PROGRAMME AND ABSTRACTS

5th CSDA International Conference on

Computational and Financial Econometrics (CFE 2011)

http://www.cfe-csda.org/cfe11

and

4th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computing & Statistics (ERCIM 2011)

http://www.cfe-csda.org/ercim11

Senate House, University of London, UK
17-19 December 2011

ERCIM
European Research Consortium
for Informatics and Mathematics

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

Birkbeck UNIVERSITY OF LONDON
http://www.bbk.ac.uk

LSE THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE
http://www.lse.ac.uk

Dear Friends and Colleagues,

We warmly welcome you to London, for the Fifth International Conference on *Computational and Financial Econometrics* (CFE 2011) and the Fourth International Conference of the ERCIM Working Group on *Computing & Statistics* (ERCIM 2011). As many of you know, this annual conference has been established as a leading joint international meeting for the interface of computing, empirical finance, econometrics and statistics, and it is endorsed by the journal of Computational Statistics & Data Analysis (CSDA).

The conference aims at bringing together researchers and practitioners to discuss recent developments in computational methods for economics, finance, and statistics. The CFE-ERCIM 2011 programme consists of over 175 sessions, 5 plenary talks and over 800 presentations. There are over 900 participants. Peer reviewed papers will be considered for publication in special issues of the journal Computational Statistics & Data Analysis and the CSDA Annals of Computational and Financial Econometrics.

The co-chairs have endeavoured to provide a balanced and stimulating programme that will appeal to the diverse interests of the participants. The international organizing committee hopes that the conference venue will provide the appropriate environment to enhance your contacts and to establish new ones. The conference is a collective effort by many individuals and organizations. The Scientific Programme Committee, the Session Organizers, the local hosting 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-ERCIM 2012 will be held at the impressive new conference centre of the historic city of Oviedo (Spain) on 1-3 December 2012. 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-ERCIM 2011 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 organize sessions and workshops during the annual WG meetings.

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

Specialized groups

Currently the ERCIM WG has approximately 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 by email at ercim@cfe-csda.org.

SCHEDULE

CFE 2011

Saturday, 17th December 2011

08:50 - 09:00	Opening (Beveridge Hall)
09:00 - 09:50	Plenary Session A (Beveridge Hall)
09:50 - 10:25	Coffee Break
10:25 - 12:30	Parallel Sessions C
12:30 - 14:00	Lunch Break
14:00 - 16:05	Parallel Sessions E
16:05 - 16:35	Coffee Break
16:35 - 18:40	Parallel Sessions F
20:00 - 21:30	Reception (Beveridge Hall)

Sunday, 18th December 2011

09:45 - 10:35	Plenary Session H (Beveridge Hall)
10:35 - 11:05	Coffee Break
11:05 - 12:45	Parallel Sessions I
12:45 - 14:15	Lunch Break
14:15 - 16:20	Parallel Sessions J
16:20 - 16:50	Coffee Break
16:50 - 18:30	Parallel Sessions K
20:30 - 23:30	Conference Dinner (Beveridge Hall)

Monday, 19th December 2011

09:05 - 10:25	Parallel Sessions L
10:25 - 10:55	Coffee Break
10:55 - 12:35	Parallel Sessions N
12:35 - 14:05	Lunch Break
14:05 - 15:45	Parallel Sessions O
15:45 - 16:15	Coffee Break
16:15 - 17:35	Parallel Sessions P
17:45 - 18:35	Plenary Session Q (Beveridge Hall)
18:35 - 18:40	Closing (Beveridge Hall)

ERCIM 2011

Saturday, 17th December 2011

09:45 - 10:05	Coffee Break
10:05 - 10:15	Opening (Beveridge Hall)
10:15 - 11:05	Plenary Session B (Beveridge Hall)
11:15 - 12:30	Parallel Sessions D
12:30 - 14:00	Lunch Break
14:00 - 16:05	Parallel Sessions E
16:05 - 16:35	Coffee Break
16:35 - 18:40	Parallel Sessions F
20:00 - 21:30	Reception (Beveridge Hall)

Sunday, 18th December 2011

08:55 - 10:35	Parallel Sessions G
10:35 - 11:05	Coffee Break
11:05 - 12:45	Parallel Sessions I
12:45 - 14:15	Lunch Break
14:15 - 16:20	Parallel Sessions J
16:20 - 16:50	Coffee Break
16:50 - 18:30	Parallel Sessions K
20:30 - 23:30	Conference Dinner (Beveridge Hall)

Monday, 19th December 2011

09:35 - 10:25	Plenary Session M (Beveridge Hall)
10:25 - 10:55	Coffee Break
10:55 - 12:35	Parallel Sessions N
12:35 - 14:05	Lunch Break
14:05 - 15:45	Parallel Sessions O
15:45 - 16:15	Coffee Break
16:15 - 17:35	Parallel Sessions P
17:45 - 18:35	Plenary Session Q (Beveridge Hall)
18:35 - 18:40	Closing (Beveridge Hall)

MEETINGS AND SOCIAL EVENTS

SPECIAL MEETINGS by invitation to group members

- CSDA Editorial Board meeting, *Jessel Room, Senate House*, Friday 16th of December 2011, 18:00-19:00.
- CSDA Editorial Board dinner, Friday 16th of December 2011, 20:00-22:30.
- ERS-IASC BoDs meeting, *Senate Room, Senate House*, Saturday 17th of December 2011, 13:00 - 14:00.

SOCIAL EVENTS

- *The coffee breaks* will take place at the McMillan Hall and Grand Lobby of the Senate House. In addition, cold drinks will be provided for the speakers at MAL B04, Birkbeck.
- *Welcome Reception, Saturday 17th 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 (McMillan, Beveridge, Crush and Chancellor's Hall). Conference registrants must bring their conference badge and any accompanying persons should bring their reception tickets in order to attend the reception.
- *Conference Dinner, Sunday 18th of December, 20:30*. The Conference Dinner will take place at the *Senate House, Beveridge Hall*. The conference dinner is optional and registration is required. You must have your Conference Dinner ticket in order to attend the conference dinner.

Addresses of venues:

- University of London, Senate House, Malet Street, London WC1E 7HU.
- Birkbeck, University of London, Malet Street, London WC1E 7HX.

Registration, exhibitors and networking activities

The registration will be located in the MacMillan Hall of the Senate House. The exhibitors will be located in the Chancellor's Hall of the Senate House. The Chancellor's Hall will be used for networking and internet access. The Room MAL B04 at Birkbeck is available for networking.

Lecture rooms

The paper presentations will take place at the Senate House and at the main building of Birkbeck (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 Beveridge Hall of the Senate House. 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 at the Beveridge Hall of the Senate House. The poster sessions will take place at the Chancellor's Hall of the Senate House.

Room	Capacity	Floor	Location	Room	Capacity	Floor	Location
Gordon	40	Ground	Senate House	Bloomsbury	50	Ground	Senate House
Bedford	50	Ground	Senate House	Beveridge Hall	500	Ground	Senate House
Woburn	70	Ground	Senate House	Torrington	50	First	Senate House
Court	70	First	Senate House	Jessel	50	First	Senate House
Senate	80	First	Senate House	S261	50	Second	Senate House
S264	50	Second	Senate House	MAL B18	44	Basement	Birkbeck Malet St
MAL B20	80	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

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. Chairs are requested to keep the sessions on schedule. Papers should be presented in the order they are listed in the programme for the convenience of attendees who may wish to go to other rooms mid-session to hear particular papers. In the case of a presenter not attending, please use the extra time for a break or a discussion so that the remaining papers stay on schedule. The PC in the lecture rooms should be used for presentations. *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. IT technicians will be available during the conference and should be contacted in case of problems. The posters should be displayed only during their assigned session. The authors will be responsible for placing the posters in the poster panel displays and removing them after the session. The maximum size of the poster is A0.

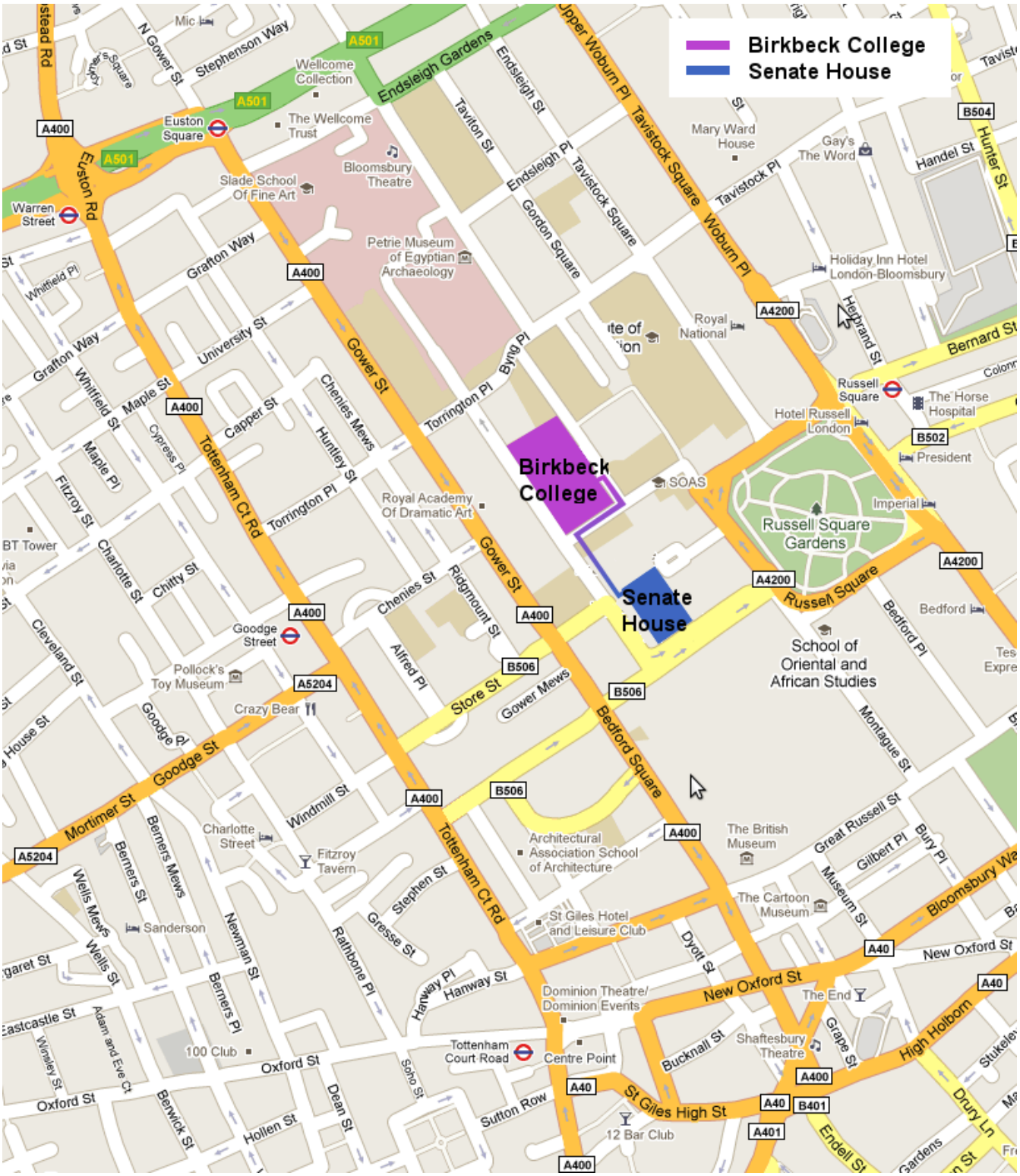
Internet

There will be wireless Internet connection in the Chancellor's Hall. You will need to have your own laptop 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 MacMillan 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 2011-2012 on the following topics:
  - Bayesian Computing, Methods and Applications.
  - Statistical Algorithms and Software in R.
  - Imprecision in Statistical Data Analysis.
  - Model Selection and High Dimensional Data Reduction.
  - Advances in Mixture Models.
  - Design Algorithms of Experiments.
  - Robust Analysis of Complex Data.

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)

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)

Cambridge University Press (URL http://www.cambridge.org/)

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

General Information **I**

Committees . . . II
Welcome . . . III
ERCIM Working Group on Computing & Statistics . . . IV
Scientific and Social Programme Schedule, Meetings and Social Events Information . . . V
Venue, lecture rooms, presentation instructions and internet access . . . VI
Map of the venue and nearby area . . . VII
Publications outlets, sponsors, exhibitors and supporting societies . . . VIII

Keynote Talks **1**

Keynote talk CFE 1 (Andrew C. Harvey, University of Cambridge, United Kingdom) Saturday 17.12.2011 at 09:00-09:50
Dynamic models for volatility and heavy tails . . . 1
Keynote talk ERCIM 1 (Alastair Young, Imperial College London, United Kingdom) Saturday 17.12.2011 at 10:15-11:05
Achieving accuracy and correctness in parametric inference . . . 1
Keynote talk CFE 2 (Dick van Dijk, Erasmus University Rotterdam, Netherlands) Sunday 18.12.2011 at 09:45-10:35
Comparing the accuracy of copula-based multivariate density forecasts in selected regions of support . . . 1
Keynote talk ERCIM 2 (Stephen Walker, University of Kent, United Kingdom) Monday 19.12.2011 at 09:35-10:25
Bayesian analysis for a stationary and transition density function . . . 1
Keynote talk CFE-ERCIM (Peter M. Robinson, London School of Economics, United Kingdom) Monday 19.12.2011 at 18:05-18:55
Parametric inference on strong dependence . . . 1

Parallel Sessions **2**

Parallel Session C – CFE (Saturday 17.12.2011 at 10:25 - 12:30) **2**

CSI01: TIME SERIES ECONOMETRICS (Room: Woburn) . . . 2
CS08: ENERGY MARKETS, CLIMATE CHANGE AND WEATHER DERIVATIVES (Room: Torrington) . . . 2
CS12: BAYESIAN NONLINEAR ECONOMETRICS (Room: Senate) . . . 3
CS17: FORECASTING FINANCIAL MARKETS (Room: Bedford) . . . 3
CS32: BEHAVIOURAL FINANCE (Room: S264) . . . 4
CS34: QUANTITATIVE RISK MANAGEMENT I (Room: Jessel) . . . 5
CS52: VAST DIMENSIONAL FINANCIAL ECONOMETRICS (Room: Bloomsbury) . . . 6
CS68: RECENT ADVANCES IN BOND PRICING (Room: Court) . . . 6
CS56: COMPUTATIONAL METHODS IN APPLIED ECONOMETRICS (Room: Gordon) . . . 7
CP02: POSTER SESSION II (Room: Chancellor's) . . . 8
CS95: FINANCIAL ECONOMETRICS I (Room: S261) . . . 9
CP01: POSTERS SESSION I (Room: Chancellor's) . . . 10

Parallel Session D – ERCIM (Saturday 17.12.2011 at 11:15 - 12:30) **12**

ES06: SMALL AREA ESTIMATION (Room: B34) . . . 12
ES20: APPLIED STATISTICS I (Room: B33) . . . 12
ES39: HIGH DIMENSIONAL DESIGN OF EXPERIMENTS (Room: B20) . . . 13
ES62: IMPRECISION IN INFERENCE (Room: B35) . . . 13
ES77: BIOSTATISTICS II (Room: G16) . . . 13
ES11: STATISTICAL MONITORING AND ITS APPLICATIONS II (Room: B18) . . . 14

Parallel Session E – ERCIM (Saturday 17.12.2011 at 14:00 - 16:05) **15**

ES09: PERSPECTIVES ON HIGH-DIMENSIONAL DATA ANALYSIS (Room: B18) . . . 15
ES15: NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT II (Room: B35) . . . 15
ES29: ADVANCES IN OPTIMAL EXPERIMENTAL DESIGN (Room: B33) . . . 16
ES31: APPLIED STATISTICS II (Room: G16) . . . 17
ES38: STATISTICS IN FUNCTIONAL AND HILBERT SPACES I (Room: B34) . . . 18
ES63: HANDLING IMPRECISION IN GRAPHICAL MODELS (Room: B20) . . . 18
ES70: STATISTICS FOR RANDOM INTERVALS AND RANDOM SETS (Room: B36) . . . 19
EP02: POSTER SESSION II (Room: Chancellor's) . . . 20

EP01: POSTERS SESSION I (Room: Chancellor's) . . . 21

Parallel Session E – CFE (Saturday 17.12.2011 at 14:00 - 16:05) 24

CSI03: RECENT DEVELOPMENTS IN ECONOMETRICS (Room: Beveridge) . . . 24
CS04: APPLIED FINANCIAL ECONOMETRICS (Room: Senate) . . . 24
CS11: MODELLING WITH HEAVY TAILS: COMPUTATIONAL ISSUES (Room: Woburn) . . . 25
CS19: LONG TERM RISKS (Room: Court) . . . 26
CS37: TRENDS, WAVES, SEASONS, CYCLES AND SIGNALS (Room: Bloomsbury) . . . 26
CS39: REAL-TIME DENSITY FORECASTING (Room: Jessel) . . . 27
CS54: TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS (Room: Torrington) . . . 28
CS93: ECONOMETRIC MODELLING AND APPLICATIONS I (Room: S264) . . . 28
CS25: CONTRIBUTIONS TO HIGH FREQUENCY DATA MODELING (Room: Gordon) . . . 29
CS15: CONTRIBUTIONS IN MODELLING AND FORECASTING FINANCIAL RISK (Room: S261) . . . 30
CS16: QUANTITATIVE RISK MANAGEMENT II (Room: Bedford) . . . 31

Parallel Session F – ERCIM (Saturday 17.12.2011 at 16:35 - 18:40) 32

ES10: CLASSIFICATION AND DISCRIMINANT PROCEDURES FOR DEPENDENT DATA (Room: S261) . . . 32
ES14: NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT I (Room: Bloomsbury) . . . 32
ES17: OUTLIERS AND CHANGE-POINTS IN TIME SERIES I (Room: Senate) . . . 33
ES18: EXTREME VALUE THEORY AND APPLICATIONS (Room: Gordon) . . . 34
ES25: MODEL VALIDATION (Room: Jessel) . . . 34
ES41: STATISTICS IN FUNCTIONAL AND HILBERT SPACES II (Room: Court) . . . 35
ES43: ORDER-RESTRICTED INFERENCE AND APPLICATIONS (Room: Torrington) . . . 36
ES37: BAYESIAN MODELLING AND COMPUTATION (Room: Woburn) . . . 37
ES58: STATISTICAL MONITORING AND ITS APPLICATIONS I (Room: S264) . . . 37
ES69: IMPRECISE PROBABILISTIC MODELING TO SOLVE STATISTICAL PROBLEMS I (Room: Bedford) . . . 38

Parallel Session F – CFE (Saturday 17.12.2011 at 16:35 - 18:40) 40

CS97: FINANCIAL ECONOMETRICS III (Room: B18) . . . 40
CS26: BAYESIAN EMPIRICAL MACROECONOMICS (Room: B34) . . . 40
CS28: PROBABILISTIC FORECASTING (Room: B35) . . . 41
CS35: MODELING AND INFERENCE ON ASSET PRICE BUBBLES (Room: B36) . . . 42
CS40: DYNAMIC MODELLING OF REALIZED COVARIANCE MATRICES (Room: B33) . . . 42
CS55: CONTRIBUTIONS IN FINANCIAL MARKET AND THE MACROECONOMY (Room: G16) . . . 43
CS76: CONTRIBUTIONS IN TIME SERIES ECONOMETRICS I (Room: B20) . . . 44

Parallel Session G – ERCIM (Sunday 18.12.2011 at 08:55 - 10:35) 46

ES02: STATISTICAL ALGORITHMS AND SOFTWARE I (Room: Gordon) . . . 46
ES16: STATISTICAL MACHINE LEARNING AND ROBUSTNESS (Room: Bloomsbury) . . . 46
ES19: ADVANCES IN ROBUST DATA ANALYSIS (Room: Woburn) . . . 47
ES21: COPULA MODELLING AND COMPUTATIONAL ANALYSIS (Room: S261) . . . 47
ES27: SEMIPARAMETRIC QUANTILE AND EXPECTILE REGRESSION (Room: Jessel) . . . 48
ES52: MIXTURE MODELS IN R (Room: Torrington) . . . 49
ES53: FREQUENCY DOMAIN ANALYSIS (Room: Senate) . . . 49
ES81: TIME SERIES MODELING AND COMPUTATION II (Room: S264) . . . 50
ES76: DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS II (Room: Bedford) . . . 50
ES65: SYMBOLIC DATA ANALYSIS (Room: Court) . . . 51

Parallel Session I – CFE (Sunday 18.12.2011 at 11:05 - 12:45) 53

CSI02: BAYESIAN ECONOMETRICS (Room: Beveridge) . . . 53
CS03: LARGE SWINGS IN MACROECONOMIC TIME SERIES (Room: B34) . . . 53
CS14: EMPIRICAL MODELLING OF FINANCIAL FRAGILITY (Room: B33) . . . 54
CS45: UNIVARIATE AND MULTIVARIATE VOLATILITY MODELS (Room: G16) . . . 54
CS46: FINANCIAL MARKET AND THE MACROECONOMY (Room: B35) . . . 55

CS53: BAYESIAN METHODS IN ECONOMETRIC AND FINANCIAL APPLICATIONS (Room: B36) . . . 55
CS61: LONG MEMORY TIME SERIES MODELS (Room: B18) . . . 56
CS64: EFFICIENT MCMC ALGORITHMS FOR BAYESIAN FINANCIAL ECONOMETRIC MODELS (Room: B20) . . . 57

Parallel Session I – ERCIM (Sunday 18.12.2011 at 11:05 - 12:45) 58
ESI03: ADVANCES IN MULTIVARIATE ANALYSIS (Room: Woburn) . . . 58
ES01: ROBUST ANALYSIS OF COMPLEX DATA SETS (Room: Senate) . . . 58
ES04: TIME SERIES MODELING AND COMPUTATION I (Room: Bloomsbury) . . . 59
ES23: ADVANCES IN DISTANCE-BASED METHODS AND APPLICATIONS (Room: Gordon) . . . 59
ES24: MIXTURE MODELS: APPLICATIONS AND EXTENSIONS (Room: Torrington) . . . 60
ES28: DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS I (Room: Bedford) . . . 60
ES30: BIOSTATISTICS I (Room: S264) . . . 61
ES33: ROBUST METHODS IN SMALL AREA ESTIMATION (Room: S261) . . . 62
ES36: PARAMETRIC AND SEMIPARAMETRIC HAZARDS MODELS AND ANALYSES (Room: Jessel) . . . 62
ES59: IMPRECISE PROBABILISTIC MODELING TO SOLVE STATISTICAL PROBLEMS II (Room: Court) . . . 63

Parallel Session J – CFE (Sunday 18.12.2011 at 14:15 - 16:20) 65
CS07: ADVANCES IN COMPUTATIONAL METHODS FOR DSGE MODELS (Room: B33) . . . 65
CS38: SIGNAL EXTRACTION AND FORECASTING (Room: B36) . . . 65
CS42: DYNAMIC CORRELATION MODELS (Room: G16) . . . 66
CS50: BAYESIAN MODEL AVERAGING IN ECONOMETRICS (Room: B34) . . . 67
CS63: COMPUTATIONAL AND ECONOMETRIC METHODS IN DERIVATIVES APPLICATIONS (Room: B35) . . . 67
CS69: MEASURING SYSTEMIC RISK (Room: B20) . . . 68
CS92: FINANCIAL ECONOMETRICS II (Room: B18) . . . 69
CP03: POSTERS SESSION III (Room: Chancellor's) . . . 69

Parallel Session J – ERCIM (Sunday 18.12.2011 at 14:15 - 16:20) 72
ES08: OPTIMAL DESIGN (Room: Bloomsbury) . . . 72
ES12: HIGH-DIMENSIONAL STATISTICS, SPARSITY AND APPLICATIONS (Room: Jessel) . . . 72
ES22: DEALING WITH RARE EVENTS: RESAMPLING-BASED METHODS (Room: S261) . . . 73
ES26: MCMC FOR ESTIMATING DIFFUSIONS (Room: Gordon) . . . 74
ES35: LONGITUDINAL DATA ANALYSIS (Room: Woburn) . . . 74
ES40: BAYESIAN NONPARAMETRICS MODELLING (Room: Senate) . . . 75
ES46: RECENT ADVANCES IN MULTI-STATE MODELS (Room: Bedford) . . . 76
ES49: NEW DEVELOPMENTS IN QUANTILE REGRESSION (Room: Torrington) . . . 77
ES56: APPROACHES TO THE TREATMENT OF IMPRECISION OF STATISTICAL DATA (Room: Court) . . . 78
ES72: NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT III (Room: S264) . . . 78
EP03: POSTERS SESSION III (Room: Chancellor's) . . . 79

