs**

Presenter: **Andrea Carriero**, Queen Mary University of London, UK

Co-authors: Todd Clark, Massimiliano Marcellino

We discuss how to make the computation of BVAR forecasts quick and easy, for example by making specific choices on the priors and by using direct rather than iterated forecasts. We then assess whether speed and simplicity have a cost in terms of decreased forecast precision with respect to more general BVAR specifications. We also address a set of other empirically relevant related issues, such as the choice of the lag length of the BVAR, whether or not to transform the variables to get stationarity, and how to deal with possible changing volatility. We obtain a large set of empirical results, but we can summarize them by saying that we find very small losses (and sometimes even gains) from the adoption of quick and easy BVAR modelling choices. This finding could therefore further enhance the diffusion of the BVAR as an econometric tool for a vast range of applications.

CS36 Room MAL 151 BEHAVIOURAL FINANCE I Chair: Gulnur Muradoglu

C250: **Venture capitalist/entrepreneur financial contracting and performance: the effects of positive and negative reciprocity**

Presenter: **Richard Fairchild**, University of Bath, UK

We analyze the economic and behavioural factors affecting venture capital/entrepreneur relationships and performance. We develop a behavioural game-theoretic model in which double-sided moral hazard problems and reciprocal behaviour interacts to affect the value-creating ability of a venture capital dyad. In economic terms, we consider double-sided ex ante effort-shirking, and double-sided ex post hold-up/renegotiation. In behavioural terms, we consider reciprocity, both positive (fairness, trust, empathy) and negative (retaliation, anger, spite). We also consider the

impact of overconfidence (in fairness) and anxiety. We analyse the conditions under which such reciprocity may enhance or reduce value-creation.

C548: **Decision avoidance and deposit interest rate setting**
Presenter: **Robert Hudson**, Newcastle University, UK
Co-authors: John Ashton, Robert Anderson

This study examines why banks offer multiple and similar deposit services with different interest rates. This question is examined both theoretically and empirically by considering the influence of decision avoidance and the use of different quasi-hyperbolic and exponential discounting functions by customers with varying degree of decision making sophistication. The model examines if the number of deposit services and the levels of interest rates are influenced by banks exploiting customers' decision avoidance and different forms of discounting the costs and benefits of future decisions. As customers will have different propensities to switch deposit accounts according to the degree they discount quasi-hyperbolically, banks can exploit these biases to enhance profits by introducing duplicate new deposit products with competitive interest rates for new customers and lowering interest rates on existing deposit services. These predictions are tested empirically; high levels of duplicate products are reported conforming to the model predictions. Further there is strong evidence of pricing heterogeneity and that this is linked to institution type, suggesting distinct costs arise as a result of consumer inertia.

C225: **The 2007-? financial crisis: a money market perspective**
Presenter: **Claudio Morana**, Universita del Piemonte Orientale, Italy
Co-authors: Nuno Cassola

The evolution of the spreads between unsecured money market rates of various maturities and central banks' key policy rates has been subject to considerable debate and controversy in relation to the worldwide financial market turbulence that started in August 2007. Our contribution to the ongoing debate on the dynamics of money market spreads is empirical and methodological, motivated by the "shocking" evidence of non-stationary behaviour of money market spreads. In fact, in our view, empirical work assessing the effectiveness of central bank policies has largely overlooked the complexity of the market environment and its implications for the statistical properties of the data. Thus, our main goal is to carefully document both the economic and statistical "fingerprint" of money market turbulence, in the framework of a novel Fractionally Integrated Heteroskedastic Factor Vector Autoregressive (FI-HFVAR) approach, carefully accounting for the persistence properties of the data.

C863: **Outliers, market portfolio risk and the estimation of betas and other risk measures for stocks**
Presenter: **Panayiotis Theodossiou**, Cyprus University of Technology, Cyprus

The issue of which estimation method is the most appropriate and whether it should be resistant to outlier returns is of great importance to financial researchers and practitioners involved in the estimation of cost of capital for individual companies, the determination of rates in regulated industries, the computation of inputs used in portfolio construction and the investigation of stock reactions to company specific events. This paper employs a mixed return regression model with a regular and an outlier component to examine the economic impact, if any, of outlier returns on the total risk of a market portfolio. The regular and outlier components of stock returns are estimated using Huber's Robust M estimation method, which is shown to be consistent with the mixed return process employed. The theoretical and empirical results show that outlier returns do not affect the market portfolio risk, thus, Huber's Robust M is an appropriate estimation method of betas and other risk measures for individual stocks.

CS40 Room MAL B20 SHORT PANEL DATA MODELS Chair: Jan F. Kiviet

C316: **Bootstrap bias-correction in dynamic linear panel data models with predetermined regressors**
Presenter: **Takashi Yamagata**, University of York, UK

A simple bootstrap-bias correction of the first-differenced (DIF) and the system (SYS) GMM estimators is proposed. Their finite sample behaviour is investigated, especially with many instruments and under the weak instrument problem. The Monte Carlo evidence shows that the bootstrap bias-correction reduces the bias of the DIF and SYS GMM estimators successfully even under many instruments and/or weak instruments problems. Particularly, when the ratio of the variance of individual effects to that of idiosyncratic errors rises, the bootstrap bias-corrected DIF GMM estimator can outperform (bootstrap bias-corrected) SYS GMM estimators in terms of root mean square errors. In addition, when the mean stationarity assumption is violated, the SYS GMM estimator becomes inconsistent while the DIF GMM estimator remains consistent. Therefore, in such situations, the bootstrap bias-corrected DIF GMM estimator can be a reliable alternative.

C483: **Bias adjustment of the profile score for spatial dynamic panel models with fixed effects and small T**
Presenter: **Geert Dhaene**, K.U.Leuven, Belgium
Co-authors: Maryam Giahi

Maximum likelihood estimates of slope parameters in linear spatial (static or dynamic) panel models with unit-specific fixed effects are subject to an incidental parameter problem. Specifically, the MLE is inconsistent as the number of spatial units, N, grows large while the number of time periods, T, remains fixed. We calculate the large N, fixed T bias of the profile score. This bias turns out to depend only on the slope parameters and, therefore, allows a re-centering of the profile score. For any T, the re-centered profile score equation is asymptotically unbiased for the slope parameters and, so, allows fixed T consistent estimation. The idea of re-centering the profile score was applied to non-spatial dynamic panel models. The approach complements recent work which suggested GMM-based fixed T consistent estimation in the same setting.

C143: **Derivation and exploitation of the limiting distribution of the LSDV estimator in dynamic panel data models**
Presenter: **Jan Kiviet**, University of Amsterdam, Netherlands
Co-authors: Maurice Bun

The limiting distribution is derived and examined of the LSDV estimator for the dynamic panel data model, in particular for the case with fixed number of time periods T and infinite number of independent cross-section units N. Due to its inconsistency the limiting distribution, though normal, is not centered at the true parameter values, and has an uncommon asymptotic variance too. By simulation it is investigated whether the additional terms in the asymptotic variance can be used to improve the precision of variance estimation in finite samples. If this is the case, that would provide a key to assessing the variance of operational bias corrected LSDV estimators, which in many cases are known to have smaller mean squared error than IV and GMM. Results obtained hencefar are encouraging, and should lead to the development of appropriate and accurate inference techniques for bias corrected LSDV.

C451: **Dealing with discontinuities in series of the monthly Dutch labour force**
Presenter: **Jan van den Brakel**, Statistics Netherlands, Netherlands
Co-authors: Sabine Krieg

The Dutch Labour Force Survey (LFS) is based on a rotating panel design. Each month a sample of addresses is drawn and data are collected

of the residing households. Households are re-interviewed four times at quarterly intervals. Recently an estimation procedure that is based on a multivariate structural time series model is adopted to produce monthly official statistics about the labour force. This approach handles problems with rotation group bias and small sample sizes in an effective way and enables Statistics Netherlands to produce timely and accurate estimates about the situation on the labour market. Currently the LFS is redesigned. Such revisions generally have a systematic effect on the outcomes of a survey. In an ideal survey transition process, the systematic effects of the redesign are explained and quantified to keep series consistent and preserve comparability of the outcomes over time. Therefore the old and the new survey are conducted in parallel for a period of six months. In this paper it is discussed how the time series modelling approach applied to the LFS is adjusted to deal with discontinuities due to the redesign of the LFS. Information from the parallel run is included as prior information in this model.

CS66 Room MAL B30 DYNAMIC FACTOR MODELLING AND FORECASTING Chair: Arvid Raknerud

C198: **Estimation of factors for term structures with dependence clusters**

Presenter: **Dennis Philip**, Durham University, UK

This paper is the first to study the importance of accounting for clustering feature between term structure maturities while forecasting yield curves. We study the term structure of US zero coupon bonds and find the presence of two maturity clusters. We identify these maturity clusters using a hierarchical clustering algorithm. The first three principal components – level, slope, and curvature – driving the two maturity clusters are loosely dependent on each other. Dependence graphs (Chi-plots and recursive Kendall plots) show that these dependencies are also time-varying. Incorporating the clustering feature into a term structure factor model, we study the forecasting performance of the model. In this, we introduce a block dynamic framework to the dynamic Nelson and Siegel model by relaxing the assumption of common factor dynamics among clusters. The block dynamic framework produces a better in-sample fit and a better out-of-sample forecast performance. We further extend the empirical application to Libor-Swap term structure and reach a similar conclusion: that clustering is a vital feature that, if present in the term structure data, cannot be ignored when estimating common factors.

C593: **A comparison of sparse methods for factor forecasting**

Presenter: **Pilar Poncela**, Universidad Autonoma de Madrid, El Salvador

Co-authors: Julio Rodriguez, Julieta Fuentes

Dynamic factor models have been applied extensively for forecasting when high dimensional datasets are available. Within this framework, instead of using all available predictors, we focus on the choice of the appropriate ones, a specification issue that remains open in the literature. We consider the variable selection process as one possible refinement to factor forecast; in particular, we propose sparse techniques as, for instance, sparse partial least squares, as a way to improve the forecast efficiency. We develop several approaches to apply the sparse partial least squares for forecasting time series. We use the Stock and Watson data base in order to compare the forecasting performance of the sparse methods to those widely used nowadays as principal component and partial least square regressions. Our preliminary results indicate that the sparse models have a better forecasting ability than the considered alternatives.

C611: **The real effects of financial shocks: evidence from a structural factor model**

Presenter: **Lucia Alessi**, European Central Bank, Germany

We carry out a semi-structural analysis aiming at estimating the macroeconomic effects of different types of financial shocks. We estimate a Structural Factor Model for the euro area, which includes more than 200 quarterly variables. By using such a wide information set we are able to: 1) identify structural shocks which a small-scale VAR would not be able to retrieve; 2) avoid any variable selection bias; 3) exploit as many variables as we need to identify the shocks, and study their responses in a unified framework. We find that the euro area economy can be well described by five structural shocks, and assume that one of these shocks has a financial nature. To achieve identification of this shock, we use a mix of short-run, long-run and sign restrictions. In particular, we identify the effects of an equity price bust, a shock to house prices, and a credit crisis. We find that: 1) a drop in the equity market has only a relatively limited effect on real GDP growth and no clear effects on inflation; 2) a decline in the housing market has more severe consequences for growth; and 3) credit shocks have macroeconomic effects, which last for about one year.

C657: **Forecasting Euro-area macroeconomic variables using a factor model approach for backdating**

Presenter: **Ralf Brueggemann**, University of Konstanz, Germany

Co-authors: Jing Zeng

We suggest to use a factor model based backdating procedure to construct historical Euro-area macroeconomic time series data for the pre-Euro period. We argue that this is a useful alternative to standard contemporaneous aggregation methods. The paper investigates for a number of Euro-area variables whether forecasts based on the factor-backdated data are more precise than those obtained with standard area-wide data. A recursive pseudo-out-of-sample forecasting experiment using quarterly data and a forecasting period 2000Q1-2007Q4 is conducted. Our results suggest that some key variables (e.g. real GDP and inflation) can indeed be forecasted more precisely with the factor-backdated data.

C860: **Multivariate stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck processes: A quasi-likelihood approach**

Presenter: **Arvid Raknerud**, Statistics Norway, Norway

Co-authors: Oivind Skare

This paper extends the ordinary quasi-likelihood estimator for stochastic volatility models based on non-Gaussian Ornstein-Uhlenbeck (OU) processes to vector processes. Despite the fact that multivariate modeling of asset returns is essential for portfolio optimization and risk management – major areas of financial analysis – the literature on multivariate modeling of asset prices in continuous time is sparse, both with regard to theoretical and applied results. This paper uses non-Gaussian OU-processes as building blocks for multivariate models for high frequency financial data. The OU framework allows exact discrete time transition equations that can be represented on a linear state space form. We show that a computationally feasible quasi-likelihood function can be constructed by means of the Kalman filter also in the case of high-dimensional vector processes. The framework is applied to Euro/NOK and US Dollar/NOK exchange rate data for the period 2.1.1989-4.2.2010.

CS79 Room Senate Beveridge Hall NUMERICAL METHODS IN QUANTITATIVE FINANCE Chair: Daniel Kuhn

C833: **Hedging electricity swing options in incomplete markets**

Presenter: **Phebe Vayanos**, Imperial College London, UK

Co-authors: Wolfram Wiesemann, Daniel Kuhn

The deregulation of electricity markets renders public utilities vulnerable to the high volatility of electricity spot prices. This price risk is effectively mitigated by swing options, which allow the option holder to buy electric energy from the option writer at a fixed price during a prescribed time period. Unlike financial derivatives, swing options cannot be assigned a unique fair value due to market frictions. We determine upper and lower

bounds on the option value by under- and over-replicating the option's cash-flow stream with standard market products such as forward contracts. We formulate these bounding problems as robust dynamic optimization models, which we solve in polynomial decision rules via constraint sampling.

C287: **Stability and sensitivity analysis of stochastic programs with second order dominance constraints**
Presenter: **Huifu Xu**, University of Southampton, UK
Co-authors: Yongchao Liu

We present stability and sensitivity analysis of a stochastic optimization problem with stochastic second order dominance constraints. We consider perturbation of the underlying probability measure in the space of regular measures equipped with pseudometric discrepancy distance. By exploiting a result on error bound in semi-infinite programming, we show under the Slater constraint qualification that the optimal value function is Lipschitz continuous and the optimal solution set mapping is upper semicontinuous with respect to the perturbation of the probability measure. In particular, we consider the case when the probability measure is approximated by empirical probability measure and show the exponential rate of convergence of optimal solution obtained from solving the approximation problem. The analysis is extended to the stationary points when the objective function is nonconvex.

C535: **Parallel algorithms for multistage stochastic programs in finance**
Presenter: **Marc Steinbach**, Leibniz Universitat Hannover, Germany

We consider multistage stochastic programming models in scenario tree form, with application examples from computational finance and energy trading. Our solution approach combining interior point methods with tree-sparse KKT algorithms is well suited for parallelization. We describe a highly efficient portable implementation running on standard multi-core PC hardware, and a software tool under development that generates application-specific sparse parallel code automatically. Computational results will be presented for multiperiod mean-risk models in portfolio optimization.

C630: **Multi-stage stochastic interest rate management**
Presenter: **Ronald Hochreiter**, WU Vienna University of Economics and Business, Austria

A major part of a banks total interest rate risk is due to the position of non-maturing deposits. In this talk, a multi-stage stochastic programming model for managing this risk factor will be shown. The uncertainty is given both in the interest rate development as well as the volume of the specific product. Different deposit products from an Austrian retail bank will be used to show the applicability of the model.

CP04 Room Senate Crush Hall POSTERS IV Chair: Nicos Koussis

C683: **Effects of economic crises after 1990 on the Turkish insurance sector**
Presenter: **Esma Gaygisiz**, Middle East Technical University, Turkey
Co-authors: Pelin Ozbek

The micro-level and macro-level determinants of the premiums in the insurance sector in Turkey are investigated with a special emphasis on the effects of the recent economic crises. The macro-level analysis (with aggregated data) using a time series analysis approach and the micro-level analysis (with disaggregated data) employing a dynamic panel data approach indicate that the premiums are pro-cyclical and very sensitive to economic volatility measures, estimated with GARCH models. Also, interestingly, some macro-variables show different effects in the macro-level analysis when it is compared with the micro-level analysis.

C845: **Health care expenditure and income in the European countries: Evidence from panel data**
Presenter: **Felipa de Mello-Sampayo**, ISCTE-IUL, Portugal
Co-authors: Sofia de Sousa-Vale

The economic relationship between health care expenditure and income using a panel of european countries observed over the 1990s is considered. In particular, the non-stationarity and cointegration properties between health care spending and income is studied. This is performed in a panel data context controlling for both cross-section dependence and unobserved heterogeneity. Heterogeneity is handled through fixed effects in a panel model. The findings suggest that health care is a necessity rather than a luxury, with an elasticity much smaller than expected.

C446: **From consumer incomes to car ages: How the distribution of income affects the distribution of vehicle vintages**
Presenter: **Anna Yurko**, State University - Higher School of Economics, Russia

The relationship between consumer incomes and ages of the durable goods consumed is studied. At the household level, it presents evidence from the Consumer Expenditure Survey of a negative correlation between incomes and ages of the vehicles owned, controlling for the size of the vehicle stock. At the aggregate level, it constructs a dynamic, heterogeneous agents, discrete choice model with multiple vehicle ownership, to study the relationship between the distribution of consumer incomes and the distribution of vehicle vintages. Two versions of the model are solved, one with the restriction of at most one vehicle per agent and one with multiple vehicle ownership. For each version of the model, the parameters are calibrated to match vehicle ownership data for 2001. The moments of the income distribution are then varied to generate predictions for mean and median ages of vehicles and the results from the two versions of the model are compared. While these are mostly similar, some of the differences are quite illuminating.

C252: **Prediction intervals for the Gaussian autoregressive processes following the unit root tests**
Presenter: **Wararit Panichkitkosolkul**, King Mongkut's University of Technology North Bangkok, Thailand
Co-authors: Sa-aat Niwitpong

Recent works have indicated that the preliminary unit root test is a useful tool for improving an accuracy of a one-step-ahead predictor for an AR(1) process. This paper extends these mentioned concepts to prediction interval for the Gaussian autoregressive processes. We propose the methods to construct the simple prediction interval based on the residual model, PI_a, and the prediction intervals following the unit root tests, PI_{fi}. The unit root tests applied in this paper consist of the augmented Dickey-Fuller test, the Phillips-Perron test, and the Elliott-Rothenberg-Stock test. In addition, an expression of the coverage probability is derived and we found that the structure of the coverage probability is independent from the parameter of a random error, but it is a function of the autoregressive parameters only. The coverage probability and length of prediction intervals are compared through Monte Carlo simulation studies. Simulation results have shown that the proposed prediction intervals have minimum coverage probabilities 0.95 for almost all situations. Furthermore, the lengths of prediction intervals PI_{fi} are shorter than that of a prediction interval PI_a when the first-order autoregressive parameter value approaches one. A comparison of the proposed methods is also illustrated by using an empirical application.

C818: **Measurement of market impact functions**

Presenter: **Virmantas Kvedaras**, Vilnius University, Lithuania

Co-authors: Danas Zuokas

The issue is considered of estimation of a conditional market impact function related to any recorded action in an order-driven financial market. A functional series expansion based on some orthonormal basis functions is used to get a flexible estimator of the underlying market impact function conditioned upon a series of market state variables. We formulate conditions and propose a consistent estimator of market impact functions in a price formation model, where a conditional expectation of a price change during a period is represented as an outcome of all market impacts related to all the events that have happened in the considered time period and measured at the last moment of the period. The issues related to an empirical implementation and testing of some key assumptions of the model are also discussed.

C865: **Empirical evidences about hourly electricity prices in some European markets**

Presenter: **Paolo Chirico**, Universita di Torino, Italy

The study originates from a work about the economic risk analysis of a windpower plant to be built on an Italian site. The annual gain distribution is drawn by means of simulations of a business year. A crucial point is the simulation of the hourly sale prices of the electricity on the basis of a suitable time-series model. The model should take into account the typical features of the hourly electricity prices: (i) seasonality, particularly with daily and weekly periodicity; (ii) mean-reversion that can be formalised by including an AR(1) relation in the model; (iii) jumps and spikes that can be due to the difficulty of the electricity supply to match the demand; (iv) volatility clustering; (v) leptokurtic distribution. Analysing the 2008-2009 Italian PUN (national common price) of the electricity an easy seasonal AR-GARCH model with t-Student standardised innovation has been detected. The same type model has been tested with electricity prices of Spain and Norway markets. The model fits well both of the data. At last a singular simplification of the model is analysed: the model can be reduced to an AR(1)-GARCH model by means of an average of seasonal differences.

C898: **Disentangling crashes from tail events**

Presenter: **Sofiane Aboura**, University of Paris Dauphine, France

The study of tail events has become a central preoccupation for academics, investors and policy makers, given the recent financial turmoil. However, what differentiates a crash from a tail event? This article answers this question by taking a risk management perspective that is based on an augmented extreme value theory methodology with an application to the French stock market (1968-2008). Our empirical results indicate that the French stock market experienced only two crashes in 2007-2008 among the 12 identified over the whole period.

Saturday 11.12.2010 14:45 - 16:50 Parallel Session I – ERCIM

ES20 Room MAL 152 MATRIX COMPUTATIONS AND MULTIVARIATE DATA ANALYSIS Chair: Kohei Adachi

E230: **The SVMMaj majorization algorithm for Support Vector Machines with and without kernels**

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

Co-authors: Georgi Nalbantov, Hoksan Yip

Support vector machines have become one of the main stream methods for two-group classification. We have proposed SVMMaj, a majorization algorithm that minimizes the SVM loss function. A big advantage of majorization is that in each iteration, the SVM-Maj algorithm is guaranteed to decrease the loss until the global minimum is reached. Nonlinearity was achieved by replacing the predictor variables by their monotone spline bases and then doing a linear SVM. A disadvantage of the method so far is that if the number of predictor variables m is large, SVMMaj becomes slow. Here, we extend the SVMMaj algorithm in the primal to handle efficiently cases where the number of observations n is (much) smaller than m. We show that the SVM-Maj algorithm can be adapted to handle this case of $n \gg m$ as well. In addition, the use of kernels instead of splines for handling the nonlinearity becomes also possible while still maintaining the guaranteed descent properties of SVM-Maj.

E549: **Acceleration of convergence for the alternating least squares iteration**

Presenter: **Michio Sakakihara**, Okayama University of Science, Japan

Co-authors: Masahiro Kuroda

The alternating least squares (ALS) method is widely used to estimate parameters in a statistical model and determine a regression curve from a data set. The iteration is based on an interesting linearization. The convergence rate of the iteration is of first order. Therefore, applying some acceleration methods to the ALS is an attractive problem. Here, we discuss the vector epsilon acceleration for the ALS iteration, and the convergence and some properties of the accelerated ALS iteration by the vector epsilon.

E110: **Exploratory factor analysis of data matrices with more variables than observations**

Presenter: **Nickolay Trendafilov**, Open University, UK

Co-authors: Steffen Unkel

A new approach for exploratory factor analysis (EFA) of data matrices with more variables p than observations n is presented. First, the classic EFA model $(n > p)$ is considered as a specific data matrix decomposition with fixed unknown matrix parameters. Then, it is generalized to a new model, called for short GEFA, which covers both cases of data, with either $n > p$ or $p \geq n$. An alternating least squares algorithm (GEFALS) is proposed for simultaneous estimation of all GEFA model parameters. As principal component analysis (PCA), GEFALS is based on singular value decomposition, which makes GEFA an attractive alternative to PCA for descriptive data analysis and dimensionality reduction. Finally, the new approach is illustrated with Thurstone's 26-variable box data.

E352: **Singular value reparameterization with its applications to least squares computations**

Presenter: **Kohei Adachi**, Osaka University, Japan

Singular value reparameterization (SVR) refers to reparameterizing unknown parameter matrices by their singular value decomposition. SVR can allow us to solve some complicated least squares problems by alternately iterating simple steps, that is, [1] the orthogonal Procrustes rotation for obtaining singular vector parameters; [2] the solving of quadratic or quartic equations for obtaining singular value parameters. The two problems are illustrated for which SVR is used. One is the rank preserving computation in which the condition number of a matrix to be obtained is constrained to be less than or equal to a prescribed constant. The other is the joint Procrustes analysis for simultaneously matching component score and loading matrices to given target matrices, where its loss function is a function of a transformation matrix to be obtained and its inverse.

ES24 Room MAL G16 NONLINEAR DIMENSIONALITY REDUCTION Chair: Pedro Delicado

E239: **Dimensionality reduction: from PCA to recent nonlinear techniques**

Presenter: **John Lee**, Universite Catholique de Louvain, Belgium

Co-authors: Michel Verleysen

Dimensionality reduction is an old and yet unsolved problem, with many applications in data visualization, knowledge discovery, and machine learning in general. Our aim in this talk will be to review several developments in the field of dimensionality reduction, with a particular focus on nonlinear methods. As an introduction, we will point out some weird properties of high dimensional spaces, which will motivate the use of dimensionality reduction. Next, we will go back in time and start our review with a short reminder about well known techniques such as principal component analysis and multidimensional scaling. Our travel into time will also bring us to visit Sammon mapping and other methods based on distance preservation. Next, we will come across self-organizing maps and auto-encoders with bottleneck neural networks. Some spectral methods such as Isomap and locally linear embedding will be reviewed as well. A glance at recent methods based on similarity preservation such as stochastic neighbor embedding will close the survey. Finally, we will try to identify the relationships between the different approaches, and say a few words about quality criteria for dimensionality reduction techniques.

E639: **Inroads into data visualization using generative topographic mapping**

Presenter: **Alfredo Vellido**, Universitat Politecnica de Catalunya (UPC), Spain

Many research fields are becoming data intensive, which is an attractive challenge for data analysts from all fields. Often, the analyst's challenge stems from the high dimensionality of data. As a result, much effort has been put into the definition of dimensionality reduction (DR) techniques. Interpretability is one common requirement for DR. One of the most intuitive forms that interpretation can take is visualization, and significant inroads have been made in defining powerful DR visualization methods. Self-Organizing Maps (SOM) artificial neural networks are a paradigmatic example of nonlinear DR methods of this kind, successfully applied to anything from neuroscience to business. Over a decade ago, SOM was redefined within a probability theory framework as a model called Generative Topographic Mapping (GTM). This offspring of SOM and mixture models belongs to the active field of Statistical Machine Learning. Importantly, its probabilistic definition is an adequate framework for the definition of many extensions. A brief overview of some of them will be provided here, including, amongst others, variations for the analysis of multivariate time series, missing data imputation, outlier detection, and hierarchical clustering.The usefulness of this family of methods will be illustrated with examples of their application, mostly to biomedical problems.

E489: **Non-linear principal manifolds for bioinformatics applications**

Presenter: **Andrei Zinovyev**, Institut Curie, France

Co-authors: Alexander Gorban

Principal manifolds is a useful tool for analysis and visualization of multidimensional data. It has been shown to be particularly suitable for analysis of bioinformatics data where high-throughput data are abundant. We make an overview of application of non-linear data approximation techniques in bioinformatics applications and underline specific requirements for them. For methods based on mapping of data from multidimensional space onto non-linear manifolds, we formulate four simple criteria of the mapping quality. These criteria are mean square error of approximation, distance structure, point neighborhood and class compactness preservation. We demonstrate that application of non-linear mappings outperforms linear ones in concrete applications. We also introduce the notion of local uncertainty of non-linear mapping and show how to visualize and use it.

E547: **Elastic and pluriharmonic graphs as universal approximants**

Presenter: **Alexander Gorban**, University of Leicester, UK

Co-authors: Andrei Zinovyev

Karl Pearson invented principal component analysis (PCA) in 1901 and found 'lines and planes of closest fit to system of points'. Elastic maps approximate data by curves and surfaces with minimization of the weighted sum of the quadratic error of approximation with bending and stretching energies. This method is based on the compromise between the accuracy of approximation and the deviation of the approximant from the ideal affine manifolds. All these methods allow the expectation/maximization splitting algorithms with minimization of quadratic functionals at each step. We consider harmonic and pluriharmonic embeddings of graphs in the dataspace as ideal approximants for the datasets of complex topology and develop general expectation/maximization algorithms for construction of elastic graphs with compromise between the accuracy of data approximation and the deviation from the pluriharmonic embedding. The method is based on the topological grammars algorithms and on the three complexity restrictions: geometric complexity (deviation from pluriharmonic embedding), topological complexity of the graph and algorithmic complexity. Various applications in genomics and systems biology are presented.

E160: **A support convexity test with application to nonlinear dimensionality reduction**

Presenter: **Pedro Delicado**, Universitat Politecnica de Catalunya, Spain

Co-authors: Adolfo Hernandez, Gabor Lugosi

Let Y be a random vector with probability distribution μ on R^d having density f with support S. A test for the null hypothesis that S is a convex set is proposed. The consistency of the test is proven and a way to approximate the null distribution of the test statistic is discussed. In addition to the intrinsic relevance of the support convexity test, we present an application of such a test for the automatic choice of tuning parameters in some nonlinear dimensionality reductions methods. In particular, we consider the transformation ISOMAP. The key point is to choose a tuning parameter such that the configuration obtained from ISOMAP is compatible with the null hypothesis of support convexity. These proposals are illustrated with simulations which show the good behavior of both the support convexity test and the automatic choice of the parameter in ISOMAP algorithm.

ES06 Room MAL B36 STATISTICAL ALGORITHMS AND SOFTWARE II Chair: Uwe Ligges

E426: **topicmodels: An R package for fitting topic models**

Presenter: **Bettina Gruen**, JKU Linz, Austria

Co-authors: Kurt Hornik

Topic models allow the probabilistic modeling of term frequency occurrences in documents. The fitted model can for example be used to estimate the similarity between documents as well as between a set of specified keywords using an additional layer of latent variables which are referred to as topics. The R package topicmodels provides basic infrastructure for fitting topic models based on data structures from the text mining package tm. The package includes interfaces to two algorithms for fitting topic models: the Variational Expectation-Maximization algorithm and an algorithm using Gibbs Sampling. The functionality available in the package will be presented as well as the use illustrated.

E383: **A Bayesian filter framework for dual use**

Presenter: **Frank Kruger**, Universitat Rostock, Germany

Co-authors: Thomas Kirste

Bayesian state space estimation is used in various application such as activity recognition, intention analysis or localization. These applications employ Bayesian filters for inference. Adjusting parameters to fight effects like particle-clinging is a time consuming process during development of standalone applications that can benefit greatly from the data manipulation and visualizing capabilities of the R environment. We designed a Bayesian filter framework that performs statistical inference on dynamic models such as KF, (H)HMM and DBNs with discrete or continuous state spaces. The main objective of this framework is the separation of model definition and filtering algorithms. The framework itself is used as back-end in various environments and programming languages and provides implementations of several Bayesian filter such as Kalman filter, particle filter and Rao-Blackwellized particle filter. A front-end for this framework benefits from the speed, portability and model specific parameter representation that empowers users to leverage the features of their preferred environment for model analysis. We developed an R front-end that combines the advantages of our implementation with the rapid development process of the R environment. The presented framework enables users to experiment and visualize results within R and use the same code for online activity recognition.

E462: **Sequential implementation of Monte Carlo tests with uniformly bounded resampling risk**

Presenter: **Axel Gandy**, Imperial College London, UK

This talk describes an open-ended sequential algorithm for computing the p-value of a test using Monte Carlo simulation. It guarantees that the resampling risk, the probability of a different decision than the one based on the theoretical p-value, is uniformly bounded by an arbitrarily small constant. Previously suggested sequential or nonsequential algorithms, using a bounded sample size, do not have this property. Although the algorithm is open-ended, the expected number of steps is finite, except when the p-value is on the threshold between rejecting and not rejecting. An extension to multiple thresholds will be discussed. The algorithm is implemented in the R-package simctest, available on CRAN.

E562: **Efficient and safe computation of the power of Monte Carlo tests**

Presenter: **Patrick Rubin-Delanchy**, Imperial College London, UK

Co-authors: Axel Gandy

Hypothesis testing is an important statistical tool that is used in many fields of science. The threshold at which to reject, or the p-value, is often computed by Monte Carlo simulation, because the distribution of the test statistic is not available in analytical form. Prime examples of this are bootstrap or resampling-based tests. With these Monte Carlo tests, there is a new risk that not enough simulations have caused us to reject or accept

the null hypothesis, when after an infinite number we would have chosen to do otherwise. For practical applications it is important to know the power (the probability of rejecting for a specific alternative) of the test. Naive simulations, using essentially a double loop, are not efficient and may lead to biased results. In this talk, we introduce an algorithm that computes the power of a Monte Carlo test in a safe and efficient way.

E682: **Beta regression in R**

Presenter: **Achim Zeileis**, Universitat Innsbruck, Austria

Co-authors: Francisco Cribari-Neto, Bettina Gruen

The class of beta regression models is commonly used by practitioners to model variables that assume values in the standard unit interval (0, 1). It is based on the assumption that the dependent variable is beta-distributed and that its mean is related to a set of regressors through a linear predictor with unknown coefficients and a link function. The model also includes a precision parameter which may be constant or depend on a (potentially different) set of regressors through a link function as well. This approach naturally incorporates features such as heteroskedasticity or skewness which are commonly observed in data taking values in the standard unit interval, such as rates or proportions. Here, the betareg package is introduced which provides the class of beta regressions in the R system for statistical computing. Some applications are discussed as well as reuse of the software in extended models such as latent class regression or model-based recursive partitioning.

ES08 Room MAL B35 QUANTILE REGRESSION AND SEMIPARAMETRIC METHODS Chair: Hee-Seok Oh

E373: **Fast nonparametric quantile regression with arbitrary smoothing methods**

Presenter: **Hee-Seok Oh**, Seoul National University, Korea (Rok)

Co-authors: Thomas Lee, Doug Nychka

The calculation of nonparametric quantile regression curve estimates is often computational intensive, as typically an expensive nonlinear optimization problem is involved. This paper proposes a fast and easy–to–implement method for computing such estimates. The main idea is to approximate the costly nonlinear optimization by a sequence of well–studied penalized least–squares type nonparametric mean regression estimation problems. The new method can be paired with different nonparametric smoothing methods and can also be applied to higher dimensional settings. Therefore, it provides a unified framework for computing different types of nonparametric quantile regression estimates, and it also greatly broadens the scope of the applicability of quantile regression methodology. This wide-applicability and the practical performance of the proposed method are illustrated with smoothing spline and wavelet curve estimators, for both uni– and bivariate settings. Results from numerical experiments suggest that estimates obtained from the proposed method are superior to many competitors.

E568: **Modeling binary regression quantiles using Dirichlet process priors**

Presenter: **Dries Benoit**, Ghent University, Belgium

Co-authors: Dirk Van den Poel

This research elaborates on the recent work on Bayesian quantile regression for dichotomous dependent variables. The frequentist approach to this type of regression has proven problematic in both optimizing the objective function and making inference on the parameters. By accepting additional distributional assumptions on the error terms, the Bayesian method, based on the Asymmetric Laplace Distribution (ALD), sets the problem in a parametric framework in which these problems are avoided. However, as the frequentist approach to binary quantile regression is semi-parametric, comparison (Bayes vs. frequentist) is somewhat awkward. The ALD method for binary quantile regression can be generalized using a Dirichlet process mixture model. By doing so, the methodology proposed now is back in a semi-parametric framework. Relaxing the assumption of ALD distributed errors, makes that the shape of the error distribution could adapt to the data and thus provide better fit compared to parametric error distributions. Moreover, this reduces the risk of model misspecification. The current research also investigates whether the data is not pushed too far by applying this approach to binary data.

E586: **Prior elicitation in mixed quantile regression models**

Presenter: **Rahim Al-Hamzawi**, Brunel, UK

Co-authors: Keming Yu

A quantile dependent prior for Bayesian quantile regression is presented. We introduce the idea of the power prior distribution employing a likelihood function that is based on the asymmetric Laplace distribution. The advantage of the method is that the prior distribution is changing automatically when we change the quantile. Thus, we have prior distribution for each quantile and the prior is proper. In addition, we propose joint prior distributions using mixture of normal representation of the asymmetric Laplace distribution. The behavior of the power prior is clearly and quite robust with different weights for power parameter. The propriety of the power prior is one of the critical issues in Bayesian analysis. Thus, we discuss the propriety of the power prior in Bayesian quantile regression.

E595: **On fractile transformation of covariates in regression**

Presenter: **Bodhisattva Sen**, Columbia University, USA

Co-authors: Probal Chaudhuri

The need for comparing two regression functions arises frequently in statistical applications. Comparison of the usual regression functions is not very meaningful in situations where the distributions and the ranges of the covariates are different for the populations. For instance, in econometric studies, the prices of commodities and people's incomes observed at different time points may not be on comparable scales due to inflation and other economic factors. In this talk we describe a method of standardizing the covariates and estimating the transformed regression function, which now become comparable. We develop smooth estimates of the fractile regression function and study its statistical properties analytically as well as numerically. We also provide a few real examples that illustrate the difficulty in comparing the usual regression functions and motivate the need for the fractile transformation. Our analysis of the real examples leads to new and useful statistical conclusions that are missed by comparison of the usual regression functions.

E651: **On multivariate generalized spatial quantiles, regression and causality**

Presenter: **Snigdhansu Chatterjee**, University of Minnesota, USA

Co-authors: Nitai Mukhopadhyay

High dimensional data routinely arises in various research areas. In many such cases traditional assumptions like multivariate Normality may not be viable, also in several applications the dimension of the observations is extraordinarily high, but the sample size may or may not be high. We propose using spatial quantiles and its generalizations, in particular, the projection quantile, for describing, analyzing and conducting inference with multivariate data. We develop multivariate quantile regression and a quantile-based extension of Granger causality, based on the generalized spatial quantiles. Some theoretical results are presented, along with examples and applications of quantile-based causality networks.

ES61 Room MAL 421 INTEGER TIME SERIES II Chair: Roland Fried

E490: **Trading activity and liquidity supply in a pure limit order book market: An empirical analysis**

Presenter: **Andreas Heinen**, Universite Cergy-Pontoise, France

Co-authors: Erick Rengifo, Stefan Frey, Joachim Grammig

The interplay between liquidity supply and demand in a pure limit order book (LOB) market is studied using a multivariate autoregressive count model. Our econometric methodology combines flexibility with computational tractability. This allows to analyze in detail the interdependence of key market events, and to conduct a broad market study based on a cross section of stocks. Contrary to the predictions of static LOB theory, large market order submissions do not tend to be associated with the processing of private information. Instead, they attract competitive liquidity supply, a result that underscores the resiliency of a pure limit order market. While dynamic LOB theory is more successful in explaining the interaction of liquidity supply and demand, we also document stylized facts which are unaccounted for, most importantly the attraction of small market orders by opposite side visible depth. In line with recent experimental evidence, our results emphasize the informational content of limit order submissions and cancellations. We also show that visible and hidden liquidity present in the book exert different effects on trading activity. The possibility to hide part of a limit order volume is shown to fit the needs of a trader who wants to process a large order. It attracts liquidity demand from the opposite market side while ample visible depth (in line with predictions of dynamic LOB theory) increases own-side liquidity demand.

E540: **Estimation and Forecasting in INBL(1,0,1,1) Model**

Presenter: **Isabel Pereira**, University of Aveiro, Portugal

Co-authors: Nelia Silva

The analysis of count processes has become an important area of research in the last two decades. In this paper we consider the simple non-negative integer-valued bilinear process INBL(1,0,1,1). The parameter estimation is addressed and the problem of forecasting integer-valued time series modeled by the INBL(1,0,1,1) process is also considered. Bayesian methodology is used to obtain parameter estimates, point predictions and confidence intervals for future values of the process. The parameter estimates and predictions thus obtained are compared with their classic counterparts. The proposed approaches are illustrated with a simulation study and applied to a real data set.

E662: **Multivariate Poisson autoregression**

Presenter: **Konstantinos Fokianos**, University of Cyprus, Cyprus

We are considering linear and log-linear models for multivariate Poisson autoregressions. The model initially is specified by assuming that the conditional distribution of each component of the vector process given the past is Poisson with a dynamic evolving intensity process. Then we assume linear and log–linear GARCH type models for the intensity process. We study some properties of these models and discuss maximum likelihood estimation. The method is illustrated to real and simulated data sets.

E375: **Statistical inference for count time series data**

Presenter: **Peter Neal**, University of Manchester, UK

Co-authors: Victor Enciso-Mora, Tata Subba Rao

Integer valued time-series occur in many different situations, and often in the form of count data. Conventional time-series models such as the standard univariate autoregressive moving-average (ARMA) process with Gaussian errors, assume that the data take values in the real numbers. Therefore these standard models are wholly inappropriate for integer valued time-series, especially for low frequency count data where continuous approximations of the discrete process are particularly unsuitable. Hence there has been considerable interest in developing and understanding time-series models which are suitable for integer valued processes. One such model is the INARMA (integer-valued autoregressive moving-average) model. This is based upon the ARMA model but is constructed to take only integer values. Whilst this model is more applicable for count data than classical ARMA models it is harder to analyse. Classical statistical inferential approaches have mainly been restricted to the simplest such model, INAR(1). However MCMC (Markov chain Monte Carlo), via data augmentation, enables us to tackle INARMA models in full generality, including order selection and inclusion of explanatory variables. This talk will introduce INARMA models and discuss a simple but very effective MCMC algorithm to analyse such models. We will also discuss extensions of the model to include order selection and explanatory variables.

E442: **Parameter estimation for integer-valued self-exciting threshold autoregressive processes**

Presenter: **Magda Monteiro**, University of Aveiro, Portugal

Co-authors: Isabel Pereira, Manuel Scotto

The class of Self-Excited Threshold Integer-Valued Autoregressive processes driven by independent Poisson-distributed random variables is presented and its properties related with the existence of a stationary distribution. Parameter estimation is addressed in both classical and Bayesian perspectives. In the classical point of view conditional maximum likelihood estimation is used whereas in Bayesian perspective it is used MCMC with the Gibbs methodology with Metropolis step. A simulation study is presented to compare both estimation approaches.

ES55 Room MAL 509 ALGORITHMS FOR DESIGN Chair: Peter Goos

E407: **An algorithm for creating d-optimal exact designs for linear models on continuous factors**

Presenter: **Bradley Jones**, SAS Institute, USA

Most exact design algorithms use a discretization of the range for continuous factors. This simplifies the search for designs but it also makes the resulting designs generally sub-optimal. Using coordinate exchange combined with a one-dimensional continuous optimizer it is possible to generate optimal exact designs in polynomial time. This talk will describe how the algorithm works and demonstrate its performance in test cases where the globally optimal design is known.

E514: **Optimal design: metaheuristics can help**

Presenter: **Kenneth Sorensen**, University of Antwerp, Belgium

Determining an optimal design is essentially an optimization problem in which an objective function (the optimality criterion: D-, A-, G-, ...) is maximized subject to some constraints (e.g. a limit on the number of rows/runs in the design matrix), and in which decision variables represent the decisions to be taken (e.g. the order of the rows/runs, the value of each element in the design matrix, ...). Given the computational complexity of many optimal design optimization problems, heuristic optimization methods are generally used, characterized by the fact that they are not guaranteed to find the best possible design matrix. Although some of the proposed heuristics achieve excellent results, seen from the field of metaheuristics - an umbrella term that encompasses a wide range of heuristic optimization concepts - these algorithms can only be described as primitive. The aim of this talk is to bridge the gap between the field of metaheuristics and the field of optimal design by presenting an overview of

recent advances and insights into these more advanced heuristics. We postulate that some relatively simple concepts can be imported into existing heuristics for optimal design to generate much better designs in less computing time.

E597: **Algorithmic construction of optimal restricted discrete choice designs**
Presenter: **Roselinde Kessels**, Universiteit Antwerpen, Belgium
Co-authors: Bradley Jones, Peter Goos

In a discrete choice experiment each respondent typically chooses the best product or service sequentially from many groups or choice sets of alternatives which are characterized by a number of different attributes. Respondents can find it difficult to trade off prospective products or services when every attribute of the offering changes in each comparison. Especially in studies involving many attributes, respondents get overloaded by the complexity of the choice task. To overcome respondent fatigue, it makes sense to simplify the comparison by holding some of the attributes constant in every choice set. A study in the health care literature where eleven attributes were allocated across three different experimental designs with only five attributes being varied motivates the algorithm we present. However, our algorithm is more general, allowing for any number of attributes and a smaller number of fixed attributes. We describe our algorithmic approach and show how the resulting design performed in our motivating example.

E753: **Sequential procedures for KL optimization using trimmed K-means**
Presenter: **Paula Camelia Trandafir**, University of Valladolid, Spain
Co-authors: Alfonso Gordaliza

An appropriate experimental design allows one to use resources in the most efficient way, employing only those empirical observations which are strictly necessary for providing sufficient and representative information about the population. A weak point is that the design always depends on unknown parameters, thus for the same model one has different experimental designs for different parameter values.To overcome this inconvenience we develop a sequential procedure, consisting in generating a considerable set of parameters in a space defined beforehand, and calculating the optimal design for each subset. In the first phase we use trimmed K-means and in the second we calculate the KL designs, which discriminate between two models with a given distribution. Thus our algorithm models the initial design and provides an efficient design for discrimination, also giving good estimates for the parameters. The use of K-means for D-optimality can be found in the literature, but we believe using trimmed K-means allows for a better fit.This procedure will be applicable to both single and multiple groups. We conclude with some applications to pharmacokinetic models, e.g. to variations of the Michaelis-Menten model, among others.

E757: **Optimal design of experiments for Markov chains**
Presenter: **Ben Parker**, Queen Mary University of London, UK
Co-authors: John Schormans, Steven Gilmour

Suppose we have a system that we can measure a fixed number of times, but at any chosen interval. Motivated by an example of probing data networks, we use the statistical principles of design of experiments to model numerical experiments that can be designed optimally. We demonstrate how to analyse the evolution of a system as a Markov Chain, and deduce its likelihood function, and hence the Fisher information matrix. From this, numerical methods guide us to best design for the experiment for different values of input parameters. We further develop our ideas to show what happens when we take into account the effect of the observations interfering with the experiment, as would always be the case with active probing. We present examples, and demonstrate how this could be useful to many fields, with particular reference to experiments on packet networks. We present a general result for showing how to optimally measure any system that evolves according to the Markov principle.

ES65 Room MAL 532 TOPICS IN BAYESIAN NONPARAMETRICS Chair: Antonio Lijoi

E644: **Single factor transformation priors for density regression**
Presenter: **Suprateek Kundu**, UNC Chapel Hill, USA
Co-authors: David Dunson

Although mixture modeling has formed the backbone of the literature on Bayesian density estimation incorporating covariates, the use of mixtures leads to some well known disadvantages. Avoiding mixtures, we propose a flexible class of priors based on a random transformation of a uniform latent variable. These priors are related to Gaussian process latent variable models proposed in the machine learning literature. For density regression, we model the response and predictor means as distinct unknown transformation functions dependent on the same underlying latent variable, thus inducing dependence through a single factor. The induced prior is shown to have desirable properties including large support and posterior consistency. We demonstrate advantages over Dirichlet process mixture models in a variety of simulations, and apply the approach to an epidemiology application.

E721: **Nonparametric model selection and multiple testing**
Presenter: **Michele Guindani**, UT MD Anderson Cancer Center, USA
Co-authors: Antonio Lijoi, Fabrizio Leisen

The use of Bayesian Non Parametric (NP) priors to model heterogeneity has experienced an increased popularity in recent years. However, despite the availability of a wide array of NP prior models, the assessment of the particular NP model to use in a particular application has received much less attention. Arguably, such a decision should be made on the basis of the clustering structure implied by the proposed NP prior. We discuss a modeling framework to compare the clustering behavior of different NP priors, with particular regard to the behavior of the two-parameters Poisson-Dirichlet processes. We illustrate the challenges and the relevance of such analysis by means of simulation experiments. Furthermore, we discuss how our modeling framework can be used to guide decisions and control false positives in a general multiple hypotheses setting.

E812: **On the use of discrete nonparametric priors for continuous data**
Presenter: **Pierpaolo De Blasi**, University of Turin and Collegio Carlo Alberto, Italy
Co-authors: Antonio Lijoi, Igor Pruenster

Most of the discrete nonparametric priors currently use, with the exception of the Dirichlet processes, are inconsistent if used to model directly continuous data. On the other hand, they generally have large enough weak support to make them suitable for hierarchical mixture modelling of continuous models. In this paper we provide sufficient condition for consistency for Gibbs-type prior and present two examples within this class which exhibit completely opposite behaviour in the limit.

E709: **A new enrichment of the Dirichlet process for multivariate random distributions**

Presenter: **Sara Wade**, Bocconi University, Italy

Co-authors: Silvia Mongelluzzo, Sonia Petrone

The precision parameter α plays an important role in the Dirichlet Process. When assigning a Dirichlet Process prior to the set of probability measures on R^k, $k > 1$, this can be quite restrictive in the sense that the variability is determined by a single parameter. The aim of this paper is to consider a generalization of the Dirichlet Process that is more flexible with respect to the precision parameter yet still conjugate. Moving from the notion of enriched conjugate priors for the natural exponential family, we construct a process that allows more flexibility in modelling uncertainty on the marginal and conditionals and describe an enriched urn scheme which characterizes this process. The resulting enriched conjugate prior can also be obtained from the stick-breaking representation of the marginal and conditionals. This allows global clustering of the marginal variable and local clustering of the conditional variables. Finally, we consider an application to mixture models that allows for uncertainty between homoskedasticity and heteroskedasticity.

E692: **Bayesian nonparametric inference for species variety**

Presenter: **Stefano Favaro**, University of Turin and Collegio Carlo Alberto, Italy

Co-authors: Antonio Lijoi, Ramses H. Mena, Igor Pruenster

Sampling problems from population which are made of different species arise in a variety of ecological and biological contexts. Such problems concern the evaluation, conditional on a sample of size n, of the species variety featured by an additional sample of size m or in estimating the so-called sample coverage. A Bayesian nonparametric methodology is proposed in order to deal with these species sampling problems.

EP03 Room Senate Crush Hall POSTERS III Chair: Cristian Gatu

E842: **A shrinkage method for weighting person and household calibration estimators**

Presenter: **Juan Francisco Munoz**, Universidad de Granada, Spain

Co-authors: Jose Miguel Contreras, Antonio Arcos

In two stages sampling, a sample of units is realized by two-stage selection from the finite population grouped into clusters. This design involves sampling from two distinct populations: the population of clusters or primary stage units, and the population of units or second-stage units. Calibration estimators for unit statistics on the combined information on cluster totals and unit totals can be defined. To calibrate on the combined auxiliary information an option for integrated weighting yields the same method of Lemaitre and Dufour. We define an alternative estimator for unit statistics that use the same available auxiliary information. The calibration estimator for the population total with unit weights satisfying calibration equation at unit level and the calibration estimator for the population total with unit weights derived from calibration equation at cluster level are obtained. Thus, our estimator is defined shrinking the unit-estimator towards the cluster-estimator in order to get a better estimate. How much improvement can this shrinkage estimator give? The answer to this question is reported through a simulation study in two real population. The data are obtained from the Household Budget Survey in Spain and from the PISA (Program for International Student Assessment) database.

E844: **Tourism and growth in European countries: An application of likelihood-based panel cointegration**

Presenter: **Sofia de Sousa-Vale**, ISCTE-IUL, Portugal

Co-authors: Felipa de Mello-Sampayo

The tourism and economic growth relationship is investigated for a panel of European countries over the period 1988-2010. The results reveal that the variables contain a panel unit root and they cointegrate in a panel perspective. The findings show that tourism enhance economic growth for some countries in the sample. The paper is organized as follows: first, we present a simple theoretical model, then we describe the sources of data and report our panel test results for unit roots, cointegration and ranks. Finally, we discuss the long run relationship equilibrium taking into account the cross dependence within the panel to conclude that tourism is important for economic growth but with a smaller magnitude than expected.

E850: **Robust modelling of extremes in linear heavy tailed models**

Presenter: **Jan Dienstbier**, Technical University of Liberec, Czech Republic

We discuss the inference of extremal properties in heavy-tailed linear models. In the literature it has been already proposed to use quantile sensitive linear regression estimators such as regression quantiles. For the sake of simplicity we deal with the estimation of the extreme value index only. We propose new methods of the estimation based on intercepts of regression quantiles and two-step regression quantiles and compare them with the older approaches. New methods are based on smooth functionals of the regression quantile process and/or the two-step regression quantile process. They are asymptotically consistent and normal. This property is based on approximations of the regression quantile process and the two-step regression quantile process. We compare the power of the methods computationally using simulations. Finally the ability of the methods is demonstrated on real data sets as Codroz data of calcium levels in Belgium Condroz region or insurance AON Re data.

E864: **Conditional symmetry models for three-way tables**

Presenter: **Maria Kateri**, University of Ioannina, Greece

Co-authors: Petros Dellaportas

We generalize the conditional symmetry model for an $I \times I \times I$ contingency table with commensurable ordinal classification variables. The new family of models possesses the usual desirable properties that relate this family to symmetry and marginal homogeneity models for three-way tables. That is, connections between the models above obey a coherent structure. The features of the new introduced models are exhibited by characteristic examples.

E171: **A stochastic model related to the Richards growth curve**

Presenter: **Patricia Roman-Roman**, Universidad de Granada, Spain

Co-authors: Francisco Torres-Ruiz

Diverse models have proved useful for describing sigmoid growth patters. Among others, the logistic and Gompertz curves have points of inflection that are always at a fixed proportion of their asymptotic values. Extending the logistic curve, Richards curve is a flexible sigmoid function widely-used for growth modelling. It has an additional parameter that can make it equivalent to the logistic, monomolecular and to other well known models like Gompertz and von Bertalanfy equations. Varying this shape parameter allows the point of inflection of the curve to be at any value between the minimum and the upper asymptote. So the Richards equation provides better fits and it is more flexible in describing various asymmetrical sigmoid patterns. In order to take into account random environmental effects, a stochastic model related to the Richards growth curve is proposed. Also, with the purpose of fitting this model to real data, we study the maximum likelihood estimation of the parameters of the model. In addition, and regarding the numerical problems for solving the likelihood equations, a strategy is developed for obtaining initial solutions for the

usual numerical procedures. Such strategy is validated by means of simulated examples. Finally, an application to real data is presented.

E576: **On estimating population proportions with calibrated estimators**

Presenter: **Sergio Martinez**, Universidad de Granada, Spain

Co-authors: Maria del Mar Rueda, Juan Francisco Munoz, Helena Martinez

The estimation of a population proportion is discussed in this paper. Proposed estimators are based upon the calibration approach, which are used to define a new estimator with desirable efficiency properties. A new class of estimators based upon the proposed calibrated estimators is also defined, and the optimal estimator into the class is achieved in the sense of minimal variance. Finally, an estimator of the population proportion based upon new calibration conditions is defined. Simulation studies are considered to evaluate the performance of the proposed calibration estimators via the empirical relative bias and the empirical relative root mean square error, and desirable results are achieved.

Saturday 11.12.2010 17:20 - 19:25 Parallel Session J – CFE

CI96 Room Senate Beveridge Hall INVITED SESSION: MODELS FOR LARGE MULTIVARIATE SYSTEMS Chair: Esther Ruiz

C348: **Quantile-based inference for multivariate dynamic models**
Presenter: **David Veredas**, Universite Libre de Bruxelles, Belgium
Co-authors: Yves Dominicy, Hiroaki Ogata

It is often the case that large multivariate system cannot be estimated easily by traditional methods such as maximum likelihood or the method of moments. It happens in situations in which is the density does not have a closed form or moments do not exist. In this talk it is shown – based on the Method of Simulated Quantiles – how to estimate the parameters by means of quantiles, which always exist.

C423: **Realized mixed-frequency factor models for vast dimensional covariance estimation**
Presenter: **Dick van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Karim Bannouh, Martin Martens, Roel Oomen

We introduce a Mixed-Frequency Factor Model (MFFM) to estimate vast dimensional covariance matrices of asset returns. The MFFM uses high-frequency (intraday) data to estimate factor (co)variances and idiosyncratic risk and low-frequency (daily) data to estimate the factor loadings. We propose the use of highly liquid assets such as exchange traded funds (ETFs) as factors. Prices for these contracts are observed essentially free of microstructure noise at high frequencies, allowing us to obtain precise estimates of the factor covariances. The factor loadings instead are estimated from daily data to avoid biases due to market microstructure effects such as the relative illiquidity of individual stocks and non-synchronicity between the returns on factors and stocks. Our theoretical, simulation and empirical results illustrate that the performance of the MFFM is excellent, both compared to conventional factor models based solely on low-frequency data and to popular realized covariance estimators based on high-frequency data.

C670: **Smooth dynamic factor analysis with an application to U.S. term structure of interest rates**
Presenter: **Siem Jan Koopman**, VU University Amsterdam, Netherlands
Co-authors: Borus Jungbacker, Michel van der Wel

We consider the dynamic factor model and show how smoothness restrictions can be imposed on the factor loadings. Cubic spline functions are used to introduce smoothness in factor loadings. We develop statistical procedures based on Wald, Lagrange multiplier and likelihood ratio tests for this purpose. A Monte Carlo study is presented to show that our procedures are successful in identifying smooth loading structures from small sample panels. We illustrate the methodology by analyzing the U.S. term structure of interest rates. An empirical study is carried out using a monthly time series panel of unsmoothed Fama-Bliss zero yields for treasuries of different maturities between 1970 and 2009. Dynamic factor models with and without smooth loadings are compared with dynamic models based on Nelson-Siegel and cubic spline yield curves. All models can be regarded as special cases of the dynamic factor model. We carry out statistical hypothesis tests and compare information criteria to verify whether the restrictions imposed by the models are supported by the data. Out-of-sample forecast evidence is also given. Our main conclusion is that smoothness restrictions on loadings of the dynamic factor model for the term structure can be supported by our panel of U.S. interest rates and can lead to more accurate forecasts.

CS20 Room MAL 151 EMPIRICAL MODELLING OF FINANCIAL MARKETS Chair: Andrea Cipollini

C213: **Portfolio choice under local industry and country factors**
Presenter: **Carlos Castro**, Universidad del Rosario, Colombia

The parametric portfolio policy approach to optimize portfolios with large numbers of assets is extended. The proposed approach incorporates unobserved effects into the portfolio policy function. These effects measure the importance of unobserved heterogeneity for exploiting the difference between groups of assets. The source of the heterogeneity is local priced factors, such as industry or country. The statistical model derived allows to test the importance of such local factors in portfolio optimization. The results suggest that local effects or return heterogeneity associated to economic sectors or geographic factors is not as straight forward to exploit financially or relevant as suggested by the extensive multivariate factor literature on the subject. Furthermore, trying to exploit industry effects rarely provides gain over simple benchmarks, both in-sample and more importantly out-of-sample. On the other hand trying to exploit country effects does provide gains over the benchmark. However, these gains may be offset by the increasing cost and risk inherent to such strategies. Finally, exploiting size, momentum, and liquidity anomalies in the cross-section of stocks provides strictly greater returns than the industry and country effects.

C126: **Asymmetries in financial returns: A local Gaussian correlation approach**
Presenter: **Bard Stove**, Norwegian School of Economics and Business Administration, Norway
Co-authors: Karl Ove Hufthammer, Dag Tjostheim

A number of studies have provided evidence that financial returns exhibit asymmetric dependence, such as increased dependence during bear markets. We introduce the use of a new measure of local dependence to study and quantify this asymmetry. This measure does not suffer from the selection bias of the conditional correlation for Gaussian data, and is able to capture nonlinear dependence, as the popular copula model approach. Analysing several financial returns, both monthly and daily data, from the US, UK, German and French market, we confirm results of asymmetry, and are able to quantify the asymmetry. Finally, we discuss possible applications to portfolio selection, risk management and the contagion effect, and point out a number of possible extensions.

C146: **Testing for contagion: a time scale decomposition**
Presenter: **Iolanda lo Cascio**, University of Palermo, Italy
Co-authors: Andrea Cipollini

The aim is to test for financial contagion by estimating a simultaneous equation model subject to structural breaks. For this purpose, we use the Maximum Overlapping Discrete Wavelet Transform, MODWT, to decompose the covariance matrix of two asset returns on a scale by scale basis. This decomposition will enable us to identify the structural form model and to test for spillover effects between country specific shocks during a crisis period. The empirical analysis is applied to four East Asian emerging stock markets.

C204: **Hedge fund return sensitivity to global liquidity**

Presenter: **Stephan Kessler**, University of St Gallen / Switzerland, UK

Co-authors: Bernd Scherer

A common latent liquidity factor, which is the driver of observable and commonly used liquidity proxies across asset classes is identified. We use two methodologies to identify the latent liquidity factor: State space modelling (SSM) and Principal Component Analysis (PCA). We find that the returns of hedge funds respond to an increase in illiquidity with statistically significant negative returns. The relative size of the liquidity factor loadings of different hedge fund indices are generally consistent with the liquidity sensitivities of the underlying strategies. The results hold up in a range of robustness tests. Finally, we find a surprisingly strong link between global risk factors and hedge fund returns, questioning the industry's claim to deliver pure manager alpha.

C124: **High frequency jump-response of asset prices to FX announcements and oral interventions**

Presenter: **Deniz Erdemlioglu**, Academie Universitaire Louvain, Belgium

Co-authors: Hans Dewachter, Jean-Yves Gnabo, Christelle Lecourt

Recent studies show that financial asset prices exhibit sudden movements called jumps. In this paper, using intradaily financial data, we investigate the association of asset price jumps with FX market news announcements and oral interventions. In particular, we examine whether the verbal statements of the US and Euro area monetary authorities are able to move financial markets and create jumps. Asset price jumps and cojumps are detected using the recent computational methods in high frequency financial econometrics. We apply our study to bond futures, index futures and exchange rates. The objective of our work is to provide an analysis to better understand the communication of central banks and policy makers with markets, which in turn might help to reduce uncertainty and hence improve conducting monetary policy both in good times and crises.

CS26 Room MAL 355 FINANCIAL MODELING Chair: Jerry Coakley

C825: **Separating the effects of idiosyncratic volatility and skewness on expected returns**

Presenter: **Ruey Yau**, National Central University, Taiwan

When investors do not diversify their portfolio, the CAPM theory predicts a positive relation between idiosyncratic volatility and expected return. A puzzling cross-sectional result is that monthly stock returns are negatively related to the previous month idiosyncratic volatilities. Using GARCH models with a flexible three-parameter distribution for stock returns, I find that diversification erodes skewness exposure; therefore preference for positive skewness may explain the under-diversification behavior of the investors. This finding suggests that idiosyncratic skewness is a crucial determinant of asset prices. Previous empirical evidence has indicated that high lagged idiosyncratic volatility predicts high expected firm skewness. Therefore, investors who have a preference for positive skewness may accept lower expected return on stocks that have experienced high idiosyncratic volatility. In this paper, I re-examine this hypothesis by modeling idiosyncratic volatility and skewness simultaneously after controlling for firm-specific factors. With this method, I separate the explanatory effects of expected skewness and idiosyncratic volatility on expected returns. I find that the hypothesis holds and the overlapping explanatory part is relatively smaller than the separate part.

C296: **The volatility asymmetry risk and expected returns**

Presenter: **Yujia Hu**, University of St. Gallen, Switzerland

A new factor, the volatility asymmetry from high frequency data, is introduced to explain the time series and cross-sectional variation of asset prices. The sensibility to this factor determines a premium for asset returns. I argue that the volatility asymmetry is a proxy for the market sentiment and I find that it has a predictive power on future expected returns. When prices overreact to bad news in respect to good news, the one day ahead return is expected to be higher. Small, growth and risky stocks are particularly subject to this risk. A heterogeneous risk structure with volatility asymmetry performs well in pricing tests, by using daily U.S. equity data for a sample period that ranges from 1983 to 2008. It consistently outperform the CAPM. This result is robust in respect to different testing procedures and cross-sections of portfolios.

C781: **Tests for abnormal returns under weak cross sectional dependence**

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

Co-authors: Jan Antell

In event studies abnormal returns are assumed to be cross sectionally independent. If the event day is common, and if the firms are from the same industry, abnormal returns may not be independent across firms. Tests for event effects are invalid under cross sectional dependence. We propose to model the dependence in abnormal returns by a spatial error model. In the model abnormal returns of firms belonging to the same sector or industry are correlated, but abnormal returns of firms belonging to different sectors or industries are uncorrelated. The spatial error model formalises weak dependence. Corrected tests for event effects under spatial dependence are derived and shown to be asymptotically normal. The correction depends on the spatial covariance matrix. A Monte Carlo study shows that moderate spatial autocorrelation causes tests for event effects to over-reject the null hypothesis of no event effect, whereas the corrected tests have the nominal level and nontrivial power. An empirical application to US stock returns around Lehman Brothers' bankruptcy illustrates the importance of correcting for cross sectional dependence in abnormal returns.

C680: **Chance-constrained financial index tracking models under GH distribution**

Presenter: **Bei Cheng**, Xi'an Jiaotong University, China

Co-authors: Zhiping Chen, Jia Liu

To reflect the fat-tail and skew feature of the risky assets' return distribution, we first adopt the multivariate generalized hyperbolic (GH) distribution to describe the return data, and establish a new financial index tracking model which uses a probability constraint to control the downside risk, and takes into account multiple market frictions. Since it can be transformed into a second-order conic program or a linear program, the derived index tracking model can be easily solved in polynomial time. By using the normal variance-mean mixture definition of GH distribution and Chapman–Kolmogorov equation, we extend our one-stage index tracking model to a multistage one in the second part of this paper. Empirical results based on the S&P 500 index and Shanghai–Shenzhen 300 Index demonstrate the efficiency and practical value of our new tracking models.

C695: **Rating migrations:The effect of history and time**

Presenter: **Huong Dang**, University of Sydney, Australia

Co-authors: Graham Partington

We investigate the effect of the passage of time and rating history on the hazard of rating migrations. A substantial number of rating history variables significantly affect the hazard of rating migrations and the results show a tendency for rating history to repeat itself. We control for economic conditions and industry and find that rating history is more important than the current rating in determining the hazard of a rating change. There are significant interactions between the time spent in a rating grade and the main effect variables. Failure to account for such interactions may lead

to substantial errors in the estimation of migration risk.

CS88 Room MAL B29 CONTRIBUTIONS TO COMPUTATIONAL ECONOMETRICS Chair: Paolo Foschi

C229: **Indirect likelihood inference**

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

Co-authors: Dennis Kristensen

Given a sample from a parametric model, let Z_n be a given finite-dimensional statistic - for example, the value of an initial estimator or a set of sample moments. We then propose to estimate the parameters of the model by maximizing the likelihood of Z_n. We call this the maximum indirect likelihood (MIL) estimator since it is based on the statistic Z_n rather than the full sample. We also propose a Bayesian version of the estimator leading to what we refer to as a Bayesian Indirect Likelihood (BIL) estimator. In most cases, the density of the statistic will be of unknown form. Simulability of the model allows us to combine simulations and nonparametric methods to obtain simulated versions of the MIL and BIL estimators. We show that the indirect likelihood estimators and their simulated counterparts are consistent and asymptotically normally distributed, with the same distribution as the (possibly infeasible) GMM estimator based on the same statistic using an efficient weight matrix. However, our likelihood-based estimators will in general enjoy better finite-sample properties relative to the GMM estimator. Monte Carlo results for several applications including a structural auction model and a DSGE model show that the simulated BIL estimator indeed have attractive finite sample properties relative to existing estimators.

C560: **Fast, unbiased and efficient importance sampling for state space models**

Presenter: **Marcel Scharth**, VU University Amsterdam, Netherlands

Co-authors: Siem Jan Koopman, Andre Lucas

We introduce a new efficient importance sampler for non-gaussian and nonlinear state space models. The idea behind our method is that numerical and Monte Carlo integration elements can be combined for maximal efficiency in simulating the likelihood for these models. We call this technique numerically accelerated importance sampling (NAIS). In order to achieve this we introduce approximating linear state space models in the context of efficient importance sampling (EIS). Not so far recognized, however, is the fact that despite its efficiency the EIS method as it has been implemented is subject to a relevant finite sample bias. Our framework solves this problem effectively. A simulation study is conducted to illustrate the efficiency and swiftness of this approach.

C744: **Efficient estimation of high dimensional factor models under cross sectional dependence**

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

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

C867: **Downdating the seemingly unrelated regressions model**

Presenter: **Stella Hadjiantoni**, University of Cyprus, Cyprus

Co-authors: Erricos Kontoghiorghes

The problem of deleting observations (downdating) from the general linear model (GLM) is considered. It is shown that the downdating problem is equivalent in updating the GLM with the imaginary deleted observations. This results in a GLM with negative defined dispersion matrix and comprising complex covariance values. Its solution is obtained by treating the problem as a generalised linear least squares problem (GLLSP). The proposed computational strategy for solving the GLLSP utilizes previous computations and does not use complex arithmetic. The generalized QR decomposition (GQRD) is the main computational tool employed. It is based on hyperbolic reflections.

C215: **Estimating utility functions - GME versus OLS**

Presenter: **Andreia Dionisio**, Universidade de Evora, Portugal

Co-authors: Cesaltina Pires, Luis Coelho

We estimate von Neumann and Morgenstern utility functions comparing the generalized maximum entropy (GME) with OLS, using data obtained by utility elicitation methods. Thus, it provides a comparison of the performance of the two estimators in a real data small sample setup. The results confirm the ones obtained for small samples through Monte Carlo simulations. Overall the results suggest that GME is an interesting alternative to OLS in the estimation of utility functions when data is generated by utility elicitation methods.

CS41 Room MAL B20 FINANCIAL VOLATILITY ESTIMATION AND FORECASTING II Chair: Francesco Audrino

C251: **Heterogeneous asymmetries and persistence in the volatility of realized volatility**

Presenter: **Dimitrios Louzis**, Athens University of Economics and Business, Greece

Co-authors: Spyros Xanthopoulos, Apostolos Refenes

We account for the presence of heterogeneous leverage effects and the persistence in the volatility of stock index realized volatility. The Heterogeneous Autoregressive (HAR) realized volatility model is extended in order to account for asymmetric responses to negative and positive shocks occurring at distinct frequencies, as well as, for the long range dependence in the heteroscedastic variance of the residuals. Compared with established HAR and Autoregressive Fractionally Integrated Moving Average (ARFIMA) realized volatility models, the proposed model exhibits superior in-sample fitting, as well as, out-of-sample volatility forecasting performance. The latter is further improved when the realized power variation is used as a regressor, while we show that our analysis is also robust against microstructure noise.

C262: **Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling**

Presenter: **Fulvio Corsi**, University of Lugano, Switzerland

Co-authors: Roberto Reno

We first propose a reduced form model in discrete time for S&P500 volatility showing that the forecasting performance of a volatility model can be significantly improved by introducing a persistent leverage effect with a long-range dependence similar to that of volatility itself. We also find a strongly significant positive impact of lagged jumps on volatility, which however is absorbed more quickly. We then estimate continuous-time stochastic volatility models which are able to reproduce the statistical features captured by the discrete-time model. We show that a single-factor model driven by a fractional Brownian motion is unable to reproduce the volatility dynamics observed in the data, while a multi-factor Markovian model is able to reproduce the persistence of both volatility and leverage effect. The impact of jumps can instead be associated with a common jump component in price and volatility. These findings cast serious doubts on the need of modeling volatility with a genuine long memory component, while reinforcing the view of volatility being generated by the superposition of multiple factors.

C264: **Estimating conditional jumps in volatility using realized-range measures**

Presenter: **Paolo Santucci de Magistris**, University of Padova, Italy

Co-authors: Massimiliano Caporin, Eduardo Rossi

Recent empirical evidence indicates that diffusive stochastic volatility and jumps in returns are incapable of capturing the empirical features of equity index returns. The most convincing evidence, indicates that volatility of financial returns is affected by rapid and large increments. Including jumps in the volatility process is therefore crucial because they allow for rapid increases (or decreases) in prices, like those observed during stock market crashes. The aim of this paper is to model and estimate the jump component in volatility, using the realized range as non parametric ex-post measure of the daily integrated volatility. Given that our focus is on the volatility jumps, we employ the robust realized-range bipower variation, that is a consistent estimator of the integrated volatility, when the price process is affected by jumps and microstructure noise. We suggest to generalize the HARV model, allowing for the presence of a conditional jump term, where the intensity and the magnitude parameters are time varying. In this way, we are able to model and identify periods with an increasing jump activity, where the unexpected jump risk cannot be hedged away and investors may demand a large risk premia.

C501: **Wavelet-based realized variance estimator**

Presenter: **Jozef Barunik**, IES FSV UK, Czech Republic

Co-authors: Lukas Vacha

The integrated volatility estimators by proposing a wavelet-based realized volatility estimator using the maximum overlap discrete wavelet transform is considered. Wavelet-based realized volatility estimator decomposes the realized volatility into several scales. Moreover, it allows us to study the energy contribution of every scale to the realized volatility. We study robustness of our estimator and compare its performance with other standard estimators under the presence of jumps as well. We also provide finite sample properties on the various sampling frequencies using Monte Carlo simulations. Finally, we illustrate usage of wavelet-based realized volatility estimator on the S&P 500 futures data for the period beginning with the year 1983 and ending by November 2009. Our empirical analysis reveals very interesting results. We show, how the scaling behavior of realized volatility changed over the studied period.

C279: **Structural breaks in realized variance during the financial crisis**

Presenter: **Matthias Fengler**, St. Gallen (HSG), Switzerland

Co-authors: Michael Vogt, Enno Mammen

The dynamics of realized variance constructed from high-frequency futures price data before and during the financial crisis 2008-2009 across five different asset classes are studied. Our concern is to detect a potential structural break with the advent of the crisis. For this purpose, we propose an additive cascade model for variance components of different time horizons, where each of the components is captured by a nonparametric link function. We investigate whether that the parametric competitor of our model prematurely identifies structural breaks which are attributable to a kink in the transmission function, which is caused by a threshold-driven shift in the behavior of financial agents.

CS13 Room MAL G16 FINANCIAL TIME SERIES MODELLING AND FORECASTING II Chair: Alessandra Amendola

C790: **Likelihood inference in non-linear term structure models: The importance of the zero-lower bound**

Presenter: **Andrew Meldrum**, Bank of England, UK

Co-authors: Martin M. Andreasen

It is shown how to use particle filtering and MCMC methods to estimate quadratic term structure models by likelihood inference from a Bayesian perspective. Estimation results show that quadratic models are better able to match bond yields in the US and the UK when compared to Gaussian affine models. We also find that it is important to account for the zero lower bound on interest rate during the recent financial crisis as an affine model may give negative interest rates and assign positive probability to forecasts with negative interest rates.

C201: **Heterogeneity in stock pricing: A STAR model with multivariate transition functions**

Presenter: **Matthijs Lof**, University of Helsinki, Finland

Stock prices often diverge from measures of fundamental value, which simple present value models fail to explain. This paper tries to find causes for these long-run price movements and their persistence by estimating a STAR model for the price-earnings ratio of the S&P500 index for 1961Q1 - 2009Q3, with a transition function that depends on a wider set of exogenous or predetermined transition variables. Several economic, monetary and financial variables, as well as linear combinations of these, are found to have nonlinear effects on stock prices. A two-step estimation procedure is proposed to select the transition variables and estimate their weights. This STAR model can be interpreted as a heterogeneous agent asset pricing model that makes a distinction between chartists and fundamentalists, where the set of transition variables is included in the agents' information set.

C660: **Entropy based tests for nonlinear dependence in time series**

Presenter: **Simone Giannerini**, University of Bologna, Italy

Co-authors: Esfandiar Maasoumi, Estela Bee Dagum

We propose some novel tests for the identification of nonlinear dependence in time series. The approach is based on a combination of a test statistic based on an entropy dependence metric, possessing many desirable properties, together with a suitable extension of resampling methods. We show how the tests can be employed in order to detect the lags at which a significant nonlinear relationship is expected in the same fashion as the autocorrelation function is used for linear time series. Some asymptotic results are derived and the power and size of the tests are assessed through simulation studies; applications to real time series are also shown.