Parallel Session K – CFE (Sunday 18.12.2011 at 16:50 - 18:30) 81
CS01: MODELLING MULTIVARIATE FINANCIAL TIME SERIES (Room: Bloomsbury) . . . 81
CS24: BAYESIAN EMPIRICAL MACROECONOMICS (Room: Woburn) . . . 81
CS21: MACRO-FINANCE INTERFACE (Room: Jessel) . . . 82
CS44: VOLATILITY ESTIMATION AND FORECASTING (Room: Court) . . . 82
CS59: VOLATILITY, HEAVY TAILS AND RISK (Room: Torrington) . . . 83
CS65: FINANCIAL MARKETS CONTAGION (Room: Gordon) . . . 84
CS70: EVALUATING FINANCIAL PERFORMANCES (Room: Bedford) . . . 84

Parallel Session K – ERCIM (Sunday 18.12.2011 at 16:50 - 18:30) 86
ESI04: SPACE-TIME MODELLING IN DISEASE MAPPING (Room: Senate) . . . 86
ES13: ROBUST METHODS FOR FINANCIAL APPLICATIONS (Room: S264) . . . 86
ES32: DIAGNOSTIC TESTS FOR INDEPENDENT AND TIME-SERIES DATA (Room: S261) . . . 87
ES44: MODELLING THE EXTREMES (Room: G16) . . . 87
ES45: MIXTURE MODELS: THEORY AND DATA ANALYSIS (Room: B36) . . . 88

ES47: SEMIPARAMETRIC MODELS WITH INCOMPLETE DATA (Room: B33) . . . 88
ES51: ADVANCES IN SOFTWARE FOR TREE MODELS (Room: B18) . . . 89
ES55: FUZZY SETS IN STATISTICS (Room: B35) . . . 90
ES74: OUTLIERS AND CHANGE-POINTS IN TIME SERIES II (Room: B34) . . . 90
ES67: HEURISTICS IN FINANCE (Room: B20) . . . 91

Parallel Session L – CFE (Monday 19.12.2011 at 09:05 - 10:25) **93**
CS02: REALIZED VOLATILITY IN APPLICATIONS (Room: Jessel) . . . 93
CS81: COMPUTATIONAL ECONOMETRICS AND APPLICATIONS I (Room: Senate) . . . 93
CS98: FINANCIAL ECONOMETRICS IV (Room: S261) . . . 94
CS48: CONTINUOUS TIME FINANCIAL MODELS (Room: Bloomsbury) . . . 94
CS57: REAL-TIME MODELLING WITH MIXED FREQUENCIES (Room: S264) . . . 95
CS91: COMPUTATIONAL ECONOMETRICS (Room: Gordon) . . . 95
CS05: CONTRIBUTIONS IN MULTIVARIATE FINANCIAL TIME SERIES (Room: Torrington) . . . 96
CS43: CONTRIBUTIONS IN DERIVATIVE PRICING (Room: Bedford) . . . 97
CS66: CONTRIBUTIONS TO BAYESIAN ECONOMETRICS (Room: Court) . . . 97
CS75: CONTRIBUTIONS IN TIME SERIES AND PANEL DATA ECONOMETRICS (Room: Woburn) . . . 98

Parallel Session N – ERCIM (Monday 19.12.2011 at 10:55 - 12:35) **99**
ES05: BAYESIAN NONPARAMETRIC PRIORS (Room: B34) . . . 99
ES07: MIXTURE MODELS (Room: B18) . . . 99
ES34: INVARIANT COORDINATE SELECTION AND DIMENSION REDUCTION (Room: B36) . . . 100
ES42: POINT PROCESSES: MODELLING AND FORECASTING (Room: G16) . . . 100
ES48: MULTIVARIATE NONPARAMETRICS (Room: B33) . . . 101
ES64: FUZZY KNOWLEDGE EXTRACTION AND ENGINEERING (Room: B35) . . . 102
ES68: MICROARRAY DATA ANALYSIS (Room: B20) . . . 102

Parallel Session N – CFE (Monday 19.12.2011 at 10:55 - 12:35) **104**
CSI04: FINANCIAL TIME SERIES MODELLING (Room: Beveridge) . . . 104
CS09: HIGH FREQUENCY DATA MODELING (Room: Court) . . . 104
CS18: BAYESIAN FINANCIAL RISK MANAGEMENT (Room: S261) . . . 105
CS20: NON LINEARITY AND BUSINESS CYCLES (Room: Senate) . . . 105
CS23: IDENTIFICATION-ROBUST INFERENCE AND LARGE MODELS (Room: S264) . . . 106
CS29: MODELLING AND FORECASTING FINANCIAL RISK (Room: Woburn) . . . 107
CS30: QUANTITATIVE ASSESSMENT OF FINANCIAL STABILITY AND MACRO-POLICIES (Room: Jessel) . . . 107
CS33: BAYESIAN ECONOMETRICS AND APPLICATIONS (Room: Bloomsbury) . . . 108
CS41: DERIVATIVE PRICING (Room: Bedford) . . . 108
CS51: FINANCIAL ECONOMETRICS MODELLING (Room: Gordon) . . . 109
CS60: ECONOMETRICS WITH R (Room: Torrington) . . . 109

Parallel Session O – ERCIM (Monday 19.12.2011 at 14:05 - 15:45) **111**
ESI01: ROBUST METHODS (Room: Beveridge) . . . 111
ES03: IMPRECISION IN STATISTICAL DATA ANALYSIS I (Room: B36) . . . 111
ES82: CLUSTERING AND CLASSIFICATION (Room: B34) . . . 112
ES83: CONTRIBUTIONS IN PARAMETRIC AND SEMIPARAMETRIC INFERENCE (Room: B33) . . . 113
ES60: COMPUTER-AIDED DATA ANALYSIS (Room: G16) . . . 113
ES75: CONTRIBUTIONS TO EXTREME VALUE THEORY AND APPLICATIONS (Room: B35) . . . 114
ES84: STATISTICAL ALGORITHMS AND COMPUTER EXPERIMENTS (Room: B18) . . . 115
ES50: COMPUTATIONAL STATISTICS (Room: B20) . . . 116

Parallel Session O – CFE (Monday 19.12.2011 at 14:05 - 15:45) **117**
CS77: CONTRIBUTIONS IN TIME SERIES ECONOMETRICS II (Room: S264) . . . 117
CS80: ECONOMETRIC MODELLING AND APPLICATIONS II (Room: Court) . . . 117
CS94: FINANCIAL APPLICATIONS (Room: Gordon) . . . 118

CS10: CONTRIBUTIONS IN APPLIED FINANCIAL ECONOMETRICS (Room: S261) . . . 119
CS06: FILTERING (Room: Bedford) . . . 120
CS31: CONTRIBUTIONS IN BAYESIAN ECONOMETRICS AND APPLICATIONS (Room: Woburn) . . . 120
CS36: FORECASTING VALUE-AT-RISK (Room: Torrington) . . . 121
CS49: CONTRIBUTIONS IN VOLATILITY ESTIMATION AND FORECASTING (Room: Jessel) . . . 122
CS47: FINANCIAL TIME SERIES (Room: Bloomsbury) . . . 123
CS67: FINANCIAL MODELING (Room: Senate) . . . 123

Parallel Session P – ERCIM (Monday 19.12.2011 at 16:15 - 17:35) **125**
ESI02: BAYESIAN NONPARAMETRICS (Room: Beveridge) . . . 125
ES61: CONTRIBUTIONS IN NONPARAMETRIC STATISTICS (Room: Woburn) . . . 125
ES73: CONTRIBUTIONS ON HIGH-DIMENSIONAL DATA ANALYSIS (Room: Jessel) . . . 126
ES79: CONTRIBUTIONS TO ROBUST ANALYSIS OF COMPLEX DATA SETS (Room: Court) . . . 126
ES54: IMPRECISION IN STATISTICAL DATA ANALYSIS II (Room: Senate) . . . 127
ES78: CONTRIBUTIONS IN APPLIED STATISTICS (Room: Bloomsbury) . . . 128

Parallel Session P – CFE (Monday 19.12.2011 at 16:15 - 17:35) **129**
CS78: STOCHASTIC VOLATILITY (Room: S261) . . . 129
CS13: BAYESIAN QUANTILE REGRESSION (Room: Torrington) . . . 129
CS79: FINANCIAL MARKETS (Room: B20) . . . 130
CS22: DENSITY FORECASTING USING REALIZED MEASURES (Room: Bedford) . . . 130
CS27: ANALYSIS OF LARGE-DIMENSIONAL DATASETS: RECENT ADVANCES (Room: Gordon) . . . 131
CS82: COMPUTATIONAL ECONOMETRICS AND APPLICATIONS II (Room: S264) . . . 131
CS62: FINANCIAL ECONOMETRICS FOR RISK MANAGEMENT (Room: B35) . . . 132
CS71: REGRESSION TREES AND STRUCTURAL BREAKS (Room: B33) . . . 132
CS58: CONTRIBUTIONS IN ECONOMETRICS AND FINANCIAL MARKETS (Room: G16) . . . 133
CS73: SHORT-TERM MACROECONOMIC FORECASTING: LESSONS FROM THE CRISIS (Room: B34) . . . 133
CS74: COMPUTER INTENSIVE METHODS IN ECONOMETRICS (Room: B36) . . . 134

135

Saturday 17.12.2011 09:00-09:50 Room: Beveridge Chair: Herman Van Dijk Keynote talk CFE 1

Dynamic models for volatility and heavy tails

Speaker: **Andrew C. Harvey, University of Cambridge, United Kingdom**

A class of nonlinear time series models designed to extract a dynamic signal from heavy-tailed observations is proposed. The signal may be the level or a measure of scale. The dynamics are driven by the score of the conditional distribution. When modeling scale, an exponential link function is employed, as in EGARCH. The unifying feature is that the asymptotic distribution of the maximum likelihood estimators is established by a single theorem that delivers an explicit analytic expression for the asymptotic covariance matrix of the estimators. The properties of the models, particularly for the volatility models which employ an exponential link function, are very general. For example, expressions for unconditional moments, autocorrelations and the conditional moments of multi-step predictive distributions can be found for absolute values of the observations raised to any power. The generality of the approach is further illustrated by consideration of dynamic models for non-negative variables. Such models have been used for modeling durations, range and realized volatility in finance. Again the use of an exponential link function combined with a dynamic equation driven by the conditional score gives a range of analytic results similar to those obtained with the new class of EGARCH models.

Saturday 17.12.2011 10:15-11:05 Room: Beveridge Chair: Ana Colubi Keynote talk ERCIM 1

Achieving accuracy and correctness in parametric inference

Speaker: **Alastair Young, Imperial College London, United Kingdom**

The two key desiderata in frequentist parametric inference are accuracy and inferential correctness. When constructing, say, a confidence set for a parameter of interest in the presence of nuisance parameters, we desire high levels of coverage accuracy from the confidence set, as well as inferential correctness, in particular in relation to key principles of inference involving appropriate conditioning. We contrast and compare ways in which such accuracy and correctness can be achieved, with focus on elucidation of the relative strengths of analytic and simulation-based approaches, but considering also objective Bayes methodologies.

Sunday 18.12.2011 09:45-10:35 Room: Beveridge Chair: Tommaso Proietti Keynote talk CFE 2

Comparing the accuracy of copula-based multivariate density forecasts in selected regions of support

Speaker: **Dick van Dijk, Erasmus University Rotterdam, Netherlands** Coauthors: Cees Diks, Valentyn Panchenko, Oleg Sokolinskiy

A testing framework for comparing the predictive accuracy of copula-based multivariate density forecasts, focusing on a specific part of the joint distribution, is developed. The test is framed in the context of the Kullback-Leibler information criterion, and using (out-of-sample) conditional likelihood and censored likelihood in order to restrict the evaluation to the region of interest. Monte Carlo simulations show that the resulting test statistics have satisfactory size and power properties in small samples. In an empirical application to daily exchange rate returns we find evidence that the dependence structure varies with the sign and magnitude of returns, such that different parametric copula models achieve superior forecasting performance in different regions of the copula support. Our analysis highlights the importance of allowing for lower and upper tail dependence for accurate forecasting of common extreme appreciation and depreciation of different currencies.

Monday 19.12.2011 09:35-10:25 Room: Beveridge Chair: Christophe Croux Keynote talk ERCIM 2

Bayesian analysis for a stationary and transition density function

Speaker: **Stephen Walker, University of Kent, United Kingdom**

Bayesian nonparametric models using mixtures can construct both arbitrary transition and stationary density functions. Stationary models have often been seen as too restrictive but this assumption can be challenged when the stationary density function and the transition density function can take an arbitrary shape. Under this set up we can see no restrictions in using a stationary model. How to construct and analyse such models will be demonstrated. An analysis on "non-stationary" looking data will also be considered.

Monday 19.12.2011 18:05-18:55 Room: Beveridge Chair: Stephen Pollock Keynote talk CFE-ERCIM

Parametric inference on strong dependence

Speaker: **Peter M. Robinson, London School of Economics, United Kingdom**

Fractional models are considered for time series which can be non-stationary or stationary, and in the latter case can be invertible with short or long memory, or non-invertible. It is not assumed a priori which of these possibilities obtain, in which case the unknown memory parameter can lie in an arbitrarily large interval, and a proof of consistency of parameter estimates (a prerequisite for proving their asymptotic normality) is especially challenging. We establish these properties for Gaussian pseudo-maximum likelihood estimates. Monte Carlo evidence on finite sample properties and empirical examples are described.

Saturday 17.12.2011 10:25 - 12:30 Parallel Session C – CFE

CSI01 Room Woburn TIME SERIES ECONOMETRICS Chair: Andrew Harvey

C493: **Alternative methods of seasonal adjustment**
Presenter: **Stephen Pollock**, University of Leicester, United Kingdom
Co-authors: Emi Mise

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

C175: **On the issue of how many variables to use when estimating common factors using the Kalman filter**
Presenter: **Esther Ruiz**, Universidad Carlos III de Madrid, Spain
Co-authors: Pilar Poncela

In the context of dynamic factor models, we analyze the properties of the common factors estimated by using the Kalman filter with finite cross-sectional and time dimensions. It is already known that, when the parameters are known, the Kalman filter generates consistent filtered and smoothed estimates of the factors as the cross-sectional dimension increases. We analyse how the uncertainty associated with the estimated factors decreases with the number of variables in the system. We show that, regardless of the dynamic dependence of the factor, if the number of variables is beyond a relativelly small number, the uncertainty only decreases marginally.

C787: **Exponential models for the spectrum of a time series**
Presenter: **Tommaso Proietti**, University of Sydney, Australia
Co-authors: Alessandra Luati

The exponential model for the spectral density of a stationary process is based on the truncated Fourier series approximation of the logarithm of the spectral density function. The approximation yields a trigonometric regression model for the log-periodogram, with log-Gamma errors. We address the issue of selecting the order and the trigonometric terms to retain by Bayesian stochastic specification search. Moreover, we consider using the logarithm of the averaged periodogram as dependant variable. We illustrate the methodology with respect to the estimation of cycles and the long memory parameter.

CS08 Room Torrington ENERGY MARKETS, CLIMATE CHANGE AND WEATHER DERIVATIVES Chair: M. Dolores Furio

C457: **Allocating allowances for free in emissions markets: Implications for new industrial installations**
Presenter: **Morgan Herve-Mignucci**, Climate Policy Initiative, Italy
Co-authors: Barbara Buchner, Valerio Micale

Over 2005-2012, most existing installations covered by the European Union Emissions Trading Scheme (EU ETS), EU flagship cap-and-trade climate policy, were allocated allowances for free based on grandfathering (historical CO2 emissions) and new entrants were to some extent allocated allowances for free based on benchmarks. Starting 2013, various product benchmarks will be used for industrial installations requiring a certain portion of allowances allocated for free (exposure to international competitiveness, carbon leakage, etc.). Similar rules are in discussion or will be in emerging emissions markets (emissions intensity product benchmarks in the AB 32 legislation in California for instance). Allocation rules have important implications for both installations' profitability and government resources. We will explore to what extent various allocation rules impact industrial installations' profitability (including potential windfall profits), technology choices and timing of investment. In order to investigate this issue, we resort to a real options approach considering a given industrial installation (cement producer) calibrated using recent technology and emissions data. We will elicit the timing option, technology choices and alternative locations for the industrial facility (EU, California or China) in order to derive policy recommendations for EU phase III benchmarks implementation and other emissions markets.

C436: **The link between the carbon market and the stock market: A policy evaluation of the EU-ETS**
Presenter: **Daniel Rittler**, University of Heidelberg, Germany

The impact of European carbon prices on the market value of companies covered by the EU-ETS is empirically investigated. For this, we make use of a comprehensive dataset considering all sectors under the EU-ETS. Furthermore, we extend the sector-specific analysis in order to capture country-specific effects. For the first commitment period the empirical results imply the existence of a loose relationship between the carbon and the stock market. Only energy stock returns are affected by the carbon price whereas this effect is predominantly driven within the period of the carbon market breakdown. In sharp contrast, for the second commitment period, the results indicate the existence of a close link between the carbon and the stock market. Most importantly, the sign and the magnitude of these effects depend on the short/long position with allowances in the individual sectors. In particular, there is a positive (negative) effect of the carbon price on the market values of companies operating in sectors characterized by a long (short) position. Moreover, the results indicate high relevance of country-specific stringency in allowance allocation. Finally, we reveal that rising carbon prices affect companies' stock returns more than decreasing prices. This result mirrors stricter allocation rules for the post Kyoto commitment period of the EU-ETS.

C907: **Forecasting spike occurrences in electricity spot prices**
Presenter: **Michael Eichler**, Maastricht University, Netherlands
Co-authors: Dennis Turk

The problem of forecasting the occurrence of extreme prices, so-called spikes, in the Australian electricity market is considered. For this purpose, we first discuss the role of the market structure and resulting implications for the occurrence of spikes. Taking into account the specific features of the market under consideration, we propose a simple model for the conditional intensity of spikes. The model compares favorably in out-of-sample forecasting to a more sophisticated competing approach based on the autoregressive conditional hazard model.

C902: **Characterization and prediction of the electricity demand in the Iberian peninsula by using nonlinear time series analysis**
Presenter: **Vivaldo M.P. Mendes**, ISCTE-IUL, Portugal
Co-authors: Diana Aldea Mendes

An accurate forecasting of the electricity load demand is essential in the operation of an electric power system, especially in deregulated electricity markets. The worldwide electric power industry has been moving into a new scenario of market deregulation. The same has been happening in the

Iberian Peninsula, and in 2007 the MIBEL- Iberian Electricity Market was opened to all consumers from Portugal or Spain, in a free competitive market. Electricity is traded as a commodity, therefore load forecasting is essential for all market players, because the market is a day-ahead process and all sale and purchase orders of electricity have to be made until 11 a.m. of the previous day of dispatch. However, forecasting techniques applied in Electricity Markets are still in their early stages of maturity. We employ an empirical analysis of electricity loads time series in a nonlinear framework. This is very important since nonlinear phenomena contain information about the structure of the series and provide understanding of the process governing the underlying system, helping to distinguish between stochastic and chaotic processes and allowing an adequate choice for the forecasting techniques.

C840: **Error correction models for electricity future prices in Europe**
Presenter: **Muriel Renault**, EDF R&D, France
Co-authors: Jeremy Froger, Isabelle Parent, Virginie Dordonnat

Cointegration is a natural modelling framework to understand the evolution of European electricity forward prices due to their nonstationarity and the influence of other prices. We first specify a classical cointegration model and we estimate both the long-run and the short-run relationships using the Engle and Granger methodology. The goodness of fit and precision (in sample and out sample RMSE) of this univariate model are satisfactory but the residuals exhibit a strong ARCH effect (Conditionnal Autoregressive Heteroscedasticity). To address the issue of the non white-noise residuals we consider a non-constant specification for the regression coefficients of the short-term equation using the linear Gaussian state-space framework. The model is estimated using the EM algorithm. After model estimation, we investigate both the evolution of the short-term coefficients through time as well as the forecasting accuracy using the Kalman filter and smoother. We also check the validity of the model specification. We present models for German and British markets. In the case of the German market, explanatory variables are the future prices of coal, gas, Brent and carbon while for the British market, only coal, gas and carbon prices are considered in the model.

CS12 Room Senate BAYESIAN NONLINEAR ECONOMETRICS Chair: Roberto Casarin

C214: **Aggregating forecast probabilities for turning point detection**
Presenter: **Francesco Ravazzolo**, Norges Bank, Norway
Co-authors: Monica Billio, Roberto Casarin, Herman K. van Dijk

A forecast probability combination approach is proposed for predicting turning points of the European business cycle using individual country business cycles. We consider Markov-Switching VAR models for the business cycle analysis and for the turning point prediction of the EU 12 countries. A suitable combination of the predictive probabilities is then used in order to obtain a European aggregate prediction of the turning points. The weights of the combination scheme belong to the unit interval and can be interpreted as posterior model probabilities. A Bayesian approach has been applied to estimate the model probabilities and to forecast the turning points. A comparison has been carried out, in terms of statistical accuracy, between our forecasts for the turning points and other forecasts which ignores country specific information.

C233: **Extending Black-Litterman: views and covariance uncertainty**
Presenter: **Daniele Bianchi**, Bocconi University, Italy
Co-authors: Carlos M. Carvalho, Roberto E. Wessels

In the early 1990s, the Black-Litterman [BL] model became one of the most prominent portfolio management applications, addressing portfolio weights regularization by relying on theoretical market equilibrium and Bayesian mixed estimation. The standard BL framework is extended by fully characterizing parameter uncertainty on both expected returns and covariances. This is done through a hierarchical model where the investor's views and market equilibrium returns are nested in the prior scheme. A Bayesian Markov-Chain Monte Carlo (MCMC) algorithm is developed for the hierarchical model. The latter is compared to the classical BL approach and a standard Normal-Inverse-Wishart based Bayesian portfolio allocation. The model performances are investigated through a simulation example and an empirical analysis on a 10-sectors portfolio. We find that fully characterizing parameters uncertainty, keeping investor's views and market equilibrium assumptions, helps to reduce risk exposure and increase risk-adjusted excess returns.

C344: **Model selection for beta autoregressive processes**
Presenter: **Luciana Dalla Valle**, University of Plymouth, United Kingdom
Co-authors: Roberto Casarin, Fabrizio Leisen

Bayesian model selection for beta autoregressive processes is considered. The choice of the priors with possible parameter restrictions is discussed and a Reversible Jump Markov-Chain Monte Carlo (RJMCMC) procedure based on a Metropolis-Hastings within Gibbs algorithm is suggested.

C513: **Money growth and inflation: a regime switching approach**
Presenter: **Gianni Amisano**, European Central Bank and University of Brescia, Germany
Co-authors: Gabriel Fagan

A time-varying transition probabilities Markov Switching model, in which inflation is characterised by two regimes (high and low inflation), is developed. Using Bayesian techniques, we apply the model to the euro area, Germany, the US, the UK and Canada for data from the 1960s up to the present. Our estimates suggest that a smoothed measure of broad money growth, corrected for real-time estimates of trend velocity and potential output growth, has important leading indicator properties for switches between inflation regimes. Thus money growth provides an important early warning indicator for risks to price stability.

CS17 Room Bedford FORECASTING FINANCIAL MARKETS Chair: Ana-Maria Fuertes

C050: **A nonparametric analysis of predictive hedge fund performance using stochastic dominance tests**
Presenter: **Marcos Sanso-Navarro**, Universidad de Zaragoza, Spain
Co-authors: Jose Olmo

Predictive performance between hedge fund strategies using stochastic dominance tests of arbitrary orders is compared. This method allows investors to compare strategies not only in terms of some risk-return trade-off measure but also in terms of the whole predictive density of returns and for different levels of investors' risk aversion. The choice of the bandwidth parameter through cross-validation methods for nonparametric conditional density estimation provides an automatic procedure to derive the optimal set of predictive factors. Our results suggest that the relevant factors differ across strategies and do not generally coincide with those selected using parametric regression methods.

C051: **Measuring the hedging effectiveness of European index futures contracts**
Presenter: **Enrique Salvador**, Universitat Jaume I de Castello, Spain
Co-authors: Vicent Arago

Linear and non-linear GARCH models are estimated to find optimal hedge ratios with futures contracts for some of the main European stock

indexes. By introducing non-linearities through a regime-switching model, we can obtain more efficient hedge ratios and superior hedging performance in both in- and out-sample analysis compared with other methods (constant hedge ratios and linear GARCH). Moreover, the non-linear models also reflect different patterns followed by the dynamic relationship between the volatility of spot and futures returns during low and high volatility periods.

C283: **The economic value of stock and interest rate predictability in the UK**
Presenter: **Kavita Sirichand**, University of Leicester, United Kingdom
Co-authors: Stephen Hall, Kevin Lee

Asset return predictability is examined by comparing the out-of-sample forecasting performance of both atheoretic and theory informed models of bond and stock returns. We evaluate forecasting performance using standard statistical criteria, together with a less frequently used decision-based criterion. In particular, for an investor seeking to optimally allocate her portfolio between bonds and stocks, we examine the impact of parameter uncertainty and predictability in returns on how the investor optimally allocates. We use a weekly dataset of UK Treasury Bill rates and the FTSE All-Share Index over the period 1997 to 2007. Our results suggest that in the context of investment decision making under an economic value criterion, the investor gains from not only assuming predictability but by modelling the bond and stock returns together.

C304: **Fama French factors and US stock return predictability**
Presenter: **Sotiria Plastira**, University of Piraeus, Greece
Co-authors: Ekaterini Panopoulou

This paper investigates whether the HML, the SMB along with the short-term reversal, the long-term reversal and the momentum factors exhibit both in-sample and out-of-sample forecasting ability for the US stock returns. Our findings suggest that these factors contain significantly more information for future stock market returns than the typically employed financial variables. We also go one step further and test whether these variables can proxy for the aforementioned factors. Our results suggest that the default spread and to a lesser extent the term spread contain important information for the evolution of the factors examined.

C959: **The impact of the recent financial crisis on Eurozone sovereign credit default swap spreads**
Presenter: **Christopher Baum**, Boston College, United States of America
Co-authors: Paola Zerilli

This study evaluates the effects of the recent financial crisis on an important class of debt instruments: sovereign issues of Eurozone borrowers. Challenges to the stability of the Euro from threats of default by several Eurozone countries have raised serious concerns and led to unprecedented policy responses. We propose to study these effects by evaluating the risk premia embedded in sovereign credit default swap (CDS) spreads during periods of financial turmoil. These instruments provide insurance to their buyers, payable in the event of default. Their spreads over riskless instruments and spreads within the Eurozone CDS universe provide direct indications of market participants' valuation of risk associated with the underlying sovereign debt. Our methodology estimates the frequency of default events in order to analyse how perceptions of default risk have affected the demand for insurance on short-term vs. long-term debt instruments.

CS32 Room S264 BEHAVIOURAL FINANCE Chair: Robert Hudson

C035: **From behavioural to emotional corporate finance: a new research direction**
Presenter: **Richard Fairchild**, University of Bath, United Kingdom

Behavioural finance and behavioural corporate finance analyse the effects of psychological biases, heuristics, and emotions on investors' and managers' decision-making and performance. A major paradigm shift has been recently established by introducing a new field of research, namely Emotional Finance. This ground-breaking approach employs Freud's theory of phantastic objects to analyse the effect of unconscious, infantile emotions on investors' decisions. That work is extended by proposing a new development, namely, emotional corporate finance. We argue that, just as investors may view investments as phantastic objects, managers may view their projects similarly. We develop a formal approach that considers the effects of managerial phantasy on the investment appraisal decision, project performance, and managerial entrapment in a losing project. Our main results are as follows: a) Managerial project-phantasy may induce a manager to mistakenly invest in value-reducing projects. b) Phantasy may lead to volatility of managerial emotions, and hence volatility of project performance. c) Phantasy may lead to project entrapment, but may result in project abandonment if the manager's project-phantasy turns to project hatred. We conclude by considering whether managerial phantasy could explain management's entrapment in the Concorde project.

C038: **The price, quality and distribution of mortgage payment protection insurance: A hedonic pricing approach**
Presenter: **Robert Hudson**, Newcastle University, United Kingdom
Co-authors: John Ashton

Mortgage payment protection insurance (hereafter MPPI) provides varying combinations of accident, sickness and unemployment insurance and is used to protect the mortgage payments of policyholders in the event of them losing their income. The product has been heavily criticised for providing poor value for money and for being associated with unhelpful sales techniques especially when sold jointly with a mortgage. Consequently in 2009 the Competition Commission ruled that MPPI should not be sold jointly with lending. We investigate the price, quality and distribution of MPPI using a hedonic pricing approach. We conclude that the prices of policies sold independently do reflect the quality of the policies in terms of their coverage and conditions. In contrast policies sold jointly are clearly more expensive for a given set of benefits and conditions. Our study has implications for both policy and methodology. From the point of view of policy it does lend support to eliminating the joint sale of the MPPI with mortgages. From the methodological viewpoint the hedonic pricing approach appears very promising for future research in insurance and financial services.

C669: **Probability of attracting FDI flows**
Presenter: **Kristina Vasileva**, Westminster Business School, United Kingdom
Co-authors: Gulnur Muradoglu, Mario Levis

The purpose is to quantify the macroeconomic, geographic, institutional and cultural factors in terms of probability of an FDI flow occurring between two countries in order to see if countries can do something to enhance their attractiveness for FDI. Using a binary dependent variable method set to be zero if there is not an FDI relationship between country i and j and 1 if there is a positive relationship, we show how the aforementioned factors increase or decrease the probability of an FDI relationship occurring between two countries. We use a probit model on a set of macroeconomic, physical, institutional and cultural factors in order to determine if they are significant predictors of the probability of one country having an FDI relationship with another country. We find results that show that, when two countries have these factors in common, the probability that they will have FDI between them significantly increases. Countries can use this knowledge in order to attract more FDI from specifically targeted countries.

C693: **Optimism and portfolio choice**
Presenter: **Jiayi Balasuriya**, Cass Business School, United Kingdom
Co-authors: Gulnur Muradoglu, Peter Ayton

Innovative measures of financial optimism are developed by defining optimism as the overestimation of the favourable outcome in an individual's future financial situation. Financial optimism is found to have a significant positive effect on risk taking behaviour. Optimistic investors choose risky portfolios over risk-free portfolios for their investments and have higher debt borrowing. We use more than 660,000 observations from the British Household Panel Survey covering the period 1991 to 2007 in our analysis. Optimistic, pessimistic and neutral respondents have significantly different demographic characteristics. Optimists are significantly younger, more likely to be male, have higher business ownership, borrow more personal debt and take on a larger mortgage than non-optimists. They also have lower accumulated financial wealth and home ownership, but have a higher average unemployment rate than people who are pessimistic or neutral towards their financial situation.

C870: **Price impact of stock splits and dispersion of beliefs**
Presenter: **Maria Chiara Iannino**, University of Vienna, Austria

This paper is an empirical investigation, around the time of stock splits, of the relation between dispersion of beliefs among investors, market reaction and future performance of splitting companies. A 3-step analysis on a sample of US splits, occurred from 1993 to 2004, shows a change in the distribution of the analysts' forecasts, in mean and dispersion, after the announcement of a split. Moreover, this change helps explaining the underreaction that these companies exhibit in the post-event period. The excess returns in the post-event window are partly explained by the differences of opinion preceding the announcement. Finally, the results are consistent with the presence of an informational content in the event. This is also confirmed by the correction of the average underestimation error in the analysts' earnings forecasts for such companies. However, the signals are noisy and not completely reliable, as the dispersion of beliefs increases consistently with an increase in the uncertainty on the value of such firms and on the quality of the information conveyed by the event.

CS34 Room Jessel QUANTITATIVE RISK MANAGEMENT I Chair: Simon Broda

C235: **A decisionmetrics approach to portfolio allocation**
Presenter: **Maria Putintseva**, University of Zurich, Switzerland
Co-authors: Stanislav Anatolyev

Sophisticated multivariate econometric models designed to describe dependencies among many assets often fail to beat simple naive models in portfolio optimization. A better performance may be exhibited by a more parsimonious model tailored to a particular economic problem such as maximization of an investor's utility. This decisionmetrics paradigm is able to eventually generate more precise portfolio weights because of fewer parameters that need to be estimated. The decisionmetrics approach is applied to the problem of portfolio optimization and we develop a direct model for evolution of the portfolio weights. We compare our procedure to a conventional method of using the popular Dynamic Conditional Correlations and Dynamic Equicorrelations models with successive computations of implying portfolio weights. Using data on 30 components of the DJIA index and 10 and 40 randomly chosen components of the S&P 500 index we show that our approach provides a substantial improvement in portfolio characteristics. Among other things our method allows us to speed up computations because the portfolio is optimized simultaneously with the model parameters, while the standard procedure is necessarily a many step one.

C406: **MARC-MARS: Modeling asset returns via conditional multivariate asymmetric regime-switching**
Presenter: **Pawel Polak**, Swiss Finance Institute, University of Zurich, Switzerland
Co-authors: Marc Paolella

A multivariate asymmetric regime-switching model for asset returns is motivated and studied. It allows for volatility clustering, excess kurtosis, asymmetry, and dynamics in the dependency between assets over time. It nests several models previously proposed in the literature, and is demonstrated to outperform them in out-of-sample exercises. A new estimation procedure is developed which is far faster than existing methods, and thus crucial for use with a large number of assets.

C303: **Augmented likelihood estimation for mixture models**
Presenter: **Jochen Krause**, University of Zurich, Switzerland
Co-authors: Marc Paolella

The maximum likelihood estimation of mixture models is well-known to suffer from the degeneracy of mixture components caused by singularities in the likelihood function. We present a new and general solution to this problem based on an augmented maximum likelihood scheme dedicated to unconditional (univariate or multivariate) mixture models. Assuming mixtures of non-degenerated components only, we derive consistent estimators that fully eliminate the degeneracy issues and also nicely cope with the common bumpiness of mixture likelihood functions. Simulations studies indicate the superiority of the new estimator compared to existing methods. Moreover, an extension to conditional mixture models of higher complexity, e.g., mixture GARCH, is straightforward and also devised.

C400: **Trading activity and public news arrival**
Presenter: **Kerstin Kehrle**, University of Zurich, Switzerland

The probability of informed trading approach is extended by introducing additional trader heterogeneity. The role of public information on the arrival of privately informed and uninformed traders is investigated. Our results show that public information increases the arrival rate of uninformed buyers and sellers. The arrival sensitivity of uninformed traders to public news is highly related to liquidity and return volatility. We find evidence that an increase in the arrival of buyers due to public news increases stock returns while an increase of sellers decreases stock returns. These results are more pronounced for privately informed traders than for uninformed traders.

C285: **Tail probabilities and partial moments for quadratic forms in multivariate generalized hyperbolic random vectors**
Presenter: **Simon Broda**, University of Amsterdam, Netherlands

Quadratic forms in random vectors have numerous applications in econometrics, and thus received considerable attention in the literature. In finance, they arise in the context of the delta-gamma approximation to the value of a portfolio of assets; in particular, computing the Value at Risk requires an expression for their cdf. The numerical methods usually employed for this purpose have thus far been constrained to the case of joint Gaussianity, a severe limitation in the heavy-tailed context of asset returns. These methods are generalized to multivariate generalized hyperbolic (MGHyp) random vectors. The MGHyp is a very flexible distribution which nests, among others, the multivariate t, Laplace, and variance gamma distributions, all of which have been used successfully for modelling financial returns data. An expression for the first partial moment is also obtained; this can be used to evaluate the expected shortfall.

CS52 Room Bloomsbury VAST DIMENSIONAL FINANCIAL ECONOMETRICS Chair: David Veredas

C061: **Common volatility in evolutionary panels**
Presenter: **Matteo Barigozzi**, London School of Economics and Political Science, United Kingdom
Co-authors: Giovanni Motta

Large panels of time series typically exhibit strong evidence of co-movement. A factor model for multivariate processes whose second order structure smoothly varies over time is considered. The evolutionary filters are factorized as the product of a scalar smooth time-varying component that captures the long run common volatility, and ARMA filters describing the short run stationary dynamics. Specific non-pervasive behaviors of returns are left in the idiosyncratic components. The common volatility is then estimated in a fully non-parametric way and its asymptotic properties are derived. The estimator is easy to implement and it works well even for small sample sizes. The performance of the methodology is illustrated by means of simulation exercises. Finally, we provide an application on data of financial volatilities taken from S&P100. Empirical results show that there is a strong evidence of a factor structure and that the estimated common volatility is an indicator of economic crisis.

C080: **Modeling vast panels of volatilities with long-memory dynamic factor models**
Presenter: **Matteo Luciani**, Universite libre de Bruxelles, Belgium
Co-authors: David Veredas

Modeling vast panels of volatilities has always been an issue of paramount importance, yet elusive. These volatilities, when observed through time, share certain stylized facts: co-movement, clustering and long-memory. Based on Dynamic Factor Models, we propose a methodology that i) disentangles between commonness and idiosyncrasies, ii) is suitable for large dimensions, and iii) explains the above-mentioned stylized facts. A throughout Monte Carlo study shows the usefulness of the approach both in terms of factor identification and parameter estimation. An application to 90 daily realized volatilities, pertaining to S&P100, from January 2001 to December 2008, evinces four findings. i) All the volatilities have long memory, more than half of them in the nonstationary range, which increases during financial turmoil. ii) Tests and criteria point towards one dynamic common factor driving the co-movements, which naturally qualifies as the unobservable market volatility. iii) This factor has larger long memory that the assets volatilities, suggesting that long-memory is a market characteristic. Indeed, the idiosyncratic components show a much smaller deal of fractional integration. v) The time-varying long-memory found in the volatilities is mainly due to the common factor, while the degree of long-memory in the idiosyncrasies is relatively stable.

C253: **TailCor: A new measure of tail correlation for vast dimensional panels of asset returns**
Presenter: **Lorenzo Ricci**, Universite libre de Bruxelles, Belgium
Co-authors: David Veredas

A quantile-based method to measure tail correlations within the elliptical class distributions (TailCor) is proposed. It differs from tail dependence in that TailCor is not based on tail asymptotic arguments, and hence can be applied to any probability level. The use of TailCor is straightforward: it is a simple function and it disentangles the contribution of linear and non-linear correlation, the latter depending on the tail index. The method is successfully tested both on simulated data and on a large panel of securities (S&P500). A Monte Carlo study reveals the goodness of the measure, both in terms of computational time and for finite samples. An empirical illustration to a large panel of securities (the constituents of S&P500) over the financial crisis illustrates the usefulness of TailCor.

C515: **Wavelet-based realized covariation theory**
Presenter: **Jozef Barunik**, IES FSV UK, Czech Republic
Co-authors: Lukas Vacha

One of the most fundamental issues in finance is research of the covariance generating process between asset returns. Demand for accurate covariance estimation is becoming more important for risk measurement and portfolio optimization than ever before. We contribute to the current literature and provide a generalized theory for realized measures of covariance which are robust to jumps as well as noise in the underlying process. We introduce wavelet-based realized covariance estimator which brings the covariance estimation into the time-frequency domain for the first time. Moreover, we also present a methodology for detecting multivariate co-jumps using wavelets. The theory is supported by the small sample study of the behavior of estimators. Finally, our estimator is used to decompose the realized covariation of the real-world data into several investment horizon covariations, individual jumps and co-jumps. We utilize this decomposition to construct an ARFIMA-type model for forecasting, and we show that our estimator of realized covariance carries over the highest information content for forecasting of realized covariations and correlations. We also study the impact of co-jumps on the forecasts.

C771: **Modelling high dimensional time-varying dependence using D-vine SCAR models**
Presenter: **Hans Manner**, University of Cologne, Germany
Co-authors: Almeida Carlos, Czado Claudia

The problem of modelling the dependence of large dimensional time series data is considered. We build high dimensional time-varying copula models by combining pair-copula constructions (PCC) for the construction of flexible copulas with stochastic autoregressive copula (SCAR) models to capture dependence that changes over time. We show how the estimation of this highly complex model can be broken down into the estimation of a sequence of bivariate SCAR models, which can be achieved by using the method of simulated maximum likelihood. Further, by restricting the conditional dependence parameter on higher cascades of the PCC to be constant, we can greatly reduce the number of parameters to be estimated without losing much flexibility. Due to this estimation framework, the model can in principle be applied in arbitrarily large dimensions. We study the performance of our estimation method by a large scale Monte Carlo simulation. An application to a large dataset of stock returns of all constituents of the Dax 30 illustrates the usefulness of the proposed model class.

CS68 Room Court RECENT ADVANCES IN BOND PRICING Chair: Florian Ielpo

C027: **Credit and liquidity risks in euro-area sovereign yield curves**
Presenter: **Jean-Paul Renne**, Banque de France, France
Co-authors: Alain Monfort

A model of the joint dynamics of euro-area sovereign yield curves is proposed. The arbitrage-free valuation framework involves five factors and two regimes, one of the latter being interpreted as a crisis regime. These common factors and regimes explain most of the fluctuations in euro-area yields and spreads. The regime-switching feature of the model turns out to be particularly relevant to capture the rise in volatility experienced by fixed-income markets over the last years. In our reduced-form setup, each country is characterized by a hazard rate, specified as some linear combinations of the factors and regimes. The hazard rates incorporate both liquidity and credit components, that we aim at disentangling. The estimation suggests that a substantial share of the changes in euro-area yield differentials is liquidity-driven. Our approach is consistent with the fact that sovereign default risk is not diversifiable, which gives rise to specific risk premia that are incorporated in spreads. Once liquidity-pricing

effects and risk premia are filtered out of the spreads, we obtain estimates of the actual "or real-world" default probabilities. The latter turn out to be significantly lower than their risk-neutral counterparts.

C037: **Fed funds futures and the federal reserve**
Presenter: **Jean-Sebastien Fontaine**, Bank of Canada, Canada

Predictive regressions based on federal funds futures provide the most accurate forecasts of the Fed's target rate. Therefore, the cross-section of futures should best characterize the Fed's response function. This has been neglected in the term structure literature, perhaps due to technical challenges. A novel affine dynamic term structure model estimated using daily futures, LIBOR and target rates from 1994 to 2007 is introduced. The results uncover substantial cyclical changes in the uncertainty surrounding target changes that are induced by the Fed's response to economic conditions. The uncertainty is lowest (highest) in tightening (loosening) cycles, especially when the economy emerges from (enters) a recession. This adds to risk premium variations over the impact of changes in the price of macro risk. The model correctly characterizes risk premium since (i) it fits interest rates more accurately and (ii) it delivers unbiased target rate forecasts that match or improve upon standard benchmarks. The marginal predictive content obtained from combining futures and LIBOR rates comes from the ability to identify hedging demands in the futures market.

C081: **An analysis of ultra long term yields**
Presenter: **Simon Dubecq**, Banque de France / CREST, France
Co-authors: Christian Gourieroux

The discounting of very long-term cash-flows is crucial for the valuation of long-term investment projects. The market prices of US government bonds with very long-term time-to-maturity are analyzed, and some statistical specificities of very long-term zero-coupon rates, that standard Gaussian affine term structure models do not account for, are emphasized. In addition, we describe and estimate three Gaussian Nelson-Siegel affine term structure models, and highlight the model characteristics, which are necessary to match the dynamics of very long-term interest rates.

C039: **Fiscal policy, default risk and euro area sovereign bond spreads**
Presenter: **Vladimir Borgy**, Banque de France, France
Co-authors: Thomas Laubach, Jean-Stephane Mesonnier, Jean-Paul Renne

An arbitrage-free affine term structure model of potentially defaultable sovereign bonds is developed to model a cross-section of six euro area government bond yield curves. We make use of the coexistence of a common monetary policy under European Monetary Union, which determines the short end of the yield curve that is common to all countries, and decentralized debt policies which drive expected default probabilities and thereby spreads at the long end. The factors of our term structure model are observable macroeconomic variables, including measures of government solvency. When applying this model to yield curves of six EMU member countries over the period January 1999 to March 2010, we find strong evidence for a break in the relationship between the fiscal variable and the default intensities in early 2008. Despite using no latent factors, our model produces an excellent fit to both yield levels and spreads. For highly indebted countries, following the break the sensitivity of spreads to the fiscal variable rises sharply.

C153: **Forward rates, monetary policy and the economic cycle**
Presenter: **Florian Ielpo**, Lombard Odier Investment Managers, Switzerland

The short end of the yield curve incorporates most of the necessary information to forecast the next moves of Central Banks, but in a biased manner. A new method is proposed to forecast the Fed and the ECB rates by correcting the swap rates for their cyclical premium, using an affine term structure model. The corrected yields offer a higher forecasting power than the yields themselves, in an out-of-sample exercise. Some of them even outperform the forecasts obtained with a well-known Factor Augmented VAR.

CS56 Room Gordon COMPUTATIONAL METHODS IN APPLIED ECONOMETRICS Chair: Christopher F. Parmeter

C082: **Cross-sectional GMM estimation under a common data shock**
Presenter: **Oleksandr Zhylyevskyy**, Iowa State University, United States of America
Co-authors: Serguey Khovansky

A GMM estimation approach is developed for a stylized cross-sectional model with a non-localized common data shock. Common shocks are often encountered in economics and finance. Thus, their implications for estimation are of substantial interest. Other researchers investigated properties of OLS estimator under localized and non-localized shocks and proposed a GMM estimator for cross-sectional data with localized shocks. However, cross-sectional GMM estimation under non-localized shocks received little attention. We fill in the gap by developing an estimation framework in this setting. We propose one- and two-step GMM estimators and prove that they are consistent and asymptotically mixed normal under specified regularity conditions. The asymptotic mixed normality of our estimators differentiates them from usual asymptotically normal GMM estimators and necessitates a further investigation of statistical inference and specification testing. We show that despite the estimators' asymptotic mixed normality, conventional Wald tests can still be employed. We also prove that the OIR test retains its usual chi-squared asymptotic distribution. We investigate finite-sample performance of the method using Monte Carlo simulations of a financial model featuring a market-wide systematic risk. The approach allows us to estimate instantaneous market volatility, average instantaneous idiosyncratic volatility, and idiosyncratic risk-premium using only a cross-section of stock returns.

C760: **Sociodynamic discrete choice on spatial networks: Role of utility parameters and connectivity in emergent outcomes**
Presenter: **Elenna Dugundji**, Universiteit van Amsterdam, Netherlands
Co-authors: Laszlo Gulyas

Social interactions and generated feedback dynamics in the adoption of various transportation mode alternatives are treated. We consider a model where a commuter's choice is directly influenced by the percentages of neighbors and socioeconomic peers making each choice, and which accounts for common unobserved attributes of the choice alternatives in the error structure. We explicitly address non-global interactions within different social and spatial network structures, combining econometric estimation with computational techniques from multi-agent based simulation, and present an empirical application of the model using pseudo-panel microdata collected by the Amsterdam Agency for Traffic, Transport and Infrastructure. We observe that the estimated utility parameters for the different sociogeographic network scenarios can generate dramatically different dynamics and thus cannot be ignored in any empirical application. However, in a hypothetical simulation experiment we find that swapping the sociogeographic networks does not significantly change the long-run outcome of the simulation, when utility parameters are held fixed. We conclude highlighting recommendations for future work.

C714: **A comparative study of the Lasso-type and heuristic model selection methods**
Presenter: **Ivan Savin**, Friedrich Schiller University Jena and the Max Planck Institute of Economics, Germany

A first comparative analysis of Lasso-type (Lasso, adaptive Lasso, elastic net) and heuristic subset selection methods is presented. Although the Lasso has shown success in many situations, it has some limitations. In particular, inconsistent results are obtained for pairwise highly correlated

predictors. An alternative to the Lasso is constituted by model selection based on information criteria (IC), which remain consistent in the situation mentioned. However, these criteria are hard to optimize due to a discrete search space. To overcome this problem, an optimization heuristic (Genetic Algorithm) is applied. To this end, results of a Monte-Carlo simulation study together with an application to an actual empirical problem are reported to illustrate the performance of the methods.

C456: **Identification and estimation of social interactions through variation in equilibrium influence**
Presenter: **Mikko Packalen**, University of Waterloo, Canada

A new, computationally intensive, method for estimating social interaction effects is presented. In our applications we solve for each considered feasible value of the endogenous social interaction parameter 489 constrained optimization problems with over 12 million unknown interaction structure parameters. The proposed approach is based on using network interaction structure induced variation in equilibrium influence to construct conditionally balanced interaction structures. As equilibrium influence is determined by the known interaction structure and the unknown endogenous social interaction parameter, interaction structures are constructed for different imputed values of the unknown parameter. Each constructed interaction structure is conditionally balanced in the sense that when it is combined with observations on the outcome variable to construct a new variable, the constructed variable is a valid instrumental variable for the endogenous social interaction regressor if the true and imputed parameter values are the same. Comparison of each imputed value with the associated instrumental variable estimate thus yields a confidence set estimate for the endogenous social interaction parameter as well as for other model parameters. We provide conditions for point identification and partial identification. We apply the method to study subjective college completion and income expectations using Add Health data and provide Monte Carlo analyses of Erdos-Renyi and small-world networks.