C693: **Approximate p-values of certain tests involving hypotheses about multiple breaks.**

Presenter: **Nikolaos Sakkas**, University of Manchester, UK

Co-authors: Alastair Hall

There has been considerable interest in developing inference techniques for econometric models that exhibit parameter variation at multiple unknown points. Leading statistics are the so-called Sup-, UDmax- and sequential tests that are used respectively to test stability versus a fixed number of breaks, stability against up to a fixed number of breaks and k breaks versus k+1 breaks. These statistics have been shown to have the same limiting distribution in linear models (with exogenous regressors) estimated by Ordinary least Squares, linear models (with endogenous regressors) estimated via Two Stage Least Squares and nonlinear regression models estimated by nonlinear least squares. The limiting distributions in question are non-standard but depend only on the number of unknown breaks, a trimming parameter, and the number of parameters, n say, in the model. This paper makes two contributions that lead to improved inference with these tests. Firstly, we improve on the accuracy of the simulated quantile functions by a factor of 10,000 replications providing more accurate critical values and provide results for extended set of values for n. Secondly, we provide response surfaces that allow for the easy yet accurate calculation of asymptotic p-values. The performance of the response surfaces is analyzed, and illustrated.

C834: **Interval autoregression: an application to volatility**

Presenter: **Javier Arroyo**, Universidad Complutense de Madrid, Spain

Co-authors: Gloria Gonzalez-Rivera

The main objective is to study the empirical behavior of several linear regression models for interval financial data. From all the interval linear regression models proposed in the literature we advocate for the use of a model that consider intervals as random sets based on an intuitive arithmetic and a metric that makes possible a least-squares estimation and ensures the coherence of the estimators. In addition, we implement some inferential tools such as correlation tests. This model along with some of the less-sophisticated others is used to analyze the relationship between two interval data that arise in finance in a natural way. They are the low-high interval, as an estimator of the volatility of an asset, and the open-close (signed) interval, as an estimator of the return of an asset. The proposed model can be understood as an interval-autoregression (iAR) with explanatory interval variables. We compare the classical VAR framework with the proposed model, and by doing so we connect with prior results regarding the predictability of volatility.

CS81 Room MAL B30 HIGH FREQUENCY AND SEASONAL DATA Chair: Ana Galvao

C324: **Day-varying structure for modeling intraday periodicity**

Presenter: **Carlin Chu**, The Chinese University of Hong Kong, Hong Kong

Co-authors: Kai-pui Lam

The periodicity appeared in an intraday volatility process is always modeled to follow an identically repeating structure. It varies in a U-shape pattern within a trading day while its value for every particular intraday interval is identical across days. This rigid day-invariant structure may hinder the potential usage of the variability of periodicity across days. A method that is capable to estimate the periodicity with the allowance of day-variability has been provided. However, the performance of the day-variant periodicity is always demonstrated to be inferior to the corresponding day-invariant version empirically. We improve their normalization method by adjusting the estimated values to fulfill the implicit constraint for the construction of daily variances from their corresponding intraday variances. For the situation that the periodicity is modeled to be day-variant, the proposed method is shown to be less susceptible to heteroskedastic errors through numerical simulations. On the other hand, our method is proven to give the same performance as Andersen and Bollerslev's method does mathematically. Furthermore, our method is tested by using 10-minute returns of NASDAQ index (3-year period) and S&P 500 index (2.5-year period). Preference on using the proposed method for short to medium forecasting horizons is supported empirically.

C558: **Common intraday Periodicity**

Presenter: **Alain Hecq**, Maastricht University, Netherlands

Co-authors: Sebastien Laurent

The intraday diurnal pattern is a dominant feature in the volatility of most high-frequency financial time series. These movements can be captured by non parametric techniques or in a parametric context with a set of dummy variables or a bunch of trigonometric functions. In the latter framework however, a large number of such variables are sometimes needed. This number further inflates when considering several assets in a multivariate modeling. It might be observed however that this intraday periodic feature is common to several series. Testing, discovering and imposing these commonalities can yield additional economic insights and might be exploited to improve parameter efficiency and forecasts accuracies. To this goal we use a reduced rank approach to study the presence of commonalities in the intraday periodic movements. The Monte Carlo simulations indicate that our proposed strategy (tests and information criteria) detects remarkably well both the number of periodic elements to be added and the existence of commonalities even in the presence of jumps. Out of thirty US stock returns observed every five minutes on the period 2000-2008 we show that there exists three sources of common periodicity and that imposing these restrictions helps to better predict future values of the returns.

C444: **Common factor estimation of dynamic models with seasonality for predicting electricity prices**

Presenter: **Damian Lopez Asensio**, Universidad Politecnica de Madrid, Spain

Co-authors: Jesus Juan Ruiz, Jaime Carpio Huertas

The aim of our work is to build multivariate dynamic models for considering the evolution of high-dimension time series, as hourly electricity prices in day-ahead markets with dimension equal to 24. Beginning with a simple model, MEWMA (Multivariate exponentially weighed moving average), our methodology allows the estimation of complex models with seasonality and deterministic effects. The application to vectors of such dimension requires a large number of parameters, which is a disadvantage. In this work, we explained a simple methodology to reduce the large number of the model parameters by performing a dynamic factor analysis. A computer implementation of the EM (Expectation - Maximization) algorithm has been made for maximum likelihood estimation of the models. The obtained models give rise to a very interesting interpretation of the dynamic of prices. The approach has been applied to the study of the behavior of three series, the French market (Powernext), the Scandinavian market (Nord Pool) and the Spanish market (OMEL). The results are consistent for all of them.

C431: **Nowcasting inflation using high frequency data**

Presenter: **Michele Modugno**, European Central Bank, Germany

A methodology to nowcast and forecast euro area inflation using data with sampling frequency higher than monthly is proposed. The nowcasting literature has been focused on GDP, typically using monthly indicators in order to produce an accurate estimate for the current and next quarter. Data with weekly and daily frequency is exploited in order to produce more accurate estimates of HICP inflation for the current month and followings months. In particular, Weekly Oil Bulletin Price Statistics of the European Commission and daily World Market Prices of Raw Materials is

used. This is produced by the European Central Bank. The data are modeled as a high frequency factor model with missing observations in a state space representation. In contrast to other existing approaches, the methodology used has the advantage of modeling all data within a unified single framework that, nevertheless, allows one to produce forecasts of all variables involved. This offers the advantage of disentangling a model-based measure of "news" from each data release and subsequently to assess its impact on the forecast revision. An illustrative example of this procedure is provided. Moreover it is shown that both the weekly oil bulletin and raw material prices are useful to improve the accuracy of overall HICP inflation forecasts. This is due to the higher accuracy in the forecasting performance of the HICP energy inflation.

C772: **Vector autoregressive models of data vintages for US output growth and inflation**

Presenter: **Ana Galvao**, Queen Mary University of London, UK

Co-authors: Michael Clements

Vector-autoregressive models of different vintage estimates of observations on a single variable are shown to be a useful vehicle for obtaining forecasts of different maturities of future and past observations, including estimates of post-revision values, and offer a way of assessing the uncertainty associated with these estimates. They also suggest a way of obtaining more accurate estimates of output gaps in real time. A VAR model that more closely reflects the seasonal nature of the release of US data is not found to offer much improvement over the unrestricted model.

CS67 Room MAL B33 QUANTITATIVE RISK MANAGEMENT II Chair: Kerstin Kehrle

C147: **An unscented Kalman smoother for volatility extraction: Evidence from stock prices and options**

Presenter: **Junye Li**, ESSEC Business School, France

The application of stochastic volatility models in statistical analysis usually faces two challenges. The first problem is that variances of relevant stochastic variables are state-dependent, and the second is that when taking into account derivatives, the pricing formula is highly non-linear. These problems make the existent linear techniques rarely applicable. This paper proposes a smoothing algorithm based on the unscented transformation and shows how it can be used for volatility extraction in stochastic volatility models. The method firstly implements a forward unscented Kalman filter and then evokes a separate backward smoothing pass by only making Gaussian approximations in the state but not in the observation space for a nonlinear system. It can result in higher-order approximation accuracy and cheaper computations. A simulation study and empirical applications with the Heston stochastic volatility model indicate that in order to accurately capture the volatility dynamics, both stock prices and options are necessary.

C206: **Seasonal effects in liquidity of NYSE and NASDAQ stocks**

Presenter: **Audrone Jakaitiene**, Institute of Mathematics and Informatics, Lithuania

Co-authors: Aistis Raudys, Joana Katina

This paper analyses seasonality patterns in stock market liquidity. We analyse from intraday to annual liquidity patters across a set of NYSE and NASDAQ stocks from 1996 till 2010. As liquidity proxy we use best bid/ask spread as well as trading volume. In our research we create empirical nonparametric models to estimate future liquidity using simple averaging, regression, neural networks and nearest neighbour estimators. We conclude that there are very interesting patterns in liquidity and this information might be used by traders for reduction of trade costs. The second conclusion is that more sophisticated methods rather simple time series analysis give more accurate results when modelling liquidity seasonality.

C390: **The public and private information sources of volatility**

Presenter: **Anne Opschoor**, Erasmus University Rotterdam, Netherlands

Co-authors: Michel van der Wel, Dick van Dijk, Nick Taylor

We study the impact of public and private information on prices and volatility in financial markets. Several papers document that public information releases (in particular [surprises in] announcements of macroeconomic variables) significantly affect both quantities. In addition the relationship between asset prices and private information is established. So far, research on the impact of private information on volatility is scarce. Using a high-frequency dataset of the 30 year U.S. Treasury bond futures we aim to bridge this gap. Furthermore, we contribute to the econometric analysis of information effects on returns and propose a model that simultaneously estimates the effects of public and private information on both prices and volatility. We find that private information variables such as order flow and bid-ask spread are highly significant explanatory variables for volatility and should not be ignored. Finally, we assess whether the market reacts to information arrivals in an asymmetric fashion by distinguishing between good and bad news.

C554: **Inference on long memory in volatility with noisy realized measures**

Presenter: **Eduardo Rossi**, University of Pavia, Italy

Co-authors: Paolo Santucci de Magistris

A well documented stylized fact is that volatility of financial returns is stationary and it is characterized by long-range dependence, or long memory. As a consequence, realized volatility series are clustered and highly predictable, given its past realizations. Therefore, it is interesting to further explore the source of long memory in volatility, in order to understand its dynamic behavior. However, no theoretical justification for the presence of long memory in integrated variance is given. We show that, when the instantaneous volatility is driven by a fractional Brownian motion, the integrated volatility has long-range dependence. In the ideal situation, where prices are observed continuously and without microstructure noise, the spectral densities of integrated and realized volatility coincide. In real world, where prices are not observed continuously, realized volatility is affected by a measurement error. Therefore, discrete sampling and market microstructure noise induce a downward bias in the semiparametric estimates of fractional integration parameter. A correction of semiparametric estimation procedures is adopted. A Monte Carlo simulation and an empirical investigation provide evidence of the effectiveness of such a correction for the usual sampling frequencies adopted in applied work.

C094: **International price discovery in stock markets - A unique intensity based information share**

Presenter: **Kerstin Kehrle**, University of Zurich, Switzerland

Co-authors: Franziska Julia Peter

A new measure for contributions to price discovery based on the autoregressive conditional intensity model is proposed. While previous studies rely on equally spaced high frequency data, we use the information conveyed by quote revision intensities to determine a market's information share. Thereby, we account for the irregular nature of the data. Moreover, in contrast to the commonly applied standard approach, which yields lower and upper bounds for information shares, our method delivers a unique measure. An empirical application to US-listed Canadian stocks supports previous evidence for the home market leadership in price discovery. In a cross sectional analysis we also confirm the positive link between liquidity and the contribution to price discovery.

Saturday 11.12.2010 17:20 - 19:00 Parallel Session J – ERCIM

ES30 Room MAL 509 APPLIED STATISTICS II **Chair: Agustin Mayo**

E299: **Using individual-level models to model spatio-temporal combustion dynamics**
Presenter: **Rob Deardon**, University of Guelph, Canada
Co-authors: Irene Vrbik, Zeny Feng, John Braun

Individual-level models (ILMs) are a class of models originally designed to model the spread of infectious disease. However, they can also be considered as a tool for modelling the spatio-temporal dynamics of fire. We consider the much simplified problem of modelling the combustion dynamics on a piece of wax paper under relatively controlled conditions. The models are fitted in a Bayesian framework using Markov chain Monte Carlo (MCMC) methods. The focus here is on choosing a model that best fits the combustion pattern.

E312: **P-value under the alternative in robustness for hypothesis testing**
Presenter: **Alfonso Garcia-Perez**, Fundacion General de la UNED, Spain

The p-value (probability of obtaining as extreme or a more extreme result than the one actually observed assuming that the null hypothesis is true) is in fact a random variable with uniform distribution over the interval [0,1] under the null hypothesis. Nevertheless, we only use one of its values (the obtained in the observed sample) in the common hypotheses testing applications. Moreover, there are several accurate approximations to compute it when we consider small sample sizes. Some authors consider also the distribution of the random variable p-value, under the alternative hypothesis, to compare between several tests because the computation of the power function is, at least, problematic. In this paper we propose to use one of its values (the obtained in the observed sample) as a way to compare between several tests. We study some of its properties and obtain very accurate approximations to compute it for small sample sizes. This methodology is especially useful for comparing among robust and non-robust tests.

E802: **Integrated population model selection**
Presenter: **Panagiotis Besbeas**, Athens University of Economics and Business / University of Kent, Greece

Integrated population modelling provides an attractive framework for the combined analysis of survey data from wild animal populations. The approach unites methodology from the areas of capture-recapture and state-space modelling, and is proving to be extremely powerful for ecological research, allowing for example for different types of data, process and observation error, heterogeneity and density dependence to be included in the modelling. In this talk we discuss model selection in integrated population modelling. We consider both fixed and variable state-space model dimensions, corresponding to assuming a known or unknown age-structure for survival respectively, and show that model selection may result in unrealistic findings in either case. We highlight issues that might arise in practice, such as observation error in a state-space model incorrectly estimated as zero, and suggest potential solutions. The talk is illustrated with real and simulated data.

E612: **Fixing e-passports using mutual information estimates**
Presenter: **Apratim Guha**, University of Birmingham, UK
Co-authors: Tom Chothia

We present an estimate of the mutual information statistic based on kernel density estimates for multivariate "mixture data", i.e. multivariate data with both discrete and continuous components. For such data with one binary and one continuous component, we present a test, based on the value of the mutual information estimate, to check if the two components are independent. We derive asymptotic properties of the mutual information estimate when the two components are independent, and also when they are not, under mild regularity conditions on the kernel. We use these methods to analyse the security of the radio frequency identifier (RFID) tags in the new e-passports. In the course of our work, we have discovered an attack that can be used to trace a particular passport, based on its response times. We establish this attack by estimating the mutual information between an indicator indicating the ownership of a passport and the response times to particular messages. More importantly, we go on to design a fixed version of the e-passport, and, using the above test, we show that this new version resolves the traceability issue.

ES47 Room MAL 421 COMPUTATIONAL ISSUES ON TIME SERIES **Chair: Cira Perna**

E450: **Inference for dynamic panels with threshold effect and endogeneity**
Presenter: **Myung Seo**, London School of Economics, UK
Co-authors: Yongcheol Shin

Estimation and inference methods are developed for dynamic panel models with threshold effects and endogeneity. We allow for the threshold point to be unknown and the threshold variable to be endogenous. Depending on whether it is endogenous or not, we propose two different estimation methods; first-differenced GMM and first-differenced two-step least squares. The latter exploits the fact that the threshold variable is exogenous to achieve the super-consistency of the threshold estimator as observed in the classical threshold regression. We provide asymptotic distributions of the estimators under different assumptions on endogeneity of the threshold variable. Testing for the presence of threshold effect is also developed. Numerical examples illustrate application of our methods.

E800: **Multivariate stochastic volatility modelling using Wishart autoregressive processes**
Presenter: **Kostas Triantafyllopoulos**, University of Sheffield, UK

A new multivariate stochastic volatility estimation procedure for financial time series is developed. A Wishart autoregressive process is considered for the volatility precision covariance matrix, for the estimation of which a two stage procedure is adopted. In the first stage conditional inference on the autoregressive parameters is developed and the second stage develops unconditional inference, based on a Newton-Raphson iterative algorithm. The proposed methodology, suitable for medium dimensional data, bridges the gap between closed-form estimation and simulation-based estimation algorithms. Two examples, consisting of foreign exchange rates data and of data from the common constituents of the Dow Jones 30 Industrial Average index, illustrate the proposed methodology; for both examples we discuss asset allocation using as performance indicator mean-variance portfolio optimization.

E700: **The conjoint effects of stochastic risks on insurance portfolio internal models**
Presenter: **Valeria D'Amato**, University of Salerno, Italy
Co-authors: Emilia Di Lorenzo, Maria Russolillo, Marilena Sibillo

The study analyses the interplay between specific market consistent assumptions for interest rates, based on stochastic simulation procedures, and the Lee Carter stochastic model for the demographic description. In particular, for getting more reliable and accurate mortality projections, we pro-

pose the semi parametric bootstrap based on the Poisson distribution: the Stratified Sampling Bootstrap, reducing the variance estimator by means of Variance Reducing Techniques (VRTs). The framework in which the question is deepened and then applied, is the complex financial system of the pension annuity portfolio. In this context interest rates future evolution interacts with the uncertainty in future lifetimes, in particularly due to the systematic effects of the longevity phenomenon. Internal models have to be used for management and Solvency Capital Requirement purposes and the right description of the future financial and demographic scenario constitutes a great issue. In the paper we propose closed formulas for the portfolio surplus calculation in stochastic hypotheses for the main risk drivers.

E797: **Stratified Sampling scheme of death causes for forecasting the survival trend**

Presenter: **Maria Russolillo**, Universita degli Studi di Salerno, Italy

Co-authors: Valeria D'Amato

The 20th Century witnessed a high decline in the mortality level of populations, namely in the more developed countries. The positive evolution of mortality implied a substantial increase in the life expectancy, stimulated in the first times by the infant and youth mortality reduction and in the last years also by the reduction in the mortality rates of the old generations. The increasing survival in higher ages associated to the smaller number of births implies more and more aged populations. A higher longevity has, in this way, direct impact on the costs of the social security public systems since there is not a counterpart in terms of financial support of the contribution of a large young generation. The aim of the research is to investigate computational methods for performing a survival analysis in the framework of the stratified sampling technique. Usually by using this technique, the population is divided up into a set of smaller non-overlapping sub-groups (strata), then do a simple random sample in each sub-group. Strata can be natural groupings, such as age ranges or ethnic origins. We propose a stratification by death causes, to reduce standard error by providing some control over variance. In particular, the main idea of obtaining an acceptable accuracy of the estimate by solving the problem of reducing the variance estimator by means of Variance Reducing Techniques (VRT's). Therefore the forecasting mortality procedure reveals to be more efficient. The empirical results are presented using a range of graphical analyses.

ES49 Room MAL B34 GOODNESS-OF-FIT AND MODEL SELECTION IN GLLVM Chair: Maria-Pia Victoria-Feser

E151: **On the behaviour of marginal and conditional Akaike information criteria in linear mixed models**

Presenter: **Sonja Greven**, Ludwig-Maximilians-Universitaet Muenchen, Germany

Co-authors: Thomas Kneib

In linear mixed models, model selection frequently includes the selection of random effects. Two versions of the Akaike information criterion, AIC, have been used, based either on the marginal or on the conditional distribution. We show that the marginal AIC is no longer an asymptotically unbiased estimator of the Akaike information, and in fact favours smaller models without random effects. For the conditional AIC, we show that ignoring estimation uncertainty in the random effects covariance matrix, as is common practice, induces a bias that can lead to the selection of any random effect not predicted to be exactly zero. We derive an analytic representation of a corrected version of the conditional AIC, which avoids the high computational cost and imprecision of available numerical approximations. An implementation in an R package is provided. All theoretical results are illustrated in simulation studies, and their impact in practice is investigated in an analysis of childhood malnutrition in Zambia.

E104: **Testing for approximate fit in multivariate discrete data (with IRT applications)**

Presenter: **Alberto Maydeu-Olivares**, University of Barcelona, Spain

Co-authors: Harry Joe

Given the large number of degrees of freedom involved in IRT applications, it is very unlikely that any model for a realistic application is not rejected by a test of exact fit. We propose a Root Mean Squared Error of Approximation (RMSEA) for multivariate multinomial data. Although the approach presented here is completely general, we focus on its application to IRT models. Asymptotic methods can be used to obtain confidence intervals for an RMSEA, and hence tests of close fit.

E659: **Goodness-of-fit for generalized linear latent variables models**

Presenter: **Maria-Pia Victoria-Feser**, University of Geneva, Switzerland

Co-authors: David Conne, Elvezio Ronchetti

Generalized Linear Latent Variables Models (GLLVM) enable the modeling of relationships between manifest and latent variables, where the manifest variables are distributed according to a distribution of the exponential family (e.g. binomial or normal) and to the multinomial distribution (for ordinal manifest variables). These models are widely used in social sciences. To test the appropriateness of a particular model, one needs to define a Goodness-of-fit test statistic (GFI). In the normal case, one can use a likelihood ratio test or a modified version that compares the sample covariance matrix to the estimated covariance matrix induced by the model. In the binary case, Pearson-type test statistics can be used if the number of observations is sufficiently large. In the other cases, including the case of mixed types of manifest variables, there exists GFI based on a comparison between a pseudo sample covariance and the model covariance of the manifest variables. These types of GFI are based on latent variable models that suppose that the manifest variables are themselves induced by underlying normal variables (underlying variable approach). The pseudo sample covariance matrices are then made of polychoric, tetrachoric or polyserial correlations. We propose an alternative GFI that is more generally applicable. It is based on some distance comparison between the latent scores and the original data. This GFI takes into account the nature of each manifest variable and can in principle be applied in various situations and in particular with models with ordinal, and both discrete and continuous manifest variables. To compute the p-value associated to our GFI, we propose a consistent resampling technique that can be viewed as a modified parametric bootstrap. A simulation study shows that our GFI has good performance in terms of empirical level and empirical power across different models with different types of manifest variables.

E376: **Approximate maximum likelihood inference in latent variable models for ordinal data**

Presenter: **Silvia Cagnone**, University of Bologna, Italy

Co-authors: Silvia Bianconcini

Latent variable models for ordinal data represent a useful tool in several fields of research in which the constructs of interest are not directly observable, so that one or more latent variables are required to reduce the complexity of the data. In these cases, problems related to the integration of the likelihood function of the model can arise since analytical solutions do not exist. One of the most used numerical approximation is based on the application of the fixed-point Gauss-Hermite (GH) quadrature. With this approximation the accuracy of estimates is strictly related to the number of quadrature points as well as the sample size. When the number of latent dimensions is large, computational considerations require that the number of quadrature points per dimension be few, but the likelihood cannot be accurately evaluated with the sparse fixed points in the latent space. We investigate the theoretical properties of the GH maximum likelihood estimators in presence of multidimensional integrals, and we propose alternative solutions (e.g. multidimensional spline functions) in order to overcome the main limitations of GH approximations.

ES69 Room MAL 532 MIXTURE MODELS I Chair: Christian Hennig

E166: **Hidden Markov model selection by Monte Carlo experiments**

Presenter: **Luca De Angelis**, University of Bologna, Italy

Co-authors: Michele Costa

In the last years several methods have been proposed to detect the optimal size of hidden Markov models. The general consensus in both theoretical and empirical literature is that all selection methods are heavily characterized by a lack of robustness which inevitably affects the reliability of the chosen model. We develop a Monte Carlo study in order to compare the most widespread methods and to evaluate their performances with respect to the a priori known structure of the simulated model. We investigate the effects related to the transition probability matrix, the sample size and the order of the Markov chain. Our results provide a novel framework able to guide in the interpretation and the use of different information criteria and log-likelihood tests and to provide useful information for hidden Markov models selection.

E575: **Mixed effect models for multivariate mixed responses**

Presenter: **Irene Rocchetti**, University La Sapienza, Italy

Co-authors: Marco Alfo

We describe a regression model for mixed bivariate responses, where association between outcomes is modeled through latent effects, accounting for both heterogeneity and dependence. In a Finite Mixture (FM) context a relevant question arises when dependence should be tested vs independence. With Gaussian random effects, dependence is represented by linear correlation and the linear predictor can be specified as depending on the correlation coefficient. When a discrete mixing distribution is employed, null correlation does not imply independence and the correlation coefficient is not an explicit model parameter. Discrete bivariate mixing distributions lead to a (almost) perfect dependence between the two random effects: each location in a margin is associated to only one location in the other one. We relax the uni-dimensionality of the standard FM model, where joint and marginal distributions are defined to be the same, and proceed to define a multidimensional latent class structure, with a possibly different number of locations in each margin, giving raise to a full (bi-dimensional) association structure. Here, we may properly test the difference between the estimated joint distribution and the corresponding independence distribution through standard χ^2-based tools. The approach will be evaluated both in a simulation study and in real data examples.

E811: **Finite mixture models for longitudinal responses with attrition: a pattern mixture specification**

Presenter: **Luciano Nieddu**, Universita LUSPIO, Italy

Co-authors: Cecilia Vitiello, Marco Alfo

Longitudinal studies often generate incomplete response data according to a missing not at random mechanism: ignoring the drop-out process can lead to biased inferences. Lately joint models for time-to-event and longitudinal responses have generated considerable interest. Different factorizations of the joint pdf. of the longitudinal process and the time-to-event process lead to Selection and Pattern Mixture Models. Shared parameter models represent an alternative appealing tool for the joint modeling of the longitudinal process and the drop-out mechanism. The longitudinal process $Y_{it}, i = 1,\ldots n, t = 1,\ldots T_i$ where n is the number of units and T_i is the number of observations for each unit, is modeled via a mixed effect models and the hazard for drop-out is assumed to depend on random effects and covariates via a proportional hazard model. Random effects are the key component in the formulation of the drop-out mechanism and in the modeling interdependence between the two processes. Therefore any distributional assumptions on the random effects is a critical issue to be addressed. Our goal is to define a simple pattern mixture model, with shared nonparametric random effects, to detect informative missingness by evaluating the effect on inferences of ignoring informative missing data, under different generating missing processes.

E498: **Choice of the model and the number of components in von Mises-Fisher mixtures**

Presenter: **Wafia Parr Bouberima**, Paris Descartes University, France

Co-authors: Mohamed Nadif, Yamina Khemal Bencheikh

Finite mixture models underpin a variety of techniques in major areas of statistics including cluster analysis. The search for an optimum number of clusters leading to the greatest separation and the choice of a convenient model are both heuristic problems. These can be treated in many ways in which several information criteria are proved of a high performance. In this work, considering mixtures of von Mises-Fisher as an adapted model for directional data, we are interested in the performance of a set of selected information criteria: Bic, Aic, Aic3, Aic4, Aicc, Aicu, Caic, Clc, Icl-Bic, Ll, Icl, Awe. When applied to many groups of simulated data, our results showed that the information criteria quality was limited for many situations depending on the size and dimension of data, and the cluster overlap degree. Obviously, these factors have an important influence on the performance of the cited criteria. Let X be a real directional data matrix, how could users estimate the number of clusters or choose the appropriate model using the information criteria with more confidence ? In this work, we propose solutions to this problem.

ES51 Room MAL 152 RECURSIVE PARTITIONING Chair: Claudio Conversano

E634: **Classification and regression trees for multivariate multi-way data**

Presenter: **Roberta Siciliano**, Universita di Napoli Federico II, Italy

Co-authors: Valerio A. Tutore, Massimo Aria, Antonio D'ambrosio

Tree-based methods allow to describe how a set of predictors (of categorical or numerical type) explain the heterogeneity or variation of one response variable (univariate tree) or more response variables (multivariate tree) of dichotomous/categorical type (classification tree) or numerical type (regression tree) on the basis of the measurements of all variables on a given sample. Either ensemble methods or decision tree selection provide a decision rule definition in order to predict the response class(es)/value(s) for new objects where only measurements of predictors are known. Methods for recursive partitioning and decision rule definition have been so far proposed for the analysis of two-way data matrix crossing objects and variables. Recent results have been developed for the analysis of three-way data, namely a cube crossing units, variables and occasions. This paper provides a general tree-based methodology for the analysis of multi-way data. The concepts of conditional impurity and consensus tree partitioning will be introduced. Recursive partitioning methods for two-way data (standard methods) and for three-way data as well as new other procedures will be shown to be special cases. Particular attention will be given to multivariate trees where the response variables describe the objects rankings as expressed by units.

E688: **Projection-based recursive partitioning for large, high-dimensional data**

Presenter: **Adalbert Wilhelm**, Jacobs University, Germany

Co-authors: Iulian Ilies

Recent work in cluster analysis has focused on designing algorithms that address the issue of ever growing data sets and provide meaningful solutions for data with high cardinality and/or dimensionality, under the natural restriction of limited resources. Within this line of research, we

propose a method drawing on the principles of projection pursuit and grid partitioning, which focuses on reducing computational requirements for large data sets without loss of performance. To achieve that, we rely on procedures such as sampling of objects, feature selection, and quick density estimation using histograms. The present algorithm searches for low-density points in potentially favorable one-dimensional projections, and partitions the data by a hyperplane passing through the best split point found. Tests on synthetic and reference data indicate that our method can quickly and efficiently recover clusters that are distinguishable from the remaining objects on at least one direction; linearly non-separable clusters are usually subdivided. The solution is robust in the presence of noise in moderate levels, and when the clusters are partially overlapping.

E382: **Treatment interaction trees: a new tree-based method to identify treatment-subgroup interactions**

Presenter: **Elise Dusseldorp**, TNO-Dutch Center for Applied Scientific Research, Netherlands

Co-authors: Iven van Mechelen

When two treatments, A and B, are available, some subgroup of patients may display a better outcome with treatment A than with B, whereas for another subgroup the reverse may be true. If this is the case, a qualitative (i.e., a disordinal) treatment-subgroup interaction is present. Such interactions imply that some subgroups of patients should be treated differently, and are therefore most relevant for clinical practice. In case of data from randomized clinical trials with many patient characteristics that could interact with treatment in a complex way, a suitable statistical approach to detect qualitative treatment-subgroup interactions is not yet available. In this presentation, we introduce a new method for this purpose, called Treatment INteraction Trees (TINT). TINT results in a binary tree that subdivides the patients into terminal nodes on the basis of patient characteristics; these nodes are assigned to one of three classes: a first one for which A is better than B, a second one for which B is better than A, and an optional third one for which type of treatment makes no difference. The method will be compared to other tree-based methods (e.g., STIMA). Results of a simulation study will be shown.

E857: **Threshold interaction detection in generalized linear models**

Presenter: **Claudio Conversano**, University of Cagliari, Italy

Simultaneous Threshold Interaction Modeling Algorithm (STIMA) has been recently introduced in the framework of statistical modeling as a tool enabling us to automatically select interactions in a Generalized Linear Model (GLM) through the estimation of a suitable defined tree structure called "trunk". STIMA integrates GLM with a classification tree algorithm or a regression tree one, depending on the nature of the response variable (nominal or numeric). Accordingly, it can be based on the Classification Trunk Approach (CTA) or on the Regression Trunk Approach (RTA). In both cases, interaction terms are expressed as "threshold interactions" instead of traditional cross-products. Compared with standard tree-based algorithms, STIMA is based on a different splitting criterion as well as on the possibility to "force" the first split of the trunk by manually selecting the first splitting predictor. This paper focuses on different specifications of the generalized linear model with threshold interaction effects defined by STIMA, which depend on the nature of the response variable, as well as on different criteria for doing model selection. Results on real and synthetic data are presented in order to compare the performance of STIMA with those of alternative methods.

ES57 Room MAL B35 ROBUSTNESS AND COMPLEXITY IN SEMI- AND NON-PARAMETRICAL STATISTICS Chair: Andreas Christmann

E105: **Local bilinear multiple-output quantile regression**

Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium

Co-authors: Marc Hallin, Zudi Lu, Miroslav Siman

Recently a new quantile regression concept – based on a directional version of Koenker and Bassett's traditional single-output one – for multiple-output regression problems has been introduced. The empirical counterpart of that concept produces polyhedral contours that cannot adapt to nonlinear or/and heteroskedastic dependencies. We propose a local bilinear version of those contours, which asymptotically recovers the conditional halfspace depth contours of the multiple-output response. We establish a Bahadur representation, along with asymptotic normality results, and provide examples.

E120: **Time varying hierarchical Archimedean copulae**

Presenter: **Ostap Okhrin**, Humboldt Universitat zu Berlin, Germany

Co-authors: Wolfgang Haerdle, Yarema Okhrin

There is increasing demand for models of time-varying and non-Gaussian dependencies for multivariate time-series. Available models suffer from the curse of dimensionality or restrictive assumptions on the parameters and the distribution. The promising class of models are the hierarchical Archimedean copulae (HAC) that allow for non-exchangeable and non-Gaussian dependency structures with a small number of parameters. In this paper we develop a novel adaptive estimation technique of the parameters and of the structure of HAC for time-series. The approach relies on a local change point detection procedure and a locally constant HAC approximation. Typical applications are in the financial area but also recently in the spatial analysis of weather parameters. We analyse the time varying dependency structure of stock indices and exchange rates. We find that for stock indices the copula parameter changes dynamically but the hierarchical structure is constant over time. Interestingly in our exchange rate example both structure and parameters vary dynamically.

E328: **Supervised invariant coordinate selection**

Presenter: **Hannu Oja**, University of Tampere, Finland

Co-authors: Eero Liski, Klaus Nordhausen

Dimension reduction plays an important role in high dimensional data analysis. Principal component analysis (PCA), independent component analysis (ICA), and sliced inversion regression (SIR) are well-known but very different analysis tools for dimension reduction. It appears that these three approaches can all be seen as the comparison of two different (scatter) matrices S_1 and S_2. The components used for dimension reduction are then given by the eigenvectors of $S_1^{-1}S_2$. In SIR the second scatter matrix is supervised and therefore the choice of the components is based on the dependence between the observed random vector and a response variable. Based on these notions, we extend the invariant coordinate selection (ICS) to be applicable also in supervised dimension reduction. The second scatter matrix S_2 is then allowed to be supervised. Several families of supervised scatter matrices are discussed, and their use in supervised dimension reduction is illustrated with examples and simulations. Also some results on the asymptotic behavior of the eigenvectors and eigenvalues of $S^{-1}S_2$ are given.

E216: **Fitting additive models with Support Vector Machines: consistency and robustness**

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

Co-authors: Robert Hable

Modern learning theory is still a fast developing topic in statistics and support vector machines (SVMs) play an important role in this area. 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. SVMs based on the combination of the classical Gaussian RBF kernel and a Lipschitz continuous loss function are

known to be (i) solutions of a well-posed mathematical problem in Hadamard's sense, (ii) computationally feasible, and (iii) universally consistent and robust nonparametric estimators for functions. The talk will treat the case that the statistician has some limited prior knowledge such that a semiparametric version of an SVM is of interest. In the talk it will be shown how such semiparametric models, in particular additive models, can be fitted by support vector machines. Results on consistency and statistical robustness will be given as well. As a special case we will treat semiparametric quantile regression based on SVMs.

ES58 Room MAL B36 FUZZY SETS IN STATISTICS Chair: M. Angeles Gil

E415: **Testing linearity for a regression model with imprecise elements: a power analysis**

Presenter: **Maria Brigida Ferraro**, Sapienza Univesity of Rome, Italy

Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez

A linearity test for a simple regression model with imprecise random elements is analyzed. The concept of LR fuzzy random variable is used to formalize imprecise random elements. The proposed linearity test is based on the comparison of the simple linear regression model and the nonparametric regression. In details, based on the variability explained by the above two models, the test statistic is constructed. The asymptotic significance level and the power under local alternatives are established. Since large samples are required to obtain suitable asymptotic results a bootstrap approach is investigated. Furthermore, in order to illustrate how the proposed test works in practice, some simulation and real-life examples are given.

E604: **Factorial analysis of variance for fuzzy data**

Presenter: **Takehiko Nakama**, European Center for Soft Computing, Spain

Co-authors: Ana Colubi, Maria Asuncion Lubiano

In many real-world applications, observations are inherently imprecise, uncertain, or linguistic, and fuzzy sets are effective in encoding them. In this study, we establish factorial analysis of variance (ANOVA) for fuzzy data. Each observation is assumed to be a fuzzy set, and factorial ANOVA determines whether to reject null hypotheses about the main effects and interactions of factors on the observed fuzzy sets. The Minkowski support function is used to obtain a metric for fuzzy sets and to transform them to Hilbert-space-valued functions. We first establish factorial ANOVA for Hilbert-space-valued data and then particularize it to the case of fuzzy data. We derive test statistics that are appropriate for testing the null hypotheses. Typically the distributions of the test statistics will be unknown, so we develop a bootstrap scheme for approximating the p-values of the observed test statistics.

E605: **FINDCLUS: fuzzy individual differences clustering**

Presenter: **Paolo Giordani**, Sapienza University of Rome, Italy

Co-authors: Henk A.L. Kiers

The ADCLUS model is a tool for overlapping clustering of two-way proximity matrices (objects X objects). In SAFC, ADCLUS has been extended to provide a fuzzy partition of the objects, that is the objects belong to the clusters with the so-called membership degrees ranging from zero (complete non-membership) to one (complete membership). INDCLUS is a generalization of ADCLUS for handling three-way proximity arrays (objects X objects X judges). Here, we propose an extension of INDCLUS capable to offer a fuzzy partition of the objects by generalizing in a three-way context the idea behind SAFC. This new model is called FINDCLUS (Fuzzy INdividual Differences CLUStering). An algorithm is provided for fitting the FINDCLUS model to the data. Finally, the results of an application to real data are discussed.

E360: **A comparison among supervised classification criteria for fuzzy random sets**

Presenter: **Gil Gonzalez-Rodriguez**, University of Oviedo, Spain

Co-authors: M. Angeles Gil, Wolfgang Trutschnig

Fuzzy random sets are useful to model experiments in which both imprecision and randomness are involved. Based on the linkage between fuzzy random sets and random elements in an appropriate separable Hilbert space established in previous works, two non-parametric supervised classification criteria for fuzzy data were introduced. Afterwards, with the aim of taking advantage of the possible relationship between a kind of "location" measure of functional data and the corresponding membership class, another method based on the estimation of a specific relative proximity of a given point to each class was considered. All these three methods depend on some key parameters, namely the bandwidth associated with the density estimation and the ball size in the first two cases and the determination, by means of a weighting average, of "class centres" in the last one. The aim of this work is to analyse a way of tuning these parameters for improving the classification criteria, as well as to analyse the behaviour of these methods both with real and simulated data.

Sunday 12.12.2010 09:00 - 10:40 Parallel Session K – CFE

CS29 Room MAL 152 MULTIVARIATE VOLATILITY MODELS II **Chair: Massimiliano Caporin**

C122: **Model based monte carlo pricing of energy and temperature quanto options**
Presenter: **Juliusz Pres**, West Pomeranian University of Technology, Poland
Co-authors: Massimiliano Caporin, Hipolit Torro

In the last few years, weather derivatives have become very popular tools in weather risk management. One of the elements supporting their diffusion was the increase in volatility observed on many energy markets. Across the several available contracts, quanto options are now becoming very popular for a simple reason: they take into account the strong correlation between energy consumption and certain weather conditions, thus allowing controlling price and weather risk at the same time. These products have higher efficiency and, in many cases, are significantly cheaper than the composition of simpler plain vanilla options. Unfortunately, the specific features of energy and weather time series do not allow using neither the analytical formulae based on the Black-Scholes pricing approach, nor other more advanced continuous time methods extending the Black-Schole one, unless under strong and unrealistic assumptions. In this study, we propose a Monte Carlo pricing framework based on a bivariate time series model. Our approach takes into account the mean and variance interdependence between temperature and energy price series. Furthermore, it includes other relevant empirical features, such as periodic patterns in mean, in variance and in correlations. The model structure allows a more appropriate pricing of quanto options compared to traditional methods.

C848: **Multivariate volatility modelling of electricity futures**
Presenter: **Diane Pierret**, UCL Louvain-la-Neuve, Belgium
Co-authors: Luc Bauwens, Christian Hafner

Deregulation of European electricity markets has led to an increasing need in understanding the volatility and correlation structure of electricity prices. We analyse a multivariate futures series of the EEX index, using asymmetric GARCH type models for volatilities and augmented DCC type models for correlations. In particular, we allow for smooth structural changes in the unconditional volatilities and correlations. It is also shown that the correlation dynamics of long term contracts is very different from that of short term contracts.

C564: **A nonparametric technique for forecasting the variance-covariance matrix**
Presenter: **Robert O'Neill**, University of Manchester, UK
Co-authors: Ralf Becker, Adam Clements

The forecasting of variance-covariance matrices (VCMs) is an important issue in finance, having applications in portfolio selection and risk management. Advances in multivariate volatility modelling are hindered by the restrictions that all forecasts of VCMs must be positive-definite and symmetric, restrictions which mean it is also difficult for models to incorporate macroeconomic information. We apply a new approach in a multivariate setting, using comparisons of variables, calculated using a realization of the VCM and macroeconomic data to produce weights using a nonparametric kernel, which are then applied to realizations of the VCM. As we take averages of symmetric, positive-definite matrices, the resulting forecast also has these properties. We show, via a comparison of our method with other models, that our non-parametric approach is able to produce forecasts of the VCM which are more accurate at small forecast horizons. We also show that macroeconomic data are critical to this improvement, providing evidence that these variables provide important information for predicting the VCM.

C455: **Variance clustering improved dynamic conditional correlation MGARCH estimators for vast dimensional systems**
Presenter: **Gian Piero Aielli**, University of Padova, Italy
Co-authors: Massimiliano Caporin

We extend the dynamic conditional correlation (DCC) model by allowing for an unknown asset clustering structure based on the dynamic parameters of the asset conditional variances. The resulting model can be estimated relying on a two-step approach: first, the clustering structure is estimated via gaussian mixture clustering of the univariate estimates of the variance dynamic parameters; then, the asset variance/correlation dynamics are estimated conditionally on the variance clustering provided by the first step. Thanks to the suggested approach - and in contrast with the traditional two-step DCC estimators - we are capable of estimating jointly the variance dynamic parameters and the correlation dynamic parameters of vast dimensional systems. We illustrate via simulations and applications to real data the advantages of the proposed approach with respect to the traditional two-step methods.

CS28 Room MAL B34 RISK MANAGEMENT **Chair: Mike So**

C478: **The economic value of dynamic hedging strategies**
Presenter: **Nathan Liu**, Feng Chia University, Taiwan

There has been growing importance placed on research in hedging performance for static and dynamic strategies. Most literature compares the statistical efficiency of different hedging strategies to decide the best strategy. However, it is still difficult to differentiate their performance for the further applications. This study tries to discuss the economic values among hedging strategies on investors' side. Under the setting of constant relative risk aversion (CRRA), the derived realized utility function is used to compare the hedging performance. In consideration of economic value of hedging strategies, this study can get more practical inference for the investors with different levels of risk aversion. There are 15 commodities used to estimate the economic values, i.e., stock index (FTSE 100, Nikkei 225 and S&P500), exchange rate (British Pond, Japanese Yen and Swiss Franc), metal (gold and silver), grain (corn, soybeans and soybean oil), soft (coffee, cotton and sugar), and energy (crude oil). The hedging strategies include original least squares (OLS), constant conditional correlation (CCC) and dynamic conditional correlation (DCC) models. In addition, this study utilizes the magnitude of basis to build the selective hedging strategy and estimates its economic value. This study helps investors decide the optimal hedging strategy.

C748: **On the estimation of extreme values for risk assessment and management: The average conditional exceedance rate method**
Presenter: **Kai Erik Dahlen**, Molde University, Norway
Co-authors: Arvid Naess, Per Bjarte Solibakke, Sjur Westgaard

For prediction of tail quantiles we have used the average conditional exceedance rate (ACER) method for prediction of extreme values, generalized to be able to accommodate heavy tailed data. The ACER method has been applied to both the logarithmic returns and to the residuals after an AR-GARCH filtering of the returns for daily electricity spot prices from Nord Pool. Since the data used are highly seasonal and displays significant serial correlation and volatility clustering, the conditional approach with the ACER method applied to the residuals is of interest here. Earlier it has

been observed that the use of the peaks over threshold (POT) method, compared to just the use of an AR-GARCH model, to predict the residual tail quantiles produces more accurate estimates when the observed price changes are extreme (as observed in the hourly spot price data from the Nordic exchange for electricity trading, Nord Pool). Following this we have applied the ACER method to the AR-GARCH filtered price changes in the Nord Pool data. These analyses have led to the observations that the ACER method is able to estimate the tail quantiles even more accurately than the POT method for these data. Where the POT method uses the generalized Pareto distribution (GPD) to fit observations over a given threshold, the ACER method fit the observations to an extreme value distribution modified to capture to some extent also sub-asymptotic behaviour. In principle, since there is no asymptotic assumption involved in the derivation of this distribution, the ACER method is able to utilize more of the observations than the asymptotic alternatives, leading to more accurate estimate of the tails. This is expected to be the case for data where the appropriate choice of the threshold in the POT method leads to a high proportion of discarded observations.

C304: **Testing separately for positive and negative jumps**

Presenter: **Janine Balter**, Saarland University, Germany

Co-authors: Stefan Kloessner

Testing for jumps is confronted with two empirical problems: existing jump tests are unable to detect gradual jumps, and their finite-sample distribution is severely affected by high volatility of volatility. To solve these problems, we consider new jump tests based on sparse sampling in combination with intradaily OHLC data, allowing separate testing for positive and negative jumps. To cope with the slow asymptotics, especially when volatility changes fast, we use the Gini coefficient of intradaily volatility estimates to determine an appropriate distribution of the Pearson system which replaces the limiting normal distribution for moderate frequencies. By extensive Monte Carlo studies, we show that these new tests have remarkably good size and power even when volatility fluctuates heavily.

C689: **The hazard-adjusted portfolio: A new capital allocation scheme from an extreme-risk management perspective**

Presenter: **Falk Laube**, University of Luxembourg, Luxembourg

Co-authors: Virginie Terraza

Recent changes in market dynamics have lead the financial industry to the brink of breakdown. The degree of co-movements between markets has increased significantly in the recent past. These strong effects of contagion and spillover between two or more markets are responsible for the disappearance of the protective diversification effect in portfolios. This paper proposes a new asset allocation scheme that complements the traditional portfolio optimization approach. Recognizing the fact that diversification no longer exists when most needed, we propose a pre-allocation step that ensures that risk metrics are representative when applied during optimization. In light of these phenomena and recognizing the existence of multiple investment horizons within single markets, we extend our extreme-risk-protection approach by multivariate multifractal analysis. Using extended versions of the MSM, we model volatility dynamics both on a time-structural and frequency domain. We use this inherent multifractal analysis to derive information about extreme risk probabilities for individual assets both in the univariate and multivariate sense. We identify common volatility cycles and quantify probabilities for joint crises and extreme co-movements as well as conditional multifractal correlations between markets. We explicitly quantify contagion by the concept of re-correlation, where assets observed to be independent correlate suddenly and only during abnormal market times and induce a chain of losses during crises. We use our approach to create a synthetic Fund of Hedge Funds and show that crisis-resilient portfolios can be constructed in a robust manner.

CS33 Room MAL G16 BAYESIAN FINANCIAL ECONOMETRICS AND RISK MANAGEMENT Chair: Richard Gerlach

C302: **Nonlinear quantile autoregressive models with exogenous variables and heteroskedasticity**

Presenter: **Cathy Chen**, Feng Chia University, Taiwan

Co-authors: Richard Gerlach

A threshold nonlinear quantile autoregression with exogenous regressors and heteroskedasticity is considered, allowing semi-parametric representation of both asymmetry and volatility clustering. As such, GARCH-type dynamics with nonlinearity are added to a nonlinear quantile time series regression model. An adaptive Bayesian Markov chain Monte Carlo scheme, exploiting the link between the quantile loss function and the skewed-Laplace distribution, is employed for estimation and inference, simultaneously estimating and accounting for heteroskedasticity, varying threshold limits and delay lags. A simulation study illustrates sampling properties of the method. Two data sets are considered in the empirical applications: modelling daily maximum temperatures in Melbourne, Australia; and exploring dynamic linkages between the futures markets in Taiwan and Singapore.

C411: **Bayesian estimation of pair copula construction models with applications to financial data**

Presenter: **Anastasios Panagiotelis**, Technische Universitaet Muenchen, Germany

Modeling dependence in financial data has important implications for risk management. A popular way to construct multivariate models with a wide variety of dependence structures and known margins is to use copulas. An approach known as Pair Copula Constructions (PCCs) uses bivariate copulas as building blocks to create multivariate copulas of potentially high dimension. These constructions are flexible and are able to capture asymmetric tail dependence, which is a common empirical feature of financial datasets. As such, PCCs have been shown to be highly competitive in the modeling of financial returns. This talk outlines current and ongoing research on PCCs with a focus on models that are applied to financial data and their estimation using Markov chain Monte Carlo (MCMC). Specific issues that shall be discussed are PCCs with time varying parameters or Markov switching that capture the dynamic nature of dependence, the impact of joint Bayesian estimation of marginal and copula parameters on Value at Risk (VaR) estimates, Bayesian estimation of PCCs with discrete margins and identifying conditional independence using Bayesian model selection to obtain a more parsimonious PCC for higher dimensional data.

C448: **Assessing the stability of conditional variance models for daily returns of the WIG index**

Presenter: **Jacek Kwiatkowski**, Nicolaus Copernicus University in Torun, Poland

We propose a method to estimate time varying adjusted probability of informed trading (PIN) and probability of symmetric order-flow shock (PSOS) proposed in the literature as measures of asymmetric information and illiquidity, respectively. Our method is an extension of an existing one to estimate time varying PIN using high-frequency transaction data. The approach is based on an asymmetric autoregressive conditional duration model of expected durations of buy and sell orders. Our empirical results indicate that daily adjusted PIN is much more stable than daily PIN and that, in contrast to PIN, adjusted PIN is negatively correlated with integrated volatility (IV). Instead, daily PSOS is the component that is positively correlated with IV. Moreover, by comparison with the daily adjusted PIN, the daily PSOS seems to be quite volatile over time.

C888: **Model selection and adaptive MCMC for Bayesian cointegrated VAR models**

Presenter: **Kannan Balakrishnan**, Boronia Capital, Australia

Co-authors: Gareth Peters, Ben Lascock, Chris Mellen

This talk presents developments in the area of Bayesian Cointegrated Vector Autoregression models utilising computationally efficient Markov chain Monte Carlo (MCMC) sampling methodology. The methodology exploits existing matrix-variate conjugacy properties in an adaptive MCMC strategies for posterior inference on the matrix of cointegration vectors. Secondly we develop a model incorporating a mixture of matrix-variate Gaussian and alpha-stable error distributions for the Error Correction Model formulation. We derive novel conjugacy results for this error structure. These results allow one to again exploit adaptive MCMC strategies for efficiently sampling the resulting posterior distribution. The alpha-stable errors allow one to model skewness and asymmetry, whilst still including as a special case the standard Gaussian CVAR model. We illustrate this model on inter-day jumps in price series. The third aspect of this talk will extend this model for the case in which the alpha-stable portion of the model does not admit a closed form matrix-variate likelihood structure, via Approximate Bayesian Computation. This is particularly relevant when the symmetry parameter is non-zero in our particular alpha-stable parameterization.

CS47 Room MAL 151 BEHAVIOURAL FINANCE II Chair: Gulnur Muradoglu

C218: **Herding in foreign direct investments**
Presenter: **Kristina Vasileva**, Cass Business School, UK
Co-authors: Gulnur Muradoglu, Mario Levis

We investigate herding in foreign direct investment outflows with data on FDI outflows from the OECD member countries towards the rest of the world. This is investigated in a bilateral country pair setting of 1867 country pairs over 25 years. The intuition and motivation for this study comes from a well documented behavioural finance phenomenon – herding in the equity markets, something that has not been investigated as such for FDI flows. We test several hypothesis: world leaders herd direct investors; regional leaders herd direct investors; portfolio investors herd FDI investors, the FDI receiving country herds further investors. We find evidence of support for herding in FDI.

C228: **The rationality of financial optimism**
Presenter: **Jiayi Balasuriya**, Cass Business School, UK
Co-authors: Gulnur Muradoglu, Peter Ayton

This study investigates whether being financially optimistic has any benefits on objective and subjective well-being. We define financial optimism as the overestimation of the favorable outcome in an individual's future financial situation. We analyze data from the British Household Panel Survey covering the period 1991 to 2007. We find that financial optimism is positively associated with current income level and lead to future increase in financial wealth but has little effect on one's subjective well-being. Our results indicate that respondents in the BHPS do give answers on the base of certain rationalities and optimism is not an illusion that respondents create to fool themselves in order to stay happy or satisfied. Additionally, factors contributing to subjective well-being are unlikely to have long-term effects and respondents' subjective feelings do change fairly frequently.

C553: **Analogical transfer of experience and the misuse of diversification**
Presenter: **Ugo Rigoni**, University of Venice, Italy
Co-authors: Massimo Warglien

When the payoff function is convex, diversification can be the wrong strategy for dealing with risky choices. We constructed two experimental settings in which it is stochastically dominant not to diversify. Nevertheless, subjects inappropriately preferred the diversification strategy. A de-biasing treatment suggests that this diversification fallacy may be interpreted as the outcome of the analogical transfer of decision heuristics learned in more usual domains.

C694: **Portfolio compositions of individual investors**
Presenter: **Belma Ozturkkal**, Bilgi University, Turkey
Co-authors: Ana Maria Fuertes, Gulnur Muradoglu

This paper studies the relation between the portfolio composition of individual investors and their demographic characteristics during the most recent crisis period. Our sample consists of 3.3 million transactions by 60,300 individual Turkish investors over the period 2008-2010. We focus our analysis on three aspects of portfolio composition: diversification, size of financial asset portfolios, and trading properties. We find that individual investors that have larger financial wealth, that trade larger volumes and that receive financial advise have more portfolio diversification. In terms of demographics, better educated, older individuals have more diversified portfolios. Finance professionals tend to be less diversified suggesting a relatively higher risk loving behaviour. Men have better portfolio diversification than women and singles have better portfolio diversification than married people whereas previous studies have suggested that women and married people are more risk averse. Furthermore, total financial wealth is higher for women and for single people although trading volume is higher for men and married people. Thus cultural roles assigned to gender and marriage must be taken into account in assessing risk taking behaviour.

CS55 Room MAL B33 ROBUSTNESS IN COMPLEX MODELS AND TIME SERIES Chair: Marco Riani

C159: **Extremes of continuous-discrete time series**
Presenter: **Kamil Turkman**, University of Lisbon, Portugal

In many applications, the primary interest is the supremum of some continuous time process over a specified period. However, data are usually available on a discrete basis only. The true continuous time maximum will be larger than the maxima of discrete versions sampled at different frequencies. If one wishes to estimate the extremes of the continuous time process based on discrete time data, an adjustment is required to allow for the effect of discrete sampling and provide a measure of how much smaller it tends to be.

C399: **Integrated banking economic capital: The forward search approach**
Presenter: **Tiziano Bellini**, Universita Degli Studi di Parma, Italy

Economic capital models are potentially powerful tools for banking risk management and for the supervisory review process. In order to fulfill this potential, they need to go beyond the modular approach that dominates Basel II and to be fully integrated. In this setting, our contribution is twofold. On the one hand, exploiting a Monte Carlo simulation framework we develop a model where both banking assets and liabilities are considered in order to estimate economic capital requirements. On the other, we apply the forward search to estimate model parameters to be used into our scenario generation environment. We exploit a regression analysis to infer the relationship between macroeconomic factors and credit risk, while a Nelson Siegel model is estimated through Kalman filter to capture the evolution of interest rates.

C466: **Sensitivity and robustness in MDS configurations for mixed-type data with applications**

Presenter: **Rosario Romera**, Universidad Carlos III de Madrid, Spain

Co-authors: Aurea Grane

Multidimensional scaling (MDS) techniques are initially proposed to produce pictorial representations of distance, dissimilarity or proximity data. Sensitivity and robustness assessment of multivariate methods is essential if inferences are to be drawn from the analysis. To our knowledge, the literature related to MDS for mixed-type data, including variables measured at continuous level besides categorical ones, is quite scarce. The main motivation of this work was to analyze the stability and robustness of MDS configurations as an extension of a previous study on a real data set coming from a panel-type analysis designed to assess the economic crisis impact on Spanish people who were in situations of high risk of being socially excluded. The main contributions of the paper on the treatment of MDS configurations for mixed-type data are: (i) to propose a joint metric based on distance matrices computed for continuous, multi-scale categorical and/or binary variables, (ii) to introduce a systematic analysis on the sensitivity of MDS configurations and (iii) to present a systematic search for robustness and identification of outliers through a new procedure based on geometric variability notions.

C887: **Robust fitting of a linear mixed model**

Presenter: **Isabel Molina**, Universidad Carlos III de Madrid, Spain

Co-authors: Betsabe Perez, Daniel Pena

Different proposals for robustification of the estimators of the variance components in a linear mixed model are discussed. Specifically, several variations of the Henderson method III, a moments method that gives explicit expressions of the estimators and which avoids the normality assumption, are introduced. Simulation studies are carried out to analyze the robustness of the different proposals. 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.

CS59 Room MAL B30 LARGE PANEL DATA MODELS Chair: Lorenzo Trapani

C552: **Two stage inference in heterogeneous panels**

Presenter: **Carolina Castagnetti**, University of Pavia, Italy

Co-authors: Eduardo Rossi, Lorenzo Trapani

Inference in a stationary panel model where slopes are allowed to be heterogeneous and common unknown factors are present are considered. A two stage estimator is proposed, based on the CCE estimator in the first stage and on a similar approach to the Interactive Effect estimator in the second stage. This affords the estimation of the common factors coefficients, and, as a by-product, it also yields another estimator for the individual slopes and for the average slope. The properties of both estimators are analysed. Building on the asymptotics derived in the paper, we also propose a test for strong cross sectional dependence arising from common factors and for slope heterogeneity.

C404: **Testing weak exogoneity in cointegrated panels**

Presenter: **Christian Gengenbach**, Maastricht University, Netherlands

Co-authors: Jean-Pierre Urbain

This paper proposes Lagrange-multiplier tests for weak exogeneity in panel error correction models with common factors. We consider tests for orthogonality between innovations and conditioning variables, tests for the absence of error correcting behavior from the marginal model, as well as a joined test. We derive the asymptotic distributions of individual specific statistics and propose group-mean type panel tests. The finite sample properties of the tests are assessed in a Monte Carlo simulation.

C290: **First-differenced non-stationary factors versus factors from first-differences**

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

Co-authors: Ekaterina Ipatova

The existing inferential theory for non-stationary panel factor models is extended by proposing a novel estimation methodology for common factors, loadings, and common components, in the context of large time series (T) and cross sectional (n) dimensions. The proposed methodology is based on extracting the non-stationary common factors by applying Principal Components (PC) to data in levels, and then use their first differences. First order asymptotics for the estimated loadings and common components is found to be the same as when the stationary factors are directly estimated using first-differenced data. Conversely, higher order terms are shown to converge to zero at a faster rate as (n,T) pass to infinity, thereby suggesting that the proposed methodology yields better finite sample properties than direct estimation from first-differenced data. The theoretical findings are investigated through a comprehensive Monte Carlo exercise, which shows that the asymptotic results are a very good approximation of the finite sample properties of the estimated loadings and common components, even for fairly small T. The theoretical findings are validated by applying the proposed methodology (and the existing ones) to the study of unemployment fluctuations in 60 sectors in the US.

CS74 Room MAL B36 REAL-TIME MODELLING WITH MIXED-FREQUENCY DATA Chair: Klaus Wohlrabe

C211: **Spurious anti-persistence and mixed-frequency data**

Presenter: **Vladimir Kuzin**, DIW Berlin, Germany

Co-authors: Boriss Siliverstovs

This study investigates the properties of the estimator of the idiosyncratic persistence parameter in mixed-frequency models. We report an unexpected and puzzling result that the estimated parameter densities are bimodal for small true values of the persistence parameter. Only in case of large persistence the estimated density is unimodal. We also show that in a small-scaled mixed-frequency factor model estimated in state-space framework for small values of the persistence parameter most of the probability mass of the estimated parameter density lies in the left negative-valued hump of the distribution, implying a non-negligible positive bias in estimated persistence. This finding provides an explanation to often observed negative correlation in the coincident indicator mixed-frequency models suggesting that such "zig-zag" pattern is spurious.

C691: **Specification Issues for mixed-frequency MIDAS models**

Presenter: **Klaus Wohlrabe**, ifo Institute for Economic Research, Germany

MIDAS models have become recently popular in forecasting macroeconomic variables using mixed-frequency data. Surprisingly the specification of such models has been neglected so far. We investigate this issue in a Monte Carlo study. We focus how the specification aspect affects forecasting performance. MIDAS models are very parsimonious independently of the number of included lags and the frequency mixture. There are three issues that intrude the specification for this model class: the choice of the weighting function, whether the weighting function should be restricted or not and the number of included lags. In empirical applications a specific MIDAS model is chosen rather on an ad hoc basis and not

grounded on standard model selection criteria. Our simulations show that standard selection criterion lead to a rather rich lag structure which does not always coincide with optimal forecasting properties. Furthermore we outline that under certain circumstances restricted weighting functions increase forecasting accuracy. The choice of the weighting function (Almon lag or Beta weights) does not affect forecasting results.

C600: **Forecast combination in discrete choice models: predicting FOMC monetary policy decisions**
Presenter: **Laurent Pauwels**, The University of Sydney, Australia
Co-authors: Andrey Vasnev

This paper provides a new methodology to combine forecasts based on several univariate discrete choice models. This is achieved primarily by combining one-step-ahead forecast of probabilities associated with each univariate model. The paper applies well-established scoring techniques for qualitative response models in the context of forecast combination. Log-scores and quadratic-scores are used both to evaluate the forecasting performance of each model and to combine the forecasted probabilities. This new methodology is applied to forecast the Federal Open Market Committee (FOMC) decisions in changing the federal funds target rate. Several of the economic fundamentals influencing the FOMC decisions are nonstationary over time and are modelled appropriately. The empirical results show that combining forecasted probabilities using scores mostly outperforms both equal weight combination and forecasts based on multivariate models.

C270: **Macroeconomic indicators: A new way to aggregate expert forecasts using regression trees**
Presenter: **Fabian Krueger**, University of Konstanz, Germany
Co-authors: Simon D. Knaus

We use regression trees to forecast several important US macroeconomic variables, based on individual-level expert forecasts from the established Survey of Professional Forecasters. Regression Trees are well-suited to meet the peculiarities of this forecast combination setup: First, through the use of surrogate splits they can handle missing data without explicitly addressing imputation. Second, they can be combined with Bagging and Boosting which act as robustification devices. Third, they are flexible enough to capture many potential characteristics of opinion pools, like differences in experts' predictive ability and various forms of bias at the individual- and group level. While statistical learning techniques have recently entered the realm of macroeconomic forecasting, our application of Regression Trees to forecast combination with missing data, which poses significant problems to traditional approaches, is novel. Our analysis is relevant for two reasons: First, the information content of individual-level expert forecasts, which have become widely available, is poorly explored at present. Our analysis sheds light on this important issue. Second, our analysis paves the way for other applications of Regression Trees to forecast combination setups involving missing data. Forecasting from macroeconomic data at mixed sampling frequencies is a natural example in this context.

CS76 Room MAL B35 IDENTIFICATION AND INFERENCE Chair: Lynda Khalaf

C471: **Finite-sample bootstrap inference in GARCH models with heavy-tailed innovations**
Presenter: **Richard Luger**, Georgia State University, USA

A simulation-based procedure is proposed for the construction of valid finite-sample confidence sets in the context of stationary GARCH models. The proposed method uses a parametric bootstrap test procedure based on the likelihood ratio statistic which is numerically inverted to build simultaneous confidence sets. The computational curse of dimensionality is overcome by the use of a generic random walk Metropolis-Hastings MCMC algorithm. Projection techniques are then exploited to produce conservative confidence intervals and confidence sets for general functions of the parameters. The approach allows for heavy-tailed innovations without any moment conditions, and remains valid whether the model is identified or not. This last feature stands in sharp contrast to Wald-type methods based on standard asymptotic theory that are either invalid when the model is not identified or perform poorly in finite samples when the model is weakly identified. A simulation study illustrates the performance of the proposed procedure in the context of a benchmark GARCH model. Finally, the new inference procedure is applied to models of daily, weekly, and monthly returns on a major stock market index.

C529: **Modelling financial data with the asymmetric generalized t-distribution**
Presenter: **John Galbraith**, McGill University, Canada
Co-authors: Dongming Zhu

The asymmetric generalized t-distribution introduced in recent work by Zhu provides a very general representation for asymmetric, skewed, and heavy-tailed data, and nests many well-known distributions. We establish additional properties of this distribution that will be of use in simulation and in modelling financial data, including a stochastic representation and general expressions for the Value at Risk and Expected Shortfall. We use the results in generalizing results in the previous literature on partially adaptive estimators and on estimating risk, and apply the downside risk measures to the problem of predicting risk of foreign exchange loss.

C468: **Inference on parameter ratios with applications to weak identification**
Presenter: **Bertille Antoine**, Simon Fraser University, Canada

We propose a new inference method, the Modified-Wald procedure, to overcome issues related to the poor behavior of Wald when identification is not completely ensured. We focus here on the multidimensional ratio of parameters when the denominator is close to singularity. The key idea is to integrate the informational content of the null hypothesis in the computation of the Wald metric. This correction preserves the computational tractability, while allowing for unbounded confidence regions when needed. A simulation exercise compares the inference properties of Wald and Modified-Wald for a bidimensional ratio. We also reconnect identification failure at the frontier of the parameter space to some asymptotic identification patterns: as an illustration, we consider the single-equation linear IV model in cases where the identifying properties of the instruments may vary.

C732: **Inference in autoregressive models around polynomial trends of unknown order, under non-stationary volatility**
Presenter: **Nikolaos Kourogenis**, University of Piraeus, Greece

The problem of inference in autoregression around a polynomial trend, under non-stationary volatility is investigated. This problem appears to be quite difficult when the non-stationary volatility is allowed to have a not-"a priori"-known asymptotic order, k, because of its involvement in the test statistics concerning the autoregressive coefficients and the coefficients of the deterministic terms. It is shown, however, that as far as the autoregressive coefficients are concerned, this difficulty can be circumvented by the use of the Eicker-White covariance matrix estimator, leading to asymptotically standard normal t-statistics. An immediate implication of this result is that the order of the autoregression can be inferred. Next, an algorithm is introduced, that given the order of the autoregression, leads to the joint determination of the polynomial trend and of k. In contrast with other relevant studies, this approach is fully applicable since the "a priori" knowledge of k is not required. Simulation results are presented for several different volatility specifications, including the cases of trending variance and of nonstationary nonlinear heteroskedasticity.

Sunday 12.12.2010 09:00 - 10:40 Parallel Session K – ERCIM

ES23 Room MAL 532 BAYESIAN NONPARAMETRICS: THEORY Chair: Sonia Petrone

E075: **On posterior rates of convergence under partially improper non parametric priors**
Presenter: **Judith Rousseau**, ENSAE and Universite Paris Dauphine, France
Co-authors: Dongchu Sun

The impact of the use of partially improper priors in non parametric models is studied. In particular we derive a general result on posterior concentration rates under priors that are improper on a part of the parameter. We apply this result in 3 different contexts: smoothing splines, regression models in large dimension and goodness of fit tests for a parametric model.

E079: **Posterior consistency of species sampling priors**
Presenter: **Jaeyong Lee**, Seoul National University, Korea (Rok)
Co-authors: Gun Ho Jang, Sangyeol Lee

Recently there has been increasing interest in species sampling priors, the nonparametric priors defined as the directing random probability measures of the species sampling sequences. In this talk, we show that not all of the species sampling priors produce consistent posteriors. In particular, in the class of Pitman-Yor process priors, the only priors rendering posterior consistency are essentially the Dirichlet process priors. Under certain conditions, we also give a set of necessary and sufficient conditions for the posterior consistency for the general species sam- pling prior. Considered examples include the normalized inverse-Gaussian process, the Poisson-Kingman partition and the Gibbs partition.

E091: **Bayesian inverse problems**
Presenter: **Harry van Zanten**, Eindhoven University of Technology, Netherlands

We investigate a Bayesian approach for dealing with nonparametric linear inverse problems. We consider a setting in which we are interested in an element μ of some Hilbert space, but we only observe a noisy version of $K\mu$, for K a known linear operator. To make inference about μ we endow it with a Gaussian process prior and compute the corresponding posterior. For this setting we are interested in the asymptotic behaviour of the posterior as the signal-to-noise ratio tends to infinity. We consider, among other things, contraction rates for the full posterior for μ, contraction rates for the marginal posterior of linear functionals of μ, and the Bernstein-von Mises phenomenon for linear functionals of μ. In particular, we explain how the behaviour of the posterior depends on the regularity of the signal μ of interest, the regularity of the prior, and the ill-posedness of the operator K.

E502: **Projective limits of Bayesian models**
Presenter: **Peter Orbanz**, University of Cambridge, UK

Bayesian nonparametric models can be regarded as Bayesian models on infinite-dimensional spaces. These infinite-dimensional distributions can be constructed from finite-dimensional ones using the projective limit approach familiar from stochastic process theory. I will discuss how this approach can be generalized considerably by applying the projective limit tools available in pure mathematics both to conditional probabilities and to the various mappings associated with a Bayesian model – the random variables associated with the model, its sufficient statistics, and the mapping to the posterior parameters. This allows us to define a nonparametric Bayesian model from parametric Bayesian equations, and to study its properties in terms of the finite-dimensional models. For example, the nonparametric Bayesian model is conjugate if the parametric models in the construction are conjugate, and the sufficient statistics of the parametric models define a sufficient statistic of the nonparametric model. I will briefly discuss for which models these constructions follow a generic recipe, and for which cases we have to expect mathematical complications.

ES26 Room MAL 509 DATA MODELING BY SHAPE RESTRICTION Chair: Ioannis Demetriou

E710: **Best L1 convex fit to univariate data: a characterization theorem and an algorithm**
Presenter: **Ioannis Demetriou**, University of Athens, Greece
Co-authors: Sotirios Papakonstantinou

If a convex function provides measurements, but convexity has been lost due to errors of the measuring process, then we address the following calculation. We smooth the data by minimizing the sum of the moduli of the errors subject to the constraints that the second divided differences of the smoothed data are nonnegative. It follows that the piecewise linear interpolant to the smoothed data is convex. For example, one may require estimating a utility function that is represented by a finite number of observations that are corrupted by random errors, in which case the criterion of convexity enters by the assumption of non-decreasing returns of utility functions. The optimization problem is a constrained L1 problem that can be expressed as a linear programming calculation. Necessary and sufficient conditions are established that allow the development of a special algorithm that is much faster than general linear programming procedures. Some numerical results are presented that illustrate the computation and present the efficacy of the method in small, medium and large data sets. Applications of the problem on real data are given in another presentation within this session. It is to be noted that suitable algorithms for the corresponding minimax and least squares problems have already been studied in the literature.

E713: **Case study: Applications of the best L1 convex approximation on certain economic data sets**
Presenter: **Sotirios Papakonstantinou**, University of Athens, Greece
Co-authors: Ioannis Demetriou

The problem of convexity runs deeply in economic theory. For example, increasing returns or upward slopes (convexity) and diminishing returns or downward slopes (concavity) of certain supply, demand, production and utility relations are often assumed in economics. Quite frequently, however, the observations have lost convexity (or concavity) due to errors of the measuring process. In this case we derive an estimate of the data by minimizing the sum of the moduli of the errors subject to the condition that the second divided differences of the smoothed data are nonnegative (that is, by assuming non-decreasing returns). It is an L1 problem that is solved by a special method, also presented by the authors within this session. This paper presents certain applications that test the convexity assumption of real economic times series derived from money demand versus the interest rate, the per capita GNP versus the infant mortality rate for the year 1995 for 149 countries and the evolution of the Gini coefficient in the U.S.A. for 50 years. The results are analyzed and the interpretation capability of the method is demonstrated.

E723: **Data fitting by divided differences and splines**

Presenter: **Evangelos Vassiliou**, University of Athens, Greece

A smooth univariate function is measured at a finite number of points and the measurements contain random errors. We seek an approximation that should be closer than the measurements to the true function values. The approximation task can be done either by a parametric or by a non-parametric approach. The first approach assumes that the underlying function is described by a functional form that depends on a set of parameters, for example by using splines functions. There are several notions of nonparametric approach, and we take the view that the errors should be corrected by making least the sum of squares of the errors by imposing a limit, q say, on the number of sign changes of the divided differences of a prescribed order r. Depending on the values of r and q, interesting properties may be revealed, like monotonicity, piecewise monotonicity, convexity, convexity/concavity, etc. This paper applies initially a divided difference method in order to smooth the data and then applies a spline regression on the smoothed data by following the shape of the smoothed data. Some examples with real and simulated data are presented in order to illustrate this method. They all suggest that the method provides usually quite suitable fits, because the user is directed by the divided difference step in locating the knots of the spline.

ES27 Room MAL 355 ANALYSIS OF SPATIAL DATA: ESTIMATION, MODELLING, AND INFERENCE Chair: Elvan Ceyhan

E152: **Estimates for geographical domains through geoadditive models in presence of missing information**

Presenter: **Chiara Bocci**, University of Florence, Italy

Co-authors: Emilia Rocco

The implementation of geoadditive models needs the statistical units to be referenced at point locations. If the aim of the study is to analyze the spatial pattern or to produce a spatial interpolation of a studied phenomenon, spatial information are required only for the sampled units. If, however, the geoadditive model is used to produce estimates of a parameter of interest for some geographical domains, the spatial location is required for all the population units. This information is not always easily available. Typically, we know the coordinates for sampled units, but for the non-sampled units we know just the areas - like blocks, municipalities, etc. - to which they belong. In such situation, the classical approach is to locate all the units by the coordinates of their corresponding area centroid. This is obviously an approximation and its effect on the estimates can be strong, depending on the level of nonlinearity in the spatial pattern and on the area dimension. We decided to investigate a different approach: instead of using the same coordinates for all the units, we impose a distribution for the locations inside each area. To analyze the performance of this approach, various MCMC experiments are implemented.

E393: **Linear approximations of individual-level models for infectious disease**

Presenter: **Grace Pui Sze Kwong**, University of Guelph, Canada

Co-authors: Rob Deardon

Individual-level models (ILMs) for infectious diseases, fitted in a Bayesian MCMC framework, are an intuitive and flexible class of models that can take into account population heterogeneity via various individual-level covariates. ILMs containing a geometric distance kernel to account for geographic heterogeneity provide a natural way to model the spatial spread of many diseases. However, in even only moderately large populations, the likelihood calculations required can be prohibitively time consuming. It is possible to speed up the computation via a technique which makes use a linearized distance kernel. Here we examine some methods of carrying out this linearization and compare the performances of these methods.

E182: **Adjusted calibration estimators for sparse spatial data**

Presenter: **Ivan Sciascia**, Universita di Torino, Italy

The technique of calibrated estimation is coupled with an algorithm of vehicles routing, building through the latter, a distance function that adjusts the estimator. The calibration and a vehicle routing algorithm are considered in this paper with the constraint of the limited availability of auxiliary variables and with the use of remotely sensed distances. Remote sensing is an interesting topic due to the development of modern GIS systems which can be useful for sample surveys with geospatial data. In a context of spatial data analysis with a constraint of low-density we propose to adjust the calibrated estimator by a distance function. The Traveling Salesman Problem (TSP) is a basic routing problem stated for visit to P cities with the shortest closed tour. Different algorithms have been developed to resolve heuristically the TSP. This article develops a distance function calculated on the shortest tour distance coming from an heuristic algorithm. The performances of the proposed estimator are evaluated using a simulation study.

ES46 Room MAL B29 COMPONENT ANALYSIS Chair: Marieke Timmerman

E284: **The generic subspace clustering model**

Presenter: **Marieke Timmerman**, University of Groningen, Netherlands

Co-authors: Eva Ceulemans, Kim De Roover

The clustering of high-dimensional data can be troublesome. In the case of high-dimensional data, a subspace clustering model can be used to achieve a proper recovery of the clusters and to obtain an insight into the structure of the variables relevant to the clustering. In such a model the objects are assigned to mutually exclusive classes in low dimensional spaces. In this paper, we present the Generic Subspace Clustering Model. As will be shown, this model encompasses a range of existing (subspace) clustering techniques as special cases. The specific properties of the model variants will be discussed. An algorithm for fitting the Generic Subspace Clustering Model is presented and its performance is evaluated by means of a simulation study. The value of the model for empirical research is illustrated with data from psychiatric diagnosis research.