C187: **Who benefits from financial development: new methods, new evidence**
Presenter: **Christopher Parmeter**, University of Miami, United States of America
Co-authors: Daniel Henderson, Chris Papageorgiou

A fresh look at the impact of financial development on economic growth is taken by using recently developed generalized kernel methods that allow for heterogeneity in partial effects, nonlinearities, and endogenous regressors. Our results suggest that while the positive impact of financial development on growth has increased over time, it is also highly nonlinear with more developed nations benefiting while low-income countries do not benefit at all. This finding contributes to the ongoing policy debate as to whether low-income nations should scale up financial reforms.

CP02 Room Chancellor's POSTER SESSION II Chair: Christodoulos Louca

C634: **Mean square analysis of delayed geometric Brownian motion**
Presenter: **Jevgenijs Carkovs**, Riga Technical University, Latvia

An algorithm for time asymptotic analysis of stochastic linear functional differential equations is proposed. The approach is based on some developments in the space of countable additive symmetric measures. The weak infinitesimal operator of this semigroup helps to find a Lyapunov-Krasovsky type quadratic functional that gives a necessary and sufficient asymptotic stability condition for the equation defined by the selected deterministic part of the analyzed stochastic equation. Moreover, substituting the solution of the such an equation as an argument of this quadratic functional, we have a stochastic process usable for Ito stochastic differential. This property permits us to derive an analogue of Ito formula for the above mentioned stochastic process and to discuss equilibrium asymptotic stochastic stability conditions for the original stochastic functional differential equation. As an example we have deduced necessary and sufficient condition for mean square decreasing of stochastic exponent given by Ito type scalar equation with delay.

C631: **Performance of Bayesian dynamic latent factor model in measuring pricing errors and forecasting returns**
Presenter: **Meltem Chadwick**, Central Bank of Turkey, Turkey

A dynamic Bayesian framework is proposed in order to obtain the exact distribution of pricing errors in the bounds of arbitrage pricing theory with the aim of observing if the usage of a dynamic model contributes to a significant reduction of the pricing errors over a static factor model. In doing so, a dynamic latent factor model, with and without time-varying factor loadings, has been utilised for the analysis of the pricing errors and we use this model to test out-of-sample and in-sample forecasting performance with respect to asset returns adopting the Fama-French data for US monthly industry returns. We observe that the pricing errors increased slightly using a dynamic factor model, when compared with the static factor model. Besides, inclusion of factors beyond the first one posed an improvement with respect to the pricing errors both for the static and the dynamic factor model. Considering the predictive performance of the dynamic factor model, it has been observed that in-sample predictive accuracy can be improved over a simple moving average model, however out-of-sample performance does not differ considerably.

C698: **On price stochastic equilibrium of adaptive single-component market**
Presenter: **Karlis Sadurskis**, Riga Technical University, Latvia
Co-authors: Maris Buikis, Jevgenijs Carkovs

A stochastic adaptive Samuel-Marshall type single-component market with randomly delayed supply is discussed. Our model is of the form of quasi-linear functional differential equations with perturbations dependent on phase coordinates and frequently switched an ergodic Markov process. The approach is based on stochastic calculus and asymptotic methods of bifurcation theory. It allows not only to advance in price equilibrium stability analysis, but also to detect a price stochastic business cycle and to prove that time delay of supply exerts significant influence on market price dynamics.

C719: **Modelling and testing threshold moving-average processes**
Presenter: **Bo Guan**, The University of Hong Kong, Hong Kong
Co-authors: Guo Dong Li, Wai Keung Li

Existing threshold models have been concentrated on autoregressive-type (TAR) models, and TAR models have been treated as a standard class of models for time series analysis; while moving-average-type (TMA) models have not drawn much attention. Although much profound progress in theoretical research has been made, there are still many significant characteristics of the TMA models which have not yet found a satisfactory interpretation. An investigation into the TMA models is undertaken in order to have a more complete picture of threshold time series. Properties of the Akaike information criterion (AIC and AICc) and the Bayesian information criterion are studied using numerical simulation, and results suggest that AICc performs better than AIC and BIC. Moreover, the asymptotic distribution of residual autocorrelations for TMA models is derived. By using sample averages in the evaluation of the asymptotic covariance matrix, it is possible to obtain standard errors of residual autocorrelations that are closer to the actual values. Simulation results demonstrate that this methodology can be applied to facilitate model diagnostic checking.

C836: **Economic trend: Identification and prediction through financial modelling**
Presenter: **Satyarth Pandey**, Institute of Management Technology of Ghaziabad, India
Co-authors: Veena Choudhary

The world is getting more and more averse towards shocks and surprises pertaining to the economic conditions of the respective nations. Information

based prior preparation and readiness towards such surprises is the need of the hour. It is understood that no information can provide a hundred percent guarantee towards such economic instabilities but still can make a nation better prepared to face them. We aim to identify and then we make an attempt to predict the trends that appear in an economy over the years through analysis of factors such as unemployment rate, industrial development rate which play an important role in governing the status an economy is in. The objective is to come up with a robust model that can provide a base for future work on this line. Through this work we also aim to overlap the paths of a developed and a developing economy, trying to find whether there exists any similarity in the economic conditions that developed countries found themselves while they were in developing phase with the current developing economies, and thus try to predict the future economic conditions of the developing economies on the basis of this analysis.

C899: **A nonlinear factor analysis for large sets of macroeconomic time series**
Presenter: **Diana Aldea Mendes**, ISCTE-IUL, Portugal
Co-authors: Vivaldo M.P. Mendes

Dynamic factor models are frequently used for empirical research in macroeconomics. Several papers in this area argue that the co-movements of large panels of macroeconomic and financial data can be captured by a relatively few common unobserved (latent) factors. The main purpose is to analyse and compare the transmission mechanism of monetary policy in the euro area based on two different approaches, in order to include information from large data sets. The first one consists of a classical linear methodology where the factors are obtained through a principal component analysis (FAVAR models), while the second approach employs a nonlinear manifold learning algorithm in order to obtain such factors (Kernel PCA). In contrast to the linear dimensionality reduction techniques, manifold learning methods provide a more powerful non-linear dimensionality reduction by preserving the local structure of the input data. We perform the dynamic inference on the nonlinear manifold by using a dynamic Bayesian network.

C855: **On moment conditions for quasi-maximum likelihood estimation of multivariate ARCH models**
Presenter: **Marco Avarucci**, Maastricht University, Netherlands
Co-authors: Eric Beutner, Paolo Zaffaroni

It is analyzed whether it is possible to derive consistency and asymptotic normality of the Gaussian quasi-maximum likelihood estimator (QMLE) for possibly the simplest VEC-GARCH model, namely the multivariate ARCH(1) model of the BEKK form, under weak moment conditions similar to the univariate case. In contrast to the univariate specification, we show that the expectation of the loglikelihood function is unbounded, away from the true parameter value, if (and only if) the observable has unbounded second moment. Despite this non-standard feature, consistency of the Gaussian QMLE is still warranted. The same moment condition proves to be necessary and sufficient for stationarity of the score, when evaluated at the true parameter value. This explains why high moment conditions, typically bounded sixth moment and above, have been used hitherto in the literature to establish the asymptotic normality of the QMLE in the multivariate framework.

C923: **Statistical methods to measure the efficiency of alternative multifactor single index portfolios**
Presenter: **Nina Kajiji**, University of Rhode Island, United States of America
Co-authors: Gordon Dash

Measuring portfolio efficiency is a process that is designed to account for possible return improvements and/or risk contractions. Recent evidence provides a synthesis and extension of contemporary findings on the properties of distance functions for measuring portfolio efficiency. The purpose is to augment the view of distance measures to include comparisons of alternative efficient sets. We analyze the axiomatic effects designed to improve the performance of simple portfolio models to replicate the allocative efficiency of the traditional mean-variance portfolio model. Several alternative versions of the Sharpe single index model are formulated. The model specifications incorporate either the factor-optimized or the Bayesian augmented measures of systematic risk. The resultant efficient sets are compared using ANCOVA and MANOVA model specifications based on Cartesian risk and return portfolio descriptors. Using the results from the Bonferroni multiple comparison test we have evidence of a near equivalence between the traditional Markowitz efficient set and the allocative portfolios produced by the Bayesian single-index portfolio optimization model. These findings are significant for model builders who wish to reduce the simplicity of nonlinear asset-liability optimization programs that include an embedded specification of the Markowitz portfolio model.

C637: **Testing the expectations hypothesis of the Czech term structure of interest rates**
Presenter: **Petr Jablonsky**, University of Economics, Czech Republic

The expectations hypothesis for the Czech interbank market in 2000 to 2011 are tested. The study is a continuation of a similar analysis performed for the Czech interbank market in the late 1990's by other authors. Even though this study was based on data for the young Czech market it proved the EH for about 50% of the interbank rates. As the Czech interbank market has developed we expected that our analyses on 2000 to 2011 data would prove the EH for most of the tested rates. To our surprise the results indicate a rejection of the EH at the Czech market during the tested period. The same results were achieved even when assuming a shorter time series up to the year 2007 excluding the data for the financial crises.

CS95 Room S261 FINANCIAL ECONOMETRICS I Chair: Kameliya Filipova

C805: **Estimation of an EGARCH(1,1)-AR(1)-M model**
Presenter: **Antonis Demos**, AUEB, Greece
Co-authors: Sofia Anyfantaki

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 employ a 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. Furthermore, the theoretical dynamic properties of a time-varying EGARCH(1,1)-M are derived. We discuss them and apply the suggested Bayesian estimation to three major stock markets.

C757: **Extreme value theory versus traditional GARCH approaches applied to financial data: A comparative evaluation**
Presenter: **M. Dolores Furio**, University of Valencia, Spain
Co-authors: Francisco J. Climent

Although stock prices fluctuate, the variations are relatively small and are frequently assumed to be normal distributed on a large time scale. But sometimes these fluctuations can become determinant, especially when unforeseen large drops in asset prices are observed that could result in huge losses or even in market crashes. The evidence shows that these events happen far more often than would be expected under the generalized assumption of normal distributed financial returns. Thus it is crucial to properly model the distribution tails so as to be able to predict the frequency and magnitude of extreme stock price returns. We follow an approach suggested previously and combine the GARCH-type models with the extreme value theory (EVT) to estimate the tails of three financial index returns DJI, FTSE 100 and NIKKEI 225 representing three important financial areas in the world. Our results indicate that EVT-based conditional quantile estimates are much more accurate than those from conventional AR-GARCH models assuming normal or Student's t-distribution innovations when doing out-of-sample estimation (within the in-sample estimation,

this is so for the right tail of the distribution of returns). The results should be useful to investors in general, since their goal is to be able to forecast unforeseen price movements and take advantage of them by positioning themselves in the market according to these predictions.

C773: **Assessing shocks to inflation expectations in a data rich environment**
Presenter: **Lucia Alessi**, European Central Bank, Germany
Co-authors: Luca Onorante

A semi-structural analysis aiming at estimating the macroeconomic effects of shocks to inflation expectations is carried out. 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 four structural shocks, and assume that one of these shocks has an expectational nature. To achieve identification of this shock, we use a mix of zero and sign restrictions. Our results confirm an important role for inflation expectations in affecting the dynamics of real and nominal variables.

C487: **A spatial analysis of international stock market linkages**
Presenter: **Hossein Asgharian**, Lund University, Sweden
Co-authors: Wolfgang Hess, Lu Liu

The severe global impacts of the recent financial crises have intensified the need to understand how country specific shocks are transmitted to other countries. Using spatial panel econometric techniques, we analyze the impact of various linkages between countries on the interdependence of their stock markets. To eliminate the effect of common trend in return data, we employ a panel specification with two spatial lags. This enables us to directly compare the spatial dependencies among neighbors with those among non-neighbors. A number of different linkages are used to define the neighborhood among countries: geographical neighborhood, similarity in industrial structure, the volume of countries' bilateral trade, bilateral foreign direct investment, convergence in expected inflation, and the stability of the bilateral exchange rate. We analyze returns of 41 national stock market indexes over a period of 16 years. Our empirical results indicate similarity in industrial structure to be the most important linkage. Other important linkages that explain stock market synchronization include geographic closeness, bilateral trade, and exchange rate stability. This is the first in-depth analysis of the underlying economic structure of financial markets interactions that relies on spatial econometric techniques.

C707: **Yield curve predictability, regimes, and macroeconomic information: An asset pricing approach**
Presenter: **Kameliya Filipova**, University of St. Gallen, Switzerland

We develop a new multivariate dynamic term structure model, which takes into account the nonlinear (time-varying) relationship between interest rates and the state of the economy. In contrast to the classical term structure literature, where nonlinearities are captured by increasing the number of latent state variables, or by latent regime shifts, in our no–arbitrage framework the regimes are governed by thresholds and they are directly linked to different economic fundamentals. Specifically, starting from a simple monetary policy model for the short rate, we introduce a parsimonious and tractable model for the yield curve, which takes into account the possibility of regime shifts in the behavior of the Federal Reserve. In our empirical analysis, we show the merit of our approach along four dimensions: (i) interpretable bond dynamics; (ii) superior out-of-sample short rate forecasting performance; (iii) accurate short end yield curve pricing; (iv) bond excess return predictability.

CP01 Room Chancellor's POSTERS SESSION I **Chair: Christodoulos Louca**

C490: **Adaptive MCMC for non-life insurance reserving via paid-incurred claims models**
Presenter: **Alice Dong**, University of Sydney, Australia
Co-authors: Gareth Peters, Mario Wuthrich, Jennifer Chan

Computationally efficient adaption strategies are developed and compared for MCMC algorithms, based on adaptive Metropolis and Robbins-Monro. These adaptive algorithms are used to facilitate estimation of important reserving quantities in non-life insurance models. This significantly extends the range of possible reserving measures and the ability to estimate these reserves based on the predictive loss distributions in sophisticated insurance models. In particular, a novel stochastic Paid-Incurred-Claims model for reserving, that allows us to combine claims payments and incurred losses information, is developed and compared to a recently developed PIC model. The extensions considered remove the conjugacy properties previously developed and therefore require efficient sampling methodologies to tackle the high-dimensions and induced posterior correlation structures from the loss reserving triangle.

C423: **Asset pricing in DSGE models - comparison of different approximation methods**
Presenter: **Jan Acedanski**, University of Economics in Katowice, Poland

Differences between asset price properties in a standard DSGE model approximated by using different methods are investigated. Among the discussed approaches there are perturbation, projection and value function iteration techniques as well as the methods specially designed for asset price approximation like the loglinear-lognormal approach. We compare basic moments of stock prices and returns and risk free rates. We also check if there are any differences between predictability of stock premium and dynamic correlations between financial variables and real economy in the model for different approximation methods. The results show that despite significant differences between the method's accuracy in many areas their implications remain very similar.

C518: **Study of the informational efficiency of Warsaw stock exchange during 2007-2009 with machine learning**
Presenter: **Marcin Ciemny**, University of Bedfordshire, United Kingdom
Co-authors: Livia Jakaite, Vitaly Schetinin

An attempt to verify the hypothesis of informational efficiency of financial markets, known as "random walk", is undertaken. The main aspects of the hypothesis are described in relation to financial crises. The hypothesis is tested on the real data of the Warsaw Stock Exchange during the financial crisis years 2007-2009 by predictive modeling based on Machine Learning (ML). In particular, Group Method of Data Handling (GMDH) has been used. The results have shown that there exists a weak form of efficiency. The use of GMDH has been shown to provide selection of a near optimal structure of predictive models that is of crucial importance for testing the informational efficiency. The GMDH has been shown capable of automatically inducing the desired model in the absence of expert knowledge. The results obtained with the GMDH models are compared with those achieved with the conventional ML methods. The GMDH models have been shown to provide a better accuracy of the predictions than a random choice, that allowed us to conclude on the weak form of informational efficiency.

C558: **On the estimation of exponential regression models: an integrated GMM approach**
Presenter: **Esmeralda Ramalho**, Universidade de Evora, Portugal
Co-authors: Joaquim Ramalho

Exponential regression models are widely used in economic applications where the variable of interest can take only nonnegative values (e.g. gravity equations for international trade, Mincerian equations for wages, production functions, Euler equations). Several estimators have been proposed for these models, that possess optimality properties under some data generating processes (DGP), e.g. Gamma and Poisson-based quasi-

maximum likelihood estimators for cases where regressors are exogenous and GMM estimators based on orthogonality conditions that include a residual function from a transformed version of the model of interest for cases of endogeneity. Building on this last approach, we propose a unifying new class of GMM estimators that nests the two lines of approach for both exogenous and endogenous regressors and, additionally, incorporates more flexible new estimators where the orthogonality conditions use spline functions of the regressors or the instrumental variables to approximate the optimal instruments. These estimators display a very promising performance in a Monte Carlo simulation study, being close to the optimal choices for a variety of DGP.

C567: **Hedonic functions, hedonic methods, estimation methods and Dutot and Jevons house price indexes**
Presenter: **Joaquim Ramalho**, Universidade de Evora, Portugal
Co-authors: Esmeralda Ramalho

Hedonic methods are a prominent approach in the construction of house price indexes. It is investigated in a comprehensive way whether or not there exists any kind of link between the type of price index to be computed (Dutot or Jevons) and the form of hedonic functions, hedonic methods and estimation methods, with a link being defined as a specific combination of price indexes, functions and methods that: (i) simplifies substantially the calculations required to compute hedonic price indexes, while other combinations, although possible, require additional assumptions and, in general, the use of bias corrections; or (ii) produces estimates of unadjusted price changes that are identical to those observed in the dwelling sample. It is found that: (i) there is a link between Dutot indexes, exponential hedonic functions and the Poisson pseudo maximum likelihood estimator, on the one hand, and Jevons indexes log-linear hedonic functions and ordinary least squares, on the other hand; and (ii) there is no link between the time dummy variable method and Jevons indexes. A Monte Carlo simulation study illustrates both the convenience of the links identified and the biases that result from overlooking them or implementing bias corrections based on invalid assumptions.

C597: **Customers' satisfaction measurement via a flexible fuzzy clustering**
Presenter: **Mohammad Ansari**, Islamic Azad University, Iran
Co-authors: Mohsen Haghighi, Mahdi Zowghi

Customer satisfaction represents a modern approach for quality in enterprises and organizations and serves the development of a truly customer-focused management and culture. The paper proposes a fuzzy clustering model for customer satisfaction measuring. The clustering problem consists of shaping classes from a set of objects, based on knowing some of their properties. The proposed unsupervised-supervised clustering methodology evaluates the satisfaction class of a set of customers in different aspects of customer focus as fuzzy vectors. The main advantages of this method are fully considering the qualitative form of customers' judgments through fuzzy theory and evaluate customers' satisfaction in different aspects of customer focus.

C839: **Hedging of discrete time auto-regressive stochastic volatility options**
Presenter: **Juan-Pablo Ortega**, CNRS/Universite de Franche Compte, France
Co-authors: Joan del Castillo

Numerous empirical proofs indicate the adequacy of the time discrete auto-regressive stochastic volatility models in the description of the log-returns of financial assets. The pricing and hedging of contingent products that use these models for their underlying assets is a non-trivial exercise due to the incomplete nature of the corresponding market. We apply two volatility estimation techniques available in the literature for these models, namely Kalman filtering and the hierarchical-likelihood approach, in order to implement various pricing and dynamical hedging strategies. Our study shows that the local risk minimization scheme developed by Follmer, Schweizer, and Sondermann is particularly appropriate in this setup, especially for at and in the money options or for low hedging frequencies.

C850: **Log-range based detection of volatility mean breaks**
Presenter: **Vasiliki Chatzikonstanti**, University of Patras, Greece
Co-authors: Ioannis Venetis

Unconditional mean volatility level breaks in the Dow Jones industrial average components is examined. Our daily data sample spans the post dotcom bubble period covering the recent financial crisis. We use the daily log-range as a volatility proxy. Log-range based volatility estimation can be powerful and convenient, as it curtails the impact of noise present in the absolute or squared log-return measures of volatility. Moreover, log-range is approximately normally distributed and robust toward microstructure effects. We employ two approaches to detect, if present, and characterize breaks. A multiple mean break model that assumes abrupt mean shifts or jumps in volatility and a smooth transition model that allows abrupt shifts, smooth shifts or a combination . In all cases, we do not employ any upper bounds on the allowed number of breaks. We find that all stock returns experienced significant multiple structural breaks in volatility whereas not all breaks are sudden. Some exhibit smooth transition over time albeit the time interval is short enough to avoid counterintuitive persisting trends in variance. Volatility persistence is purged once volatility breaks are accounted for. This result is enhanced further with the adoption of smooth transition changes at least for some of the breaks.

C910: **Measuring spillovers: An application to the stock markets**
Presenter: **Jilber Urbina**, Universitat Rovira i Virgili, Spain
Co-authors: Nektarios Aslanidis, Oscar Martinez

Stock market contagion are tested on the basis of the 2008-2009 global financial crisis. Using Forbes and Rigobon methodology and correcting for heteroskedasticity bias, we find strong evidence in favor of interdependence instead of contagion. Additionally, we measure volatility spread among countries and summarize it into a volatility spillover index to provide a measurement of such interdependence. Our spillover index is based on the forecast error variance decomposition (fevd) for a VAR model at h-step ahead forecast, and we construct it using both the orthogonalized fevd and the generalized fevd (Gfevd); both of them provide similar results, but the generalized version is easier to handle when a data set with more than 6 variables is involved; this is true since the Gfevd does not depend on the restrictions imposed by the Cholesky decomposition. This fact makes it attractive when economic theory does not fit well with variables relationship. An R package to enable the calculation of Generalized fevd functions that are not yet available neither in R nor in other software is developed.

C964: **Revealing market's animal spirits of the Euro-area sovereign debt crisis using a generalised loss function: The role of fiscal rules and fiscal institutions.**
Presenter: **Emmanuel Mamatzakis**, US & UNIPI, Greece

The underlying market's preferences for Euro-zone sovereign bonds, as depicted by the difference between the spread over swaps and Credit Default Swaps (CDS), are examined. Our sample covers those Euro-zone member states most at risk of default namely; Greece, Portugal, Ireland, Spain and Italy. Moreover, a generalised flexible loss function is employed so as to reveal the behaviour of market participants. The results show that market's preferences over Euro-zone sovereign debt have shifted towards pessimism post the Emergency Financing Mechanism (EFM). Having derived market's preferences over the Euro-zone sovereign debt crisis, we examine the impact of fiscal policy institutions and fiscal rules on those preferences for the period from third quarter of 2008 to second quarter of 2011. The empirical evidence shows that there is a clear relationship between fiscal rules-institutions and market's preferences with the direction of causality running from the former to the latter.

Saturday 17.12.2011 11:15 - 12:30 Parallel Session D – ERCIM

ES06 Room B34 SMALL AREA ESTIMATION **Chair: Domingo Morales**

E217: **Mean squared error estimation of cancer risk predictions using area level models in disease mapping**
Presenter: **Tomas Goicoa**, Universidad Publica de Navarra, Spain
Co-authors: Lola Ugarte, Ana F. Militino, Jaione Etxeberria

Health agencies have promoted research on space-time evolution of cancer mortality or incidence risk to better understand the disease. However, this information is becoming incomplete as mortality or incidence risk for future years are required for efficient planning and coordination of public health programs and clinical services. As mortality records are only available after a few years later, predictions are necessary to supply cancer mortality risks for years to come. Some research in disease mapping provides predictions of disease rates in small areas, however the spatial dependence is not included in the models. A space-time P-spline model is considered to model the spatio-temporal distribution of risks and to forecast risk values for future years. This model explicitly assumes a space-time interaction effect. The spatial and temporal dependence are modelled through B-spline bases for space and time. Forecasting will be carried out by extending the B-spline basis in the time dimension. The mean squared error of the forecast risks will be provided exploiting the mixed model formulation of the P-spline model. Results will be illustrated with different cancer types in Spain from 1975 to 2008. Forecast values will be provided up to 2011.

E136: **Multinomial-based small area estimation of labor force indicators**
Presenter: **Maria Jose Lombardia**, University of Coruna, Spain
Co-authors: Esther Lopez-Vizcaino, Domingo Morales

Unemployment is nowadays a major problem. Small area estimation of labour force characteristics like totals of employed or unemployed people and unemployment rates is considered. Small area estimators of these quantities are derived from a multinomial logit mixed model with independent random effects on the categories of the response vector. In addition, the incorporation of the time effect to improve the predictive power of the small area estimators is studied. The mean squared errors are estimated both by explicit formulas and by a parametric bootstrap method. Simulation experiments designed to analyse the behaviour of the introduced estimators have been carried out. Finally, an application to real data from the Spanish Labour Force Survey in Galicia (region in the Northwest of Spain) is given.