E340: **Clusterwise SCA-P**

Presenter: **Kim De Roover**, K.U. Leuven, Belgium

Co-authors: Eva Ceulemans, Marieke Timmerman, Patrick Onghena

Numerous research questions in educational sciences and psychology concern the structure of a set of variables. For instance, one can study the structure of a number of variables, measured on multiple occasions for different subjects. In that case, one may wonder whether the same structure is underlying the data of all subjects. Obviously, the crucial question is how such data have to be analyzed to find out whether and in what way the structure of the variables differs across the persons. A number of principal component analysis techniques exist to study such structural differences, for instance, simultaneous component analysis (SCA). However, these techniques suffer from important limitations. Therefore, we propose a novel modeling strategy, called Clusterwise SCA-P, which solves these limitations and which encompasses several existing techniques as special cases. Clusterwise SCA-P partitions the subjects into a number of clusters and specifies a simultaneous component model per cluster. Also, it allows for between subject differences in the variances and the correlations of the cluster specific components, to offer additional insight into interindividual differences. The value of the model for empirical research is illustrated.

E087: **Comparison of several methods for handling missing data in principal component analysis**

Presenter: **Joost Van Ginkel**, Leiden University, Netherlands

Co-authors: Henk Kiers, Pieter Kroonenberg

Principal component analysis is a widely used statistical technique for determining subscales in test and questionnaire data. As in any other technique, missing data may complicate computations in principal component analysis, and may bias its results. A possible solution for handling missing data in principal component analysis is multiple imputation in combination with Generalized Procrustes analysis. Earlier research has shown that this approach recovers the results from principal component analysis well. In the current study, the performance of Generalized Procrustes analysis is further studied, and is compared with two other methods that can deal with missing data in principal component analysis: maximum likelihood principal components analysis, and the EM algorithm. Simulations are carried out to study how well each method performs under different circumstances.

E334: **How to select a model in multi-way and multi-level component analysis: The CHull procedure**

Presenter: **Eva Ceulemans**, Katholieke Universiteit Leuven, Belgium

Co-authors: Henk Kiers, Marieke Timmerman

Principal component analysis (PCA) is a popular method for summarizing two-mode data. The last decades, several techniques have been proposed that generalize PCA to more complex types of data, like three-mode data and multi-level data. Applying these techniques in practice is often not straightforward, as many data-analytic choices, like preprocessing, model selection, and visualization of results, become rather complicated. In this paper, we focus on the model selection problem. Specifically, we discuss recent work on the CHull procedure. This procedure looks for the component solution with the best balance between fit and complexity, where complexity is expressed via the number of free parameters. Simulation results on the performance of the CHull procedure will be presented.

ES50 Room MAL 421 PARTIAL LEAST SQUARES METHODS I Chair: Vincenzo Esposito Vinzi

E337: **Correlated component regression: an alternative to PLS-regression**

Presenter: **Jay Magidson**, Statistical Innovations, USA

PLS regression (PLS-R) utilizes K orthogonal components to predict one or more dependent variables, each component defined as a linear function of P predictors. We propose an alternative to PLS-R, called Correlated Component Regression (CCR), which eliminates the orthogonality requirement, and typically results in fewer components and more reliable predictions. CCR is an ensemble approach that involves sequential application of the Naïve Bayes rule. Similar to PLS-R, CCR extends traditional regression modeling to apply to high dimensional data where the number of predictors P may exceed the number of cases N (P » N). CCR yields K correlated components, weights associated with the first component providing an ensemble averaging of direct effects for the predictors, and weights for each additional component providing an ensemble averaging of partial effects which yield improved prediction by including suppressor variable and moderator effects of predictors not contributing to the first component. In addition, a step-down algorithm is proposed to reduce the number of predictors to any desired number P* < P. Results with real and simulated data suggest that when one or more suppressor variables are included among the predictors, CCR outperforms PLS-R as well as penalty approaches such as lasso and Elastic Net.

E481: **Correlated component PLS-type regression with variable selection feature for large datasets**

Presenter: **Laura Trinchera**, SUPELEC, France

Co-authors: Vincenzo Esposito Vinzi, Arthur Tenenhaus, Michel Tenenhaus

The analysis of landscape matrices, i.e. matrices having more columns (variables) than rows (observations), is a challenging task in several domains. Two different kinds of problems arise when dealing with landscape matrices. The first refers to computational and numerical problems. The second deals with the difficulty in assessing and understanding the results. Partial Least Squares (PLS) methods are classical feature extraction tools that work in the case of high-dimensional data sets. Since PLS methods do not require matrices inversion or diagonalization, they allow us to solve computational problems. However, results interpretation is still a hard problem when facing with very high-dimensional data sets. Nowadays interest is increasing in developing new PLS methods able to be, at the same time, a feature extraction tool and a feature selection method (i.e. a variable selection method). Here a new PLS-type algorithm including variable selection and correlated components will be presented. The use of correlated components instead of orthogonal components allows us to take into account so called suppressor variables, i.e. variables having no direct effect on the response variables but improving prediction by suppressing irrelevant variation in the lower-order components. This is of main importance in order to obtain predictive variable selection.

E663: **Generalized multiple-set Kernel CCA**

Presenter: **Arthur Tenenhaus**, SUPELEC, France

A problem in applied statistics is to study relationships between several blocks of variables observed on the same set of individuals. Typical examples are found in large variety of fields such as bioinformatics, sensory analysis, marketing, food research, chemometrics, where the common general objective is to extract pattern(s) of one block related to pattern(s) of others blocks. To study such a kind of relationships between blocks, the starting point of this communication is the regularized generalized canonical correlation analysis (RGCCA). Within the RGCCA framework, all blocks are not necessary fully connected and this allows RGCCA to include a remarkably large number of well-known methods as particular cases. However, RGCCA captures linear relations between connected blocks and to recover nonlinear relations, the Kernel Generalized Canonical Correlation Analysis (KGCCA) is proposed. Searching for a fixed point of the stationary equation related to KGCCA, a monotonic convergent algorithm is obtained. One of the distinct advantage of KGCCA is that with a single algorithm all particular cases of RGCCA are automatically kernelized.

E314: **Regularized generalized canonical correlation analysis**

Presenter: **Michel Tenenhaus**, HEC Paris, France

Regularized generalized canonical correlation analysis (RGCCA) is a generalization of regularized canonical correlation analysis to three or more sets of variables. It constitutes a general framework for many multi-block data analysis methods. It combines the power of multi-block data analysis methods (maximization of well identified criteria) and the flexibility of PLS path modeling (the researcher decides which blocks are connected and which are not). Searching for a fixed point of the stationary equations related to RGCCA, a new monotone convergent algorithm, very similar to the PLS algorithm proposed by Herman Wold, is obtained. Finally, a practical example is discussed.

ES07 Room MAL B20 FUZZY STATISTICAL ANALYSIS Chair: M. Brigida Ferraro

E577: **How to compare two quasi-continuous histograms**

Presenter: **Olivier Strauss**, LIRMM, France

A histogram is a straightforward and useful tool for representing the random phenomenon underlying a real observation process. In image processing, histograms are often used to compare density functions underlying two sets of pixels via a sampled version of statistical distances (e.g the Kolmogorov variational distance). In short, comparing two histograms consists generally in summing up a distance between the normalized values associated to each bin of each histogram. This comparison is often not robust enough to ensure the reliability of the process it involves. One of the reasons of this lack of robustness is the high sensitivity of the histogram to both bin width and reference interval. To low this influence, many authors have proposed to replace the crisp partition by a fuzzy partition. This replacement leads to confusion in the accumulated values (i.e. each observations can vote for more than one bin). Thus, the comparison of the two histograms cannot be rigorously achieved by summing up bin-to-bin comparisons. In this article, we propose two different extensions of classical histograms dissimilarity measures accounting for the distributed nature of the accumulation process. One of these extensions coincides with a classical cross-bin comparison method. The other leads to an imprecise dissimilarity measure.

E690: **An approach to the median of a random fuzzy set**

Presenter: **Beatriz Sinova**, Universidad de Oviedo, Spain

Co-authors: Maria Angeles Gil, Ana Colubi, Gil Gonzalez-Rodriguez, Stefan Van Aelst

To avoid the bad influence that "extreme data" or data changes have on the estimation of the Aumann expected value for random fuzzy sets (a usual problem working with means of random elements) a measure with a more robust behaviour is defined. It is inspired by the real concept of median, intended to be the value which is the 'nearest' one, in terms of a mean Euclidean distance, to all the values of the variable; actually, the approach to the median as the middle position value is not applicable in this setting because of the lack of a universally accepted total ordering on the space of fuzzy values. The suggested measure will be based on a 1-norm distance assuring the existence of the previous minimum. After the analysis of some of its basic properties, inherited from those for the real-valued case, a comparative study with other central tendency measures, involving theoretical results about the breakdown point as well as some simulations, is developed.

E821: **A fuzzy least-squares regression based on a concept of distance with application to Economics and Finance**

Presenter: **Concepcion Roldan**, University of Jaen, Spain

Co-authors: Juan Martinez-Moreno, Antonio Roldan

Least-squares technique is well-known and widely used to determine the coefficients from observations based on a concept of distance. Traditionally, the observations consist of pairs of exact values. However, in many real-life problems, the independent or explanatory variable can be observed precisely (for instance, the time) and the dependent or response variable is usually described by approximate values, such us "about 300 pounds" or "approximately 500$", instead of exact values. In this communication we present a new technique to obtain fuzzy regression models that consider triangular fuzzy numbers in the response variable. The procedure is theoretically supported, solves linear and non linear problems, is easy to compute in practice and may be applied in different contexts. The usefulness of the proposed method is illustrated using simulated and real-life examples.

E628: **Comparative study of Markov chain models and probabilistic fuzzy systems**

Presenter: **Rui Jorge Almeida**, Erasmus University Rotterdam, Erasmus School of Economics, Netherlands

Co-authors: Nick Verbeek, Uzay Kaymak

A probabilistic fuzzy system consists of a set of rules whose antecedents are fuzzy conditions, the consequents are possible fuzzy conclusions and there is a stochastic mapping between the antecedent and the consequents. The rules specify a probability distribution over a collection of fuzzy sets that partition the output domain. They also convey linguistic information. Therefore, such a system deals both with linguistic uncertainty as well as probabilistic uncertainty. A Markov chain model can be defined through a set of realized states and a set of transition probabilities that depend on the past state realizations. This type of system leads to a probabilistic interpretation of the uncertainty. In this work we study the relationship between Markov chain models and probabilistic fuzzy systems. Specifically we compare the parameters of the stochastic mapping between the antecedent and the consequents of a probabilistic fuzzy system with the transition probabilities of Markov chain models. Furthermore, we discuss the parallelism between the interpretation of Markov chain models and probabilistic fuzzy systems, through a series of experiments with synthetic and real data.

ES53 Room Senate Beveridge Hall ROBUST METHODS IN ECONOMETRICS Chair: Kris Boudt

E275: **Model-free estimation of large variance matrices**

Presenter: **Filip Zikes**, Imperial College London, UK

Co-authors: Karim M. Abadir, Walter Distaso

A new method for estimating large variance matrices is introduced. Starting from the orthogonal decomposition of the sample variance matrix, we exploit the fact that orthogonal matrices are never ill-conditioned and therefore focus on improving the estimation of the eigenvalues. We estimate the eigenvectors from just a fraction of the data, then use them to transform the data into approximately orthogonal series that we use to estimate a well-conditioned matrix of eigenvalues. Our estimator is model-free: we make no assumptions on the distribution of the random sample or on any parametric structure the variance matrix may have. By design, it delivers well-conditioned estimates regardless of the dimension of problem and the number of observations available. Simulation evidence show that the new estimator outperforms the usual sample variance matrix, not only by achieving a substantial improvement in the condition number (as expected), but also by much lower error norms that measure its deviation from the true variance.

E280: **Outlier resistant GARCH modeling**

Presenter: **Yohan Chalabi**, ETH Zurich, Switzerland

Co-authors: Diethelm Wuertz

Generalized autoregressive heteroskedastic (GARCH) models are nowadays widely used to reproduce stylized facts of financial time series and play an essential role in risk management and volatility forecasting. Although these models have been extensively studied, numerical problems may arise in the estimation of the parameters in the presence of outliers. To overcome this limitation, we provide new robust estimators for the standard GARCH model with the weighted trimmed likelihood estimator (WTLE). We suggest a fast implementation and explain how the additional robust parameter can be estimated. An extensive Monte-Carlo simulation is performed to compare our approach to recently introduced robust GARCH estimators.

E303: **The Gaussian rank correlation estimator: Efficiency and robustness properties**

Presenter: **Jonathan Cornelissen**, K.U.Leuven, Belgium

Co-authors: Kris Boudt, Christophe Croux

We propose the Gaussian Rank correlation as a robust and efficient correlation estimator. Its asymptotic breakdown point is 12.4 %. At the Gaussian distribution, it is asymptotically as efficient as the Pearson correlation coefficient. Moreover, the correlation matrix based on the proposed estimator is positive semidefinite and the estimator is computationally simply, also in high dimensions. A simulation study confirms the good efficiency and robustness properties of the proposed estimator with respect to the popular Kendall and Spearman correlation measures. An application illustrates the empirical usefulness of the estimator.

E419: **Econometric model selection with more variables than observations**

Presenter: **Jurgen Doornik**, University of Oxford, UK

Several algorithms for indicator saturation are compared and found to have low power when there are multiple breaks. A new algorithm is introduced, based on repeated application of an automatic model selection procedure (Autometrics) which is based on the general-to-specific approach. The new algorithm can also be applied in the general case of more variables than observations. The performance of this new algorithm is investigated through Monte Carlo analysis. The relationship between indicator saturation and robust estimation is explored. Building on the results of Soren Johansen and Bent Nielsen, the asymptotic distribution of multi-step indicator saturation is derived, as well as the efficiency of the two-step variance. Next, the asymptotic distribution of multi-step robust estimation using two different critical values (a low one at first) is derived. The asymptotic distribution of the fully iterated case is conjectured, as is the asymptotic distribution of reweighted least squares based on least trimmed squares, called RLTS here. This allows for a comparison of the efficiency of indicator saturation with RLTS. Finally, the performance of several robust estimators and the new approach is studied in the presence of a structural break. When there are many irrelevant regressors in the model, the robust estimators break down while the new algorithm is largely unaffected.

CS17 Room MAL B34 DYNAMIC PANEL MODELLING Chair: Michael Binder

C768: **In search for a robust method for estimating dynamic panel data models of capital structure**

Presenter: **Yongcheol Shin**, Leeds University, UK

Co-authors: Viet Dang, Minjoo Kim

Existing studies have used the dynamic, partial adjustment model of leverage to test the trade-off theory against alternative theories. The challenge of this approach is to consistently estimate both the speed of leverage adjustment and the long-run relationship underlying target leverage in short panels with individual fixed effects. This paper conducts empirical and Monte Carlo simulation analysis to evaluate the performance of all existing methods for estimating such a complex model. Results show that traditional least-squares estimators are fundamentally flawed while instrumental variable and GMM approaches are unreliable in several important settings. The paper considers alternative methods based on either analytical or bootstrap bias corrections. It shows that the latter estimator obtains the most plausible and robust results and outperforms the other estimators in terms of accuracy and efficiency. Importantly, this method can appropriately estimate target leverage, toward which adjustment is found to take place at a moderate speed, a finding generally consistent with the dynamic trade-off theory.

C823: **Likelihood based panel analysis of growth models**

Presenter: **Michael Binder**, Goethe University Frankfurt, Germany

Co-authors: Susanne Brock

We re-consider panel based estimation of models of economic growth in settings where the time dimension of the data set analyzed is short. We argue that estimation of growth models by Generalized Methods of Moments (GMM), as is widely popular for such data sets, can be rather problematic, and show that a Quasi Maximum Likelihood (QML) estimation approach not only overcomes the problems incurred by the GMM estimators, but is also readily implementable when data features such as country-specific time trends, intertemporally heteroscedastic errors and/or simple measurement error are present. Using data from the Penn World Tables, we document that estimation results for growth models obtained under GMM can differ sizably from those obtained under QML, and that for data generating processes typically encountered in cross-country growth analyses the QML estimator does perform well.

C669: **Forecasting Swiss Inflation and GDP with a small global VAR model**

Presenter: **Katrin Assenmacher**, Swiss National Bank, Switzerland

Co-authors: Daniel Geissmann

We assess the forecast ability of a global VAR model including Switzer-land, the euro area, Japan and the US for Swiss output and inflation. We find that the GVAR performs as well as a simple autoregression or random-walk model. Imposing cointegrating restrictions improves forecasts, especially at long-run horizons. By contrast, if the interest is in predicting the domestic economy, modelling the feedback to the foreign variables is less important. Finally, during exceptional periods like the financial crisis, country models with exogenous variables may be useful as non-model information can be entered in a simple way.

C827: **Endogenous regime-switching in global vector autoregressions**

Presenter: **Marco Buchmann**, Goethe University Frankfurt, Germany

Co-authors: Michael Binder

We develop a Markov-Switching Global Vector Autoregressive model (MSGVAR), bringing together and extending two modelling approaches: (i) Global Vector Autoregressions ideally suited to modelling interdependencies and spillovers across countries; and (ii) Markov-Switching models allowing for endogenous switches between a finite number of regimes. Particularly appealing about our MSGVAR model is that model dynamics in a tractable manner become a function of the complete set of regime-constellations across countries, in addition to the spillovers present in existing GVAR models. We develop an EM algorithm based computational procedure for estimation and solution of MSGVAR models. We use our MSGVAR model to shed light on cross-country financial and goods market interdependencies during the financial crisis, how these crisis interdependencies differ(ed) from those during non-crisis periods, and how regime dependency of spillovers across countries affects the forecasting performance of GVARs.

C354: **Estimating financial distress likelihood: gaining additional insight via a dynamic nonlinear approach**

Presenter: **Konstantinos Konstantaras**, University of the Aegean, Greece

Co-authors: Costas Siriopoulos

Employing earnings shortfall as a financial distress indicator, we formulate a dynamic nonlinear model, implementing Wooldridge's conditional maximum likelihood estimator and accounting for potentially endogenous covariates. Likewise, we not only achieve a significant improvement in consistency and classification accuracy over static approaches, but we also manage to understand better the evolution of the financial distress process. In particular, we find that the higher the initial period's positive performance and the lower the leverage, the greater the chance that a company enters financial distress further down the road, possibly due to either managerial overconfidence or debt-imposed discipline by company stakeholders.

CS84 Room MAL 151 THE ECONOMETRICS OF COMMODITY AND FINANCIAL MARKETS Chair: David Ardia

C184: **Simulated scenarios of conventional oil production**

Presenter: **Vlasios Voudouris**, London Metropolitan University, UK

Co-authors: Dimitrios Stasinopoulos, Robert Rigby

The ACEGES (Agent-based Computational Economics of the Global Energy System) model is an agent-based model of conventional oil production. The model accounts for four key uncertainties, namely Estimated Ultimate Recovery (EUR), estimated growth in oil demand, estimated growth in oil production and assumed peak/decline point. This work provides an overview of the ACEGES model capabilities and an example of how it can be used for long-term scenarios of conventional oil production. Because the ACEGES model has been developed using the Agent-based Computational Economics (ACE) modelling paradigm, the macro-phenomenon of interest (world oil production) grows from sets of micro-foundations (country-specific decision of oil production). The simulated data is analyzed in GAMLSS (Generalised Additive Models for Location Scale and Shape). GAMLSS is a general framework of modelling where the response variable (oil production) can have a very general (up to four parameters) distribution and all of the parameters of the distributions are modelled as linear or smooth function (e.g., Penalised Beta (Regression) Spline) of the explanatory variable (e.g., time). From a methodological perspective, ACEGES and GAMLSS are applied to help leaders in government, business

and civil society better understand the challenging outlook for energy through controlled computational experiments.

C377: **Dynamic causal linkages between the east Asian economies and the US stock market**
Presenter: **Hyunjoo Kim**, Jonkoping International Business School, Sweden

An empirical study in the dynamic causal relationships between each of national stock market of the east Asian economies (Hong Kong, Singapore, Korea, and Taiwan)and the U.S. stock market is presented. This paper complements the existing studies by analyzing the dynamic causal relationship between the east Asian stock markets and the U.S. stock market at different time scales by employing wavelet analysis. The main empirical insight is that the causal relationship is strongest at the finest time scale, whereas the relationship is less and less apparent at longer time horizon. Unlike the other studies which focus on the contagion effect during the financial crisis periods, the empirical evidences of the current study indicate that the U.S. stock market Granger-causes almost all the east Asian stock markets regardless of the crisis and non-crisis period at the finest time scale. The impulse response functions show that the response of Hong Kong stock market is greater than the other east Asian stock markets. The impact of the shock to the U.S. market is greater during the financial crisis started in the year 2007 than the non-crisis period (post-Asian financial crisis started in the year 1997).

C754: **Are market makers uninformed and passive? Signing Ttrades in the absence of quotes**
Presenter: **Michel van der Wel**, Erasmus University Rotterdam / CREATES, Netherlands
Co-authors: Albert Menkveld, Asani Sarkar

We develop a new likelihood-based approach to sign trades in the absence of quotes. It is equally efficient as existing MCMC methods, but more than 10 times faster. It can deal with the occurrence of multiple trades at the same time, and noisily observed trade times. We apply this method to a high-frequency dataset of the 30Y U.S. treasury futures to investigate the role of the market maker. Most theory characterizes him as an uninformed passive liquidity supplier. Our results suggest that some market makers actively demand liquidity for a substantial part of the day and are informed speculators.

C133: **Informed and uninformed traders at work: evidence from the French market**
Presenter: **Fabrizio Ferriani**, University of Bologna, Italy

The aim is to empirically asses the impact of informed and uninformed transactions on market prices. Informed trades are associated with institutional agents while uninformed trades are executed on behalf of retail investors. An ordered probit model is estimated using high-frequency data from Euronext Paris, where the traders' identities at transaction level is taken into account. The results show that when the category of trader is different on the two sides of the market, stock prices follow the direction indicated by institutional agents. There is no significant effect when the type of agent is the same on both market sides. Since traders' identities are concealed in Euronext Paris, the last part of the paper discusses the informational content provided by observed market variables. Institutional trading is more common in periods of low price changes and high frequency of transactions. Moreover it increases throughout the day and decreases in the first part of the continuous auction. Price variations show that informed agents are usually able to trade at better price conditions. Finally, the results confirm that informed traders always act as initiators of market transactions.

CS38 Room MAL B30 MODELLING AND FORECASTING FINANCIAL MARKETS Chair: Alena Audzeyeva

C200: **Dynamic linkages between stock markets: the effects of crises and globalization**
Presenter: **Malgorzata Doman**, Poznan University of Economics, Poland
Co-authors: Ryszard Doman

The aim is to verify the opinion about growing level of the world stock market interdependencies and describe the changes in the dynamics of the linkages. We focus on the possible effects of globalization and differences between crisis and non-crisis periods. Our analysis is based on the return series of 12 selected stock indices from the period 1995-2009. To model the dependencies and compare their strength, both over time and across pairs of markets, we apply a Markov switching copula model and make use of a formal sequential testing procedure based on the model confidence set methodology.

C480: **Robust risk-return analysis of international portfolios**
Presenter: **Vitali Alexeev**, University of Tasmania, Australia
Co-authors: Francis Tapon

We investigate the two main routes investors can use to capitalize on the advantages of international diversification: i) to invest in foreign markets indirectly, that is, to invest in US multinational corporations with a large presence abroad; ii) to invest directly in foreign markets by purchasing Exchange Traded Funds (ETFs) which are listed on the New York Stock Exchange (NYSE) and based on foreign stock indexes, as well as stocks of foreign companies listed as American Depositary Receipts (ADRs) or as regular stocks on the NYSE. We consider a problem of portfolio selection using a covariance matrix robust estimator to capture the large unconditional variance caused by occasional high volatility periods. We track the composition of robust and classical Markowitz portfolios for the period from 1988 to 2009. We find that the resulting robust portfolios exhibit more stable weight compositions and reduced rebalancing costs.

C175: **Predicting the daily covariance matrix for market timing**
Presenter: **Wei Liu**, City University, UK
Co-authors: Ana-Maria Fuertes, Elena Kalotychou

This study evaluates the forecasting ability of various covariance estimation approaches using statistical and financial criteria. The economic loss function chosen to compare their performance is the quadratic utility function implicit in mean-variance asset allocation. A dynamic optimal-weight portfolio strategy is deployed which is based on one-day-ahead co-variance matrix forecasts from a model-free EWMA approach and three competing MGARCH formulations. We extend each of these estimators with realized covariances in order to exploit intraday returns. The analysis is applied to a portfolio of three indices, NASDAQ 100, Russell 2000 and the CRB commodity index. For all four estimators there are significant economic gains in switching from daily to intradaily returns which are robust to transaction costs. A time varying return-driven rebalancing approach yields significant economic value over daily rebalancing but performs similarly to the simpler monthly rebalancing approach. The MGARCH forecasts result generally in worse portfolios than those obtained from the rolling EWMA estimator.

C624: **The Dynamics of Correlations in Asset Allocation**
Presenter: **Gang Zhao**, Cass Business School, City University, London, UK, UK
Co-authors: Elena Kalotychou, Sotiris Staikouras

The value of correlation dynamics in mean-variance asset allocation is assessed. A correlation-timing framework is deployed with conditional

correlation models competing against static allocation strategies. We address the extent to which the superior statistical properties of multivariate conditional correlation models translate into enhanced investment performance as gauged by utility metrics and risk-adjusted returns. Using sector data in three major markets, we find that correlation timing is rewarding. Moreover, dynamically optimized portfolios based on multivariate conditional correlation models bring substantial economic gains that are not eroded by transaction costs. Risk-averse investors are willing to pay up to 1200bps per annum for switching from a static to a nonparametric dynamic correlation strategy based on the RiskMetrics approach; and a further 2000 bp to reap the benefits of a richer parametric correlation specification such as the DCC.

C831: **Aggregational Gaussianity and barely infinite variance in crop prices**

Presenter: **Antonios Antypas**, University of Piraeus, Greece

Co-authors: Phoebe Koundouri, Nikolaos Kourogenis

We aim at reconciling two apparently contradictory empirical regularities of financial returns, namely the fact that the empirical distribution of returns tends to normality as the frequency of observation decreases (aggregational Gaussianity) combined with the fact that the conditional variance of high frequency returns seems to have a unit root, in which case the unconditional variance is infinite. We show that aggregational Gaussianity and infinite variance can coexist, provided that all the moments of the unconditional distribution whose order is less than two exist. The latter characterises the case of Integrated GARCH (IGARCH) processes. Finally, we discuss testing for aggregational Gaussianity under barely infinite variance. Our empirical motivation derives from crop prices, while our results are relevant for financial returns in general.

CS80 Room MAL G16 MACROECONOMETRIC APPLICATIONS Chair: Ana-Maria Fuertes

C156: **Lagoons and lakes: Macroeconomic effects of the Chilean pension reform**

Presenter: **Romulo Chumacero**, University of Chile / Central Bank of Chile, Chile

A DSGE model with overlapping generations and heterogeneous agents calibrated to replicate business-cycle characteristics of the Chilean economy as well as life-cycle profiles for cohorts of the population is used. The model is used to assess the effects of the recent reforms to the Chilean pension fund system that provides widespread subsidies to retirees that satisfy certain characteristics. It is shown that this subsidy tends to crowd out the contribution density of potential retirees.

C285: **Measuring Euro Area monetary policy transmission in a structural dynamic factor model**

Presenter: **Matteo Luciani**, Sapienza University of Rome, Italy

Co-authors: Matteo Barigozzi, Antonio Maria Conti

A Structural Dynamic Factor model on a large panel of Euro Area quarterly series is estimated. We find that the Euro Area business cycle is driven by four common shocks, among which we identify the monetary policy shock. Impulse responses show neither a price puzzle nor a delayed overshooting puzzle. We find evidence of asymmetric responses to a monetary policy shock with respect to Greek GDP, German and Italian CPI, Italian and Spanish unemployment. Finally, the Euro's introduction is found to account for a weaker GDP response, a stronger CPI reaction and an increased homogeneity in the transmission of shocks across countries.

C106: **A data oriented tool for identifying monetary policy SVARs**

Presenter: **Matteo Fragetta**, University of Salerno, Italy

There is an ongoing debate on how to identify monetary policy shocks in SVAR models. Graphical modelling exploits statistical properties of data for identification and offers a data based tool to shed light on the issue. We conduct a cross-country analysis, considering European Monetary Union (EMU), Japan and US. We obtain some important results. The information set of the monetary authorities, which is essential for the identification of the monetary shock seems to depend on availability of data in terms of higher frequency with respect to the policy instrument (US and Japan). Moreover, there is not yet a widespread consensus on whether the European Monetary Union should be considered as a closed economy. Our results indicate that EMU official interest rate depends on the US federal funds rate.

C573: **Estimating persistent and transitory monetary shocks in a learning environment**

Presenter: **Juan Angel Lafuente**, University Jaume I, Spain

Co-authors: Rafaela Perez, Jesus Ruiz

An estimation method for persistent and transitory monetary shocks using the monetary policy modelling is proposed. The contribution is twofold: a) we propose a state-space representation for the time evolution of monetary shocks that allow us the use of the Kalman filter as the optimal signal extraction mechanism, and b) it offers the possibility to perform a maximum likelihood estimation for all the parameters involved in the monetary policy. We present empirical evidence for the US to show the potential applicability of our estimation method. Our findings show that the evidence of a regime change in US monetary policy making from 1980 to 2001 is weak. However, September eleven and the recession that started in the second quarter of 2001 emerge as events that probably led to regime shifts in US monetary policy making. Interestingly enough, we show that the use of a Taylor rule with time-varying responses in accordance with a Markov-swicthing setting leads to similar qualitative findings.

C646: **Macroeconomic forecasting using kernel ridge regression**

Presenter: **Peter Exterkate**, Erasmus University Rotterdam, Netherlands

Co-authors: Patrick J.F. Groenen, Christiaan Heij

Nonlinear and high-dimensional relations can be estimated using kernels, which form a popular and well-established tool in the machine learning community. They allow modelling and estimating nonlinear relations by mapping the observed predictor variables nonlinearly into a high-dimensional space, where prediction takes place. To avoid overfitting a penalty term is added. This form of kernel ridge regression has been found to perform very well in many different contexts. Despite these attractive features, this method is not commonly used in time-series applications. In this paper, we study the forecasting performance of kernel ridge regression. We present Monte Carlo evidence for the adequacy of this method for prediction purposes. Empirical results confirm that kernel regression produces more accurate forecasts for two key U.S. macroeconomic variables than traditional methods do, especially when forecasting at longer horizons.

CS85 Room MAL 355 CONTRIBUTIONS TO FINANCE I Chair: Marc Paolella

C217: **Endogenous dynamics of financial markets**

Presenter: **Miguel Caldas**, Universidade Nova de Lisboa, Portugal

Co-authors: Joao Amaro de Matos

The aim is to define a theoretical framework that will allow for a reliable simulation of the dynamical behavior of financial markets. This will in

turn allow for the identification of endogenous cycles and other coherent dynamical structures occurring in financial markets' time series. In order to do so, a dynamical model exhibiting four different feedback mechanisms was created (in opposition to the single feedback loop usually adopted in the available models). Our model implements feedback through a dynamical update of agents' beliefs, the presence of consistent behavior bias, the influence of investors' sentiment on investors' risk preferences and the effect of borrowing constrains enforced through a margin restriction mechanism. The model also allows for external money flows, being therefore able to model the dynamical effect of liquidity shocks. The overall model contains all the dynamic, non-linear and feedback features necessary to account for the observed dynamic behavior of financial markets.

C305: **High frequency FOREX market transaction data handling**

Presenter: **Shaimaa Masry**, University of Essex, UK

Co-authors: Monira Aloud, Alexandre Dupuis, Richard Olsen, Edward Tsang

The foreign exchange market generates millions of daily tick data, often referred to as high frequency data (HFD), as a result of market participants decisions. By analyzing these data, one could reveal many of the market properties. However, HFD may possibly contain observations that are not reliable in terms of actual market activity. We manipulate a real dataset storing the full transaction history of more than 40,000 traders on an account level for 2.25 years. Prior to exploring the data to discover basic mechanisms of the market dynamics, we perform a cleaning procedure to remove erroneous or misleading observations from the set that would affect the validity of any forthcoming results. Validating the adopted cleaning methodology, we can confirm a clean transaction dataset allowing for a better understanding of the financial market.

C519: **Investment options with debt financing and differential beliefs**

Presenter: **Nicos Koussis**, Frederick University, Cyprus

Co-authors: Spiros Martzoukos

The impact of differential information between debt and equity holders on firm value, optimal capital structure, the timing of investment and other variables such as the credit spreads is investigated. The model explains the existence of debt constraints when debt holders have unfavorable beliefs about the volatility or growth of the firm. Within this framework, equity holders' volatility choice prior and after the exercise of the investment option is also analyzed which is affected by the stickiness of debt holders beliefs.

C594: **Option pricing using fuzzy and neuro-fuzzy inference systems**

Presenter: **Leandro Maciel**, State University of Campinas, Brazil

Co-authors: Rosangela Ballini

In recent decades, option pricing has become the focus of risk managers, policymakers, traders and more generally all market participants, since they find valuable information in these contracts. Most of the research efforts have attempted to relax some of the restrictive assumptions underlying the traditional models. A number of alternative approaches to account for the misspecifications of traditional methodologies have been proposed, especially applications of non-linear models. This paper proposes the pricing and hedging performance on spot U.S. Dollar contract, traded at Brazilian derivatives markets, using fuzzy and neuro-fuzzy models, for the period from January 2000 to December 2007. A fuzzy ruled-based system was built with a family of conditional if-then statements whose consequent are quadratic polynomial functions of the antecedents. Moreover, the same model was composed with the aid of fuzzy neurons, i.e., forming an adaptive neuro-fuzzy inference system. The antecedent membership functions were obtained with fuzzy c-means and subtractive clustering algorithms as a comparison. The performance of these models was compared with the Black model. The models were evaluated according to the traditional error measures and statistical tests in different degrees of moneyness. Results suggested that a neuro-fuzzy model proposed outperforms the others methodologies, mainly in deep out-of-the-money options.

C697: **On credit risk transfer between states and financial institutions before and after government interventions**

Presenter: **Yves Stephan Schueler**, University of Konstanz, Germany

Co-authors: Adrian Alter

This study analyses credit risk transfer from financial institutions to countries during the financial crisis. For this, we consider sovereign CDS from several EU countries (France, Germany, Italy, Ireland, Netherlands, Portugal, and Spain) together with a selection of bank CDS from these states. Data is analyzed from June 2007 to September 2010 and divided in five sub-periods in order to account for structural breaks: three during the subprime crisis but before government bailout programs, and two after the rescue packages. Firstly, we research on the long-run and short-run relations between the time series within a VECM/VAR framework. Secondly, we use Multivariate GARCH techniques for highlighting the correlation structure change. We show that cointegration relationships and short-run dependencies between sovereigns and financial institutions emerge due to bailout programs. Furthermore, our analysis reveals an increase in the correlation between public and private institutions after the aid packages. Our results should be valuable for regulators, portfolio managers, and credit protection issuers for gaining a better understanding of what are the implications of private-to-public risk transfer.

CS72 Room MAL B36 EMPIRICAL MODELLING OF FINANCIAL INTERMEDIARIES Chair: Andrea Cipollini

C142: **Stress testing credit risk: The great depression scenario**

Presenter: **Simone Varotto**, University of Reading, UK

By using Moody's historical corporate default histories we explore the implications of scenarios based on the Great Depression for banks' economic capital and for existing and proposed regulatory capital requirements. By assuming different degrees of portfolio illiquidity, we then investigate the relationship between liquidity and credit risk and employ our findings to estimate the Incremental Risk Charge (IRC), the new credit risk capital add-on introduced by the Basel Committee for the trading book. Finally, we compare our IRC estimates with stressed market risk measures derived from a sample of corporate bond indices encompassing the recent financial crisis. This allows us to determine the extent to which trading book capital would change in stress conditions under newly proposed rules. We find that, typically, banking book regulation leads to minimum capital levels that would enable banks to withstand Great Depression-like events, except when their portfolios have long average maturity. We also show that although the IRC in the trading book may be considerable, the capital needed to absorb market risk related losses in stressed scenarios can be more than twenty times larger.

C875: **Stress testing Czech households**

Presenter: **Petr Jakubik**, Charles University in Prague, Czech Republic

The economic impact of the current global economic downturn on the household sector is investigated. Household budgets can be negatively affected by declines in nominal wages and increases in unemployment. We empirically test this effect for the Czech economy. As a result of the lack of individual data on the Czech household finances, micro data are simulated. Our analysis clearly points out that there is a significant additional decline in consumption related to an increase in household default rates and unemployment. We find that potential household insolvencies have important implications for the financial system as well as for the aggregate economy.

C158: **Modelling credit risk for innovative firms: the role of innovation measures**

Presenter: **Costanza Torricelli**, University of Modena and Reggio Emilia, Italy

Co-authors: Chiara Pederzoli, Grid Thoma

The aim is to build a logit model for the estimation of the probability of default (PD) of innovative firms. To this end we use public accounting data (AMADEUS business directory) and we investigate the advantages of including specific measures of innovation as explanatory variables in estimating the PD. To define an innovative company we use patent data. While not inventions are patented, patenting activities have increased significantly in the last decade in terms of larger company patent portfolios and larger share of firms applying for patents in many different technologies (OECD Patent Compendium, 2008). On the other hand, patents can be considered a highly objective data source over time and they provide very detailed information regarding the invention and its inventors. The definition of innovative companies adopted here is sufficiently comprehensive, including all European firms that have filed at least one patent in one of the major patent office such as EPO, USPTO or WIPO. Then, relying on the AMADEUS business directory we integrated dataset with demographic and accounting information, such as sector activity, ownership, balance sheet, profit and loss account, firm's legal status such as active or bankrupt.