E182: **A sensitivity analysis to the hyperprior specification in disease mapping and small area models**
Presenter: **Enrico Fabrizi**, Universita Cattolica del S. Cuore, Italy
Co-authors: Fedele Pasquale Greco, Carlo Trivisano

Sensitivity to prior specification for variance components is studied in a class of models widely used in spatial disease mapping accounting for both spatially structured and unstructured heterogeneity. Notoriously, in disease mapping priors on the variance components sensibly affect inferences, in particular when the object of inference is a functional of the ensemble of area-specific parameters. We propose to use the Generalized Inverse Gaussian (GIG) distribution. We show that a set of conditions on the GIG parameters implies a prior distribution with finite moments on the relative risks, a property not shared by many popular alternative priors for the variance components regardless of the parameters choice. This property allows us to easily incorporate available prior information, to give equal a priori importance to structured and unstructured heterogeneity and to control for the amount of shrinkage in the map. We compare inferences obtained using the GIG prior to those derived under alternatives currently popular in the literature using a simulation exercise and the analysis of real data sets. The method may be extended to all small area models that make use of a log transformation such as those typical of poverty mapping.

ES20 Room B33 APPLIED STATISTICS I **Chair: Paula Camelia Trandafir**

E588: **Test for accuracy in ranking in moving extreme ranked set sampling**
Presenter: **Mohammad Al-Saleh**, Yarmouk University, Jordan
Co-authors: Asma Ababneh

Ranked set sampling (RSS) is a sampling technique suitable for situations when the units can be ranked (with respect to the variable of interest) by judgment without actual measurement. To obtain such a sample, m sets of size m each from the population of interest are drawn; the elements within each set are ranked by judgment from smallest to largest. From the ith set, take for actual quantification the element (judgment) ranked as the ith order statistic. The obtained sample consists of m independent order statistics. One problem with the RSS procedure is the requirement that the ranking should be done by judgment or at a negligible extra cost. Thus, it is usually hard to believe that the ranking is perfect. Error in ranking is usually unavoidable especially with large set size. Moving Extreme Ranked Set Sampling (MERSS) is an important variation of RSS. In MERSS, we take for actual measurement the judgment maximum of random samples of sizes $1, 2, \ldots, m$. Unlike RSS, MERSS allows for an increase of the set size without introducing too much ranking error. Testing for error in ranking should be done before using the MERSS for inference. In this paper, we consider the problem of testing whether judgment ranking is as good as actual ranking. Our contribution is to provide methods to test for error in ranking in MERSS. We consider several nonparametric tests that are based on the distance between the actual and the judgment ranking of the obtained data. The null and the alternative distributions of the test statistics are derived. Synthetic as well as real data sets are used for illustration.

E653: **Optimal target allocation proportion for correlated binary responses in a two-treatment set up**
Presenter: **Paula Camelia Trandafir**, Universidad Pública de Navarra, India
Co-authors: Saumen Mandal, Atanu Biswas

Optimal allocation designs for the allocation proportion are obtained for a two-treatment clinical trial, in the presence of possible correlation between the proportion of successes for two treatments. Possibility of such type of correlation is motivated by some real data. It is observed that the optimal allocation proportions highly depend on the correlation. We also discuss completely correlated set up where the binomial assumption cannot be made.

E494: **Coherent forecasting for discrete-valued time series data with application to infant sleep status data**
Presenter: **Raju Maiti**, Indian Statistical Institute, India
Co-authors: Atanu Biswas

A time series of discrete data which is the sleep status of an infant over two hours observed in a minute-by-minute way is considered. We develop the theory of coherent forecasting of the discrete time series based on Pegram's model, MTD model and logistic regression, and apply them to the infant's sleep status data. The properties are also studied by extensive simulation.

ES39 Room B20 HIGH DIMENSIONAL DESIGN OF EXPERIMENTS Chair: Davide Ferrari

E575: **Optimizing experiments with mixtures**
Presenter: **Michele Forlin**, European Centre for Living Technology, Italy

Design of experiments often is used for learning the relations between experimental inputs and the system response. This task is sometimes challenging, especially when the complexity of the system increases or when reliable a priori information is missing. From this perspective, design of experiments overlaps with the field of stochastic search and optimization, and the related inference tasks greatly benefit from computational methodologies to drive the experimentation. We propose a computational approach to intelligently navigate the experimental space and identify optimal formulations when dealing with experiments with mixtures (i.e. experiments in which factors are proportions of different components). The approach is based on a variant of genetic algorithm, which includes statistical modeling and explores efficiently the design space by a sequence of experiment generations. Statistical models are inferred and updated at each generation of experiments and the obtained information is incorporated into the crossover operator of the genetic algorithm. This device is shown to boost the basic genetic algorithm, so that the number of experiments required to identify optimal solutions (i.e. best formulation of components proportions) is dramatically reduced.

E589: **High dimensional design of experiments: The combination of evolution and statistical models**
Presenter: **Debora Slanzi**, University Ca' Foscari, Italy
Co-authors: Irene Poli

In several research areas, such as biology, chemistry or material science, experimentation is complex and high dimensional, extremely expensive and time-consuming, so an efficient plan of experimentation is essential to achieve good results and avoid unnecessary waste of resources. In this work we address the high dimensional design of experiments by developing a sequential procedure based on the evolutionary paradigm, where the information from an initial set of data is detected and processed by a class of statistical models. This computational procedure derives an experimental design in which a small number of trials will produce a set of *intelligent data*, i.e data that contain the relevant information on the system under study. By simulating several high dimensional problems, we demonstrate that this procedure is an efficient tool to identify the factor components of the optimizing experimental design deriving only a small percentage of the whole experimental space. Moreover, modeling is shown to play a fundamental role in the evolutionary process making the search in high dimensional space efficient and able to identify not only the relevant factor components but also the factors interactions affecting the system response.

E611: **Improving complex experiments by co-information composite likelihood optimization**
Presenter: **Matteo Borrotti**, University of Ca'Foscari, Italy
Co-authors: Davide De March, Davide Ferrari

An adaptive procedure for improving the response outcomes of complex combinatorial experiments is proposed. New experiment batches are chosen by minimizing the co-information composite likelihood objective function, which is derived by coupling importance sampling and composite likelihood principles. We show convergence of the best experiment within each batch to the globally optimal experiment in finite time, and carry out simulations to assess the convergence behavior as the design space size increases. The procedure is tested as a new enzyme engineering protocol in an experiment with a design space size of order 10^7 .

ES62 Room B35 IMPRECISION IN INFERENCE Chair: Maria Brigida Ferraro

E150: **Testing partial inclusion of the mean of a random interval in a fixed interval**
Presenter: **Ana Belen Ramos Guajardo**, European Centre for Soft Computing, Spain
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez, Maria Angeles Gil

In many real-life situations involving interval data it is useful to determine if the mean of these data is included in a given interval with certain degree or not. For instance, in order to analyze the quality of a specific product from the perceptions of an expert, it could be interesting to check if the expected value of these perceptions is contained in previously fixed limits with a given inclusion degree. The variables considered to model experimental interval data are called random intervals. Concerning the inclusion of the mean of a random interval in an interval, the total inclusion and the empty intersection (or null inclusion) have already been analyzed. The aim is to develop a test for the partial inclusion of a random interval in a given closed interval by using both asymptotic and bootstrap techniques. In addition, the performance of the bootstrap test will be shown by means of simulation studies. To conclude, an example involving real data is presented to illustrate the use of the proposed test.

E483: **Extension of Ruspini's formulation of evidential reasoning to evidence fusion based on conditioning**
Presenter: **Takehiko Nakama**, European Center for Soft Computing, Spain
Co-authors: Enrique Ruspini

Ruspini's probability-theoretic formulation of evidential reasoning provides a mathematically sound approach to evidence fusion. Using his framework, we rigorously formulate the process of combining partial evidential bodies based on conditioning. Epistemic logics, which are effective in dealing with both the state of a system and the state of knowledge about it, are utilized to establish the formulation. Evidential bodies are represented by probability spaces that reflect their epistemic states and uncertainties. We derive a formula for combining partial evidential bodies based on conditioning and examine situations where the formula can be properly used.

E279: **Fitting parametric link functions in a regression model with imprecise random variables**
Presenter: **Maria Brigida Ferraro**, Sapienza Univesity of Rome, Italy

A regression model for imprecise random variables has been introduced in our previous works. The imprecision of a random element has been formalized by means of the fuzzy random variable (FRV). In detail, a particular case of FRVs characterized by a center, a left and a right spread, the LR family (LR FRV), has been considered. The idea is to jointly consider three regression models in which the response variables are the center, and two transformations of the left and the right spreads in order to overcome the non-negativity conditions of the spreads. Response transformations could be fixed, as we have done so far, but all inferential procedures, such as estimation, hypothesis tests on the regression parameters, linearity test etc., are affected by this choice. For this reason we consider a family of parametric link functions, the Box-Cox transformation model, and by means of a computational procedure we will look for the transformation parameters that maximize the goodness of fit of the model.

ES77 Room G16 BIOSTATISTICS II Chair: Ayse Ulgen

E536: **Next generation association studies: In search of low frequency and rare variants affecting complex traits**
Presenter: **Ioanna Tachmazidou**, The Wellcome Trust Sanger Institute, United Kingdom
Co-authors: Eleftheria Zeggini

There is growing interest in the role of rare variants in multifactorial disease etiology, and increasing evidence that rare variants are associated with complex traits. Single SNP tests are underpowered in rare variant association analyses, so locus-based tests must be used. In the analysis of the association of rare variants and disease, there is a loss of power due to genotype misspecification. SNP quality scores are available for sequencing data, but they are rarely accounted for in rare variant analyses. In addition, the 1000 Genomes reference set contains variants with MAF as low as .01, which makes the imputation of rare variants now possible. A probability distribution for the genotype at each variant may be estimated using the imputation method of choice. We propose methods for rare variant analyses that take advantage of the extra information contained in quality scores derived from sequencing and probability distributions resulting from imputation.

E863: **Capturing the time-varying drivers of an epidemic with particle Markov Chain Monte Carlo algorithms**
Presenter: **Joseph Dureau**, London School of Economics, France
Co-authors: Konstantinos Kalogeropoulos, Marc Baguelin

Epidemics are often modelled using non-linear state-space models such as the "Susceptible-Infected-Removed" (SIR) system of ODEs, and extensions thereof. Inference is typically based on various observation regimes that include partial and noisy observations. As a consequence, the likelihood function is generally intractable, posing a challenging estimation and computational problem. In our work we consider stochastic extensions to the popular SEIR model, with parameters evolving in time, to capture unknown influences of changing behaviors, public interventions, seasonal effects, etc. Our models assign diffusion processes for the time-varying parameters, such as the geometric Brownian motion or smoother process such as integrated diffusions. Our inferential procedure is based on Markov Chain Monte Carlo methods (MCMC), and in particular on the particle MCMC algorithm, suitably adjusted to accommodate the features of this challenging non-linear stochastic model. We also implemented and compared various recently developed inference methodologies such as the Maximum likelihood via Iterated Filtering algorithm, and alternative MCMC schemes including hybrid Monte Carlo. We discuss the pros and cons of the different algorithms and illustrate their performance on simulated data. Moreover, we elaborate on the robustness and flexibility of the adopted model using real data from the 2009 influenza epidemic in London.

E793: **Sample size and asymptotic relative efficiency when using composite endpoints**
Presenter: **Moises Gomez Mateu**, Universitat Politecnica de Catalunya, Spain
Co-authors: Guadalupe Gomez Melis, Urania Dafni

In randomized clinical trials, composite endpoints are often used to assess the efficacy of a new treatment. The decision on whether to expand the composite endpoint T_1 with an additional endpoint T_2 is controversial. To assess the difference in efficiency between using logrank test Z based on T_1 or logrank test Z_* based on the composite endpoint $T_* = min\{T_1, T_2\}$, some authors base the strategy on the behaviour of the asymptotic relative efficiency of Z_* versus Z. Given that both tests Z and Z_* are asymptotically N(0,1) under the null H_0 and H_0^*, respectively, and asymptotically normal $N(\mu,1)$ and $N(\mu*,1)$ under a sequence of contiguous alternatives to the null hypothesis, their $ARE = (\mu_*/\mu)^2$. The goal of this work is to check if the usual interpretation of the asymptotic relative efficiency as the reciprocal ratio of sample sizes needed to attain a given power for a significance level is fulfilled when the two null hypothesis are not the same. To do so, we simulate the joint distribution of (T_1 ,T_2) by means of Frank's copula for different marginal distributions, association degrees, probabilities of observing T_1 and T_2 and anticipated treatment effects with respect to each endpoint.

ES11 Room B18 STATISTICAL MONITORING AND ITS APPLICATIONS II Chair: Rebecca Killick

E379: **Likelihood based estimation of the log-variance function with a change point**
Presenter: **Jib Huh**, Duksung Women's University, Korea RoK

Consider that the variance function or its νth derivative in a regression model has a change/discontinuity point at an unknown location. To use the local polynomial fits, the log-variance function, which breaks the positivity, is considered. The location and the jump size of the change point are estimated based on a one-sided kernel weighted local-likelihood function, which is provided by the χ^2-distribution of squared residuals. The whole structure of the log-variance function is then estimated using the data sets split by the estimated location. Asymptotic results of the proposed estimators are described. Numerical works demonstrate the performances of the methods with simulated and real examples.

E393: **On the modeling and estimation of the US health process and healthy life expectancy**
Presenter: **Ying Yang**, Tilburg University, Netherlands
Co-authors: Anja De Waegenaere, Bertrand Melenberg

The increase in life expectancy (LE) is an important phenomenon in an aging population. An important question is whether the increase in LE is associated with an increase in life years in good health (healthy life expectancy -HLE-). A contribution to the health and longevity literature is made in two ways. Firstly, we model the dynamics of self-assessed health, based on the Lee-Carter framework. We can use our model to predict the future HLE, and to quantify the (health) longevity risk. We find increasing trends in both LE and HLE, where the latter increases faster than the former, but with larger uncertainties. Moreover, the males' LE and HLE are lower than the females', but they increase faster, although with larger uncertainties. Secondly, we include observed variables, such as GDP, to model the health dynamics of various age groups. This leads to a substantial improvement in the model fit, where a part of the time trend in health (younger age groups) or even the full time trend (oldest age group) can be attributed to trends in the observed variables. The model extended with observed variables is very helpful to predict future LE and HLE under various scenarios.

E815: **Monitoring the mean value in the contaminated normal family of distributions**
Presenter: **Fernanda Otilia Figueiredo**, Faculdade de Economia da Universidade do Porto and CEAUL, Portugal
Co-authors: M. Ivette Gomes

The contaminated normal family of distributions is very useful for modeling data sets with heavy tails, for instance, in the areas of biometrics, biology, clinical chemistry and medicine, among others. In this study we propose the use of the total median statistic to estimate and monitor the mean value of such processes. This statistic is related to the bootstrap sample and to the sample of the order statistics. Several simulation studies allow us to conclude that the total median is more efficient than the sample mean, and also more robust to small deviations in the underlying data distribution. We consider several types of contaminated normal distributions and we compare the efficiency of the mean and the total median estimators for the process mean value. Finally,we implement control charts based on the total median statistic and we analyze its performance.

Saturday 17.12.2011 14:00 - 16:05 Parallel Session E – ERCIM

ES09 Room B18 PERSPECTIVES ON HIGH-DIMENSIONAL DATA ANALYSIS Chair: Ejaz Ahmed

E060: **Estimation by projection on confidence regions**
Presenter: **Pierre Alquier**, Universite Paris 7 (and other affiliation: CREST), France

A general method to build estimators in the case of high-dimensional data is discussed. The idea is basically to first build large confidence regions for the parameters, and then to perform successive projections on these confidence regions. The first step usually relies on concentration of measure inequalities. The well-known LASSO and the Dantzig selector may be seen as particular cases of this method. Other cases of interest include the correlation selector, and a new algorithm to select blocks of features. An application of this last algorithm to genomic data (CGH data) will be presented.

E131: **Clustering and classification of high-dimensional data via modified t-factor analyzers**
Presenter: **Paul McNicholas**, University of Guelph, Canada
Co-authors: Jeffrey Andrews

A family of twelve mixtures of modified t-factor analyzers is introduced. Modified factor analyzers were recently introduced within the Gaussian context but utilization of the multivariate t-distribution presents a more flexible modelling paradigm. The application of these models to high-dimensional data is discussed; including implementation in parallel via a coarse grain (MPI) approach. This novel family of models is applied within three paradigms: model-based clustering; model-based classification; and model-based discriminant analysis. Parameter estimation is carried out using an AECM algorithm and the BIC is used for model selection.

E360: **Mixing generalized ridge estimators**
Presenter: **J. Sunil Rao**, University of Miami, United States of America
Co-authors: Hemant Ishwaran

High dimensional regression problems naturally involve correlated predictors - in part due to an artifact of dimensionality. Given that generalized ridge estimators (GRR) were originally designed for situations that are ill-conditioned due to correlated predictors, we study them in the case where the number of predictors exceeds the sample size. We describe a novel geometric interpretation for GRR and show that GRR possesses a stability property in correlated settings. However, GRR is constrained to lie in a low dimensional subspace which limits its usefulness. To overcome this, we introduce a mixing GRR procedure using easily constructed exponential weights and establish a finite sample minimax bound for this procedure. We discuss the various terms appearing in this bound and what implications they have for practical implementation of the mixing procedure. We then describe two efficient algorithms and study their performance using a wide range of examples.

E499: **Regularization in finite mixture of regression models with diverging number of parameters**
Presenter: **Abbas Khalili**, McGill University , Canada
Co-authors: Shili Lin

Feature selection has become a fundamentally important problem in recent statistical literature. Often, in applications many variables are introduced to reduce possible modeling biases. The number of introduced variables thus depends on the sample size, which reflects the estimability of the parametric model. The problem of feature selection in finite mixture of regression models is considered when the number of parameters in the model can increase with the sample size. We propose a penalized likelihood approach for feature selection in these models. Under certain regularity conditions, our approach leads to consistent variable selection. We carry out a simulation study to evaluate the performance of the proposed approach under controlled settings. A real data on Parkinson's disease is also analyzed. The data concerns whether dysphonic features extracted from the patients' speech signals recorded at home can be used as surrogates to study PD severity and progression. Our analysis of the PD data yields interpretable results that can be of important clinical values. The stratification of dysphonic features for patients with mild and severe symptoms lead to novel insights beyond the current literature.

E519: **Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series**
Presenter: **Yulia Gel**, University of Waterloo, Canada
Co-authors: Peter Bickel

A "large p – small n" problem in a time series framework is addressed. Properties of banded regularization of an empirical autocovariance matrix of a time series process are considered. Utilizing the banded autocovariance matrix enables us to fit a much longer AR(p) model to the observed data than typically suggested by AIC, while controlling how many parameters are to be estimated precisely and the level of accuracy. We present results on asymptotic consistency of banded autocovariance matrices under the Frobenius norm and provide a theoretical justification on optimal band selection using cross-validation. Remarkably, the cross-validation loss function for banded prediction is related to the conditional mean square prediction error (MSPE) and, thus, may be viewed as an alternative model selection criterion. The proposed procedure is illustrated by simulations and application to predicting sea surface temperature (SST) index in the Nino 3.4 region.

ES15 Room B35 NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT II Chair: Carmen Cadarso

E218: **Software developments for non-parametric ROC regression analysis**
Presenter: **Maria Xose Rodriguez Alvarez**, Complexo Hospitalario Universitario de Santiago, Spain
Co-authors: Javier Roca-Pardinas, Carmen Cadarso-Suarez

The ROC curve is a fundamental technique in the characterization of the accuracy of continuous diagnostic tests. Regression approaches of either the test results (induced methodology) or the ROC curve itself (direct methodology) have become the usual methods for the assessment of covariate effects on the ROC curve. Recently, new non-parametric estimators for the conditional ROC curve have been proposed, based on both induced and direct modelling. We introduce a user-friendly software, called npROCregression, that implements these new approaches. The estimation procedure of both methodologies implies a high computational cost, especially as bootstrap methods are used for inference purposes. As a result, the programming language selected to implement these approaches was Fortran. However, to facilitate the use in practice, R was chosen as the user interface program. The software offers numerical and graphical output for the estimated conditional ROC curve, jointly with other summary measures of the accuracy, such as the area under the curve (AUC) or the generalized Youden index (YI). The software also provides the thresholds values based on the YI criterion and the criterion of equal sensitivity and specificity.

E226: **New method for evaluating the degree of exposure to radiation**
Presenter: **Pedro Puig**, Universitat Autonoma de Barcelona, Spain
Co-authors: Joan Francesc Barquinero

The main objective of biological dosimetry in radioprotection is to quantify the dose received in individuals who have been exposed to ionizing radiation. This is essential for predicting the derived health consequences; for both early effects, such as vomiting, skin injuries or haematopoietic depletion; and late effects like the development of cancer. Nowadays, the most widely used and accepted method is the analysis of the induced chromosome aberrations, in particular the analysis of the frequency of dicentrics observed in metaphases from peripheral blood lymphocytes. The $r-$th order univariate Hermite distributions are proposed to model the number of dicentrics in Biological Dosimetry. These families of distributions are introduced from Compound Poisson process modelling. To quantify the dose of a possible exposure to radiation, it is necessary to establish previous dose-effect calibration curves irradiating peripheral blood lymphocytes to a certain number of doses. Regression models appropriate for analyzing the number of dicentrics as a function of doses of radiation are presented, and an example of application is also given.

E290: **Geoadditive survival models for the identification of geographical patterns in coronary heart disease**
Presenter: **Francisco Gude**, Complexo Hospitalario Universitario de Santiago de Compostela, Spain
Co-authors: Mar Rodriguez-Girondo, Thomas Kneib, Carmen Cadarso-Suarez

Patients admitted to hospital with acute coronary syndrome (ACS) have a high risk of death or re-hospitalisation within the following months. In Spain, such patients admitted to district hospitals have traditionally been transferred to tertiary centres for investigation and treatment. Due to the large numbers involved and scarcity of tertiary centre beds, this results in long delays in transferring from district hospitals and may lead to differences in prognosis among different residence areas. In 2006, regional healthcare authorities of Galicia (Spain) developed a strategy to improve access to percutaneous intervention for ACS patients. The objective of this study was to determine geographical inequalities in survival of patients admitted to the coronary care unit of the Hospital Clinico Universitario de Santiago de Compostela (Galicia, Spain) with a diagnosis of ACS between years 2003 and 2010. We used structured geoadditive survival models controlling simultaneously for spatial dependency and possible nonlinear and time-varying effects of known cardiovascular risk factors. We also investigate how these spatial patterns have changed over time. In particular, we wish to assess whether this healthcare strategy led to decrease short- and long-term mortality in patients with ACS.

E325: **Multivariate methods for the integration of omics data of different types and different nature**
Presenter: **Alex Sanchez-Pla**, Universitat de Barcelona, Spain
Co-authors: Ferran Reverter, Esteban Vegas, Jose M. Oller, M. Carme Ruiz de Villa

As developments in high throughput technologies and bioinformatics have become more common and accessible, it is becoming usual to take different simultaneous approaches to study the same problem. In practice this means that different sets of data of different types (expression, proteins, metabolites...) may be available for the analysis of the same problem, highlighting the need for methods and tools to use them in a combined way. We discuss the application of two multivariate statistical approaches to facilitate the combined analysis of bio-molecular information: Multiple factorial analyses (MFA) designed to analyze sets of observations decribed by several blocks of variables and Kernel methods such as Kernel PCA Biplot and Kernel CCA that have emerged as powerful tools to explore biological data through dimensionality reduction. Using a public multi-omics dataset we show how these techniques can be used to perform reduction dimension and then visualize data of one type useful to explain those from other types. Whereas this is more or less straightforward when we deal with two types of data, it turns out to be more complicated when the goal is to visualize simultaneously more than two types particularly when one of these data types consists of biological knowledge.

E357: **Partial additive beta-binomial model for bounded outcome scores**
Presenter: **Inmaculada Arostegui**, Universidad del Pais Vasco (UPV/EHU), Spain
Co-authors: Vicente Nunez-Anton

Bounded outcome scores (BOS) are responses that are restricted to finite intervals. A BOS can be recoded from 0 to n, and then, it can be considered as grouped data for a dichotomous outcome, where correlation within individuals exists. Therefore, such data can be analyzed using logistic regression with random effects, as is the case in the beta-binomial regression (BBR) model. We propose an extension of the BBR model to a partial additive beta-binomial (PABB) model. Continuous covariates are incorporated into the model using smooth functions based on P-splines, whereas the discrete variables are transformed to dummy variables that are included in a parametric form. Both models were applied to the Short Form – 36 (SF–36) questionnaire where both, continuous and discrete covariates, were available. We fitted the PABB model using cubic splines and second order differences in the penalization parameter. Model fit was tested with the AIC criterion. The PABB was superior to the BBR model at least for one dimension of the SF–36. The PABB model is a powerful analytical technique for BOS, especially when there are continuous covariates available and a flexible approach may be required for better fitting purposes.