C145: **Forecasting financial distress in emerging markets**

Presenter: **Andrea Cipollini**, University of Modena and Reggio Emilia, Italy

Co-authors: George Kapetanios, George Chortareas

We employ a dynamic probit model to forecast in sample and out-of-sample financial distress in a number of emerging market countries. For this purpose we use a pooled panel dataset of financial stress indicators constructed by the IMF (which do take into account both financial markets variables refelecting stress in the banking sector, as well as in the currency market). Given the presence of overlapping data, generating autocorrelation in the forecast errors, we use, when turning our focus on in sample forecasting, Maximum Likelihood together with a robust estimator of the covariance matrix of the parameters. Then, we employ the iterative procedure to produce dynamic out-of-sample forecasts and we compare the predictive performance of the pooled (dynamic) model with various competing models. Preliminary analysis shows an important role played by contagion effects from US financial market distress to the one affecting emerging markets countries.

C183: **The risk-taking channel in Colombia**

Presenter: **Martha Lopez**, Banco de la Republica - Colombia, Colombia

Co-authors: Fernando Tenjo, Hector Zarate

The recent financial crisis has brought to the forefront the need of a better understanding of the transmission mechanisms of monetary policy. The main step forward in this direction has drawn on work aimed at stressing the role of the financial sector in this transmission. Particular emphasis has been place on how policy actions impact risk perceptions and attitudes of banks and other financial institutions, leading to shifts in the supply of credit. We use a data set from the Credit Register from Colombia that contains detailed information on individual bank loans during the period 2000:I-2008-IV to investigate the impact of monetary policy stance on the risk-taking behavior of banks. Using econometric models of duration, the present paper finds a significant link between low interest rates and banks' risk taking based on evidence from Colombia. Lower interest rates raise the probability of default on new loans but reduce that on outstanding loans. Furthermore, this channel of policy transmission depends on some bank, loan and borrower characteristics, as well on macroeconomic conditions such as the rate of growth of the economy.

CS75 Room MAL B35 RESOURCE ECONOMETRICS Chair: Lynda Khalaf

C333: **The environmental Kuznets curve**

Presenter: **Michael Gavin**, Carleton University, Canada

Co-authors: Lynda Khalaf, Jean-Thomas Bernard, Marcel-Cristian Voia

We consider an empirical estimation of the Environmental Kuznets Curve (EKC), with a focus on assessing the existence of the curve and confidence sets estimation of the turning point where relevant. Various econometric methods are considered, reflecting the implications of persistence in the data. Our results underscore the importance of: (i) correcting for endogeneity, and (ii) disaggregating our panel of countries geographically. We also show that reliance on the delta method in deriving confidence sets can lead to spurious decisions.

C435: **Air pollution, spillovers and U.S. state productivity**

Presenter: **Neophyta Empora**, University of Cyprus, Cyprus

Co-authors: Theofanis Mamuneas, Thanasis Stengos

This paper measures the effect of pollution and pollution spillovers on the Total Factor Productivity growth among the 48 contiguous U.S. states, for the period 1965-2002. We use sulphur dioxide (SO2) as a measure of air pollution. We examine the relationship between TFP growth and pollution using a semiparametric smooth coefficient model that allows us to directly estimate the output elasticity of pollution without imposing any restriction on the functional form of the production function. Following the same methodology, we also account for the effect of interstate pollution spillovers on state level productivity growth. The results indicate that there is a positive relationship between a state's "own pollution" and TFP growth whereas in the case of the spillover pollution the relationship with TFP growth is negative. Furthermore this relationship is nonlinear.

C561: **Kernel-smoothed P-values**

Presenter: **Patrick Richard**, Universite de Sherbrooke, Canada

Although personal computers' capacities have dramatically increased during the past decades, some simulation-based inference methods remain too costly to be widely usable because it is hard, if not impossible, to obtain a large enough number of simulated statistics in a reasonable time. Some examples are: 1) applying the bootstrap to test hypothesis in models estimated by methods requiring simulations; 2) using the bootstrap when the standard error of the statistic is itself obtained with the bootstrap; 3) applying the Maximized Monte Carlo procedure over a large set of nuisance parameters values; 4) subsampling procedures might not allow large numbers of simulated statistics to be computed. These techniques are often useful within the context of energy and resources economics modeling. A recent work suggests that bootstrap P-values based on a nonparametric kernel estimate of the simulated statistics' distribution may allow some accuracy gains when the number of simulations is small. I use extensive simulation evidence to illustrate how such kernel-smoothed P-values may provide more accurate inferences in a wide variety of situations. Practical examples in the field of energy and resource economics are provided. Some attention is also given to the problem of optimal bandwidth selection.

C606: **Estimation of the Hotelling rule for oil and for coal under stochastic investment opportunities**

Presenter: **Johnson Kakeu**, Georgia Tech, USA

This empirical paper uses market capitalization of oil and coal companies and proved reserves to investigate the role of in-ground natural resources in risk diversification. A formal relationship between the return on a unit of the resource in the ground and the stock market return of the mining

company is derived. The data is then used to estimate a stochastic version of the Hotelling rule of exhaustible resource exploitation and uses it to infer on the riskiness of investment in nonrenewable resources. This work builds on a theoretical resource model which use an intertemporal capital asset pricing approach to derive the stochastic version of the Hotelling rule which forms the basis for the performed estimations. We rely on an econometric approach for estimating stochastic diffusion processes and another method for estimating ratio parameters. The empirical results suggest that holding oil and coal reserves as assets can constitute a form of insurance against adverse long-run market fluctuations.

C730: **Independent factor autoregressive conditional density model**

Presenter: **Alexios Ghalanos**, Cass Business School, UK

Co-authors: Giovanni Urga, Eduardo Rossi

An Independent Factor Autoregressive Conditional Density (ACD) model using independent Component Analysis is proposed. It extends recent approaches in multivariate modelling with higher conditional moments and orthogonal factors. Using the Generalized Hyperbolic (GH) as conditional distribution, conditional co-moments and weighted density for use in portfolio and risk management applications are derived. The empirical application, using MSCI equity index iShares data, shows that ACD modelling with a GH distribution allows to obtain tail dependent risk measures, such as Value at Risk (VaR) estimates, which improve on those obtained from higher moment time-invariant conditional densities.

CS78 Room MAL B33 COMPUTATIONAL ECONOMETRICS AND FINANCIAL TIME SERIES Chair: Alessandra Amendola

C641: **Modelling realized covariances and returns**

Presenter: **John Maheu**, University of Toronto, Canada

Co-authors: Xin Jin

New dynamic component models of realized covariance (RCOV) matrices based on recent work in time-varying Wishart distributions are proposed. The specifications are linked to returns for a joint multivariate model of returns and covariance dynamics that is both easy to estimate and forecast. Realized covariance matrices are constructed for 5 stocks using high-frequency intraday prices based on positive semi-definite realized kernel estimates. The models are compared based on a term-structure of density forecasts of returns for multiple forecast horizons. Relative to multivariate GARCH models that use only daily returns, the joint RCOV and return models provide significant improvements in density forecasts from forecast horizons of 1 day to 3 months ahead. Global minimum variance portfolio selection is improved for forecast horizons up to 3 weeks out.

C484: **A machine learning approach to cross-section analysis and asset management**

Presenter: **Ludovic Cales**, Universite Paris-1 Pantheon-La Sorbonne, France

Co-authors: Monica Billio, Dominique Guegan

The aim of this paper is to study the cross-sectional effects present in the market using a new framework based on graph theory and machine learning. Within this framework, we model the evolution of a dynamic portfolio, i.e. a portfolio whose weights change over time, as a function of cross-sectional factors where the predictive ability of each cross-sectional factor is described by a variable, called the "level of expertise" of the factor. Practically, the models permit us to measure the effect of different cross-section factors on a given dynamic portfolio. Associated to a regime switching model, we are able to identify phases during which the cross-sectional effects are present in the market. Another interesting application of such framework is the combination of models. Indeed, in the same way, we can estimate the level of expertise of each model. Then, we can use them to combine the models' forecasts of the optimal portfolio and produce a more robust forecast. We present some empirical applications in the US and European markets in order to support the models presented.

C526: **Wavelet-based functional confidence intervals for contaminated financial data**

Presenter: **Rosaura Fernandez Pascual**, University of Jaen, Spain, Spain

Co-authors: M. Dolores Ruiz Medina

The classical statistical approaches that represent the stochastic behaviour of the price of an underlying asset are extened. The statistical analysis of financial data involves working with massive data sets that present a complicated spatiotemporal structure. We propose a fractional model from a Black-Scholes equation that adequately represents key features related to the statistical issues of heavy tails and long-range dependence observed in time series of market data sets. The least-squares linear extrapolation and filtering problems are considered from contaminated observations. In particular, the discrete wavelet domain is considered for weakness the spatial dependence structure. This paper concentrates on studying the suitable conditions for deriving the asymptotic normality of the functional estimator. Under such conditions, functional confidence intervals are also defined. A simulation study illustrates the methodology presented.

C851: **Analyzing and forecasting volatility using wavelets and nonlinear time series analysis**

Presenter: **Andrew Clark**, Thomson Reuters, USA

Multiresolution Analysis (MRA) via wavelets is used to understand stock and commodity market volatility at various time horizons. An analysis of the different horizons via the use of wavelet coefficients shows that shorter time horizons – anywhere from 1 day to approximately 3 weeks – is very noisy and has a significant level of determinism. That is, there is the possibility of forecasting each of these wavelets. Longer time horizons are characterized by lower levels of noise and larger wavelet coefficients – typically 1 order of magnitude or more versus short dated coefficients. These larger coefficients, indicative of a higher energy level, demonstrate the strong influence of longer dated traders on the market in most market conditions. The longer dated wavelets also show signs of determinism. Forecasts are made for all the wavelets and good volatility forecasts going out 5–15 days business days is demonstrated. The same techniques are shown to generate good forecasts of volatility curves.

C372: **Semiparametrically estimate vector Multiplicative Error Model using empirical likelihood method**

Presenter: **Hao Ding**, The Chinese University of Hong Kong, Hong Kong

Co-authors: Kai Pui Lam

In previous work on estimating vector Multiplicative Error Model (vMEM), we find that Generalized Method of Moments (GMM) is sensitive to moment conditions, especially the first moment condition. When data contain outliers the pre-assumed first moment condition (expectation of innovations equals a vector of one) may not be satisfied. This phenomenon motivates us to find a method which is more robust meanwhile has fewer assumptions than GMM: Empirical Likelihood (EL) method proposed by Owen. We reduce the impact of outliers by ranking each pair of innovations, and using the ranks to construct weighted mean and covariance matrix which serve as constraints for empirical likelihood function. This EL method avoids imposing assumptions on innovations and has the following advantages: first of all it retains the efficiency of Maximum Likelihood (ML) method which assumes innovations follow parametric margins linked by copula function, secondly it does not require the estimation of optimal weight matrix as in GMM and thirdly it is more robust than ML method and GMM when dealing with contaminated data. In the simulation study, we compare these three methods in two cases of model mismatch: data with outliers and heavy tailed noise. The results show that EL method outperforms GMM and ML method.

Sunday 12.12.2010 11:10 - 12:50 Parallel Session L – ERCIM

EI83 Room Senate Beveridge Hall INVITED SESSION: SMALL AREA ESTIMATION Chair: Domingo Morales

E071: **New developments in small area estimation**

Presenter: **Danny Pfeffermann**, University of Southampton, UK and Hebrew University of Jerusalem, Israel, UK

Small area estimation (SAE) is a topic of great importance due to the growing demand for reliable small area statistics such as means and proportions, even when only very small samples or no samples are available for some of these areas. Interest in small area estimation methods has further increased in recent years due to the tendency of many European countries to base future censuses on administrative records in conjunction with small area estimates, used for testing, correcting and supplementing the administrative data. In this talk I shall focus on the following new developments in SAE (time permitting): 1- The Small Area estimation problem; 2- Estimation of prediction MSE; 3 - Bayesian methods; 4- Benchmarking; 5- Errors in covariates 6- Robustness: Outliers, MBD estimation, M-quantiles, Penalized spline regression 7- Prediction of ordered area means; 8- Other developments.

E174: **Small area prediction for count data**

Presenter: **Wayne Fuller**, Iowa State University, USA

Co-authors: Emily Berg

Small area procedures for a two-way table of counts, where the estimated counts are from complex samples are considered. Modeling is difficult because distributions may differ by area, estimates of means may be correlated with estimates of variances, the parameter space is restricted, mean models are nonlinear, and benchmarking may be required. Some aspects of small area prediction for such situations are considered and illustrated with data from the Canadian Labour Force survey.

E545: **Hierarchical Bayesian small area estimation for diurnal data with application to recreational fisheries**

Presenter: **Jay Breidt**, Colorado State University, USA

Co-authors: Daniel Hernandez-Stumpfhauser, Jean Opsomer

In this application, we are interested in obtaining predictions of the daily distributions of the departures of recreational anglers along the coasts of the United States, as a function of the type of fishing trip, its location and time of year. In order to reflect the circular nature of the departure times, we model them as projected bivariate normal random variables. We propose a new latent hierarchical regression model which makes it possible to incorporate covariates and allows for spatial prediction and inference, using Bayesian methods and Markov chain Monte Carlo. We investigate a number of issues related to model specification, model selection and computational efficiency. The approach is applied to a large dataset collected by the US National Oceanic and Atmospheric Administration.

ES39 Room MAL 152 LONGITUDINAL DATA ANALYSIS Chair: M. Carmen Pardo

E150: **Penalized mixed models for longitudinal data**

Presenter: **Maria Durban**, Carlos III University of Madrid, Spain

Penalized splines are a well established method for smoothing Gaussian and non-Gaussian data. The aim of this talk is to show how the mixed model representation of penalized splines is a flexible method, useful in the analysis longitudinal data, in particular for the estimation of subject-specific curves. These models relax the linearity assumption in the population mean, and allow to fit smooth individual departures from the mean. The method is computationally efficient, it is based on the mixed model representation of penalized splines, and it is easily implemented in standard statistical software. We illustrate the use of the models through the analysis of data from children suffering from leukemia.

E321: **A hidden Markov model for informative dropout in longitudinal response data**

Presenter: **Robin Henderson**, Newcastle University, UK

We adopt a hidden state approach for the analysis of longitudinal data subject to dropout. Motivated by two applied studies, we assume subjects can move between three states: stable, crisis, dropout. Dropout is observed but the other two states are not. During a possibly transient crisis state both the longitudinal response distribution and the probability of dropout can differ from the stable state. We adopt a linear mixed effects model with subject-specific trajectories during stable periods and additional random jumps during crises. We place the model in the context of Rubin's taxonomy and develop the associated likelihood. The methods are illustrated using the two motivating examples.

E463: **Gene-by-gene interaction analysis for correlated phenotypes in genetic association studies**

Presenter: **Taesung Park**, Seoul National University, Korea (Rok)

Co-authors: Sohee Oh

Most complex biological phenotypes are often affected by multiple genes and environmental factors. Thus, the investigation of gene-gene and gene-environment interactions can be essential in understanding the genetic architecture of complex traits. Many different methods have been proposed to analyze gene-gene interactions in genetic association studies. Among them, the multifactor dimensionality reduction (MDR) method is known to have the advantages in examining high-order interactions and detecting interactions and has been widely applied to detect gene-gene interactions in many common diseases. However, the current MDR analysis can handle only single phenotype. We extend MDR to handle the clustered phenotypes such as correlated phonotypes or repeatedly measured phenotypes. The proposed MDR analysis is applied to a large scale data from a genome-wide association study in Korea.

E434: **Influence measures for GEE**

Presenter: **Maria del Carmen Pardo**, Complutense University of Madrid, Spain

The generalized estimating equations (GEE) to analyze longitudinal data, have been widely applied to a wide range of medical and biological applications. In order to make regression a useful and meaningful statistical tool, emphasis should be put not only on inference or fitting but also on diagnosing potential data problems. Most of the usual diagnostics for linear regression models have been generalized for GEE. However, influence measures based on the volume of confidence ellipsoids are not available for the GEE analysis. We present an extension of such measures valid for correlated measures regression analysis via GEEs. The proposed measures are illustrated by an analysis of epileptic seizure count data arising from a study of prograbide as an adjutant therapy for partial seizures.

ES41 Room MAL B20 EVOLUTIONARY MODEL-BASED DESIGN OF EXPERIMENTS Chair: Irene Poli

E582: **Adaptive design optimization of combinatorial experiments via interaction information**
Presenter: **Davide Ferrari**, University of Modena and Reggio Emilia, Italy
Co-authors: Davide De March, Matteo Borrotti

We study an adaptive procedure for selecting design points corresponding to an optimal outcome in a sequence of experiment batches. The design variables are qualitative factors with many levels, typically involving a combinatorially large search space due to the potential presence of relevant interactions. For each experiment generation, the strategy entails importance sampling, where the importance weights are chosen by optimizing an empirical approximation of the interaction information, a multivariate extension of the usual mutual information of two variables, which measures the amount of synergy bound up in a set of variables. The convergence properties of the adaptive procedure are studied by means of asymptotic analysis and numerical simulations. The method exhibits a desirable global behaviour as the best solution in each batch of experiments converges in probability to the optimal design. Under proper conditions, the optimal design is attained within a relatively small number of experiments.

E673: **The evolutionary Bayesian network design for combinatorial synthetic proteins**
Presenter: **Debora Slanzi**, University Ca' Foscari, Italy
Co-authors: Irene Poli, Davide De Lucrezia, Giovanni Minervini

The design of synthetic proteins requires the search in high dimensional and complex space. The combinatorial nature of proteins, the non-linear interactions among their components and the complexity of the optimization process can infact make unfeasible the exhaustive analysis of the whole experimental space. In this work we address the high-dimensional combinatorial experimental design of synthetic proteins by adopting an evolutionary procedure guided by an adaptive Bayesian network, which identifies in a sequential way the best solutions of the system. We build the procedure in a way that design and laboratory experimentation interact, leading the search toward the optimality region of the space. In developing this approach we model the experimental results to understand the underlying process in term of components and interactions among them, and then we derive inference on the components, which a-posteriori have maximum probability of occurrence given the achieved interaction structure. The comparison with a genetic algorithm approach to design and analyze the synthetic proteins space is developed and the evolutionary Bayesian network based procedure shows a better performance in finding set of proteins which are similar to natural existing target proteins.

E764: **Comparison of different algorithms for optimizing the pseudo–likelihood surface in copula models**
Presenter: **Matteo Borrotti**, University of Bologna, Italy
Co-authors: Enrico Foscolo

In order to maximize the log–pseudo–likelihood in copula models, gradient–based algorithms are regarded as standard tools. Nevertheless, they seem to fail when the number of dimensions and copula parameters turns out to be high. Moreover, an initial vector of values for the parameters are required in the optimization procedure. When poor initial guesses are given, local optima are produced and biased estimates for the parameters are usually returned. In order to avoid these drawbacks, we design experiments applying a population based, general–purpose stochastic search technique, namely the Adaptive Particle Swarm Optimization (APSO). The APSO is compared with other optimization algorithms in order to assess the performance of this bio–inspired procedure in such a problem. Several simulation studies are carried out with one–parameter and two–parameter families of Archimedean copulas. Constructions of asymmetric multivariate copulas are also considered for testing the proposed method. The APSO provides unbiased estimates for the copula parameters with reasonable computational efforts in comparison with other approaches.

E771: **The optimal design of building envelopes via particle swarm optimization**
Presenter: **Daniele Orlando**, European Centre for Living Technology, Italy
Co-authors: Silvio Giove, Irene Poli

The Particle Swarm Optimization (PSO) is a stochastic bio-inspired optimization technique, based on populations of particles moving in large spaces in order to optimize a specified function. This technique has been developed in the optimization literature mainly for problems described by continuous variables, with few contributions to categorical variables problems. In this paper we propose an innovative methodology (Q-PSO) where the updating rule is derived adopting probabilistic attractive measures based on evolutionary principles. We apply this procedure to an engineering optimization problem concerning the facade building design. More specifically we derive an optimal design of a building envelope, in order to minimize the emissions of carbon dioxide produced, due to the energy spent for heating, cooling and lighting. We then develop a simulation study to evaluate this optimization procedure and derive comparisons with more conventional approaches. In this simulation study the proposed approach shows a very good performance in achieving the assigned target.

ES10 Room MAL 509 OPTIMAL DESIGN FOR NONLINEAR MODELS Chair: Ben Torsney

E212: **Optimal design for additive partially nonlinear models**
Presenter: **Stefanie Biedermann**, University of Southampton, UK
Co-authors: Holger Dette, David Woods

We develop optimal design theory for additive partially nonlinear regression models, showing that Bayesian and standardised maximin D-optimal designs can be found as the products of the corresponding optimal designs in one dimension. A sufficient condition under which analogous results hold for D_s-optimality is derived to accommodate situations in which only a subset of the model parameters is of interest. To facilitate prediction of the response at unobserved locations, we prove similar results for Q-optimality in the class of all product designs. The usefulness of this approach is demonstrated through an application from the automotive industry where optimal designs for least squares regression splines are determined and compared with designs commonly used in practice.

E698: **Bayesian L-optimal exact design of experiments for biological kinetic models**
Presenter: **Steven Gilmour**, Queen Mary University of London, UK
Co-authors: Luzia Trinca

Data from experiments in steady-state enzyme kinetic studies and radiological binding assays are usually analyzed by fitting nonlinear models developed from biochemical theory. Designing experiments for fitting nonlinear models is complicated by the fact that the variances of parameter estimates depend on the unknown values of these parameters and Bayesian optimal exact design for nonlinear least squares analysis is often recommended. It has been difficult to implement Bayesian L-optimal exact design, but we show how it can be done using a computer algebra package to invert the information matrix, sampling from the prior distribution to evaluate the optimality criterion for candidate designs and implementing an exchange algorithm to search for candidate designs. These methods are applied to finding optimal designs for the motivating applications in

biological kinetics, in the context of which some practical problems are discussed. A sensitivity study shows that the use of a prior distribution can be essential, as is careful specification of that prior.

E491: **Optimal design in probit choice models**
Presenter: **Rainer Schwabe**, Otto-von-Guericke University, Magdeburg, Germany
Co-authors: Ulrike Grasshoff, Heiko Grossmann, Heinz Holling

Conjoint analysis is a popular tool in marketing research. Stated choice experiments are performed to evaluate the influence of various options on the consumers' preferences. The quality of the outcome of such experiments heavily depends on its design, i.e. on which questions are asked. In discrete choice experiments the potential answers of customers are usually assumed to follow a multinomial logistic model. It is a well-known phenomenon that those model assumptions are unrealistic, if the choice alternatives are too similar ("red bus, blue bus" paradox). In applications the presentation of similar alternatives is therefore avoided. The concept of design optimization leads, however, to (nearly) identical alternatives. Hence, no meaningful optimal choice sets can be obtained for the multinomial logistic model, which makes this model doubtful. To avoid this pitfall a probit approach is suggested, which allows for modeling the similarity of the alternatives by a dependence structure of the underlying random utilities and which seems therefore to be more realistic. For this new probit choice model optimal designs are obtained by numerical optimization, which support the theoretical considerations.

E793: **The EM algorithm and optimal designs for mixtures of distributions**
Presenter: **Jesus Lopez-Fidalgo**, University of Castilla-La Mancha, Spain
Co-authors: Raul Martin, Mercedes Rodriguez-Hernandez

Maximum Likelihood Estimates (MLE) for a model with mixture of distributions is usually an unaffordable task from a computational point of view, even for simple cases when the number of distributions is known. The EM algorithm is frequently used in this context to approximate the MLE. The Fisher information matrix for the marginal log-likelihood of the observations should be used here. Pure EM estimates are computed and compared to the MLE. Some comparisons of the two information matrices are also performed. Finally, optimal designs are computed for a mixture of normal distributions with modeled means throughout an explanatory variable.

ES60 Room MAL B29 HYBRID MODELS IN UNCERTAINTY AND DECISION Chair: Giulianella Coletti

E336: **On an implicit assessment of fuzzy volatility in the Black and Scholes environment**
Presenter: **Andrea Capotorti**, University of Perugia, Italy
Co-authors: Gianna Figa-Talamanca

Parameter estimation in option pricing models is imprecise by definition. Assuming a Black and Scholes (BS) dynamics for price changes, the crucial parameter is the volatility, the estimation of which is debatable. While pure statisticians suggest the use of sample standard deviation of the asset returns, practitioners claim that market prices for options written on the risky asset are more informative of the current market believes. Recently, several authors have furthermore suggested to explicitly include imprecision in parameters estimation by grading their admissibility through membership functions. In most approaches the support and the shape of the membership functions of the fuzzy parameters are chosen "a priori" . In this work we support the idea of vague-imprecise estimations and we suggest a methodology to obtain, in the context of BS model, the membership of the volatility parameter through implicit information on the risky asset price as well as on market prices for options written on the risky asset. To this aim we profit from the interpretation of membership functions as coherent conditional probabilities and consequent extension principle. Tentatively, we also investigate the possibility of reconciling both sources of information (asset price and options' prices) by minimizing a proper discrepancy measure.

E739: **From qualitative probability to qualitative possibility**
Presenter: **Barbara Vantaggi**, Univ. La Sapienza, Italy

Given two partitions of the sure event with suitable logical conditions between them (weakly logical independent partitions), we proved that any probability distribution P on one of them induces a possibility measure on the algebra A spanned by the other one: in fact, the upper envelope of all the coherent extensions of P to the algebra A is a possibility. Our aim is to study the same problem in a qualitative setting: we consider the so called comparative probabilities (characterized by a condition, which is necessary but not sufficient for the representability by a numerical probability) and the binary relations representable by a probability. The goal is to study the comparative relation induced in the other algebra. Moreover, the results are extended to a family of random variables with finite support, we take in consideration a preference relation agreeing with either an expected valued or an expected utility. We study both the qualitative and numerical decision model.

E746: **Possibilistic graphical models in a coherent setting**
Presenter: **Davide Petturiti**, University of Perugia, Italy
Co-authors: Giuseppe Busanello

Graphical models are an important tool for representing independencies or associations among random variables in probability and statistics. In this paper we study the representation by means of an acyclic directed graph (DAG) of the independence model induced by a coherent conditional possibility. Our starting point is the general axiomatic definition of T-conditional possibility (where T is any t-norm) that, together with the concept of coherence, allows to manage partial assessments (i.e. defined on arbitrary sets of conditional events) and their extension to any new events of interest. In this framework a suitable notion of (conditional) independence has been introduced which allows to avoid the usual counterintuitive situations related to logical dependence highlighted by the other definitions present in the literature. The study of graphoid properties showed that the independence model is not necessarily closed with respect to symmetry and for this reason a proper separation criterion is needed (namely L-separation criterion). We consider the graphical representation of the independence model induced by a coherent T-conditional possibility for $T = \min$ and T strict and study some relevant properties. Finally, we deal with the inference process by means of DAG.

E666: **Inference with hybrid information in a coherent setting**
Presenter: **Giulianella Coletti**, University of Perugia, Italy

We deal with the specific problem of finding the most probable element among those of interest (e.g those of a database), when probabilistic and fuzzy information is available. We refer to the interpretation of membership as a coherent conditional probability, regarded as a function of the conditional events. In this frame the problem is amenable to a Bayesian setting. We analyze the problem when the available "prior probability" and "likelihood" are related to sets of events different from those of interest. This reduces to a problem of joint propagation of fuzzy and probabilistic information, keeping the consistency with coherent conditional probability. Firstly the global coherence of "prior probability" and "likelihood" assessments must be checked: the coherence of each assessment does not assure the global one. Then we need to coherently extend the latter to the new events of interest. Thus, we need to update the probability distribution of the relevant variable, given a "fuzzy event" that means a

claim concerning the relevant fuzzy property. In general this (posterior) probability is not unique, so we can compute the probability intervals and then their upper and lower envelopes. We study the problem of choosing "the most probable" element, when the aforementioned intervals overlap.

ES70 Room MAL 532 MIXTURE MODELS II Chair: Luca Tardella

E274: **Modelling heterogeneity in meta-analysis of diagnostic studies using nonparametric mixture models**

Presenter: **Dankmar Boehning**, Univesity of Reading, UK

Co-authors: Walailuck Boehning, Heinz Holling

Meta-analysis of diagnostic studies are considered using the Lehmann family which has been suggested recently. The Lehmann family has a simply parameter which is identifiable in each study and can be estimated nonparametrically. This parameters represents the diagnostic accuracy involved in each study. Its estimate is considered as a summary statistic containing the study-specific diagnostic information and will be the basis of any further analysis. The approach focuses on modelling heterogeneity of this parameter by means of a nonparametric mixture model with a special normal mixture kernel. A general theory building upon the nonpametric maximum likelihood estimator of the mixing distribution can be developed. Several examples will illustrate the approach.

E273: **Mixture models for Bayesian estimation of survival probabilities in acute myocardial infarction**

Presenter: **Francesca Ieva**, Politecnico di Milano, Italy

Co-authors: Alessandra Guglielmi, Anna Paganoni, Fabrizio Ruggeri

Studies of variations in health care utilization and outcomes involve the analysis of multilevel clustered data. Those studies quantify the role of contributing factors (patients and providers) and assess the relationship between health-care processes and outcomes. We present a case-study, considering first a Bayesian hierarchical generalized linear model with a Normal random effect, later mixtured by a Dirichlet Process, to analyze MOMI2 (Month MOnitoring Myocardial Infarction in MIlan) data on patients admitted with ST-elevation myocardial infarction to one of the hospitals of the Milan Cardiological Network. Both clinical registers and administrative databanks are used to predict survival probabilities. Bayesian non-parametrics is considered in order to take advantage of the "in-built" clustering they provide. The major aim of this work consists of the comparison of the health care providers performances, together with the assessment of the role of patients' and providers' characteristics on survival outcome. Posterior estimates are computed for both regression and random-effects parameters (grouping factor is represented by hospital of admission) through MCMC algorithms. A preliminary step to choice covariates in a Bayesian fashion is performed. Some issues about model fitting are discussed through the use of predictive tail probabilities and Bayesian latent residuals.

E335: **Variational Bayes approximations for the model-based clustering and classification of longitudinal data**

Presenter: **Sanjeena Subedi**, University of Guelph, Canada

Co-authors: Paul McNicholas

Existing model-based clustering approaches to the analysis of longitudinal data are reviewed and discussed, with particular emphasis on an approach that utilizes a Gaussian mixture model with a Cholesky-decomposed component covariance structure. We build on this approach by modelling the mean in addition to the covariance structure, and by taking a different approach to parameter estimation and model selection. A Bayesian framework for parameter estimation is introduced as an alternative to the expectation-maximization framework, and we deviate from the traditional Bayesian information criterion approach to model selection. More specifically, variational approximations are used for the estimation of the parameters as well as the number of components in the mixture model. Our approach is illustrated on real gene expression time course data within both the model-based clustering and classification frameworks.

E418: **Tree selection for hierarchical mixture of experts models**

Presenter: **Ludger Evers**, University of Glasgow, UK

Mixture-of-experts models are a generalisation of mixture regression (or classification) models allowing for locally adaptive mixing weights. Hierarchical Mixture of Experts (HME) models employ a hierarchical tree of logistic regression models for the mixing weights. This tree structure needs to be chosen appropriately. However due to the mixture nature of the model, the canonical tests typically used for model selection cannot be used for tree selection. The talk proposes to use generalisations of the dispersion score test instead. These test statistics have tractable asymptotic distributions and thus can be used in fast algorithms for fitting hierarchical mixture-of-experts models. In addition, some of these tests turn out to be equivalent to well-known tests for adding quadratic interaction effects to the regression equation of one of the experts.

ES66 Room MAL 421 PARTIAL LEAST SQUARES METHODS II Chair: Vincenzo Esposito Vinzi

E803: **Predictive clustering using local partial least squares.**

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

Co-authors: Brian McWilliams

Predictive modelling where both covariates and responses are high-dimensional is becoming a popular task in many data mining applications. In some applications, for example machine translation, each set of variables is described as a "view" of the data, and several multi-view learning methods have been proposed for both regression and clustering tasks. Partial Least Squares (PLS) is a useful model in this setting. It performs "global" dimensionality reduction in both views of the data by assuming the existence of latent factors that are highly predictive. In some applications, it may be more realistic to assume that there are unobserved clusters in the data, for which a "local" PLS model would provide a better fit, yielding improved overall predictive performance. We present a novel framework for predictive clustering where observations are grouped based on their predictive ability by fitting local PLS regression models which are evaluated using a predicted residual (PRESS) statistic. We also present a method of automatically tuning the number of clusters supported by the data. We assess our method using simulated data and compare it with state-of-the-art multi-view learning methods. Our results demonstrate that the clustering and predictive ability obtained by the proposed local PLS model is superior.

E626: **Nonlinear and interaction effects between formative constructs: Two-step PLS path modelling mode B**

Presenter: **Alba Martinez-Ruiz**, Technical University of Catalonia, Spain

Co-authors: Tomas Aluja-Banet

A Two-Step PLS Path Modelling Mode B procedure is implemented to estimate nonlinear and interaction effects among formative constructs. A Monte Carlo simulation study is carried out in order to provide empirical evidence of its performance. The procedure preserves the convergence properties of PLS mode B. The procedure offers a way to build proper indices for linear, nonlinear and interaction terms, all of which are unobservable, and to estimate the relationships between them. Linear, nonlinear and interaction effects are underestimated. Accuracy and precision increase by increasing the sample size. Significant nonlinear and interaction effects and an increase in the predictability of models are detected with medium

or large sample sizes. The procedure is suited to estimate nonlinear and interaction effects in structural equation models with formative constructs and few indicators.

E500: **About the influence of quantification in PLS-PM for costumer satisfaction analysis**

Presenter: **Giorgio Russolillo**, CNAM - Laboratoire Cedric & Chaire de Statistique Appliquee, France

Due to the need to formalize models relating latent concepts, Partial Least Squares Path Modeling has been widely used in marketing research for the quantitative analysis of customer satisfaction. However, in marketing applications latent concepts are expressed as a synthesis of variables that cannot be measured *strictu sensu*. Typically, in fact, the consumer is asked to express the level of agreement to a statement, or a judgment about particular characteristics of the offered product or service, choosing one out of a set of ordered response levels. Variables observed in such a way, however, cannot be considered numerical, as they are not measured on an interval scale. Nearly always, in order to directly obtain quantitative values, the interviewer asks the interviewee to associate the agreement level to one of the values on a certain scale (e.g. 1-10 or 1-100). As a matter of fact, this procedure implies an *a priori* quantification of non-metric variables, which follows two rules: I) Quantifications are different for each level and equally spaced and II) Metric is the same for all the variables in the model. Through a sensitivity study, we investigate how the choice of different quantifications affects model quality and parameter estimation.

E717: **Scoring ordinal indicators in PLS path modeling**

Presenter: **Simona Balzano**, University of Cassino, Italy
Co-authors: Giovanni Camillo Porzio, Laura Trinchera

The PLS Path Modeling (PLS-PM) approach to the estimation of Structural Equation Models (SEM) is addressed. More specifically, we consider the quite frequent case of manifest variables expressed on an ordinal scale. This is commonly treated with standard continuous variable procedures, yielding incoherence with respect to the hypotheses underlying PLS-PM. After a review of the existing literature, in this work we propose to integrate ordinal logistic regression within PLS-PM in order to get a continuous scoring criterion for ordinal categories. As well known, in PLS-PM we can incur in two different kind of models, depending upon the causal relationships between manifest and latent variables (defining the so-called measurement model): one (called reflective scheme) is defined as a set of simple regression models, each with ordinal response variable and continuous predictor, the other (called formative scheme) is defined by a multiple regression model with ordinal predictors. We will address our main interest to the former scheme. We present an application of the proposed procedure on student evaluation of teaching data. Comparing results between classical PLS-PM estimation and our proposed algorithm, we suggest a way to detect differences in the perception of the ordinal scale among different groups of units.

Sunday 12.12.2010 14:45 - 16:25 Parallel Session N – CFE

CS24 Room MAL 152 MCMC APPLICATIONS IN FINANCE Chair: Giovanni Barone-Adesi

C367: **Improving ARMA-GARCH forecasting via partial exchangeability**

Presenter: **Petros Dellaportas**, Athens University of Economics and Business, Greece

Co-authors: Leonardo Bottolo

We exploit the partial exchangeability structure of the parameters of many AR-GARCH models to borrow strength for univariate variance forecasting. We adopt a challenging reversible jump MCMC scheme which models the parameters as a finite mixture of normals with unknown number of components. We generalise existing modeling structures by assuming that the component means follow a multivariate normal density. We discuss in detail the careful choice of prior parameters, the construction of the reversible jump algorithm that requires jumps between covariance matrices of different dimensions and the use of an adaptive regional MCMC algorithm. We test our methodology with stock returns from an S&P100 dataset and we find that our forecasts are more robust and offer better forecasting power when compared with those of standard AR-GARCH models.

C650: **Markov switching models and time series clustering applied to the CDS market**

Presenter: **Luisa Scaccia**, Universita di Macerata, Italy

Co-authors: Rosella Castellano

We exploit Markov Switching models to capture the effect of an event on security prices. We apply this methodology to a set of historical series on Credit Default Swap (CDS) quotes subject to multiple credit events, such as reviews for downgrading and downgradings, to check if market anticipates the announcements and to evaluate how long they lag behind the market. Daily returns (within 80 days windows around the event) are assumed to depend on an unobserved 'regime' in which the return generating process is at time t. We consider a model with two regimes, representing normal and abnormal market conditions and characterized by different means and volatilities of the returns. Furthermore, we propose a model-based clustering of time series to group CDS series with similar temporal dynamics, and estimate the appropriate grouping of CDS series along with model parameters. Switching between regimes occurs at different times in the different groups, while series in the same cluster share the same dynamic pattern and transition probabilities between regimes. Information from all series in the cluster is used to estimate the dynamic pattern, exploiting the attractiveness of pooling short time series. Model parameters and the number of clusters are estimated through MCMC methods.

C787: **Zero variance Markov chain Monte Carlo for Bayesian estimators**

Presenter: **Reza Solgi**, Swiss Finance Institute at the University of Lugano, Switzerland

Co-authors: Antonietta Mira, Daniele Imparato

Variance reduction techniques are essential to Monte Carlo simulations; they can achieve the same level of precision in a shorter time and using a lower budget. Several variance reduction techniques have been proposed since the beginning of popularity of Monte Carlo simulations, and recently in the physics literature, a novel approach has been proposed to build control variates for MCMC simulation. This approach is called the zero variance principle, and can reduce the variance of a Monte Carlo estimator extremely. In this research we investigate the properties of zero variance estimators. Conditions for unbiasedness of the them are derived. A central limit theorem is also proved under regularity conditions. The potential of the new idea is illustrated with real applications to Bayesian inference for probit, logit and GARCH models. In these applications the variance has been reduced dramatically (from hundreds to tens of thousands of times).