ES29 Room B33 ADVANCES IN OPTIMAL EXPERIMENTAL DESIGN Chair: Jesus Lopez Fidalgo

E288: **Adaptive designs for dose-response studies**
Presenter: **Nancy Flournoy**, University of Missouri, United States of America

The field of adaptive designs has recently exploded. Random stopping rules, sample size recalculation methods are well developed, and much is now known about adaptive randomization in the context of group comparisons. Adaptive treatment allocation methods in the dose-response setting are considered. The term "dose" can be replaced by any stimulus, such as a force applied to break a material or the difficulty of a test item. The framework is limited to binary responses for which the probability of response increases with "dose". The experimental goal is to estimate or select a quantile, or target, "dose". An important determinant of an adaptive design's behavior (Bayesian and frequentist) is whether they have long- or short-memory, referring to their use of past information to allocate a treatment to the current patient. Recent theoretical findings are reported. Then effect of memory on performance is described with an explicit example comparing a group up-and-down design with the continual reassessment method. Short memory designs are shown to excel in terms of convergence rates and reduced variance of predicted target dose with small sample sizes.

E119: **Nature-inspired metaheuristic algorithms for generating optimal experimental designs**
Presenter: **Weng Kee Wong**, UCLA, United States of America
Co-authors: Shin-Perng Chang, Weichung Wang, Ray-Bing Chen

Particle swarm optimization (PSO) is a relatively new, simple and powerful way to search for an optimal solution. The method works quite magically and frequently finds the optimal solution or a nearly optimal solution very quickly after a couple of iterations. There is virtually no assumption required for the method to perform well and the user only needs to input a few easy to work with tuning parameters. Using nonlinear regression models, we demonstrate that once a model and an optimality criterion are specified, PSO can generate many types of optimal designs quickly, including minimax optimal designs where effective algorithms to find such designs have remained elusive to date.

E151: **Optimal design for parameters of correlated processes**
Presenter: **Milan Stehlik**, Johannes Kepler University in Linz, Austria

The information approach to parameters of stochastic processes is discussed. This will enable us to define the information functionals for optimal design of parameters. We will discuss the recent results on continuity of correlation and illustrate its applications for financial and risk processes.

E152: **Optimal designs for the random effect logistic regression model with covariance structure**
Presenter: **Juan M. Rodriguez-Diaz**, University of Salamanca, Spain
Co-authors: Mayte Santos-Martin, Chiara Tomassi

The random-effect logistic regression model with covariance structure will be analyzed from the point of view of optimality. The analytical expression of the Fisher information matrix will be studied, and some results about the equivalence with the information matrix of the linearized model will be produced. This matrix depends on some integrals, and therefore some integral approximations are provided. The optimality will be studied using different optimality criteria (*D*-, *A*-, *c*-).

E579: **A maximum entropy sampling approach to adaptive design for spatial processes**
Presenter: **Henry Wynn**, London School of Economics, United Kingdom
Co-authors: Noha Youssef

Maximum Entropy Sampling (MES) is an approach to design which uses Shannon results in various ways to avoid the difficult inversion and integration which is necessary with a full Bayesian approach to spatial sampling. A challenge is to make the methods truly adaptive, that is dependent on previous response values, not just observation sites. We use a hierarchical inverse Wishart prior distribution for covariances combined with a cheap empirical Bayes method to update the Ψ parameter of the Wishart. The method is compared to a direct method which simply fits appropriate kernels and uses thresholding. The conclusion is that one can adapt to changes in local smoothness by taking more observations where the variability in the response is greater. The approach leads to a more general discussion of Bayesian learning and in particular the role of information theory.

ES31 Room G16 APPLIED STATISTICS II Chair: Agustin Mayo-Iscar

E622: **Intra-Day robust exchange rate forecasting**
Presenter: **Tiziano Bellini**, Universita Degli Studi di Parma, Italy

Price forecasting is very relevant for market players and is necessary to develop bidding strategies or negotiation skills in order to maximize trading benefits. The occurrence of outliers can lead to model parameter estimation biases, invalid inferences and poor forecasts. The detection of atypical observations, in addition, is crucial in the case of regime shift. Intraday prices availability allows improving traditional day-ahead forecasting. Then, relying on a general outlier detection method, we develop a forecasting approach which relies on prices available at different time grids. Concentrating on a currency database, we investigate how highest frequency data help forecasting lowest frequency prices and vice versa.

E767: **Robust methods for analysis of multivariate grouped data in R**
Presenter: **Valentin Todorov**, Unido, Austria

The routine use of robust statistical methods in a wide area of application domains is unthinkable without the computational power of today's personal computers and the availability of ready to use implementations of the already theoretically available algorithms. A unified computational platform organized as common patterns, which are called statistical design patterns in analogy to the design patterns widely used in software engineering, is available in the R programming language and environment. This framework provides most of the popular robust multivariate estimators for multivariate location and covariance, principal components and outlier identification. The framework is extended with algorithms targeting data sets which consist of two or more samples as in testing equality of means and in discriminant analysis: standard and robust Hotelling T^2 and Wilks' Lambda statistics; discriminant analysis based on robust covariance matrices as well as on projection pursuit approach; discrimination in high dimensions ($n \ll p$) using robust PLS and SIMCA methods. Particular attention is given to the graphical visualization methods providing means for better understanding the data structure, selecting the relevant variables and choosing the most suitable classifier. All considered methods are available in the R packages *robustbase*, *rrcov* and *rrcovHD*.

E593: **Robust estimates of the generalized loggamma distribution**
Presenter: **Alfio Marazzi**, University of Lausanne, Switzerland
Co-authors: Victor Yohai, Claudio Agostinelli

The generalized loggamma family of distributions is widely used to model highly skewed positive data. It depends on three parameters characterizing location, scale and shape and includes many common models, such as the normal model, the log-Weibull model, the log-exponential model, and the usual Gamma model. Usually, the parameters are estimated by means of the maximum likelihood principle, which provides fully efficient estimates when the observations adhere to the model but is extremely sensitive to the presence of outliers in the sample. We propose a family of robust estimates based on two steps: (a) an initial estimate which minimizes a τ scale of the differences between empirical and theoretical quantiles ($Q\tau$ estimate); (b) a weighted maximum likelihood estimate (WML), where the weights are based on a disparity measure between the model density and a kernel density estimate. The $Q\tau$ estimate can be considered as a robust version of quantile distance estimate. The WML combines both asymptotic full efficiency under the model and a high degree of robustness under outlier contamination.

E599: **Comparating methods for robust elliptical clustering, including the robust improper ML estimator**
Presenter: **Christian Hennig**, UCL, United Kingdom
Co-authors: Pietro Coretto

A simulation study will be presented that compares five different methods for finding elliptical clusters (four of them based on a Gaussian cluster prototype), including some new developments regarding the implementation and tuning of Hennig and Coretto's robust improper ML estimator for Gaussian mixtures, a method for robust model-based clustering based on adding a mixture component with an improper constant density over the real line. The other methods are an ML-estimator for Gaussian mixtures, Gaussian mixtures with uniform component over the convex hull of the data, the tclust method based on a well-known trimming, and mixtures of t-distributions. Implementation and tuning of these methods is reviewed as well, including some new developments. The main focus of the paper will be on the decisions that need to be made when designing comparative simulation studies in clustering. Apart from the choice of model setups from which to generate data, a key issue is the measurement of quality. Assuming that in clustering we are not mainly interested in parameter estimators but rather in the grouping of the points, it is necessary to define the "true clusters" and "true outliers" of the simulated data generating process. This is by no means trivial, particularly when comparing methods that estimate different underlying models.

E828: **Power distribution and dynamic range in PCM music signals**
Presenter: **Pietro Coretto**, Universita degli Studi di Salerno, Italy
Co-authors: Francesco Giordano

An audio waveform is digitally represented as a sequence of values proportional to the amplitude measured at fixed frequency with a certain bit precision. The latter is the so called PCM-quantization which is at the base of the so called "Red Book" standard (IEC60908) that defines the CDDA (compact disc digital audio). When the Red Book was introduced in 1980, it was soon considered a big step forward because of its higher dynamic range (DR) capabilities if compared with its old-days analog competitors. Quality of a recording is directly proportional to its DR. The

DR is a measure of how a given source is able to reproduce sound pressure variations, and as a such is a measure of power variations. Despite the DR capabilities of digital recordings there is a trend in compressing the DR at the mastering stage in order to achieve higher levels of perceived loudness. While there is no statistical literature on the subject, it is crucial to measure the DR consistently. We propose a measure of DR which is based on a nonparametric estimate of the power distribution of the stochastic component of the audio wave. While the stochastic component is obtained by filtering the signal estimated via nonparametric regression, its power distribution is approximated by a subsampling scheme. We derive both optimality and consistency results for estimated quantities, and we propose a statistic which is able to consistently detect DR compression in most situations.

ES38 Room B34 STATISTICS IN FUNCTIONAL AND HILBERT SPACES I Chair: Gil Gonzalez-Rodriguez

E351: **Functional PLS versus functional PCR through simulated data and chemometric applications**
Presenter: **Ana M. Aguilera**, University of Granada, Spain
Co-authors: Manuel Escabias, Cristian Preda, Gilbert Saporta

Multivariate calibration techniques as linear regression were extended over the last few years to find the underlying relationship between a scalar response variable and a functional predictor variable where realizations are curves observed at a finite set of points. In most cases the observed curves are functions of time. In chemometric applications such as spectroscopy, we often have observations of the spectrum of chemicals as a function of wavelength. The estimation of the functional parameter associated with functional regression models is affected by two major problems. First, the dimension of the curves usually exceeds the sample size. Second, the set of variables associated with the observed points are highly correlated (multicollinearity). In order to solve these problems principal component regression (PCR) and partial least squares (PLS) regression were also extended to the functional domain. B-splines expansions of the observed curves are considered in this work to estimate these models. A comparative study of the performance of these dimension-reduction techniques to estimate the functional parameter and to forecast the response is developed with different sets of simulated data and chemometric applications.

E362: **Comparison of two methods based on 3-dimensional position data**
Presenter: **Christian Ritz**, University of Copenhagen, Denmark
Co-authors: Emil Olsen

In biomechanical studies of equine motion inertial measurement units (IMUs) provide a relatively accurate, mobile, and cheap alternative to motion capture cameras and allow for continuous data collection of consecutive gait cycles. The method relies on a combination of accelerometers, gyroscopes and magnetometers, thus giving a full set of movement parameters and yielding 3-dimensional functional data. The underlying experimental design has a hierarchical structure with repeated runs per limb per horse, resulting in correlated functional data. Knowledge of accuracy and precision of these data is crucial to the clinical applications and decisions. The objective was to evaluate the accuracy of position estimates of limb mounted IMUs relative to a well-established gold standard procedure by means of a functional data approach. Several functional data analytic methods that extend univariate procedures for comparing two treatment groups will be reviewed. In particular, we will consider approximate ANOVA-like F-tests for comparing two functional data models. How to incorporate the hierarchical structure of the experiment will also be discussed. We will also propose an extension of the Bland-Altman plot for visually evaluating two measurement methods based on functional data.

E549: **Extending induced ROC methodology to the functional context**
Presenter: **Vanda Inacio**, Lisbon University, Portugal
Co-authors: Wenceslao Gonzalez-Manteiga, Manuel Febrero-Bande, Francisco Gude, Carmen Cadarso-Suarez

The receiver operating characteristic (ROC) curve is the most widely used measure for evaluating the discriminatory performance of a continuous diagnostic test. It is well known that, in certain circumstances, the marker's discriminatory capacity can be afected by factors, and several ROC regression methodologies have been proposed to incorporate covariates in the ROC framework. Until now, these methodologies are only developed in the case where the covariate is univariate or multivariate. We extend ROC regression methodology to the case where the covariate is functional, rather than univariate or multivariate. To this end and for the most appropriate analysis of a given data set, semiparametric and nonparametric estimators are proposed. A simulation study is performed to assess the performance of the proposed estimators. Methods are applied and motivated by a metabolic syndrome study in Galicia (NW Spain).

E614: **Factor models and variable selection in high dimensional regression**
Presenter: **Alois Kneip**, University of Bonn, Germany
Co-authors: Pascal Sarda

Linear regression problems where the number of predictor variables is possibly larger than the sample size are considered. The basic motivation is to combine the points of view of model selection and functional regression by using a factor approach: it is assumed that the predictor vector can be decomposed into a sum of two uncorrelated random components reflecting common factors and specific variabilities of the explanatory variables. It is shown that the traditional assumption of a sparse vector of parameters is restrictive in this context. Common factors may possess a significant influence on the response variable which cannot be captured by the specific effects of a small number of individual variables. We therefore propose to include principal components as additional explanatory variables in an augmented regression model. We give finite sample inequalities for estimates of these components. It is then shown that model selection procedures can be used to estimate the parameters of the augmented model, and we derive theoretical properties of the estimators. Finite sample performance is illustrated by a simulation study.

E626: **Statistical inference on the second order structure of functional data**
Presenter: **Victor Panaretos**, Ecole Polytechnique Federale de Lausanne, Switzerland
Co-authors: David Kraus

The problem of conducting inferences related to the second order structure of functional data is considered. Our study is motivated by the biophysical problem of understanding the effect of basis composition on the flexibility properties of short DNA curves called minicircles. We formulate the basic problem as a two-sample testing problem, and formalise the notion of flexibility through operators related to tensor products of the data curves. Two types of operator are considered, the covariance and the dispersion operator, the latter generalising the notion of covariance. The infinite-dimensional nature of the data is seen to result in the testing problem being an ill-posed inverse problem. We develop and study tests motivated by the Karhunen-Loeve expansion, employing appropriately spectrally truncated versions of the Hilbert-Schmidt distance of the empirical versions of the operators under study.

ES63 Room B20 HANDLING IMPRECISION IN GRAPHICAL MODELS Chair: Antonio Salmeron

E299: **Building decision trees from a nonparametric predictive inference perspective**
Presenter: **Andres Masegosa**, University of Granada, Spain
Co-authors: Joaquin Abellan, Rebecca M. Baker, Frank P.A. Coolen, Richard J. Crossman

The application of Nonparametric Predictive Inference for multinomial data (NPI-M) to classification is presented. The NPI-M applied on a dataset can produce a non-closed set of probabilities about a variable in study. An approximation of this model, called A-NPI-M, gives us a closed and convex set of probabilities. We have used these two models into a known scheme for building classification trees using imprecise probabilities and uncertainty measures, so far used only with the Imprecise Dirichlet model (IDM). We prove that with the procedure using the NPI based models, we can obtain similar accuracy via smaller trees than with the procedure used with the IDM. This implies a lower level of over-fitting and the obtaining of rules for classification representing a greater portion of the data. In our experimental study we have used similar procedures with classical split criteria to compare results. The datasets used in this study has the common characteristic that the class variable has a known number $K \geq 3$ of cases or categories, an important aspect of the application of the NPI based models.

E334: **The imprecise sample size Dirichlet model**
Presenter: **Serafin Moral**, University of Granada, Spain
Co-authors: Andres Cano, Manuel Gomez-Olmedo, Andres Masegosa

Bayesian approaches to inference can be strongly dependent on prior information when sample sizes are small. Imprecise probability theory has shown that it is possible to obtain sensible and meaningful inferences by considering as prior information a set of densities. The best known imprecise model for inference about a multinomial distribution is the Imprecise Dirichlet Model (IDM) which assumes as prior information the set of all Dirichlet distributions with a fixed equivalent sample size S. The IDM estimates the probability of any event as a probability interval. It satisfies a set of rationality properties and shows a good behavior when a sample with direct observations of the multinomial values is available. However, it has been shown to be too cautious when the multinomial values are observed through an indirect process giving rise to a likelihood function. It has also been shown to be useless to decide about independence statements by generalizing the Bayesian scores. In an alternative model, which is more informative with indirect observations and to decide about independence, is proposed. This model considers as prior information all the Dirichlet distributions with a sample size belonging to an interval $[S_1, S_2]$ and with a uniform vector of parameters α.

E404: **Data clustering using hidden variables in hybrid Bayesian networks**
Presenter: **Antonio Fernandez**, University of Almeria, Spain
Co-authors: Jose Antonio Gamez, Rafael Rumi, Antonio Salmeron

The use of the NB and TAN classifiers for unsupervised data clustering where the class variable is hidden is proposed. The feature variables can be discrete or continuous, as the conditional distributions are represented by MTEs. The optimal number of clusters and its probability distribution are estimated using a data augmentation technique. In order to add a new cluster in the model, a subset of the existing clusters is randomly chosen, and they are one by one split in two; finally, we choose to add the split cluster that causes the best likelihood in the model. For those records in the database wrongly assigned to a cluster, re-assign them to the cluster that causes the best likelihood in the model. In the TAN model the tree structure is re-learned in every iteration, in which the calculation of the conditional mutual information between two features, given the hidden variable, is needed. In the first step of the procedure, this conditional mutual information is approximated by the mutual information between the two feature variables, since at this step there are no values for the hidden variable.

E124: **Estimating CG-PDGs from incomplete data using an EM approach**
Presenter: **Antonio Salmeron**, Universidad de Almeria, Spain
Co-authors: Jose Antonio Gamez, Jens Dalgaard Nielsen

Conditional Gaussian probabilistic decision graphs (CG-PDGs) have been recently introduced as an extension to the discrete PDGs, by including continuous variables. The underlying assumption is that the joint distribution of the discrete and continuous variables is Conditional Gaussian. The estimation of CG-PDGs from incomplete data is considered. It is shown how a structural EM approach can be followed, that converges to the maximum likelihood solution. The existing algorithm for discrete PDGs is extended by introducing the necessary operators for splitting and merging nodes taking into account the restrictions imposed by the presence of continuous variables.

ES70 Room B36 STATISTICS FOR RANDOM INTERVALS AND RANDOM SETS Chair: Thierry Denoeux

E927: **On the estimation of a multiple linear regression model for interval data**
Presenter: **Marta Garcia-Barzana**, University of Oviedo, Spain
Co-authors: Ana Colubi, Erricos John Kontoghiorghes

Recently, a multiple linear regression model for interval data based on the natural interval-arithmetic has been proposed. Interval data can be identified with 2-dimensional points in $\mathbb{R} \times \mathbb{R}^+$, since they can be parametrized by its mid-point and its semi-amplitude or spread, which is non-negative. The population parameters can be expressed in terms of the covariance matrix of the involved random intervals and the sign of the regression coefficients, which only regards to the information of the mid-points of such random intervals. The least squares estimation becomes a quadratic optimization problem subject to linear constraints, which guarantee the existence of the residuals. Standard numerical methods are used to solve this optimization problem. Alternative estimators are discussed. Simulations have been performed in order to assess the consistency and the bias of the estimators. Furthermore, a real-life example is considered.

E569: **On p-boxes and random sets**
Presenter: **Sebastien Destercke**, Durham University, United Kingdom
Co-authors: Enrique Miranda, Matthias Troffaes

Pairs of lower and upper cumulative distribution functions, also called p-boxes, arise naturally in many problems of statistical inference, including expert elicitation for instance in cases where bounds are specified on the quantiles, and non-parametric statistical testing for instance when performing a Kolmogorov-Smirnov test. In addition, p-boxes on arbitrary total preordered spaces have been proven to be particularly useful in multivariate analysis. Random sets arise naturally in statistical problems where the observations themselves are set-valued. Their connection to p-boxes is well known in the discrete case. In this contribution, we explore the mathematical relationship between p-boxes on arbitrary totally preordered spaces and random sets. While doing so, we also study their limit behaviour, and arrive at convenient closed form expressions for calculating lower and upper expectations for p-boxes, and a particlar class of random sets, under suitable regularity conditions.

E859: **Bootstrap confidence sets for the Aumann mean of a random closed set**
Presenter: **Raffaello Seri**, Universita degli Studi dell Insubria, Italy
Co-authors: Christine Choirat

The objective is to develop a reliable method to build confidence sets for the Aumann mean of a random closed set as estimated through the Minkowski empirical mean. The literature on confidence regions for the mean of random sets is quite limited: Seri and Choirat (2004) introduced the topic and used the Central Limit Theorem to build a confidence set for the Aumann mean; Jankowski and Stanberry (2011) based their mean and the related confidence set on the oriented distance function. In this paper, we give a definition of confidence set for the Aumann mean generalizing the one in Seri and Choirat (2004) and then we describe a bootstrap algorithm to compute it. Then we investigate the accuracy of the procedure both theoretically, using functional Edgeworth expansions, and empirically, through a simulation study.

E545: **Autoregressive conditional models for interval-valued time series data**
Presenter: **Ai Han**, Chinese Academy of Sciences, China
Co-authors: Yongmiao Hong, Shouyang Wang

An interval-valued observation in a time period contains more information than a point-valued observation in the same time period. Interval forecasts may be of direct interest in practice, as it contains information on the range of variation and the level of economic variables. Moreover, the informational advantage of interval data can be exploited for more efficient econometric estimation and inference. This paper is the first attempt to model interval-valued time series data. We introduce an analytical framework for stationary interval-valued time series processes. We formally define the probabilistic space and fundamental concepts concerning random intervals. To capture the dynamics of a stationary interval time series process, we propose a new class of autoregressive conditional interval (ACIX) models with exogenous variables and propose estimation methods. We derive the consistency and asymptotic normality of the proposed estimators and exploit the relationships between them. Simulation studies show that the use of interval time series data can provide more accurate estimation for model parameters in terms of the mean squared error criterion. In an empirical study on asset pricing, we find that when return interval data is used, some bond market factors, particularly the default risk factor, are significant in explaining excess stock returns, even after the stock market factors are controlled in regressions. This differs from previous findings in the literature.

E761: **Hypothesis testing of regression parameters in a linear model for interval-valued random sets**
Presenter: **Angela Blanco-Fernandez**, University of Oviedo, Spain
Co-authors: Ana Colubi, Gil Gonzalez-Rodriguez

The statistical analysis of linear regression problems for interval-valued random sets has been addressed in the literature by means of different linear models based on interval arithmetic. Recently, a flexible model has been formalized, and the least-squares estimation of its parameters has been solved through a constrained minimization problem guarateeing the coherency with the interval arithmetic. Once the estimation process is done, some inferential studies for the linear model may be developed. In the interval scenario exact parametric methods are not feasible yet for inferential studies. This is due to the lack of realistic parametric models to describe the distribution of the interval random sets which have been shown to be widely applicable in practice. In this work, hypothesis testing of the regression parameters of this flexible linear model are presented, and an asymptotic procedure to solve the tests is proposed. The technique is illustrated by means of its application over a real-life example, and it is reinforced by running some simulation studies.

EP02 Room Chancellor's POSTER SESSION II **Chair: Klea Panayidou**

E755: **Explained variation for non-Hodgkin's lymphoma survival: A review and comparison**
Presenter: **Refah Alotaibi**, Newcastle, United Kingdom
Co-authors: Robin Henderson, Malcolm Farrow

Survival analysis is widely used in clinical and epidemiologic research to model time until event data. The purpose of many medical survival studies is to predict the process of the outcome of a disease. We investigate predictive accuracy, especially for non-Hodgkin's lymphoma (NHL) based on a study of 1391 patients followed for 10-years (1992 to 2002). In this area physicians usually use the international prognostic index (IPI) in to order to identify low and high risk patients. We consider whether it is possible to improve on the IPI. Four R^2-type measures are used to assess IPI performance for predictive purposes and how it compares with a prognostic index obtained from the data to hand. The measures are: a) the Nagelkerke information gain R^2 routinely given in R, b) an R^2- measure based on the Brier Score, c) the Schemper and Henderson survival process measure and d) the survival process measure explained rank variation based on Stare's paper. We review and compare these measures through simulation and an analysis of the NHL data. A practically important question is whether there are particular subsets of patients who are more easy or difficult to predict than others. We adopt Tian's methods to study this aspect.

E796: **Goodness-of-fit for the Moran-Downton exponential distribution**
Presenter: **Virtudes Alba-Fernandez**, University of Jaen, Spain
Co-authors: M.Dolores Jimenez-Gamero, Inmaculada Barranco-Chamorro

Bivariate and multivariate exponential distributions are applied widely in several areas such as reliability, queueing systems or hydrology. Among this kind of distributions, the Moran-Downton exponential distribution is a good choice, specially for modelling several hydrologic proccesses. In this work a test for testing goodness-of-fit test for the Moran-Downton exponential distribution is proposed. The test statistic exploits the analytically convenient formula of the characteristic function of this distribution and compares it with the empirical characteristic function of the sample. Large sample properties of the proposed test are studied. The finite sample performance is studied numerically. An application to a real data set is also included.

E781: **Modeling the effect of therapies by using diffusion processes**
Presenter: **Patricia Roman-Roman**, Universidad de Granada, Spain
Co-authors: Francisco Torres-Ruiz

An interesting problem in treatment protocols is to model the effect of a therapy in cancer xenografts. This problem can be addressed by considering a diffusion process that models the tumor growth and a modified process that includes, in its infinitesimal mean, a time function to model the effect of the therapy. For the case of a Gompertz diffusion process, when a control group and one or more treated groups are observed, a methodology to estimate this function is known. It has been applied to infer the effect of cisplatin and doxorubicin+cyclophosphamide in breast cancer xenografts. This methodology can be extended for other diffusion processes. However, it has an important restriction: it is necessary that a known diffusion process fits adequately the control group. We propose to consider a stochastic process for a hypothetical control group, so both real control and treated groups can be modeled by modified processes. Thus, the comparison between models allows us to estimate the real effect of the therapy. Finally, this methodology is applied to infer the abovementioned effects in breast cancer and the robustness of the procedure against the choice of stochastic model for the hypothetical control group is displayed.

E885: **An imputation method for mixed-type data using nonlinear principal component analysis**
Presenter: **Giancarlo Manzi**, Universita degli Studi di Milano, Italy
Co-authors: Pier Alda Ferrari, Alessandro Barbiero, Nadia Solaro

Imputation of missing data has always represented a problem for researchers from every field. A biased imputation may strongly affect research findings, leading to wrong conclusions. Furthermore, an imputation method may be satisfactory with certain types of data, but unsatisfactory with others. Recently, an imputation method based on the recursive use of nonlinear principal component analysis has been proposed for the specific case of constructing a composite indicator from ordinal data. This method alternates the nonlinear principal component analysis with the nearest neighbour method in order to detect the most appropriate donor for the imputation of missing categories of incomplete objects. In this work, we present an extension of this imputation method to the general case of constructing a composite indicator from mixed-type data, that is, data including both quantitative and qualitative variables. The problem of the evaluation of the most suitable distance between objects to be used in the

nearest neighbour method, together with the choice of the set of the most appropriate donors, is considered. The proposed procedure is implemented using the R software. A simulation study is also implemented to test the relative performance of the proposed method when compared with other methods used in similar situations.

E868: **Learning tree or forest graphical model structures**
Presenter: **Edmund Jones**, University of Bristol, United Kingdom
Co-authors: Vanessa Didelez

Bayesian learning of Gaussian graphical model structures is highly computer-intensive, because of the large number of possible graphs and the need for time-consuming approximations on non-decomposable graphs. One possibility is to restrict attention to forests, which are graphs that have no cycles, or trees, which are connected forests. It is believed that most real-world graphs are sparse, so the restriction to forests can be justified by the notion that sparse graphs are locally tree-like. It is shown how this notion can be made rigorous using random graph theory. Even with the restriction to trees and only 15 or 20 nodes, checking all the graphs is still too time-consuming. MCMC and other methods can be used to approximate the posterior distribution by moving through the space of possible graphs. For forests, the simplest move is to add or remove an edge, and for trees the simplest move is to reposition an edge. The poster will present systems for storing forests and trees so that these moves can easily be chosen uniformly at random, which is desirable to achieve good mixing. After each move, the information that is stored is updated locally.

E867: **Efficient Bayesian analysis by combining ideas from merge and reduce and meta-analysis**
Presenter: **Leo Geppert**, TU Dortmund University, Germany

Bayesian statistics have seen a rise in popularity in recent years. When dealing with a very large number of observations the standard MCMC methods cannot be used as the computational cost and necessary memory become prohibitive. We aim to find more efficient MCMC methods by combining ideas from the methods *merge and reduce* and *meta-analysis*. *Merge and reduce* is well-known to computer scientists. It has been used for a number of applications with very large amounts of data. *Meta-analysis* on the other hand is used in statistics to compare the results from different surveys and combine them to get an overall result. In our setting, we partition the data into different blocks, run an MCMC algorithm on each of them and only keep a model that characterises the MCMC sample. Two models on the same hierarchical level are combined as soon as possible. We propose two different approaches: a "summary value approach" that can be used when it is sufficient to calculate a few summary values from the MCMC sample and a "sample approach" in which both MCMC samples are merged and then reduced to one new sample.

E851: **The Bayes premium in the collective risk Poisson-Lindley and Exponential model with different structure functions**
Presenter: **Maria del Pilar Fernandez-Sanchez**, University of Granada, Spain
Co-authors: Agustin Hernandez-Bastida

In actuarial science premium calculating problems constitute an interesting issue. The Bayes premium let join the prior information and claim experience and, accordingly, it is a very adequate approximation to the problem in question. We consider a collective risk model in which the primary distribution is a Poisson-Lindley distribution and with an Exponential distribution as the secondary distribution. Under a Bayesian point of view, the parameters of interest of the problem can be estimated using our state of knowledge about them. Here, we shall consider the Gamma density as the natural choice for the parameter of the secondary distribution while Beta distribution is a common prior distribution for the parameter of the primary distribution. We propose two prior distributions as an alternative to the Beta distribution, namely, the triangular and standardized two-sided power and analyze the consequences of using these proposed distributions instead of the Beta distribution in the value of the collective and Bayes premiums. To compare, with a numerical analysis, the results obtained from the three structure functions we rank the distributions in terms of their relative entropy with respect to the Uniform distribution. The premiums obtained from this ranking are compared and it is observed that they maintain the same position as in the relative entropy of the three structure functions.

E849: **Statisticians and bibliometric laws**
Presenter: **Silvia Salini**, University of Milan, Italy
Co-authors: Francesca De Battisti

Starting from the scientific production of Italian statisticians, we will try to study the laws of bibliometric distributions in statistics. Does the scientific productivity of statisticians follow Lotka's law? Is it true that a high level of productivity is achieved by few authors? Lotka's law, first presented in 1926, is really still the law of reference? It states that the number of authors making n contributions is about $1/n^2$ of those making one; and the proportion of all contributors, that make a single contribution, is about 60 percent. A lot of authors in the past have studied the accordance with Lotka's law in various fields and some variations and adaptations of it. For bibliometric variables, what are the dimensions that can modify it: the reference years, the average of products for author, etc? Is it possible by a simulation study to identify a bibliometric law for the production of scientific researchers? Does the scientific production of statisticians, which, based on our results, does not follow Lotka's law, follow other laws of distribution? Are the citations, the number of core journals, the distribution of topics and words in a text coherent with the known distribution laws (Bradforda's law, Zips's law, etc.)?

E963: **Empirical comparison of robust portfolios' investment effects**
Presenter: **Bartosz Kaszuba**, Wroclaw University of Economics, Poland

Numerous research concerning application of estimation methods, other than classic methods in the portfolio theory, frequently reveal better qualities of the newly-proposed methods. Regrettably, the comparison is usually made for a small group of methods, portfolios are created on the basis of a selected group of assets and, first of all, the comparison methods are frequently considerably different. The purpose is to verify the investment effects achieved by portfolios created using robust methods. The research suggests a subsampling method, where instead of a random selection of the rate of return, shares were drawn, which were then used to create rolling portfolios, whereupon the out-of-sample performance of these portfolios was analysed. The research were conducted on the basis of actual data from the Warsaw Stock Exchange Index, whereas the portfolios under consideration were five-element portfolios with short selling constraints. The comparison was made using affine equivariant robust estimators of mean and covariance matrix (S, MM, MVE, MCD, SDE), pairwise covariance matrix (Spearman, Kednall, QC, OGK, I2D), as well as robust portfolio estimators (M, S, LTS, LMS, LAD). Moreover, a comparison between properties of portfolios created using other methods, such as DCC-GARCH, CCC-GARCH, Shrink, Ledoit-Wolf, CVaR and Student was made.

EP01 Room Chancellor's POSTERS SESSION I Chair: Klea Panayidou

E638: **Structural equation models based on covariance vs partial least squared to develop spatial indices**
Presenter: **Maria Luisa Rodero Cosano**, Universidad de Cordoba, Spain
Co-authors: Carlos Ramon Garcia Alonso

Bayesian Networks (BN) have been found useful in designing indicators when enough expert knowledge exists on the domain under study. In this framework, there are two methodological approaches based on Structural Equation Models (SEM): i) covariance-based and ii) variance-based (Partial Least Squares, PLS). This research compares the results obtained by both approaches (in addition to a factor analysis) in the development of a synthetic spatial index for deprivation. The BN includes the following domains: education, income, employment, housing, utilities and structures

and, finally, health care. Variable values have been obtained from statistical sources. Results are: i) three spatial-based models for the deprivation index and ii) the index for each spatial unit (770 municipalities in Andalusia, Spain). Both the structure of the models and their explained variance are analysed. The PLS approach (exploratory model) is flexible and allows us to improve the BN. On the other hand, the covariance-based model (more restrictive) confirms that the former model is reliable. Results for each spatial unit have been spatially analyzed (autocorrelation scores) to determine their degree of spatial agreement. Both SEM-based models show a high degree of agreement (kappa index and Intraclass Correlation Coefficient) but the factor analysis does not.

E655: **Linear estimation based on covariance using uncertain observations featuring random delays and packet dropouts**
Presenter: **Josefa Linares-Perez**, Granada University, Spain
Co-authors: Raquel Caballero-Aguila, Aurora Hermoso-Carazo

In networked systems, the measured system output is transmitted to a data processing centre producing the signal estimation. The unreliable network characteristics can produce uncertain observations (outputs containing noise only) and, moreover, eventual random transmission delays and/or packet dropouts can occur due to different causes, such as random failures in the transmission mechanism, accidental loss of some measurements or data inaccessibility at certain times. Networked systems simultaneously featuring these three types of random uncertainties in the measurements are considered. On the one hand, it is assumed that each real observation may not contain the signal to be estimated, uncertainty which is modelled by Bernoulli random variables representing the presence or absence of the signal in the observations. On the other, it is assumed that one-step delays and packet dropouts occur randomly in the transmission; this situation is modelled by introducing, in the measurement model, sequences of Bernoulli variables whose values indicate if the current measurement is received, delayed or dropped-out. Under the assumption that the signal evolution model is unknown, recursive least-squares linear estimation algorithms are derived by an innovative approach, requiring only the covariances of the processes involved in the observation equation and the uncertainty probabilities.

E702: **Distributed fusion filter for systems with markovian delays**
Presenter: **Maria Jesus Garcia-Ligero**, Granada, Spain
Co-authors: Aurora Hermoso-Carazo, Josefa Linares-Perez

Due to its extensive application, the signal estimation problem using measurement data derived from multiple sensors has received great attention. The fusion of the information provided by the different sensors is generally processed by two different methods: centralized and distributed filters. The computational advantages provided by the distributed filter lead us to consider this method versus the centralized. On the other hand, in wireless communication networks an unavoidable problem is the existence of errors during the transmission, which can lead to delays in the arrival of the measurements. These delays have been usually modeled by independent Bernoulli random variables. However in real communication systems, current time delays are usually correlated with previous ones; a reasonable way to model this dependence is to consider the random delay as a homogeneous Markov chain. The signal estimation problem from multiple sensors with delays modeled by homogeneous Markov chains is addressed. In this context, assuming that state-space model is unknown, we derive local least-squares linear estimators for each sensor using the information provided by the covariance functions of the processes involved in the observation equations. The distributed fusion filter is obtained as linear combination of the local filters using the least- squares criterion.

E706: **Recursive estimation algorithm from measurements with upper-bounded random delays**
Presenter: **Aurora Hermoso-Carazo**, Universidad de Granada,, Spain
Co-authors: Raquel Caballero-Aguila, Josefa Linares-Perez

The least-squares linear estimation problem using covariance information in discrete-time linear stochastic systems is analyzed by assuming bounded random observation delays. This is a realistic assumption, for example, in networked control systems and sensor networks, where random time delays usually occur in data transmission. A new model to describe such random delays is constructed by including in the system description some sequences of Bernoulli random variables, whose parameters are known and represent the delay probabilities. Conditions about these Bernoulli variables guarantee that the largest time delay is upper-bounded. Based on this new model, a recursive algorithm, including the computation of predictor, filter, and fixed-point smoother, is obtained by an innovate approach. To measure the performance of the estimators, the estimation error covariance matrices are also calculated. The proposed estimators do not need to know if a particular measurement is delayed or updated, and only depend on the delay probabilities at each sampling time. Moreover, the derivation of the estimation algorithm does not require full knowledge of the state-space model generating the signal to be estimated, but only information on the covariance functions of the processes involved in the observation equation.

E723: **Steps of a recursive partitioning algorithm**
Presenter: **Betul Kan**, Anadolu University, Turkey
Co-authors: Berna Yazici

Recursive Partitioning is a useful tool in data mining. It is based on the splitting of the dependent variables into subgroups. It is very common in practice to construct a recursive splitting model; however, each algorithm is with its own unique details. In this study, to construct a regression tree a recursive partitioning algorithm is used. The idea is to develop a tree which is a combination of subgroups where the data is modelled in each partition. There are stopping rules generated in the algorithm when it is infeasible to continue splitting. First, the algorithm is run to obtain the relation between one independent variable and the response variable. Additively, the algorithm is modified for the cases when two independent variables exist. Basically, the algorithm treats each observation as a potential splitting point, so called knot. Function $G()$ searches for potential subregions with the corresponding splitting point. Function $F()$ provides one sibling and calculates some criteria for the possible partitions. The procedure is repeated in each of the children nodes, until there are enough observations otherwise the recursion stops. The stopping rule is set to a fixed number of observations that is the minimum sample size of the nodes.

E751: **Software in R and computational methods for branching processes**
Presenter: **Manuel Molina**, University of Extremadura, Spain
Co-authors: Manuel Mota, Alfonso Ramos

Inside the general setting of stochastic modeling, branching processes are widely used as appropriate mathematical models describing the probabilistic behavior of systems whose components after a certain life period reproduce and die. Branching process theory has become an active research area of interest and applicability to such fields as biology, demography, ecology, epidemiology, genetics, population dynamics, and others. They have especially played a major role in modeling general population dynamics. We are particularly interested in the study of stochastic models describing the behavior of sexual reproduction populations, where females and males coexist and form couples. To this end, several classes of two-sex branching processes have been studied and some results about their probabilistic and inferential theories established. In order to check the accuracy concerning the proposed models and their derived results, it is necessary to implement specific software and to derive appropriate computing techniques. We will present the software in R we have developed. Also, we will provide a computational method to determine the highest posterior density credibility sets for the main parameters involved in the probabilistic model. By way of illustration, we will include some applied examples in population dynamics.

E752: **Approximate Bayesian computation methods for branching models in genetic context: application to X-linked genes**

Presenter: **Manuel Mota**, University of Extremadura, Spain
Co-authors: Miguel Gonzalez, Cristina Gutierrez, Rodrigo Martinez

Branching Processes have been shown to be useful models in the field of Population Genetics. In particular, some Multitype Bisexual Branching Processes can explain accurately some aspects related to the extinction-fixation-coexistence of some alleles of sex-linked genes. In this work, we study one of those processes as an appropriate model to investigate such issues for X-linked genes. This process, denominated X-linked Bisexual Branching Process, seems to be too complicated for an in-depth theoretical inferential study. Even computational techniques like MCMC methods could be difficult to apply due to problems of knowing the likelihood function. In this context, an interesting alternative to these methods is the application of likelihood-free inference techniques, often known as Approximate Bayesian Computation (ABC) methods. Specifically, assuming the number of females and males of each phenotype in several generations is available, a rejection inference algorithm is proposed to obtain some estimates for the posterior densities of the main parameters. Also, a discussion about the choice of the prior distributions, the metric or the tolerance is carried out. Finally, some extensions of the basic ABC rejection algorithm are investigated.

E754: **Goodness-of-fit tests for the multivariate skew normal distribution**

Presenter: **Simos Meintanis**, University of Athens, Greece
Co-authors: Kostas Fragiadakis

Goodness-of-fit tests for the family of skew-normal distributions in arbitrary dimension have been proposed in the literature. These tests are based on the moment generating function or on the cumulative distribution function of suitably standardized data. Although a comparison of those tests has been made in univariate case, this is not the case in higher dimensions. A wide comparison in multivariate case is presented in order to study the behavior of those tests with real and simulated data.

Saturday 17.12.2011 14:00 - 16:05 Parallel Session E – CFE

CSI03 Room Beveridge RECENT DEVELOPMENTS IN ECONOMETRICS Chair: Stefan Mittnik

C564: **Forecasting the implied volatility surface dynamics for CBOE equity options: Predictability and economic value tests**
Presenter: **Massimo Guidolin**, Manchester Business School, MAGF, United Kingdom
Co-authors: Alejandro Bernales

It is investigated whether the dynamics in the volatility surface implicit in the prices of individual equity options traded on the CBOE contains any exploitable predictable patterns. In particular, we examine the possibility that the dynamics in the volatility surface implicit in S&P 500 index options may be associated and forecast subsequent movements in the implied volatility surface characterizing individual equity options. We find a strong relationship between equity and S&P 500 index option implied volatility surfaces. In addition, we discover a remarkable amount of predictability in the movements over time of both equity and stock index implied volatilities. We show that the predictability for equity options is increased by the incorporation in the model of recent dynamics in S&P 500 implied volatilities. Similarly, when we examine the economic value of these predictability patterns (by proposing and simulating trading strategies that exploit our 1-day head forecast of implied volatilities), we report that delta-hedged and straddle portfolios that take trade on the entire implied volatility surface and across all contracts examined produce high risk-adjusted profits which are maximum for the model that takes into account the feedback from past market implied volatility changes to subsequent dynamics in individual equity options implicit volatilities.

C729: **Estimation and inference for impulse response weights from strongly persistent processes**
Presenter: **Richard Baillie**, Michigan State University, United States of America
Co-authors: George Kapetanios

The problem of estimating impulse response weights (IRWs) and their appropriate confidence intervals, from processes that may be strongly dependent, is considered. We compare three main approaches for estimating IRWs, namely, QMLE, a two step estimator that uses a semi parametric estimate of the long memory parameter in the first step, and, finally, an estimator from fitting an autoregressive approximation. We show that the parametric bootstrap is valid under very weak conditions, including non Gaussianity. This allows making inference on IRWs from possibly strongly dependent processes. We also propose, and theoretically justify, a semi-parametric sieve bootstrap based on autoregressive approximations that can be used for inference on IRWs. We find that estimates of IRWs based on autoregressive approximations and also confidence intervals of IRWs based on the sieve bootstrap generally have very desirable properties, and are shown to perform well in a detailed simulation study. The paper also includes an extensive and detailed empirical application on inflation and real exchange rate data.

C960: **Simulation-based predictive analysis for 3 key 21-st century issues**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Peter de Knijff, Lennart Hoogerheide, Koene van Dijk

Using technological advances from the 'computational revolution', a novel simulation based Bayesian procedure is presented that can be used for efficient and robust posterior and predictive analysis of some key issues in the 21-st century. The issues analyzed are possible effects of education on earned income observed in microeconomic data; risk evaluation of extreme losses in financial data; temporal causality in networks as measured using functional brain-imaging; and tandem repeat patterns in DNA data from persons in Africa, Asia and Europe. The results indicate a small step towards: better understanding of policy effects observed in the education-income issue; improved insight in risk of extreme events; possible applications for the assessment of brain changes due to normal aging and neurological disease; and more accurate analysis of migration patterns from early history as well as modern times.

CS04 Room Senate APPLIED FINANCIAL ECONOMETRICS Chair: Christopher Baum

C132: **Equity premia and state-dependent risks**
Presenter: **Michel Normandin**, HEC Montreal, Canada
Co-authors: Mohammed Bouaddi, Denis Larocque

For the first time in the literature, the CCAPM is extended to establish the empirical relations between equity premia and state-dependent consumption and market risks. These relations are derived from a flexible, yet tractable, mixture distribution admitting the existence of two regimes, rather than the usual normal distribution. Focusing on the market return, we find that the consumption and market risks are priced in each state, and the responses of expected equity premia to these risks are state dependent. Extending to various portfolio returns, we show that the responses to downside consumption risks are the most important, are almost always statistically larger than the responses to upside consumption risks, and are much larger for firms having smaller sizes and facing more financial distresses.

C142: **A unifying approach to the empirical evaluation of asset pricing models**
Presenter: **Francisco Penaranda**, Universitat Pompeu Fabra, Spain
Co-authors: Enrique Sentana

Two main approaches are commonly used to empirically evaluate linear factor pricing models: regression and SDF methods, with centred and uncentred versions of the latter. We show that unlike standard two-step or iterated GMM procedures, single-step estimators such as continuously updated GMM yield numerically identical values for prices of risk, pricing errors, Jensen's alphas and overidentifying restrictions tests irrespective of the model validity. Therefore, there is arguably a single approach regardless of the factors being traded or not, or the use of excess or gross returns. We illustrate our results with the Lustig-Verdelhan currency returns.

C627: **Testing uncovered interest rate parity using libor**
Presenter: **Muhammad Omer**, University of Groningen, Netherlands
Co-authors: Jakob de Haan, Bert Scholtens

Uncovered Interest Parity (UIP) over short-term horizons is tested using LIBOR interest rates for a wide range of maturities. For the short-term horizon, UIP is rejected due to frictions; however, to date no study addresses whether UIP holds if frictions are minimal. As some of the frictions arise when individual domestic debt markets are assumed homogenous, LIBOR provides a setup where most of the known frictions responsible for the failure of UIP are absent. LIBOR is available in ten currencies and fifteen maturities. We have used seven currency specific LIBOR rates, each with fourteen maturities. To estimate the UIP relationship we have used panel unit root and cointegration techniques. We find a positive slope coefficient indicating that UIP holds for the short-term horizon, when market specific heterogeneity is controlled. Also, we have found support for the traditional negative slope coefficients reported in literature. Moreover, our result highlights that the inferences based on one or two maturities on generalization, may become misleading. The use of several maturities helped us to identify when UIP holds and therefore, we argue for use of extended maturities in drawing inferences. Furthermore, our estimates imply that the speed of adjustment of exchange rate toward its long-run

equilibrium is proportional to the maturity of the underlying instrument. This finding does not support the efficient market hypothesis and we expect further deliberation on this issue, going forwards.

C471: **Fitting an unobserved components model to the VLCC tanker sector**
Presenter: **Anna Merika**, Deree College, Greece
Co-authors: Andreas Merikas

The maritime industry is highly volatile and freight rate determination is essential for investment decisions. We focus on building an unobserved components model to explain freight rate determination in the VLCC sector during 1993(1)-2011(7).We claim that the increase in market concentration observed post 1993 has an impact on the formation of freight rates in the industry. Moreover, we substantiate a positive relationship between freight rates and market concentration over the period examined. Thus, we claim that among other factors, expectations regarding the future course of concentration are critical for forecasting the direction of freight rates.

C800: **Quantifying the estimation error in market risk measures: Delta method vs. re-sampling techniques**
Presenter: **Carl Lonnbark**, Umea University, Sweden

The standard measure of market risk is the Value at Risk (VaR). In fact, regulators impose on financial institutions to meet capital requirements based on VaR. Internal approaches are allowed but institutions are penalized by having to set aside more capital if the VaR is exceeded too often. Another measure that is gaining increasing popularity is the expected shortfall (ES). VaR and ES estimates are often reported as if they were known constants. Such important economic variables should reasonably be reported with some measure of precision, which may also help to improve the design of the penalty system mentioned above. One popular approach to deal with parameter uncertainty in case of finite samples is to employ re-sampling techniques, which has obvious disadvantages in terms of computing time. In this respect the so-called delta-method is preferable, but, being an asymptotic approximation, it comes with other disadvantages. In a quite detailed simulation study we compare the two approaches for the popular estimation technique known as filtered historical simulation. We find quite substantial differences. An application for the major stock market indices of the world is included.

CS11 Room Woburn MODELLING WITH HEAVY TAILS: COMPUTATIONAL ISSUES Chair: Wojtek Charemza

C098: **Computational problems for multivariate stable laws**
Presenter: **John Nolan**, American University, United States of America

There are now reliable computational methods for working with univariate stable distributions. However, there is a need in finance, economics, and other fields for multivariate models that deal with heavy tails. This is a challenging problem due to the complexity of the possible dependence structures and the lack of explicit formulas for multivariate stable densities. We will briefly review existing methods and then describe new theoretical results and methods for computing with multivariate stable laws. The problems are formidable for higher dimensions, except for a few special cases, e.g. elliptically contoured stable laws.