C883: **Block bootstrap and long memory**

Presenter: **Fotis Papailias**, Queen Mary, University of London, UK

Co-authors: George Kapetanios

We consider the issue of block bootstrap methods in processes that exhibit strong dependence. The main difficulty is to transform the series in such way so that, implementation of the above techniques can provide an approximation to the true distribution of the desired test statistic. The bootstrap algorithm we suggest consists of the following operations: given $x_t \sim I(d)$, 1) estimate the long memory parameter and obtain $\widehat{d}$, 2) difference the series $\widehat{d}$ times, 3) apply the block bootstrap on the above and finally, 4) cumulate the bootstrap sample $\widehat{d}$ times. Repetition of steps 3 and 4 for a sufficient number of times, results to a successful estimation of the distribution of the test statistic. Furthermore, we establish the asymptotic validity of this method. Its finite-sample properties are investigated via Monte Carlo experiments and the results show that it can be used as an alternative to the sieve *AR* bootstrap for fractional processes.

CS27 Room MAL 151 MULTIVARIATE VOLATILITY MODELS I Chair: Massimiliano Caporin

C096: **Ranking multivariate GARCH models by problem dimension**

Presenter: **Massimiliano Caporin**, University of Padova, Italy

Co-authors: Michael McAleer

Several Multivariate GARCH (MGARCH) models have appeared in the literature. The two most widely known and used are the Scalar BEKK and the DCC models. Some recent research has begun to examine MGARCH specifications in terms of their out-of-sample forecasting performance. We provide an empirical comparison of a set of MGARCH models, namely BEKK, DCC, Corrected DCC (cDCC), CCC, Exponentially Weighted Moving Average, and covariance shrinking, using the historical data of 89 US equities. Our study considered both direct and indirect model comparison methods, and contribute to the literature in several directions. First, we consider a wide range of models, including the recent cDCC model and covariance shrinking. Second, we use a range of tests and approaches for direct and indirect model comparison, including the Weighted Likelihood Ratio tests. Third, we examine how the model rankings are influenced by the cross-sectional dimension of the problem. Results suggest that model ordering is partially influence by problem dimension and that dynamic correlation models are generally preferred given their flexibility.

C234: **Modeling long memory in conditional variance and dependence through dynamic bivariate t copula**

Presenter: **Pawel Janus**, VU University Amsterdam, Netherlands

We propose new models for modeling long memory in conditional variance and conditional dependence. We adopt the copula approach that allows marginal densities of financial returns be modeled separately from their dependence structure. The novelty of the paper concerns the fractionally integrated extension of the baseline short memory dynamics in the generalized autoregressive score modeling framework. The introduced models for time-varying variance and bivariate dependence are designed to account for heavy-tails in returns, for tail dependence between two marginal series and for long memory in the processes. A Monte Carlo study shows that long memory model outperforms the short memory alternatives when the temporal dependence in the dependence process has a slow rate of decay. Extensive empirical investigation for equity data documents the

relevance of the new long memory dynamics for both variance as well as dependence. The resulting variance and dependence processes are related to high-frequency data based estimators and perform the best among other competing standard models.

C714: **Conditional correlation models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations**
Presenter: **Cristina Amado**, University of Minho and NIPE, Portugal
Co-authors: Timo Terasvirta

We investigate the effects of careful modelling the long-run dynamics of the volatilities of stock market returns on the conditional correlation structure. To this end we allow the individual unconditional variances in Conditional Correlation GARCH models to change smoothly over time by incorporating a nonstationary component in the variance equations. The modelling technique to determine the parametric structure of this time-varying component is based on a sequence of specification Lagrange multiplier-type tests. The variance equations combine the long-run and the short-run dynamic behaviour of the volatilities. The structure of the conditional correlation matrix is assumed to be either time independent or to vary over time. We apply our model to pairs of seven daily stock returns belonging to the S&P 500 composite index and traded at the New York Stock Exchange. The results suggest that accounting for deterministic changes in the unconditional variances considerably improves the fit of the multivariate Conditional Correlation GARCH models to the data.

C516: **The conditional autoregressive wishart model for multivariate stock market volatility**
Presenter: **Bastian Gribisch**, Christian-Albrechts-University Kiel, Germany
Co-authors: Roman Liesenfeld, Vasyl Golosnoy

We propose a Conditional Autoregressive Wishart (CAW) model for the analysis of realized covariance matrices of asset returns. Our model assumes a generalized linear autoregressive moving average structure for the scale matrix of the Wishart distribution allowing to accommodate for complex dynamic interdependence between the variances and covariances of assets. In addition, it accounts for symmetry and positive definiteness of covariance matrices without imposing parametric restrictions, and can easily be estimated by Maximum Likelihood. We also propose extensions of the CAW model obtained by including a Mixed Data Sampling (MIDAS) component and Heterogeneous Autoregressive (HAR) dynamics for long-run fluctuations. The CAW models are applied to time series of daily realized variances and covariances for five New York Stock Exchange (NYSE) stocks.

CS32 Room MAL B35 ECONOMIC AND FINANCIAL FORECASTING I Chair: Ana-Maria Fuertes

C181: **Long-horizon return regressions with historical volatility**
Presenter: **Natalia Sizova**, Rice University, USA

Long-horizon regressions of future stock returns on past volatility yield R-square values of more than 72 percent at 10-year horizons. For the same horizons, the predictability of volatility itself is close to zero. This puzzling combination of a higher predictability of returns with a lower predictability of volatility cannot be easily explained within existing econometric frameworks. As a solution, we suggest accounting for the long-memory property of volatility and offer a suitable econometric framework with long-range dependent predictive variables. Once we establish this framework, we apply it to test predictability in NYSE/AMEX returns.

C371: **Forecasting customer behaviour in a multi-service financial organisation: A profitability perspective**
Presenter: **Alena Audzeyeva**, Keele University, UK
Co-authors: Barbara Summers, Klaus Reiner Schenk-Hoppe

This paper proposes a novel approach to the estimation of Customer Lifetime Value (CLV). CLV measures give an indication of the profit-generating potential of customers, and provide a key business tool for the customer management process. Existing approaches show unsatisfactory performance in multi-service financial environments because of the high degree of heterogeneity in customer behaviour. We propose an adaptive segmentation approach which involves the identification of *neighbourhoods* using a similarity measure defined over a predictive variable space. The set of predictive variables is determined during a cross-validation procedure through the optimisation of rank correlations between the observed and predicted revenues. Future revenue is forecast for each customer using a predictive probability distribution based on customers exhibiting similar behavioural characteristics in previous periods. The model is developed and implemented for a UK retail bank; it is shown to perform well in comparison to other benchmark models.

C131: **Modeling the dynamics of temperature with a view to weather derivatives**
Presenter: **Eirini Konstantinidi**, University of Piraeus, Greece
Co-authors: Gkaren Papazian, George Skiadopoulos

The accurate specification of the process that the temperature follows over time is a prerequisite for the pricing of temperature derivatives. To this end, a horse race of alternative specifications of the dynamics of temperature is conducted by evaluating their out-of-sample forecasting performance under different evaluation metrics and forecast horizons. An extensive dataset of the daily average temperature measured at different locations in Europe and U.S. is employed. We find that a developed principal components model and a combination forecasts model perform best in U.S. and Europe, respectively. Point forecasts for popular temperature indices are formed, as well. The results have implications for the pricing and trading of the fast growing class of temperature derivatives, as well as for forecasting temperature.

C890: **Parameter uncertainty in portfolio selection: Shrinking the inverse covariance matrix**
Presenter: **Apostolos Kourtis**, Essex University, UK
Co-authors: Raphael Markellos, George Dotsis

We demonstrate that estimation risk in the inverse sample covariance matrix is of prime importance for mean-variance portfolio optimisation. In this context, we derive novel shrinkage estimators of the inverse covariance matrix on the basis of statistical and economic objective functions. These estimators are applied to the computation of mean-variance and global minimum variance portfolio weights. Our performance analysis shows that the developed portfolio rules are superior to competing approaches under most scenarios.

CS46 Room MAL B36 FINANCIAL MARKET AND MACRO DYNAMICS Chair: Willi Semmler

C609: **The instability of the banking sector and macrodynamics: Theory and empirics**
Presenter: **Willi Semmler**, New School, USA
Co-authors: Stefan Mittnik

The issue of local instability of the macroeconomy arising from a banking sector that is considered here as a wealth fund that accumulates capital

assets, can heavily borrow and pay bonuses is investigated. We presume that the banking system faces not only loan losses but is also exposed to a deterioration of its balances sheets due to adverse movements in asset prices. In contrast to previous studies that use the financial accelerator – which is locally amplifying but globally stable and mean reverting – our model shows local instability and globally multiple regimes. Whereas the financial accelerator leads, in terms of econometrics, to a one-regime VAR we demonstrate the usefulness of a multi-regime VAR (MRVAR). The method of a MRVAR estimate can usefully be applied whenever one has several regimes to which the economy could possible move. We present a model that gives rise to multiple regimes and estimate the model with a MRVAR. We also undertake an impulse response study with an MRVAR which allows us to explore regime dependent shocks. We show that the shocks have asymmetric effects depending on the regime the economy is in.

C859: **Are bank lending shocks important for economic fluctuations?**

Presenter: **Jorn Inge Halvorsen**, Norwegian School of Management, Norway

Co-authors: Dag Henning Jacobsen

We analyze the importance of bank lending shocks on real activity in Norway and the UK, using structural VARs and based on quarterly data for the past 21 years. The VARs are identified using a combination of sign and short-term zero restrictions, allowing for simultaneous interaction between various variables. We find that a negative bank lending shock causes output to contract. The significance of bank lending shocks seems evident as they explain a substantial share of output gap variability. This suggests that the banking sector is an important source of shocks. The empirical analysis comprises the Norwegian banking crisis (1988-1993) and the recent period of banking failures and recession in the UK. The results are clearly non-negligible also when omitting periods of systemic banking distress from the sample.

C655: **Forecasting output using interest rate spreads: exploratory analysis using wavelets**

Presenter: **Marco Gallegati**, Universita Politecnica delle Marche, Italy

Co-authors: James Bernard Ramsey, Willi Semmler

The objective of this section is to verify the benefits from using wavelets for the analysis of the forecasting power of various financial indicators (real interest rate, term and credit spreads) for economic activity. We recognize the possibility that: i) the relationships between economic variables may well be scale dependent; ii) while we may wish to observe, the aggregate variable x, it is most likely that we will instead observe $x* = x + e$, and iii) most economic and financial time series exhibit quite complicated patterns like sudden and abrupt changes, jumps and volatility clustering. Wavelets multiresolution decomposition analysis can provide an effective solution to these problems. The key insight is that wavelets provide a complete decomposition of the original series into several components. This separation is of great benefit in analyzing data with unknown structure. The results from wavelet analysis, to be regarded as *exploratory*, show that allowing for different time scale of variation in the data can provide a fruitful understanding of the complex dynamics of economic relationships between variables, richer than the one obtained using standard methods and aggregate data.

C734: **Business cycle synchronization and volatility transmission: The real-financial sector nexus**

Presenter: **Mehmet Zahid Samancioglu**, Middle East Technical University (METU) / Central Bank of Republic of Turkey (CBRT), Turkey

Co-authors: Esma Gaygisiz

Global crisis in financial sector, which transformed into one of the longest and deepest recession in many countries, emphasized the interrelation between financial markets and real economic activity. In this paper we extract business cycles of several countries, including G7 countries, in a robust manner using Kalman filter and then calculate the business cycle clock of each country, which shows where lies a country in a business cycle phase and at what speed a business cycle revolves. Upon calculation of business cycle clocks, we create two indexes of synchronization for each country: phase synchronization index which indicate whether the business cycle conditions of a country is in line with other countries and speed synchronization index which shows the co-movement of the cycle of a country with other countries. These measures of synchronizations are well defined and reveal more information about synchronizations of business cycles compared to other measures such as correlations of several growth parameters among countries. Moreover we investigate the dynamical relationships between synchronization of business cycles and the transmission of volatility in financial markets. Volatility transmission in financial markets is measured using asymmetric GARCH-BEKK model. Using binary models such as logistic model we find that in several countries, including U.K, Canada and France, periods of high volatility in financial markets can be matched by high synchronization period of business cycle with U.S. For instance in U.K. high levels of synchronization index can match 70 percent of high volatility periods.

CS69 Room MAL G16 NEW DEVELOPMENTS ON GARCH MODELS II Chair: Dimitra Kyriakopoulou

C620: **A dynamic conditional correlation model with factorial hidden Markov representation**

Presenter: **Philippe Charlot**, Aix-Marseille University and GREQAM, France

The goal of this paper is to present Markov-Switching Multivariate GARCH model with dynamic conditional correlation (DCC). We suggest a slightly different approach than the classical Markov-Switching. Indeed, our specification is based on an extension of the Hidden Markov model (HMM) called Factorial Hidden Markov model (FHMM). The FHMM generalize the HMM 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 element of the conditional correlation matrix is linking to a hidden Markov chain and allows them to switch independently from others. In other words, in our model each element of the correlation matrix has its own switching dynamic while in the classical Markov-switching DCC, all the elements of the correlation matrix have the same switching dynamic. This Factorial Markov-switching DCC model is applied to exchange rate data. Results show that our model has an interesting explanatory power by providing more flexibility in the correlation process.

C344: **Comparison of Markov switching GARCH modellings**

Presenter: **Arnaud Dufays**, Universite catholique de Louvain-la-Neuve, Belgium

Co-authors: Luc Bauwens

The GARCH modelling has proved to be a reliable candidate to model the volatility of time series. However many empirical studies reported a strong persistence in the conditional variances for return series. Also previously shown is that estimating a GARCH model on a sample displaying a structural change in the unconditional variance creates an integrated GARCH effect. Those two findings indicate a potential source of misspecification and suggest elaborating more sophisticated models. Here two different Markov switching GARCH Bayesian algorithms respectively a sequential Monte Carlo (SMC) algorithm and a Markov chain Monte Carlo (MCMC) algorithm are compared. It also introduces a particle MCMC (PMCMC) algorithm which combines advantages of the SMC and the MCMC methods. It is shown that the PMCMC algorithm performs better than the two existing ones. An empirical study has been conducted to determine the advantages and weaknesses of each algorithm. Finally an application of the three methods on very long time series (i.e the daily Dow Jones from 1920 to 2009 and the S&P500 from 1950 to 2009) is shown.

C190: **Asymptotic properties of GARCH-X processes**

Presenter: **Heejoon Han**, National University of Singapore, Singapore

The GARCH-X process in which the covariate is generalized as a fractionally integrated process I(d) for $-1/2 < d < 1/2$ or $1/2 < d < 3/2$ is considered. We investigate the asymptotic properties of this process, and show how it explains stylized facts of financial time series such as the long memory property in volatility, leptokurtosis and IGARCH. If the covariate is a long memory process, regardless that it is stationary or nonstationary, the autocorrelation of the squared process of the model generates the long memory property in volatility by following the trend commonly observed in real data. The asymptotic limit of the sample kurtosis of the GARCH-X process is larger than that of the GARCH(1,1) process unless the covariate is antipersistent. We also analyze the effect of omitting the covariate that is nonstationary as well as persistent. Our analysis shows that, if the relevant covariate is omitted and the usual GARCH(1,1) model is fitted, then the model would be estimated approximately as the IGARCH. This may well explain the ubiquitous evidence of the IGARCH in empirical volatility analysis.

C300: **Asymptotic 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 [1991, Econometrica] have not been fully explored. 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 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.

CS71 Room MAL B30 ECONOMETRICS OF ELECTRICITY MARKETS Chair: Luigi Grossi

C233: **Modelling electricity forward markets by ambit fields**

Presenter: **Almut Veraart**, Aarhus University, Denmark

Co-authors: Ole E. Barndorff-Nielsen, Fred Espen Benth

A new modelling framework for electricity forward markets, which is based on ambit fields, is proposed. The new model can capture many of the stylised facts observed in energy markets. One of the main differences to the traditional models lies in the fact that we do not model the dynamics, but the forward price directly, where we focus on models which are stationary in time. We give a detailed account on the probabilistic properties of the new model and we discuss martingale conditions and change of measure within the new model class. Also, we derive a model for the spot price which is obtained from the forward model through a limiting argument.

C437: **Nonlinear specifications and forecasting of daily electricity prices**

Presenter: **Dipeng Chen**, London Business School, UK

Co-authors: Derek Bunn

Daily power prices have a fine structure and are known to be driven by an interaction of fundamental, behavioural and stochastic factors. Furthermore, there are plausible reasons to expect the functional form of price formation to be nonlinear in these factors. Markov regime switching has been widely advocated to capture some aspects of the nonlinearity, but its specification is open to concerns of overfitting and unobservability in the underlying states, with the result that is effectiveness for forecasting in practice is controversial. We compare several extensions and alternative regime switching formulations, including logistic specifications of the underlying states, logistic smooth transition, and a structural finite mixture regression. We discuss both in sample fit and out of sample day-ahead forecasting.

C454: **Forecasting the residual demand function in electricity auctions**

Presenter: **Matteo Pelagatti**, Universita degli Studi di Milano-Bicocca, Italy

Co-authors: Lisa Crosato

In the scientific literature on optimal bidding strategies in electricity auctions it is generally assumed that the bidder knows the residual demand function in each auction she participates. The residual demand function of an operator maps all possible electricity prices to the total electricity demand not satisfied by the other competing bidders at that price. In real electricity auctions the residual demand function is a step function since operators are allowed to present a finite number of quantity-price pairs, while theoretical models approximate it with everywhere differentiable functions. Thus, the optimal bidding strategies proposed in literature remain approximately optimal if instead of the generally unknown real residual demand function one substitutes a good prediction of it. In this paper we propose two approaches for the prediction of smooth (differentiable) approximations of real residual demand functions (and of their derivatives). The first approach is based on time series forecasts of the principal components extracted from the residual quantities at a fixed grid of prices. The second is based on reduced rank regression techniques applied to the sampled residual quantities at the same grid of prices. The proposed techniques are then applied to the Italian (hourly) electricity auctions in the four years 2005-2008.

C574: **Forecasting zonal and national volatility structures of the Italian electricity wholesale market**

Presenter: **Luigi Grossi**, University of Verona, Italy

Co-authors: Angelica Gianfreda

Volatility analysis, modeling and forecasting have several applications in risk management. Since electricity is a commodity with special features not shared with other assets nor commodities, it is important to understand its risk properties and the dynamics of spot price variability. This understanding can help, for instance, firms in formulating their supply schedules, to evaluate real investments using asset pricing techniques, but also consumers to hedge against risk of price spikes or risk associated with physical delivery. Electricity markets exhibit very high volatility, and only few contributions considered the Italian case. Therefore, the volatility of Italian electricity wholesale prices is now studied and forecasted at two levels: a zonal one considering the availability of 24 hourly prices and a national one considering instead the single national price. Then traded volumes are explored as a possible cause generating volatility in the Italian market. We perform volatility forecasting on zonal and national basis accounting for this factor to verify if this explanatory variable helps forecasting models to perform better and to investigate if zonal volatility reductions produce reduced national volatility. Hence, this paper would provide the first forecasting analysis of the Italian Electricity spot volatility considering at the same time the influence of traded volumes.

CS61 Room MAL B33 BAYESIAN ECONOMETRICS AND APPLICATIONS III Chair: Teruo Nakatsuma

C306: **Bayesian risk assessment with threshold mixture extreme value models**

Presenter: **Teruo Nakatsuma**, Keio University, Japan

A new Markov chain Monte Carlo (MCMC) method to estimate a threshold mixture model of financial losses based on the generalized Pareto distribution (GPD) is proposed. In the threshold mixture model we analyze here, small losses are supposed to follow a log-normal or gamma distribution, which we refer to as a base distribution, but large ones that exceed a threshold are supposed to follow the GPD. This model is designed to take into account a stylized fact that the loss distribution tends to be positively skewed and has a heavy right tail. It is also consistent with the extreme value theory on exceedances over a high threshold. Bayesian MCMC approach enables us to simultaneously estimate the threshold along with parameters in the base distribution as well as the GPD. Our new method explicitly takes care of discontinuity in the posterior density of the threshold to obtain a workable proposal distribution for the Metropolis-Hastings algorithm. As a result, it takes less time in computation and can achieve a good mixing property of the sample path. As an application, we apply it to financial loss data and compare threshold mixture models by a Bayesian model selection procedure.

C202: **Generalized extreme value distribution with time-dependence using the AR and MA models in state space form**

Presenter: **Jouchi Nakajima**, Duke University, USA

Co-authors: Tsuyoshi Kunihama, Yasuhiro Omori, Sylvia Fruhwirth-Schnatter

A new state space approach is proposed to model the time-dependence in an extreme value process. The generalized extreme value distribution is extended to incorporate the time-dependence using a state space representation where the state variables either follow an autoregressive (AR) process or a moving average (MA) process with innovations arising from a Gumbel distribution. Using a Bayesian approach, an efficient algorithm is proposed to implement Markov chain Monte Carlo method where we exploit a very accurate approximation of the Gumbel distribution by a ten-component mixture of normal distributions. The methodology is illustrated using extreme returns of daily stock data. The model is fitted to a monthly series of minimum returns and the empirical results support strong evidence for time-dependence among the observed minimum returns.

C189: **Bayesian Semiparametric Modelling of Volatility**

Presenter: **Eleni-Ioanna Delatola**, University of Kent, UK

Co-authors: Jim Griffin

The area of Stochastic Volatility (SV) models has grown rapidly. SV models capture the unobserved time-varying volatility of the log-returns using stochastic processes. The distribution of the innovation of the level of returns is often assumed to be normal or of some other parametric form, which does not allow for some features of the data to be modeled. An offset mixture model has been proposed in the literature that approximates the distribution in question. In our work, we extend it to a nonparametric mixture model and develop efficient computational methods. We analyze the stock returns of General Motors in order to illustrate these methods.

C513: **A Semiparametric Bayesian approach to the analysis of volatilities and value at risk in financial time series**

Presenter: **M. Concepcion Ausin**, Universidad Carlos III de Madrid, Spain

Co-authors: Pedro Galeano, Pulak Ghosh

Financial time series analysis deals with the understanding of data collected on financial markets. Several parametric distribution models have been entertained for describing, estimating and predicting the dynamics of financial time series. Alternatively, this article considers a Bayesian semiparametric approach. In particular, the usual parametric distributional assumptions of the GARCH-type models are relaxed by entertaining the class of location-scale mixtures of Gaussian distributions with a Dirichlet process prior on the mixing distribution, leading to a Dirichlet process mixture model. The proposed specification allows for a greater flexibility in capturing both the skewness and kurtosis frequently observed in financial returns. Furthermore, it is also possible to obtain predictive distributions for the Value at Risk (VaR), which has become the most widely used measure of market risk for practitioners. Through a simulation study, we demonstrate the performance of the proposed semiparametric method and compare results with the ones from a normal distribution assumption. We also demonstrate the superiority of our proposed semiparametric method using real data from the Bombay Stock Exchange Index and the Hang Seng Index.

Sunday 12.12.2010 15:15 - 16:55 Parallel Session N – ERCIM

EI82 Room Senate Beveridge Hall INVITED SESSION: FUZZY SETS IN STATISTICS Chair: Ana Colubi

E179: **Fuzzy predictive distributions**

Presenter: **Reinhard Viertl**, Vienna University of Technology, Austria

In Bayesian analysis for stochastic model $X \sim f(\cdot|\theta)$, $\theta \in \Theta$ the a-posteriori predictive density $p(\cdot|D)$ and its values $p(x|D)$ are defined by the integral $p(x|D) = \int f(x|\theta)\pi(\theta|D)d\theta$ for all $x \in M_X$, where M_X is the observation space of the stochastic quantity X, and $\pi(\cdot|D)$ the a-posteriori density of the parameter θ. In case of fuzzy information, i.e. fuzzy data $D^\star$ and fuzzy a-priori density $\pi^\star(\cdot)$ the corresponding a-posteriori density becomes fuzzy. Therefore the generalization of the above integral for fuzzy valued functions is necessary. The first idea is to use the extension principle from fuzzy set theory. But by applying this, non-intuitive results are obtained. Therefore the mathematical concept of integrating δ-level functions is used. The resulting generalization of predictive densities are so-called fuzzy predictive densities which are explained. Moreover the calculation of generalized probabilities based on fuzzy probability densities, so-called fuzzy probabilities, will be presented.

E625: **Fuzzy scale as an alternative to Likert's ones from a statistical perspective**

Presenter: **Maria Angeles Gil**, University of Oviedo, Spain

Co-authors: Gil Gonzalez-Rodriguez

Likert scales or associated codings are often used in connection with opinions/valuations/ratings, and especially with questionnaires with a pre-specified response format. A guideline to design questionnaires allowing free fuzzy-numbered response format will be given, the fuzzy numbers scale being very rich and expressive and enabling to describe in a friendly way the usual answers in this context. A review of some techniques for the statistical analysis of the obtained responses is enclosed and a real-life example will be used to illustrate the application.

E769: **Maximum likelihood from uncertain data in the Dempster-Shafer framework**

Presenter: **Thierry Denoeux**, Universite de Technologie de Compiegne, France

Recent years have seen a surge of interest in methods for managing and mining uncertain data. Uncertain data arise in many applications due to limitations of the underlying equipment (e.g., unreliable sensors or sensor networks), use of imputation, interpolation or extrapolation techniques (to estimate, e.g., the position of moving objects), partial or uncertain responses in surveys, etc. Usual models involve intervals, probability distributions or fuzzy sets for modeling data uncertainty. In this talk, we introduce a more general model based on belief functions, and we propose an approach that allows uncertainty to be expressed on any continuous or discrete attribute, in any learning problem based on a parametric statistical model. The method is based on an extension of the EM algorithm, called the evidential EM algorithm, which allows us to estimate parameters in parametric statistical models based on uncertain data. We demonstrate the application of this algorithm for handling clustering and classification problems in which attributes and/or class labels are uncertain.

ES67 Room MAL 355 DATA ANALYSIS II Chair: Klea Panayidou

E153: **Alternative versions of the RESET test for binary response index models: A comparative study**

Presenter: **Joaquim Ramalho**, Universidade de Evora, Portugal

Co-authors: Esmeralda Ramalho

Binary response index models may be affected by several forms of misspecification, which range from pure functional form problems (e.g. incorrect specification of the link function, neglected heterogeneity, heteroskedasticity) to various types of sampling issues (e.g. covariate measurement error, response misclassification, endogenous stratification, missing data). We examine the ability of several versions of the RESET test to detect such misspecifications in an extensive Monte Carlo simulation study. We find that: (i) the best variants of the RESET test are clearly those based on one or two fitted powers of the response index; and (ii) the loss of power resulting from using the RESET instead of a test directed against a specific type of misspecification is very small in many cases. We introduce also supremum-type RESET statistics, which our Monte Carlo results show to be particularly useful in cases where one of the two best standard RESET versions exhibits substantially lower power than the other. In particular, the supremum RESET test based only on one and two fitted powers displays a very promising behaviour.

E640: **A robust Bayesian stopping rule for sequential trials**

Presenter: **Stefania Gubbiotti**, Sapienza Universita' di Roma, Italy

Co-authors: Pierpaolo Brutti, Fulvio De Santis

Dealing with sequential clinical trials in a Bayesian context, we denote by θ the parameter of interest representing treatment effect and we start formalizing pre-experimental information through a prior probability distribution on θ. By iteratively applying Bayes theorem, the posterior distribution is derived after each observation is collected. Hence we monitor the posterior probability that θ exceeds a minimally clinical relevant threshold as the sample size sequentially increases. Then the trial is terminated with success when it is larger than a given cutoff; otherwise, the treatment is declared ineffective. In this setting the sample size is a random variable associated to the chosen stopping rule. We show by simulation that its expectation is smaller than the non sequential optimal sample size. Moreover, we consider the issue of robustness with respect to the prior specification. We define a robust Bayesian stopping criterion that is, in general, more conservative than the non robust one. However, we show that, working sequentially, we can save observations even though a robust stopping rule is considered. More precisely, we evaluate the critical level of robustness we can afford to have a sample size that reaches the non sequential non robust optimal one.

E782: **Spatially dependent functional data: city planning through local telephonic time patterns**

Presenter: **Valeria Vitelli**, Politecnico di Milano, Italy

Co-authors: Fabio Manfredini, Paolo Tagliolato, Simone Vantini

We illustrate the case study that stimulated our research in the analysis of functional data spatially distributed on a spatial lattice. Data are measures along time (every 15 minutes for two weeks) of the use of the Telecom mobile phone network across a lattice covering the area of Milan (Italy). Aim of the analysis is the segmentation of the area into districts characterized by homogeneous telephonic patterns. One might expect that spatial dependence among functional data, when carefully exploited, could help identification of relevant features. We perform a data-driven dimensional reduction of functional data while accounting for spatial dependence. At each iteration, a random tessellation of the area is generated and local representatives of functional data are computed. Dimensional reduction is then achieved through projections over relevant and interpretable treelets (e.g. a data-driven wavelets). Corresponding scores provide, at each iteration, the local importance of each selected treelet function to the overall local phone traffic. Finally, scores provided by different iterations at the same location are pooled to obtain a better estimate of the local importance of each selected treelet and a measure of its uncertainty. Data are courtesy of "Convenzione di ricerca DiAP – Telecom Italia", Politecnico di

Milano (Italy).

E852: **Extremes of two-step regression quantiles**
Presenter: **Jan Picek**, Technical University of Liberec, Czech Republic
Co-authors: Jan Dienstbier

The contribution deals with estimators of extreme value index based on two-step regression quantiles in the linear regression model. Two-step regression quantiles can be seen as a possible generalization of the quantile idea and as an alternative to regression quantiles. We derive the approximation of the tail quantile function of errors. We consider a class of smooth functionals of the tail quantile function as a tool for the construction of estimators in the linear regression context. Pickands, maximum likelihood and probability weighted moments estimators are illustrated on simulated data.

ES31 Room B34 ROBUST METHODS AND OUTLIER DETECTION IN FINANCIAL DATA Chair: Aurea Grane

E170: **Identification of outlying volatilities in high-dimensional financial data sets**
Presenter: **Michele La Rocca**, University of Salerno, Italy
Co-authors: Pietro Coretto, Giuseppe Storti

In financial data the need for modeling dynamic inter-dependencies among several assets would in principle require the estimation of an overwhelming number of parameters. As a consequence, it is usually necessary to introduce severe, often untested, homogeneity constraints on the dependence structure of the data. Recent literature has provided evidence in favor of the hypothesis that financial returns are characterized by cluster-wise dependence structures and this information could be used to specify more parsimonious models. Cluster identification, however, can be seriously affected by outlying observations. The paper discusses a robust model based clustering procedure for high dimensional financial data sets. The key assumption of this work is that the cross-sectional distribution of volatilities in a given market is a finite mixture of distributions. Each mixture component represents a group of assets that share similar risk behavior. The addition of a 'noise-component' allows identification of clusters of outlying observations which cannot be classified in any group. The procedure does not involve any arbitrary threshold choice for outliers and it is able to deliver components with a clear cut economic interpretation. An application to all the S&P500 assets will be discussed.

E268: **Jump robust two time scale covariance estimation and realized volatility budgets**
Presenter: **Kris Boudt**, K.U.Leuven-Lessius University College, Belgium
Co-authors: Jin Zhang

We first propose a jump robust covariance estimator that is also robust to asynchronicity and microstructure noise. We then introduce the realized volatility budget for topdown attribution of volatility components caused by each portfolio assets. These risk budgets provide the portfolio manager insights into the portfolio risk concentration.

E400: **Robust estimation of optimal portfolio allocation parameters**
Presenter: **Fabrizio Laurini**, University of Parma, Italy, Italy
Co-authors: Luigi Grossi, Giacomo Scandolo

The main goal of the optimal portfolio allocation of security shares is the estimation of a vector of weights, that represent the proportion of the shares owned by a financial investor. The classical tool for determining portfolio weights has recently lost its feature of "optimal" tool because it can lead to financially irrelevant optimal portfolios. Among the reasons of this drawback a relevant role is played by the influence of extreme returns and the use of the variance as a measure of risk. Other risk measures could be used, with the most famous being the Value at Risk and the Expected Shortfall. We discuss the problem of statistical robustness of the asset allocation method, and show that the latter is not robust, with few extreme returns leading to "sub-optimal" portfolios. We study the degree of robustness of optimization methods based on several measures of risk. We achieve this target by introducing a robust estimator based on the forward search in the maximization procedure. We show that our method outperforms the maximum likelihood estimator. We also compare our method with other robust estimators which have high breakdown point, but suffer from masking effect and are far from efficient.

E066: **Outliers in GARCH models and the estimation of risk measures**
Presenter: **Aurea Grane**, Universidad Carlos III de Madrid, Spain
Co-authors: Helena Veiga

The impact of additive level outliers on the calculation of risk measures, such as minimum capital risk requirements, is considered, and four alternatives of reducing the estimation biases of these measures are compared. The first three proposals proceed by the detection and correction of outliers before estimating these risk measures with the GARCH(1,1) model, while the fourth procedure fits a Student's t-distributed GARCH(1,1) model directly to the data. The former group includes the authors' proposal, a detection procedure based on wavelets with hard- or soft-thresholding filtering, and a well known method widely used in the literature. The first results, based on Monte Carlo experiments, reveal that the presence of outliers can bias severely the minimum capital risk requirement estimates calculated using the GARCH(1,1) model. The message derived from the second results, both empirical and simulations, is that outlier detection and filtering generate more accurate minimum capital risk requirements than the fourth alternative. Moreover, the detection procedure based on wavelets with hard-thresholding filtering provides a very good performance for attenuating the effects of outliers and generating accurate minimum capital risk requirements out-of-sample, even in quite volatile periods.

ES45 Room MAL 421 TIME SERIES MODELS FOR MARKET RISK PREDICTION Chair: Giuseppe Storti

E351: **Time-varying vine copulas**
Presenter: **Giorgia Rivieccio**, Parthenope Universirty of Naples, Italy

Empirical researches in financial literature have shown evidence of a skewness and a time conditioning in the univariate behaviour of stock returns and, overall, in their dependence structure. The inadequacy of the elliptical and, in general, symmetrical multivariate constant model assumptions, when this type of dependence occurs, is an almost stylized fact. Beyond these characteristics, recent studies have highlighted a dynamic whole/tail dependence which changes over time. This paper provides a new approach for modeling multivariate financial asset returns based on time-varying vine copulas. A dynamic multivariate model is proposed, based on a decomposition of d-dimensional density into a cascade of bivariate conditional and unconditional copulas, with two desirable features: a flexible construction, which allows for the free specification of $d(d-1)/2$ copulas and easy computation in presence of high dimensional data.

E413: **Modeling and forecasting volatility subject to changes of regime**

Presenter: **Edoardo Otranto**, University of Sassari, Italy

Co-authors: Giampiero M. Gallo

The visual inspection of time series of high frequency based measures of volatility reveals a slow moving changing level of underlying average volatility by subperiods. This is particularly true in the past ten years, with the aftermath of the tech bubble, the 2001 recession, the low level of volatility in mid decade and then the explosion of uncertainty following the subprime mortgage crisis. Such an empirical regularity challenges traditional econometric approaches to modeling and forecasting, since it is difficult to get rid of autocorrelated residuals. The Multiplicative Error Model is no exception: it models the conditional expectation of the variable of interest (realized volatility) capturing persistence with a GARCH-type dynamics. In this paper we extend this class of models to accommodate the presence of regimes by inserting some threshold parameters and a Markov switching dynamics, allowing us to address the issues of a slow moving average level of volatility and of a different dynamics across regime. We apply the model to realized volatilities measured on a number of different indices and we gauge the usefulness of such an approach by the improved properties of the residuals and by the increased forecasting performance.

E854: **On the treatment of small and medium enterprises' in Basel 2**

Presenter: **Juri Marcucci**, Bank of Italy, Italy

Co-authors: Gaetano Chionsini, Mario Quagliariello

In the Basel 2 Capital Accord, exposures towards small and medium enterprises (SMEs) benefit from a more favourable treatment with respect to large corporates. Under the internal-rating based (IRB) approach, the capital reduction increases linearly from 0 to 20% with sales going from 50 to 5 million euro. The main argument for this special treatment is that SMEs, while riskier, are typically more affected by idiosyncratic shocks than by systemic factors. Studies on the relationship between firm size and asset correlation are rare and do not provide clear-cut results. In this paper, using Italian data, we study to what extent corporate default probabilities (PDs) are affected by a common systemic factor – proxied by macroeconomic developments – and, more importantly, we test whether the influence of such factors is less marked for SMEs. Our results show that, under normal cyclical fluctuations, Basel 2 treatment is right: indeed Italian SMEs are less cyclical than larger firms. However, we do find that these results do not hold in times of abnormally severe recession, such as the current one, when the PDs of all firms seem much more correlated with systemic factors, regardless firm size. We also calibrate the threshold that separates SMEs from larger corporates and find that such threshold is much smaller than the one defined by the Accord. This is probably due to the particular features of the Italian economy characterized by many SMEs and micro firms.

E086: **Exponential smoothing models in portfolio risk measurement**

Presenter: **Andrea Silvestrini**, Bank of Italy, Italy

Co-authors: Giacomo Sbrana

It is dealt with multivariate exponential smoothing models which provide a pragmatic approach to volatility forecasting in portfolios composed of several assets. We suggest a method that allows to infer the elements of the estimated covariance matrix straightforwardly, without imposing the same degree of smoothness on all its elements. This is empirically relevant to tackle the issue of dimensionality, which frequently arises in this context. In addition, we provide an example dealing with Value-at-Risk (VaR) calculation, prediction and backtesting evaluation of an equally weighed portfolio composed of CDS indexes. Preliminary empirical results show that the proposed method is at least as accurate as the calibration procedure suggested by Riskmetrics.

ES12 Room MAL B20 STATISTICAL SIGNAL EXTRACTION AND FILTERING II Chair: DSG Pollock

E315: **Band-limited stochastic processes in discrete and continuous time**

Presenter: **David Stephen Pollock**, University of Leicester, UK

Discrete-time ARMA processes can be placed in a one-to-one correspondence with a set of continuous-time processes that are bounded in frequency by the Nyquist value of π radians per sample period. It is well known that, if data are sampled from a continuous process of which the maximum frequency exceeds the Nyquist value, then there will be a problem of aliasing. However, if the sampling is too rapid, then other problems will arise that will cause the ARMA estimates to be severely biased. The paper reveals the nature of these problems and it shows how they may be overcome. It is argued that the estimation of macroeconomic processes may be compromised by a failure to take account of their limits in frequency

E603: **Calibration of hidden Markov diffusion models: A comparison of direct likelihood maximization and EM**

Presenter: **Franz Konecny**, BOKU - University of Natural Resources and Applied Life Sciences, Vienna, Austria

This study was motivated by working on a conceptual stochastic rainfall-runoff model, formulated in continuous-discrete state space form. In our applications some system parameters are unknown and need to be estimated in a off-line manner, To perform ML parameter estimation, one could try to perform a direct maximization of some particle approximation of the log-likelihood. But in the high-dimensional parameter case it can be difficult to properly scale the parameter increments used to perform direct maximization. A popular alternative method is the expectation-maximization (EM) algorithm. This algorithm is combined with a nonlinear smoother, such as particle smoother or unscented Kalman smoother. A comparative study of these methods, applied to a conceptual rainfall-runoff model, is presented.

E722: **Time-varying autoregressive models: Theory and speech applications**

Presenter: **Patrick Wolfe**, Harvard University, USA

In many speech signal processing tasks and other engineering applications of time series analysis, nonstationary data are analyzed by first windowing the signal of interest to obtain empirically stationary segments, and then by fitting autoregressive models independently to each segment. Here we show how this framework extends naturally to the time-varying autoregressive setting, with the resultant partial correlation coefficients corresponding directly to parameters of acoustic models of the human vocal tract. We then provide related parametric tests for signal stationarity, as well as methods for ensuring that the associated parameter estimates allow for pointwise (i.e., frozen-time) stability of the corresponding signal reconstruction. Finally, we illustrate the developed techniques in the context of several forensic speech applications.

E738: **A stochastic frontier model with time-varying vectorautoregressive inefficiency**

Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy

Co-authors: Ulrich Woitek

An extension to a recent approach to consider time delays in the adjustment process of efficiency is proposed. We allow the dynamic adjustments towards the technological frontier to follow a vector autoregressive model with time varying parameters. Focussing on the two main regional aggregates for Italy - Centre-North and Mezzogiorno - over the observed period 1979-2003, we apply this method to identify the channels through which core and non-core public infrastructures affect regional differences in per capita GDP.

ES13 Room MAL 509 SMALL AREA ESTIMATION Chair: Maria Dolores Ugarte

E167: **Estimation and parameter choice in nonparametric mixed effects models**

Presenter: **Maria Dolores Martinez-Miranda**, University of Granada, Spain

Co-authors: Maria Jose Lombardia, Stefan Sperlich, Wenceslao Gonzalez-Manteiga

It is formulated a nonparametric one-way model which allows to analyze data with special correlation structure as those in small area statistics, among other relevant problems like longitudinal and clustered data. To estimate population parameters as the mean function, the classical approach for small area prediction is based on parametric (linear) mixed models. The flexibility of the nonparametric modeling can play an important role in exploring longitudinal/clustered data, and also it offers new perspectives for various problems typically faced in small area statistics. Among the nonparametric approaches the kernel methods as Local Polynomial Smoothers are intuitive and simple exhibiting nice theoretical and practical properties which make them very popular in a wide range of statistics problems. The generalized estimating equation introduces kernel-weights to take into account only observations in a neighborhood which is controlled by the bandwidth or smoothing parameter, and then a parametric model is assumed only locally. Different ways to introduce the kernel-weights provide different estimators exhibiting different theoretical and practical properties. Several of these estimators are presented and explored, and also the problem of choosing the smoothing parameter is addressed. We propose to use resampling methods to solve the bandwidth choice problem. Bootstrap approximations of the Mean Squared Error provide simple local bandwidth selectors for the nonparametric estimators.

E331: **Model based estimates of households in poverty**

Presenter: **Philip Clarke**, Office for National Statistics, UK

Co-authors: Nargis Rahman, Alan Taylor, Kevin McGrath, Denise Silva

The Office for National Statistics (ONS) is preparing now for the first publication of small area estimates (at middle layer super output area, MSOA) of proportion of households whose equivalised income is under 60% of national median income. Equivalised household income is used to allow for household size and the given threshold is the poverty definition. Small area estimation methods can be used to estimate proportions of a binary variable (in this case of household income under this threshold value) by modelling under a logistic transformation. The data at the household level is transformed into a binary variable; 1 if household in poverty and 0 otherwise. The underlying model is a two level model with $y_id\ Bin(1, p_id)$, $y_id = p_id + e_id$, $\text{logit}(p_id) = X_d + u_j$. Here y_id is the survey variable for household i in the estimation area d, j is the sampling area, p_id is the expected probability of y_id, X_d is a vector of values for area d of a set of covariates, u_j is the area level (sample area) residual assumed to have expectation 0 and variance σ^2u and e_id is the individual within area residual, with expectation 0 and variance σ^2e. This paper discusses the results of modelling these poverty estimates.

E449: **Small area estimation of official crime statistics**

Presenter: **Bart Buelens**, Statistics Netherlands, Netherlands

Co-authors: Jan van den Brakel, Virginie Blaess

Statistics Netherlands conducts an annual crime survey producing statistics on themes as victimization, public safety and satisfaction with the police. The survey as conducted until 2008 is known as the National Safety Monitor (NSM), and reached approximately 750 respondents in each of the 25 police districts. Sufficiently precise statistics can be produced at the national and district levels. In 2008, the survey was redesigned, changing the questionnaires, the modes of data collection and the sampling scheme. In order to quantify discontinuities due to this redesign, the NSM was conducted at reduced scale in parallel with the new survey. The sample size of the reduced NSM is too small to produce precise police district level statistics. Small area estimation methods using random effects models are applied to improve the precision. Model selection measures are utilized to identify suitable covariates from administrative registers and from the police register of reported offenses, as well as from past editions of the full scale NSM survey and from the new crime survey. Of the eight key target variables that are considered, those that are factual variables about crimes and offenses benefit more from the model based approach than variables pertaining to attitudes or opinions.

E479: **Spatial robust small area estimation for business statistics**

Presenter: **Timo Schmid**, University of Trier, Germany

Co-authors: Ralf Muennich

Small area estimation is becoming more and more important in survey sampling due to a growing demand for reliable small area statistics. Application in business statistics in general suffer from very skewed distributions or outliers which violate against classical assumption in small area modelling. Thus, it seems necessary to include robustification methods into these models. Two such robust small area methods are the robust EBLUP estimator for linear mixed models and the robust M-quantile approach. Additionally to outliers, spatial dependencies (e.g. similar industry segments) may often occur. This leads to generalizations of the given methods by spatial modeling which is available for the M-quantile method. The paper gives an overview of the recently used robust small area methods and presents a spatial extension of the robust EBLUP estimator. A comparative study will be performed within a Monte-Carlo study on realistic business data. The performance of the estimators will be shown by different scenarios in the context of business applications.

ES16 Room MAL B29 MATRIX COMPUTATIONS AND STATISTICS Chair: Paolo Foschi

E365: **Solving linear regression problems using LSQR**

Presenter: **David Titley-Peloquin**, University of Oxford, UK

Co-authors: Xiao-Wen Chang, Chris Paige

A linear least squares (LS) problem is often solved in an attempt to recover the parameter vector $\hat{x}$ in a linear model $b = A\hat{x} + v$, where v is a random noise vector following a multivariate normal distribution with mean 0 and covariance matrix $\sigma^2 I$. We consider the convergence of the iterative algorithm LSQR for solving large sparse LS problems, when the data arise from the above linear model. Specifically, we present interesting convergence behaviour of the errors $\|x_k - \hat{x}\|_2$ and $\|x_k - x_{LS}\|_2$, where x_k is the k-th iterate of LSQR, $\hat{x}$ is the parameter vector, and x_{LS} is the true LS solution. Based on this, we argue that one should use stopping criteria specifically designed to recover $\hat{x}$, and we propose such criteria that can be verified efficiently at every iteration of LSQR. We illustrate our results with several numerical examples.

E388: **Exact Cross-Validation for k-NN in binary classification, applications to passive and active learning**

Presenter: **Tristan Mary-Huard**, AgroParisTech / INRA, France

Co-authors: Alain Celisse

A closed form expression of the LpO risk estimator for kNN is derived, leading to a fast and exact computation of LpO in binary classification. It is first used to study the LpO risk minimization strategy for choosing k_p in the passive learning setting. The impact of p on the choice of k and the

LpO estimation of the risk are inferred. In the active learning setting, a procedure is proposed that selects new examples using a LpO committee of kNN classifiers. The influence of p on the choice of new examples and the tuning of k at each step is investigated. The behavior of k_p is shown to be different from what is observed in passive learning.

E422: **Dependent structure detection by blockwise risk**

Presenter: **Dongik Jang**, Seoul National University, Korea (Rok)

Co-authors: Hee-Seok Oh

This study considers the detection of time or spatial dependent structure changes of one-dimensional and two-dimensional scattered data. General approaches for selecting changing points of a time series, which only concern mean changes, are inefficient for finding the dependent structure changes of data, and harmonic analysis based on frequency such as Fourier and wavelets transform have restriction that data should be regular grid, equally space, observation. In this study, we propose detection method of dependent structure changes based on blockwise risk estimation over different region. Since smoothing parameters λ's of smoothing spline regression control the smoothness of function, a small λ is proper in high frequency region, on the other hand, a large λ is adequate in low frequency area. The main idea of proposed changing region detection method is to split the observation domain that have maximum derivative of difference in smoothing parameters of adjacent blocks. The simulation examples show effective results to segment the different area that have significantly different frequency.

E762: **Computational methods for estimating the two-way error-components regression model with heteroscedastic errors**

Presenter: **Paolo Foschi**, University of Bologna, Italy

Co-authors: Petko Yanev, Erricos J. Kontoghiorghes

Numerical strategies for computing the GLS estimator in regression models where the disturbances have a two-way error component structure are considered. The model is treated as a general linear model and the estimation problem is formulated as a generalized linear least squares problem, which is solved by means of a sequence of QR decompositions. This provides a computationally efficient and numerically stable strategy for estimating this panel data model. In the development of the algorithm, the usual homoschedastic assumptions on the individual, time and idiosyncratic effects are relaxed. Thus, contrary to the usual spectral decomposition approaches, the proposed method allows for any kind of heteroscedasticity. Numerical results are presented and the computational efficiencies of the proposed algorithms are analyzed.

ES14 Room MAL 532 MIXTURE MODELS III **Chair: Dankmar Boehning**

E428: **An over-dispersion test for zero-modified count data with covariates**

Presenter: **Fazil Baksh**, University of Reading, UK

Co-authors: Dankmar Boehning

The Poisson distribution is commonly used to model zero-modified count data obtained in a range of disciplines including epidemiology, public health and biology. In particular, capture-recapture studies concerned with estimating size of hidden or difficult to measure populations, such as number of drug users within a region, give rise to zero-truncated count data. Over-dispersion, which occurs when the variance of the observed counts is greater than the mean, is important to detect as it and can lead to biased inference and poor estimation of variability. As a result, a simple yet powerful test for over-dispersion for truncated count data was recently proposed. A key feature is that, unlike existing score and likelihood ratio procedures, this test is more generally applicable as it does not depend on knowledge of the alternative distribution. However, it is well known that over-dispersion is often a result of sampling from different sub-populations. The test is therefore now extended and evaluated for studies with covariate information and shown to be uniformly more powerful than existing test procedures over a range of alternative distributions.

E464: **Population size estimation based upon discrete mixtures of bivariate, conditional independent Poisson distributions**

Presenter: **Rattana Lerdsuwansri**, The University of Reading, UK

Co-authors: Dankmar Boehning

The Lincoln-Petersen approach is a classical capture-recapture model used to estimate the size of an elusive target population. The proposed model is developed to extend the Lincoln-Petersen estimator relying on a two-source situation and binary source/listing variables to non-binary source/listing variables. We consider a bivariate count variable where counts are used to summarize how often a unit was identified from source/list 1 and source/list 2. The mixture model is presented to model unobserved population heterogeneity by assuming independence for a homogeneous component (similar to the axiom of local independence in latent class analysis). The EM algorithm is discussed for maximum likelihood estimation. The model is selected on the basis of Bayesian Information Criterion (BIC). A simulation study was conducted to assess the performance of the proposed estimator. As an application, the number of methamphetamine users in Bangkok in the year 2002 was examined.

E607: **Alternative Bayesian analysis of Markov Chain behavioral effects in capture-recapture experiments.**

Presenter: **Danilo Alunni Fegatelli**, Sapienza Universita di Roma, Italy

Co-authors: Luca Tardella

In the context of capture-recapture experiments we consider a flexible model framework recently proposed which allows to incorporate in the analysis the possible presence of behavioral effects on the capture probabilities of each unit during the sequence of trapping stages. In the original work the conditional likelihood approach has been proposed. However, in the simplest case of behavioural model it is known that with positive probability the maximum (conditional) likelihood estimator (CMLE) yields degenerate estimates for which the the probability of never being captured during the experiment is one and, consequently, the population size estimate is unbounded. We review the pathological output of the CMLE in the more general model framework and derive the conditions under which such pathologies occur. To overcome the problem we investigate alternative Bayesian estimator under different noninformative prior distributions and verify their comparative merits in terms of efficiency and interval estimate coverage with a simulation study. Some extensions of the model framework to incorporate heterogeneity of individual probability through the use of finite mixture distributions will be discussed.

E778: **Extension of Chao's lower bound estimator to include covariate information**

Presenter: **Alberto Vidal-Diez**, University of Reading, UK

Co-authors: Dankmar Boehning

Capture-recapture (CRC) methods aim to estimate the size of populations difficult to target, i.e number of grizzly bears in the Yellowstone ecosystem, number of drug users in Bangkok, etc...Some CRC estimates have been extended to use covariate information in order to model heterogeneity in the probability of capturing a unit. We will present the extension of Chao's lower bound estimate to incorporate covariate information. This estimate is robust because it uses only units captured once or twice, and it can cope with some level of heterogeneity. We will provide confidence intervals and we will compare its performance against other parametric and non parametric estimators by simulation studies. Finally we will present several case studies with real data from multiple fields.

Sunday 12.12.2010 17:20 - 19:00 Parallel Session P – CFE

CS90 Room MAL 151 EMPIRICAL VALIDATION OF AGENT-BASED MODELS Chair: Moritz Mueller

C307: **Growing networks of inventors: Time trends of local and global partner search strategies**
Presenter: **Moritz Mueller**, ETH Zurich, Switzerland
Co-authors: Michael Konig, Christian Pich, Sidonia von Proff

We study the formation of innovation networks in which inventors find potential collaborators through local search strategies (to co-inventors of their current co-inventors) and global search strategies (to any inventor existing in the network). We investigate how long-term trends, such as the advance of information and communication technologies, affect the innovation system by changing the way in which innovators search for collaborators. For this purpose we extend a theoretical network formation model and estimate it on the inventor network constructed from all US patents in the years from 1975 to 2009. Estimates are obtained by the Method of Simulated Moments, which allows us to match theoretical distributions of network statistics with their observed counterparts. In a novel approach, we use Martingale convergence techniques to show the consistency and asymptotic normality of our estimates. Our results show that local search strategies become more prevalent over time, supporting the hypothesis that inventors increasingly rely on local information to find collaboration partners in the presence of an ever-growing complexity of their environment.

CS83 Room MAL B36 MODEL SELECTION AND BAYESIAN ECONOMETRICS Chair: Alessandra Amendola

C808: **Variable selection in industry sector bankruptcy prediction**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Luca Sensini

Business failure is widely studied within corporate finance and different models have been proposed in literature to predict the risk of firm bankruptcy and insolvency. In spite of the large amount of empirical findings, significant issues still remain unsolved. Here, we develop dynamic statistical models for bankruptcy prediction of Italian firms in the limited liability sector using annual balance sheet information. Several issues involved in default risk analysis are investigated, such as the structure of the data-base, the sampling procedure and the selection of financial predictors. In particular, we focus on the variable selection problem and compare the performance of modern selection techniques based on shrinkage with that of traditional variable selection methods. The predictive accuracy of the proposed default risk models has been evaluated at various horizons by means of different accuracy measures. The empirical findings give evidence in favour of the proposed approach.

C569: **A simulation-based Bayes' procedure for robust prediction of trading strategies**
Presenter: **Lukasz Gatarek**, Erasmus University Rotterdam/ Tinbergen Institute, Netherlands
Co-authors: Herman van Dijk, Lennart Hoogerheide

We propose a new simulation method to estimate the cointegration model with nonnormal disturbances in nonparametric Bayesian framework in order to present a robust prediction of some alternative trading strategies. We apply the theory of Dirichlet processes to estimate the distribution of disturbances in form of infinite mixture of normal distributions. The simulation algorithm based on Dirichlet process priors is confronted with the standard method based on singular value decomposition and encompassing prior. We test the methodology with both simulated and true dataset evaluating our technique in context of predictive accuracy. In empirical exercise we apply the method to statistical arbitrage - pairs trading strategy.

C525: **Efficient Bayesian inference for stochastic volatility models**
Presenter: **Gregor Kastner**, Johannes Kepler University, Linz, Austria, Austria
Co-authors: Sylvia Fruhwirth-Schnatter

Bayesian inference for stochastic volatility (SV) models using an efficient MCMC approach is considered. Our method builds on the popular approximation of the log chi-squared distribution by a mixture of normal distributions which allows to sample the latent volatilities simultaneously. We introduce several improvements. First, rather than using standard forward-filtering-backward-sampling to draw the volatilities, we apply a sparse Cholesky factor algorithm to the high-dimensional joint density of all volatilities. This reduces computing time considerably because it allows joint sampling without running a filter. Second, we consider various reparameterizations of the augmented SV model. Under the standard parameterization, augmented MCMC estimation turns out to be inefficient, in particular for the volatility of volatility parameter in the latent state equation. By considering a non-centered version of the SV model, this parameter is moved to the observation equation. Using MCMC estimation for this transformed model reduces the inefficiency factor considerably, particularly, for the volatility of volatility parameter.

C245: **Kalman filter and structural change: An application to time-varying import and export models**
Presenter: **Omorogbe Asemota**, Kyushu University, Japan
Co-authors: Saeki Chikayoshi

Structural change is endemic in Japan's import and export models, yet the conventional econometric method of estimating the models assumed the stability of the model over time. The Kalman filter has been an invaluable tool for estimating the time dependent parameters in a model and for detecting the time of change for many years but only in recent years have advances in computer processing and software allowed the elements of the Kalman filter to be explored in real depth. We first discussed the use of Kalman filter to estimate changing parameters and after showing through the method of rolling regression that the parameters are time dependent, the Kalman filter estimation method is employed to estimate the import and export models of Japan. Applications are also presented showing how the auxiliary residuals can be used to detect and the time of structural breaks and outliers in the model.

CS89 Room MAL B30 CONTRIBUTIONS TO FINANCIAL ECONOMETRICS Chair: Christos Savva

C384: **A summary statistic for cross-sectional dependence in large datasets**
Presenter: **Natalia Bailey**, Queen Mary, University of London, UK
Co-authors: George Kapetanios

Consistent estimation of the variance-covariance matrix of a large dimensional dataset is strongly affected, among other things, by the underlying structure of the data in question. In this paper we focus our attention on the identification of the strength of such a structure, when it exists. We claim that all information needed for this purpose is contained in the column-sum norm of the variance-covariance matrix. This approaches infinity at rate N^a, where $0 < a < 1$. The strength of structure can then be determined by the value of a; a low value (a tends to 0) is indicative of weak

interdependence between the elements of the dataset while a large value (a tends to 1) corresponds to a strong inter-relationship. On this basis, we construct a summary statistic as a means of quantifying this parameter a. Our estimator is consistent for relatively small N/T ratios. For estimation purposes we favour non-linear optimisation. The accuracy of our statistic is checked by use of the non-parametric bootstrap approach. An application to S&P500 stock market data, in the context of building our Sectoral Concentration Indicator (SCI), is also incorporated.

C457: **Locally stationary diffusion processes with structural breaks**

Presenter: **Bonsoo Koo**, London School of Economics and Political Science, UK

A model of nonstationary diffusion processes with breaks in the drift or diffusion function is proposed. The standard diffusion process is not consistent with either regime shifts or jumps in many financial time series. By incorporating structural breaks, our model is more consistent with actual stylized facts of financial data. We propose estimators of the location and size of structural breaks in the drift or the diffusion function of a class of locally stationary diffusion processes, where the structural breaks are located at given periods of time. We establish asymptotic theory for the proposed estimators under the long span assumption on the time horizon; we do not require infill asymptotics. We investigate the impacts of major economic events such as Black Wednesday on the magnitude of the change in level and volatility of daily US dollar/UK sterling exchange rates. In addition, we analyse daily short-term interest rates where we focus on the implications of such events for option pricing. Finally, we evaluate the finite-sample performance of our estimators via Monte Carlo experiments.

C235: **Testing for co-jumps in a panel of high frequency financial data: an extreme-value based approach**

Presenter: **Yin Liao**, Australian National University, Australia

Co-authors: Heather Anderson

This paper proposes a new approach to test for common intraday jumps in a panel of high frequency financial data. We utilize intraday first-high-low-last price extreme values to construct a novel estimator for the cross variance structure of a large panel of high frequency financial data, and then provide a extreme-value based test statistic to determine if this covariance is partially attributable to common large discrete movements (co-jumps). We study the finite sample behavior of our extreme-value based test using Monte Carlo simulation, and find that it is more powerful than the existing return-based co-jump test for volatility measured at the same sampling frequency. When applied to a panel of high frequency data from the Chinese mainland stock market, our extreme-value based test identifies more common jumps than the return-based test in this emerging market.

C869: **A factor approach to realized volatility and market microstructure noise**

Presenter: **Alev Atak**, Queen Mary University of London, UK

The aim is to consider high dimensional multivariate realized volatility models for large dimensional datasets and also address the solution for noise problem coming out of volatility estimation in the presence of market microstructure effects. Standard models, where prices are contaminated with stochastically independent noise, are unable to explain the behavior of realized covariance as the sampling frequency increases. We aim to extend the current analytic methods to the construction and assessment of volatility forecasts for continuous-time volatility models to the empirically important case of market microstructure noise via factors and principal component methodology. The main contribution of this paper is to propose a novel way of conducting realized volatility, where integrated volatility takes a linear factor structure, facilitating the estimation of volatility factors while getting rid of the noise. These factors capture the market microstructure problem when applied to a large dimension of individual return series in a stock market. Monte Carlo study illustrates the finite sample properties of the proposed method and we provide the application to S&P data.

CS86 Room MAL B29 CONTRIBUTIONS TO FINANCE II Chair: Marc Paolella

C745: **Measuring misspecification bias in term structure models of commodity prices**

Presenter: **Hiroaki Suenaga**, Curtin University, Australia

Stochastic dynamics of commodity prices and pricing of financial contracts have long been studied in the field of financial economics. A common approach is to specify the stochastic dynamics of the underlying assets and derive from the suggested model valuation formulas of various derivative contracts. This conventional approach of modelling term structure of commodity prices, however, is intrinsically subject to approximation bias, because, as shown by the theory of storage, the equilibrium path of the spot and futures prices cannot be expressed in reduced form for a commodity with significant storage cost and strong seasonality. This study compares conventional term-structure models with an alternative modelling approach in which the variance of futures returns, rather than futures prices, is specified directly by flexible non-parametric functions so that the model is capable of replicating non-linear price dynamics of storable commodities with strong seasonality. Empirical applications of the model to three commodities (gold, natural gas, corn) illustrate that the conventional term-structure models are subject to misspecification bias of considerable size and lead to hedging strategies that are much less effective than the strategy based on the suggested model of futures returns.

C509: **The operation of hedge funds and regulatory reforms**

Presenter: **Raphaele Chappe**, New School For Social Research, USA

Co-authors: Willi Semmler

Ponzi finance describes situations where operating cash flow is insufficient to cover either principal or interest payments. This paper shows a model that can be used to describe situations in asset management where the use of leverage to generate above-average returns can result in some hedge funds finding themselves, willingly or unwillingly, embroiled in a Ponzi financing scheme. To explain the income and wealth process we use, as stochastic processes, Brownian motions: a process for wealth, a process of mean reverting asset returns, and an interest rate process. The model starts with investors turning over funds to a hedge manager and paying a fee defined as a percentage of the invested wealth. We run a simulation of the impact of the sudden decline of inflow of funds on the hedge funds, ultimately leading to insolvency. The hedge fund industry has been opaque, with little regulatory oversight, creating opportunities for such Ponzi financing, as we saw recently with the Madoff scandal. Greater transparency is necessary on the part of hedge funds to even begin to understand systemic risk implications. In this context, we assess the new regulatory framework enacted by the Obama administration, and the proposals by the European Commission.

C361: **Financial crisis and chaos control**

Presenter: **Ali Sanayei**, Young Researchers Club, Qazvin Islamic Azad University, Iran

Co-authors: Fraydoon Rahnamay Roodposhti, Taghi Torabi

Most mathematical models of economics are nonlinear and sensitive to social and political conditions. In addition, nonlinearity and sensitivity to initial conditions lead us to deal with a chaotic behavior. In the chaotic growth models, the economy follows nonlinear dynamics which are self-generating and never die down. Interestingly, chaotic behaviors could explain fluctuations in the economy and financial markets which appear to be random and unpredictable. Meanwhile, financial crisis could be interpreted as a chaotic behavior and controlling chaos therefore plays an important role to manage crisis. In this work, we focus our attention on a mathematical model of a self-developing market economy whose development is

characterized by spontaneous growth of capital and its movement in the technology space in response to differences in profitability. This model is a nonlinear system of PDEs that describes the formation of social wealth and it could exhibit crisis as a chaotic attractor. Consequently, the main aims of this work are controlling chaos and finding a practical method to stabilize the chaotic system. We will also find the best external adjustable control parameter to control the chaos and propose some solutions to adjust this parameter in order to manage crisis.

C726: **Exchange rate exposure of UK nonfinancial companies : a quantile regression approach**

Presenter: **Dilek Ulu**, University of Essex, UK

We apply a quantile regression approach to examine the relation between exchange rates and stock returns for a large sample of non-financial UK firms over the period 2000 to 2010. Contrary to the traditional methods used in previous research, quantile regression is more informative especially in the presence of outliers, non-normal error terms and different levels of exposure to exchange rate risks in the lower and upper tails of the distribution. To cope with the potential heteroscedasticity problem, the variance covariance matrix is estimated using the design matrix bootstrap technique which is robust under nonstandard distributional settings. Using both trade-weighted and bilateral exchange rates, the empirical results indicate that the quantile exposure coefficient is highly significant in both tails with negative coefficient for overperforming and positive coefficient for underperforming firms and is less significant/ insignificant around the median. These are in sharp contrast to the insignificant OLS estimates. Furthermore, the level of exposure has a progressive pattern as we move from median to both tails of the conditional distribution. The more firms perform better or worse than average, the more exchange rate exposure they have.

Sunday 12.12.2010 17:20 - 19:00 Parallel Session P – ERCIM

ES36 Room MAL B33 RECENT ADVANCES IN SMALL AREA ESTIMATION Chair: Isabel Molina

E338: **Two area-level time-space models for small area estimation**

Presenter: **Domingo Morales**, Universidad Miguel Hernandez de Elche, Spain

Co-authors: Yolanda Marhuenda, Isabel Molina

Time and space correlations are sources of information that might improve the quality of EBLUP estimates in small area estimation problems. Two area-level time-space models are introduced together with the corresponding algorithms to calculate REML estimates of model parameters and the EBLUP estimates of population parameters. Mean squared error of the EBLUPs are estimated by a parametric bootstrap method. Some simulations are carried out to investigate the proposed methodology. An application to the estimation of poverty indicators with data from the Spanish Living Conditions survey is given.

E485: **Performance of small area estimators under selected sampling designs**

Presenter: **Ralf Muennich**, University of Trier, Germany

Co-authors: Jan Pablo Burgard

In the European year of combating poverty and social exclusion, special attention is paid on poverty measurement and the European Laeken indicators. In recent years there was a rising interest in comparing estimates on regional levels rather than national level. Applying classical estimation methods on regional level, in general, does not yield reliable figures. In order to provide such estimates small area estimation methods may be applied. A data source in Europe which plays an important role in providing the information for estimating poverty indicators or measures for social exclusion is the European Survey on Living and Income Conditions (EU-SILC). The sampling designs vary from country to country and include sophisticated multi-stage designs. The paper aims at investigating the impact of the survey design on small area estimates of some selected Laeken Indicators and related measures. Special attention will be paid on estimating in the presence of highly skewed income variables. The study will be performed by a large-scale Monte-Carlo simulation based on realistic data with different survey designs. The estimators of interest cover design-based, EBLUP, and Pseudo-EBLUP methods.

E766: **Estimation of poverty indicators for domains with unit-level auxiliary information under unequal probability sampling**

Presenter: **Risto Lehtonen**, University of Helsinki, Finland

Co-authors: Ari Veijanen

Recent developments for the estimation of poverty indicators for population subgroups or domains and small areas are presented. The indicators include at-risk-of poverty rate, relative median at-risk-of poverty gap, quintile share ratio and the Gini coefficient. We concentrate on poverty gap for domains. Direct estimators do not use auxiliary information. Various synthetic estimators that use auxiliary data are introduced. Composite estimators are linear combinations of direct and indirect estimators. A design-unbiased direct estimator can be inefficient, and more efficient indirect synthetic estimators are often biased. Composite estimators offer a compromise for many situations. We also discuss the relative performance of the estimators under outlier contamination. Linear mixed models with area-specific random terms are fitted for the income variable underlying the indirect estimators. Unit-level auxiliary data are incorporated into the estimation procedure. We discuss the case of unequal probability sampling. By using quality measures (bias and accuracy), the relative performance of the estimators is assessed with Monte Carlo simulation experiments based on data from statistical registers of Statistics Finland. Research has been conducted in the framework of the AMELI project (Advanced Methodology for European Laeken Indicators) and is supported by European Commission funding from the Seventh Framework Programme for Research.

E814: **Non-parametric bootstrap mean squared error estimation for small area means, quantiles and poverty indicators**

Presenter: **Monica Pratesi**, University of Pisa, Italy

Co-authors: Stefano Marchetti, Nikos Tzavidis

Small area estimation is traditionally concerned with the estimation of small area averages and totals. More recently emphasis has been also placed on the estimation of the quantiles of the small area distribution function. Examples include the estimation of key percentiles of the income distribution and the percentage of individuals/households below a threshold in a small area. In parallel to point estimation, Mean Squared Error (MSE) estimation is an equally crucial and challenging task. Moreover, while MSE estimation for small area averages can be performed analytically, analytic MSE estimation for small area distributional estimates is complex. We focus on the use of the M-quantile small area model and in particular, in employing this model for estimating three types of small area quantities namely, averages, quantiles and poverty indicators. We present a non-parametric bootstrap MSE estimator that can be used for estimating the MSE of the above mentioned small area target parameters. Our approach is based on extending a bootstrap quantiles estimator to the small area estimation context. The proposed MSE estimator is evaluated in a series of simulation. We present results from the application of the proposed MSE estimator to real income data from the EU-SILC in Italy.

ES44 Room MAL B20 MEASURING FORECAST AND EXPECTATION Chair: Anna Staszewska

E195: **Heuristic model selection for leading indicators in Russia and Germany**

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

Co-authors: Peter Winker

Business tendency survey indicators are widely recognized as a key instrument for business cycle forecasting. The leading indicator property of business expectations is assessed with regard to forecasting industrial production in Russia and Germany. For this purpose, vector autoregressive (VAR) models are specified based on information criteria and estimated to construct forecasts. As the potential number of lags included is large, we compare full-specified VAR models with subset models obtained using a Genetic Algorithm enabling "holes" in multivariate lag structures. The problem is complicated by the fact that a structural break and seasonal variation of indicators have to be taken into account. These models allow for a comparison of the dynamic adjustment and the explanatory power of the leading indicator for both countries. The results exhibit marked differences in dynamics and forecasting performance between Russia and Germany.

E283: **Constructing optimal path-wise bootstrap prediction bands with threshold accepting**

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

Co-authors: Peter Winker

Typically, prediction bands for path-forecasts are constructed point-wise, while inference relates to the whole forecasted path. In general, no closed form analytical solution is available for path-wise bands in finite samples. We consider a direct construction approach based on bootstrapped pre-

diction bands. The resulting highly complex optimization problem is tackled using the local search heuristic threshold accepting with an efficient local updating procedure. Simulation results demonstrate that the method delivers the global optimum for small scaled problems, when the global optimum can be obtained by full enumeration. For larger synthetic problem instances, a comparison with point-wise and asymptotic bands is provided. Finally, a real application demonstrates the practical implications of using an appropriate tool for generating the prediction bands.

E585: **The role of forecasts in interest rate pass-through**
Presenter: **Victor Bystrov**, University of Lodz, Poland
Co-authors: Anindya Banerjee, Paul Mizen

Much of the literature on interest rate pass through assumes banks set retail rates by observing current market rates, but we argue that banks use expectations about future rates to inform retail rate setting. If expectations - captured by forecasts of future rates - are important, the empirical specifications of many previous studies that omit them could be misspecified. Including forecasts requires deliberation of a large array of information and alternative forecasting models. In this paper we use large datasets and evaluate alternative forecasts before including them in a model of retail rate adjustment in five European countries. We find a significant role for forecasts of future interest rates in determining interest rate pass-through.

E671: **Policy interactions in a monetary union: an application of the OPTGAME algorithm**
Presenter: **Dmitri Blueschke**, Klagenfurt University, Austria
Co-authors: Viktoria Blueschke-Nikolaeva, Reinhard Neck

We present a small stylized nonlinear multi-country macroeconometric model of the Euro Area economy for analysing the interactions between fiscal (governments) and monetary (European Central Bank) policy makers in the European Economic and Monetary Union, assuming different objective functions of these decision makers. Several dynamic game experiments are run for different information patterns and solution concepts. The OPTGAME algorithm is used to calculate the solutions; it calculates approximate cooperative Pareto-optimal solutions and noncooperative Nash and Stackelberg equilibrium solutions under both open-loop and feedback information patterns. We determine possible advantages of commitment and of cooperation in terms of the minimal values of the objective functions of the players.

ES15 Room MAL B34 MIXTURE MODELS IV Chair: Fazil Baksh

E556: **Semi-supervised variable selection for model-based clustering**
Presenter: **Jeffrey Andrews**, University of Guelph, Canada
Co-authors: Paul McNicholas

A new and versatile variable selection technique based upon within-group variance is introduced as a fast alternative to more established methods. Developed with high-dimensional data in mind, this technique is versatile in the sense that it can be implemented with both classification and clustering paradigms, and can even be used as a catalyst for more complicated variable selection procedures. After the method is introduced, semi-supervised variable selection (SSVS) will be illustrated in a model-based clustering framework using the MCLUST and PGMM families of Gaussian mixture models. The performance of SVSS will be compared with that of the popular clustvarsel package for the R software. The use of SSVS as a catalyst for the clustvarsel algorithm will also be illustrated.

E608: **Conditional exact tests for three classes of mixture models for categorical data.**
Presenter: **Davide Di Cecco**, University of Bologna, Italy

We propose conditional exact tests based on sufficient statistics to compare three classes of mixture models: mixtures of i.i.d. sequences, mixtures of Markov chains, and of reversible Markov chains. We just consider sequences of categorical data. Mixtures of i.i.d. sequences are characterized in terms of exchangeability as those processes for which the number of occurrences of each state is a sufficient statistic. Mixtures of Markov chains and mixtures of reversible Markov chains have been characterized in terms of partial exchangeability as the classes of recurrent processes for which the number of directed, and undirected respectively, transition counts are a sufficient statistic. Since we deal with discrete data, constructing the exact conditional distribution of the sufficient statistics is a matter of combinatorics, and it is essentially related to two problems of computational graph theory: find out the Eulerian Orientations of an undirected graph, and find out all the graphs consistent with given degrees of the vertices (degree sequence problem). Even if these graph problems are intractable in the general case, we will show that, when the state space is not too large, they are manageable. As particular cases we can test Markovianity and reversibility of a single sequence.

E776: **Hidden Markov Model averaging for binary classification in a variational Bayesian framework**
Presenter: **Stevenn Volant**, AgroParisTech, France
Co-authors: Marie-Laure Martin-Magniette, Stephane Robin

We consider a mixture of two populations, one has a known distribution (class of interest) and the other an unknown one. Several models have been developed to fit this unknown distribution. We propose an alternative based on a mixture of Gaussian distributions. The inference is done within a Bayesian framework and our aim is to infer the posterior probability of belonging to the class of interest. To this end, it makes no sense to estimate the mixture component number since each mixture model provides more or less relevant information to the posterior probability estimation. By computing a weighted average (named aggregated estimator) over the model collection, Bayesian Model Averaging (BMA) is one way of combining models in order to account for information provided by each model. The aim is then the estimation of the weights and the posterior probability for one specific model. In this work, we derive approximations of these two quantities from the variational theory. To perform our method, we consider that the data are dependent and hence we consider a Hidden Markov Model. A simulation study is carried out to evaluate the accuracy of the estimators in term of classification.

E572: **An entropy criterion in mixtures of linear regressions**
Presenter: **Susana Faria**, University of Minho, Portugal
Co-authors: Gilda Soromenho

Finite mixture models are a well-known method for modelling data that arise from a heterogeneous population. Entropy-type measures for the heterogeneity of data have been used for a long time. In a mixture model context, entropy criterions can be used to measure the overlapping of the mixture components. In this paper we study an entropy-based criterion in mixtures of linear regressions to measure the closeness between the mixture components. We show how an entropy criterion can be derived based on the Kullback-Leiber distance, which is a measure of distance between probability distributions. To investigate the effectiveness of the proposed criterion, a simulation study was performed.

Authors Index

Author index