C230: **Modeling and simulation with tempered stable laws**
Presenter: **Mark M. Meerschaert**, Michigan State University, United States of America

A tempered stable Lévy process smoothly transitions between a stable Lévy motion at early time, and a Brownian motion at late time. It is the basic building block in the CGMY model from finance, used to represent the "semi-heavy" tails seen there. Tempered stable laws have also become popular in applications to geophysics, where power laws are ubiquitous. The tempered stable process arises as a random walk limit, where the power law jumps are exponentially cooled, so that they have moments of all orders. A specific scheme for constructing these random walks will be described. The transition densities of this process solve a fractional variant of the diffusion equation. A special case is the tempered stable subordinator, which is related to time-fractional diffusion. A brief introduction to these fractional calculus models will be provided. Tempered stable random variables can be simulated by an exponential rejection method, which will be described in detail. One method for parameter estimation computes the conditional MLE for the largest observations, assuming a tempered power law model, similar to the Hill estimator. Some examples will be given, using data from geophysics. Extensions to higher dimensions will be discussed.

C295: **Fast calculation of PDFs of multi-factor Levy processes with exponentially decaying tails**
Presenter: **Marco de Innocentis**, University of Leicester, United Kingdom
Co-authors: Svetlana Boyarchenko, Sergei Levendorskii

Very fast numerical realizations of the inverse Fourier transform -parabolic and hyperbolic inverse Fourier transform- are suggested, and applied both to calculation of pdfs of several standard families of multi-factor Levy processes with exponentially decaying Levy densities. New realizations are several times faster than the standard realizations if time increment is not small and/or the process is of infinite variation. For processes of finite variations, especially for Variance Gamma processes, and for small time increments, new realizations are thousand times faster. For time increments of order a day, the standard realizations do not work even in one-factor case, whereas the new realizations give results with the relative error of order a hundredth of a percent and better in a small fraction of a second, if the number of factors is not large.

C378: **Heavy tailed time series: estimation and numerical issues for dependent observations**
Presenter: **Svetlana Makarova**, University College London, United Kingdom
Co-authors: Wojciech Charemza, Christian Francq, Jean-Michel Zakoian

Some practical problems related to computations and applications of the quasi marginal maximum likelihood estimator of alpha stable stationary process are tackled. Different estimators of the long run covariance matrix are compared for their finite-sample accuracy. In particular, a size distortion of the statistics based on these matrices (e.g. Student $t-$ratios) for various alpha-stable distributions under the changing strength of the dependences is investigated. The theoretical and simulation results are accompanied by the empirical study of the world stock market indices. For the stock market returns the market inefficiency has been evaluated tested through the maximum likelihood estimation of the characteristic parameter (alpha) and long-run covariance matrices.

C777: **On the distribution of European sovereign bond returns: Empirical evidence**
Presenter: **Christian Lau**, Martin Luther University Halle-Wittenberg, Germany
Co-authors: Christian Gabriel

Due to recent turbulences in the European debt capital market, one might ask for consequences for investors. We examine which distribution assumptions hold for European government bond returns in the period from 1999 to 2011. Returns of government bonds of several European countries, inside and outside the EMU, and the USA are analysed. We fit the data to Gaussian, Student's t, skewed Student's t and stable distributions. For the majority of the data we find that the Gaussian distribution is not flexible enough to describe government bond returns, since they are skewed and heavy-tailed. The skewed Student's t and the stable distribution fit the European government bond returns best. The results are robust to a variety of goodness of fit statistics.

CS19 Room Court LONG TERM RISKS Chair: Dominique Guegan

C762: **Historical risk measures as predictors on several markets**
Presenter: **Wayne Tarrant**, Wingate University, United States of America

The efficacy of using different risk measures as predictors across several different stock market indices and classes of commodities is observed. We use both the Value at Risk and the Expected Shortfall on each of these data sets, noting the differences in what the two measures purport to tell us. We also consider several different durations and levels for historical risk measures. Through our statistical results, we make some recommendations for a robust risk management strategy that involves using multiple historical risk measures. We contrast this to the use of spectral risk measures, explaining why the use of multiple risk measures is preferable.

C514: **Coherent risk measure in the long run, an operational risk application**
Presenter: **Ghassen Rahoui**, Universite Paris I Pantheon - Sorbonne - AON, France
Co-authors: Dominique Guegan, Bertrand Hassani

The quantification of the long term risk exposure has generated a large number of theoretical frameworks. Works on risk expected utility enabled the modelling of agent behaviour towards risk, which was firstly introduced through the moment-based risk measures. Since then the axiomatic approach to risk measures has expanded resulting in the emergence of diverse theoretical frameworks with a variety of interconnections and properties. Some of these late measures fulfil the coherence property and present an interesting alternative to the usual Value-at-Risk risk measure framework. We are particularly interested in spectral risk measures involving a weighted average of loss quantiles and distortion risk measures based in the deformation of loss objective probability distribution. We compare the axiomatic construction of these risk measures, explicit their respective properties and fill the gap between the two formulations through an equivalence relation detailed in the late part of the paper. Finally, based on these two frameworks we suggest an alternative long term coherent risk measure with controlled tail risk.

C512: **Operational risk: a long-term modeling**
Presenter: **Bertrand Hassani**, Universite Paris I Pantheon - Sorbonne - AON, France
Co-authors: Dominique Guegan, Ghassen Rahoui

In their last released paper, the Basel Committee for Banking Supervision provides precise guidelines to model operational risks. Nevertheless, in our opinion, these guidelines are not reflecting the reality of operational risks behaviour in the long term. Therefore, we propose the following solutions to deal with this particular kind of risks. First, to be compliant, we have to use five-year information sets. However, we do not know if the oldest incidents are still worthwhile to model a loss distribution, and vice versa, we do not know if such a set is sufficient, so we suggest a dynamic approach allowing more flexible models. Second, taking into account dependencies, we provide a robust and conservative way to deal with dependencies between the Basel matrix cases, using Vine Copulas. Third, we propose solutions to deal with the autocorrelation in a severity chronicle, improving the traditional Loss Distribution Approach accuracy. And last but not least, we propose to change the risk measure pertaining to a capital charge, and suggest replacing the traditional Value-at-Risk at 99.9% by a consistent expected shortfall.

C511: **Market risk aggregation using pair-copulas**
Presenter: **Fatima Jouad**, Universite Paris 1 / AXA GIE, France
Co-authors: Dominique Guegan

The advent of the Interval Model Approval Process (IMAP) within Solvency II and the desirability of many insurance companies to gain approval has increased the importance of such topics as risk aggregation in determining overall economic capital level through the calculation of a risk measure, the VaR(99.5%). The most currently used approach for aggregating risks is the variance-covariance matrix approach. Actually, linear correlations are used to describe the dependency. Although being relatively simple and well-known, linear correlations fail to model every particularity of the dependence pattern between risks. In this paper, we propose a nested copula-based model for aggregating market risks in order to calculate the economic capital needed to withstand both expected and unexpected future losses. This capital will be determined by computing a daily VaR(99.5%), a yearly VaR(99.5%) and a long-term VaR(99.5%), i.e. 40-year-VaR(99.5%).

C313: **Alternative modeling for long term VaR**
Presenter: **Dominique Guegan**, University of Paris 1 Pantheon - Sorbonne, France
Co-authors: Xin Zhao

A new modeling to compute financial risk in long term is proposed. Particularly, we suggest a dynamic approach to estimate Value-at-Risk and Expect shortfall in long term, which takes into account the persistence phenomena observed in financial markets, links with volatility clustering feature of financial series and also concerns with the occurrence of extreme events in a relative long time span. We firstly work in an univariate setting, we show the performance of the dynamic approach by estimating VaR from 1-day to 1-year with five selected stock markets' return series. Other risk measures will be considered later. We will also extend the work to a multivariate framework.

CS37 Room Bloomsbury TRENDS, WAVES, SEASONS, CYCLES AND SIGNALS Chair: Stephen Pollock

C280: **Time-varying trend of financial volatilities and its correlation with macroeconomic variables**
Presenter: **Ray Chou**, Academia Sinica, Taiwan
Co-authors: Norden Huang, Dan Li

A new method of fitting time-varying trend of volatility using the Empirical Mode Decomposition (henceforth EMD) of the Hilbert-Huang Transform is introduced. The method is highly adaptive in fitting nonlinear and nonstationary structures in trend and has been proven successful in dealing with nonlinear and nonstationary time series in many fields, e.g., physics, engineering, biomedical among others. Unlike the difficulty of trend identification in traditional time series models, the trend can be defined unambiguously by EMD. For example, the lowest frequency component can be defined as the "trend" component of the series which is monotonic, or with unique extrema over a chosen time span. Other cyclical components and seasonal patterns can also be identified together with other smooth components of higher frequencies. All of these components have different frequencies. However, being adaptive, frequencies of individual components are not fixed and are time-varying. We apply the method to volatility of nine international stock indices using daily high-low ranges of stock prices. Preliminary analysis shows promising results in that the trend component is capable of capturing the slow moving components in volatility indices. Correlations among low frequency components and various macroeconomic variables are also explored.

C311: **New innovative 3-way Anova a-priori test for direct vs. indirect approach in seasonal adjustment**
Presenter: **Enrico Infante**, EUROSTAT, Luxembourg
Co-authors: Dario Buono

The Seasonal Adjusted Series of an aggregate series can be calculated by a seasonal adjustment of the aggregate ("Direct Approach") or by

aggregating the seasonally adjusted individual series ("Indirect Approach"). A third way, the so called "Mixed Approach", is to adjust together those series that present the same kind of seasonality, and then re-aggregate them for the final aggregate series. In order to understand which series demonstrate the same kind of seasonality a test based on a three way Anova is utilised. This test could be seen as a generalisation of the Moving Seasonality Test for more than one series. Additionally, an exemplary case study is presented. An automatic procedure to choose the correct approach before obtaining results would be helpful. The literature to date has only focused upon a comparison between the results achieved by the different approaches. The problem of which approach to use is, however, not easy to solve when the growth rates are close to zero, and the values of the different approaches present different signs. As such, it is sought to set out a priori strategy for the development of an effective seasonal adjustment of the aggregate.

C663: **Singular spectrum analysis for separating trends from seasons and cycles**
Presenter: **Anatoly Zhigljavsky**, Cardiff University, United Kingdom

The method of time series analysis and forecasting, called singular spectrum analysis, is reviewed. This method is very useful for separating trend from seasonality and other periodic components. It is also useful for monitoring structural stability of time series and for detecting causality between different series. The method has some similarities to ESPRIT, MUSIC and other subspace-based techniques of signal processing and is also related to the methods of structural low-rank approximation in linear algebra. Real-life examples will be shown to demonstrate the capabilities of the method.

C854: **Short time series and seasonal adjustment**
Presenter: **Riccardo Gatto**, Eurostat, Luxembourg
Co-authors: Gian Luigi Mazzi

Seasonal adjustment of short time series is a major problem in official statistics. Change in survey methodology, development of new surveys or new indicators, new member states in institutions like the European Union are only few examples of the occurrence of the short time series problem. Not always is it possible to wait a sufficient time length to start producing seasonal adjusted data. Evaluation of the rate of decay in data quality while series are shortening can help in answering important questions like which seasonal adjustment algorithm to use in the presence of short time series or how long a series must be in order to start the production of seasonal adjusted data. In this work an evaluation of the changes in the quality performances of two different and widely used programs for seasonal adjustment, X-12-Regarima and TramoSeats, when the length of time series is progressively reduced, is carried out. The comparisons are carried out by using both an updated version of existing quality indicators and new indicators developed by the authors. The comparisons are run over a wide array of EU/Euro area time series.

C837: **SUTSE models and multivariate seasonal adjustment**
Presenter: **Filippo Moauro**, Eurostat, Luxembourg
Co-authors: Tommaso Proietti

EUROSTAT has been always involved in the field of seasonal adjustment (SA): recent efforts went to the release of the ESS-Guidelines for harmonization of SA practices among European economic indicators, as well as the development of a new tool, DEMETRA+, based on the leading algorithms for SA, i.e. TRAMO&SEATS and X-12-ARIMA. This paper exploits an alternative methodology based on SUTSE models for multivariate SA in line with the ESS-Guidelines. It is discussed how to simultaneously perform SA estimates for a top aggregate of an economic indicator and their sub-components and how to assure consistency within the entire system of SA figures. The empirical applications are devoted to the euro area industrial production and unemployment series.

CS39 Room Jessel REAL-TIME DENSITY FORECASTING Chair: Francesco Ravazzolo

C192: **Combinations for turning point forecasts**
Presenter: **Roberto Casarin**, University Ca' Foscari of Venice, Italy

New forecast combination approaches for predicting the turning points of a business cycle are proposed. Different models and combination strategies are considered. The combination schemes account for the forecasting performance of the models and give possibly better turning point forecasts. We consider turning point forecasts generated by autoregressive (AR) and Markov-Switching AR models, which are commonly used for business cycle analysis. In order to account for the parameter uncertainty, we consider a Bayesian approach to both estimation and prediction and compare, in terms of statistical accuracy, the individual models and combining turning point forecasts for the US and Euro area business cycles.

C201: **Testing for equal conditional predictive ability of real-time density forecast methods**
Presenter: **Christian Kascha**, University of Zurich, Switzerland
Co-authors: Francesco Ravazzolo

The performance of common tests for conditional predictive ability is investigated by means of a Monte Carlo study and an empirical application. In particular, the study examines the effect of various user choices that have to be undertaken on the test's size and power. Also the effect of different loss functions is examined for density forecasts and special care is taken to account for the real-time nature of economic data. We thus provide a guide for practitioners on how to implement tests for conditional predictive ability in such a context.The application to real-time US macroeconomic data confirms the empirical relevance of our findings.

C326: **Boostrapping forecast densities**
Presenter: **Andrea Monticini**, Catholic University (Milan), Italy
Co-authors: Francesco Ravazzolo

The problem of selecting the auxiliary distribution to implement the density forecast based on the wild bootstrap for stochastic processes featuring heteroscedasticity of unknown form is considered. Asymptotic refinements are nominally obtained by choosing a distribution with second and third moments equal to 1. We show that this stipulation may fail in practice, due to the distortion imposed on higher moments. We use a new class of two-point distributions. The results are illustrated by Monte Carlo experiments.

C418: **Nowcasting GDP in real-time: A density combination approach**
Presenter: **Knut Are Aastveit**, Norges Bank, Norway
Co-authors: Karsten Gerdrup, Anne Sofie Jore, Leif Anders Thorsrud

We use U.S. real-time vintage data and produce combined density nowcasts for quarterly GDP growth from a system of three commonly used model classes. The density nowcasts are combined in two steps. First, a wide selection of individual models within each model class are combined separately. Then, the nowcasts from the three model classes are combined into a single predictive density. We update the density nowcast for every new data release throughout the quarter, and highlight the importance of new information for the evaluation period 1990Q2-2010Q3. Our results show that the logarithmic score of the predictive densities for U.S. GDP increase almost monotonically as new information arrives during the quarter. While the best performing model class is changing during the quarter, the density nowcasts from our combination framework is always

performing well both in terms of logarithmic scores and calibration tests. The density combination approach is superior to a simple model selection strategy and also performs better in terms of point forecast evaluation than standard point forecast combinations.

C542: **Short-Term inflation projections: A Bayesian vector autoregressive approach**
Presenter: **Luca Onorante**, European Central Bank, Germany
Co-authors: Domenico Giannone, Michele Lenza, Daphne Momferatou

We construct a large Bayesian Vector Autoregressive model (BVAR) for the Euro Area that captures the complex dynamic interrelationships between the main components of the Harmonized Index of Consumer Price (HICP) and their determinants. The model is estimated using Bayesian shrinkage. We evaluate the model in real time and find that it produces accurate forecasts. We use the model to study the pass-through of an oil shock and to study the evolution of inflation during the global financial crisis.

CS54 Room Torrington TOPICS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Martin Wagner

C506: **Factor-augmented error-correction model: Structural analysis and forecasting**
Presenter: **Igor Masten**, University of Ljubljana, Slovenia
Co-authors: Anindya Banerjee, Massimiliano Marcellino

The implications of dynamic factor models for the analysis of vector error-correction models are discussed. Starting from the dynamic factor model for non-stationary data we derive the dynamic factor model in error-correction form or the factor-augmented error correction model (FECM), and its moving-average representation. The latter is used to discuss identification schemes of structural shocks with a focus on the distinction between permanent and transitory structural innovations to dynamic factors. The structural FECM thus offers a generalization of the classical identification schemes based on long-run restrictions to the case of large data panels. The performance of the FECM and the structural FECM relative to standard error-correction models is tested with two empirical exercises using a macro panel of US data. The first is a comparison of forecasting precision, while the second is an analysis of the role of stochastic trends in business cycle fluctuations.

C277: **Forecast combination based on multiple encompassing tests in a macroeconomic DSGE-VAR system**
Presenter: **Robert Kunst**, University of Vienna, Austria
Co-authors: Mauro Costantini, Ulrich Gunter

The benefits of forecast combinations based on forecast-encompassing tests relative to simple uniformly weighted forecast averages across rival models are studied. For a realistic simulation design, we generate data by a macroeconomic DSGE-VAR model. Assumed rival models are four linear autoregressive specifications, one of them a more sophisticated factor-augmented vector autoregression (FAVAR). The forecaster is assumed not to know the true data-generating model. The results critically depend on the prediction horizon. While one-step prediction hardly supports test-based combinations, the test-based procedure dominates at prediction horizons greater than two.

C526: **Correlation of implied default risk**
Presenter: **Jan Mutl**, Institute for Advanced Studies, Austria
Co-authors: Leopold Soegner

The correlation of implied credit risk is investigated. Credit default swaps, firm specific data and industry data from the US market are used to estimate spatial correlation models. Estimates show that the spatial correlation of credit default spreads is substantial and highly significant. We also demonstrate how these models can be used to price newly issued credit default swaps and to increase performance compared to standard panel regression settings.

C248: **A fixed-b perspective on the Phillips-Perron tests**
Presenter: **Martin Wagner**, Institute for Advanced Studies Vienna, Austria
Co-authors: Tim Vogelsang

Fixed-b asymptotic theory is extended to the nonparametric Phillips-Perron (PP) unit root tests. It is shown that the fixed-b limits depend on nuisance parameters in a complicated way. These non-pivotal limits provide an alternative theoretical explanation for the well-known finite sample problems of PP tests. It is also shown that the fixed-b limits depend on whether deterministic trends are removed using one-step or two-step approaches. This is in contrast to the asymptotic equivalence of the one- and two-step approaches under a consistency approximation for the long run variance estimator. Based on these results we introduce modified PP tests that allow for fixed-b inference. The theoretical analysis is cast in the framework of near-integrated processes which allows us to study the asymptotic behavior both under the unit root null hypothesis as well as for local alternatives. The performance of the original and modified PP tests is compared by means of local asymptotic power and a small finite sample simulation study.

CS93 Room S264 ECONOMETRIC MODELLING AND APPLICATIONS I Chair: Giuseppe Storti

C623: **Analyzing managers' sales forecasts**
Presenter: **Bert de Bruijn**, Erasmus University Rotterdam, Netherlands
Co-authors: Philip Hans Franses

The analysis and evaluation of sales forecasts of managers, given that it is unknown how they constructed these forecasts, is considered. The goal of this study is to find out whether these forecasts can be considered as rational. To examine this rationality, we argue that one has to approximate how the managers could have generated the forecasts. We describe several ways to construct these approximate expressions. The analysis of a large set of managers' forecasts for sales of pharmaceutical products leads to the conclusion that there are various avenues for improving managers' forecasts.

C656: **Factor vector autoregressive estimation of heteroskedastic persistent and non persistent processes subject to structural breaks**
Presenter: **Claudio Morana**, Universita di Milano Bicocca, Italy

The fractionally integrated heteroskedastic factor vector autoregressive (FI-HF-VAR) model is introduced. The proposed approach is characterized by minimal pretesting requirements and simplicity of implementation also in very large systems, performing well independently of integration properties and sources of persistence, i.e. deterministic or stochastic, accounting for common features of different kinds, i.e. common integrated (of the fractional or integer type) or non integrated stochastic factors, also featuring conditional heteroskedasticity, and common deterministic break processes. The proposed approach allows for accurate investigation of economic and financial time series, from persistence and copersistence analysis to impulse responses and forecast error variance decomposition. Monte Carlo results strongly support the proposed methodology.

C889: **Recent advances of econometrics tools for policy analysis at Eurostat**
Presenter: **Rosa Ruggeri Cannata**, Eurostat, Luxembourg
Co-authors: Gian Luigi Mazzi, Filippo Moauro

Ongoing activity at Eurostat on econometric tools for increasing the relevance of short term statistics is presented. Three lines of methodologies are here mentioned: the framework for nowcasting and density forecast, both implying a general linear regression as reference model. Also with the aim to increase data timelines, it moves the framework concerning coincident indicators which, in this case, focuses on bridge- and factor-models and the intensive application of the LARS algorithm. A second line of research concerns the construction of high frequency indicators of economic activity mainly based on data available at lower frequency through state space models. We present Euro-MIND, the monthly coincident indicator for the euro area, and all its most recent extensions. The last part provides the recent efforts towards the system of coincident turning point indicators for the business cycle, growth cycle and acceleration cycle, through univariate and multivariate Markov-Switching models. Discussion of all the methodologies is complemented by most significant applications to data concerning the euro area. Main conclusion is that all the mentioned instruments efficiently complement traditional official statistics in the desired direction, offering rapid instruments to policy analysis.

C806: **A new class of indirect estimators and bias correction**
Presenter: **Stelios Arvanitis**, Athens University of Economics and Business, Greece
Co-authors: Antonis Demos

A set of indirect estimators is defined based on moment approximations of the auxiliary ones. The introduction of these is motivated by reasons of analytical and computational facilitation. We provide results that describe higher order asymptotic properties of these estimators. We extend this set to a class of multistep indirect estimators that have zero higher order bias. We conclude that some currently used methodologies for bias correction are "numerical" approximations of the suggested one.

C832: **Test for change in the parameters of a diffusion process based on a discrete time sample**
Presenter: **Ilia Negri**, University of Bergamo, Italy
Co-authors: Yoichi Nishiyama

Testing on structural change has been an important issue in statistics. The parameter change problem is very popular in regression models and time series models due to the fact that such kind of model can be used to describe change occurring in financial and economical phenomena. As diffusion processes are widely used to model financial variables such as asset prices, interest or exchange rates where structural change can occur, we present a procedure to test if parameters change in an ergodic diffusion process has taken place. Our test is based on a discrete time sample and we study the asymptotic properties of the proposed test statistic, as the time of observation increases. We prove that the limit distribution of the test is a functional of independent Brownian Bridges and thus the test is asymptotically distribution free. Moreover, the test is consistent for any fixed alternative. We illustrate the performance of the proposed test statistic by applying it to some simulated trajectories of solutions of stochastic differential equations, both when the asymptotic assumption can be considered reached and also in the case when asymptotic conditions are not reached. The next natural task is to apply the procedure to real data.

CS25 Room Gordon CONTRIBUTIONS TO HIGH FREQUENCY DATA MODELING Chair: Massimiliano Caporin

C336: **Probability of informed trading and volatility for an ETF**
Presenter: **Paola Paiardini**, Queen Mary, University of London, United Kingdom
Co-authors: Dimitrios Karyampas

It is common knowledge in the microstructure literature that the order arrivals contain important information to determine subsequents price movements. Over the years, there have been developed several methods to extract this information from order flow. However, things become more difficult when we need to measure the order flow in a high frequency scenario. In a framework where trading takes place in milliseconds, trading time loses its meaning. A new procedure to estimate the Probability of Informed Trading (PIN), based on the volume imbalance is used: Volume-Synchronized Probability of Informed Trading (VPIN). Unlike the previous method, this one does not require the use of numerical methods to estimate unobservable parameters. We also relate the VPIN metric to volatility measures. However, we use most efficient estimators of volatility which consider the number of jumps. Moreover, we add the VPIN to a Heterogeneous Autoregressive model of Realized Volatility to further investigate its relation with volatility. For the empirical analysis we use data on an exchange traded fund (SPY).

C708: **Macroeconomic news effects on the stock markets**
Presenter: **Barbara Bedowska-Sojka**, Poznan University of Economics, Poland

A growing literature has documented the significance of macroeconomic news announcements in price formation process. Routinely the announcements considered in the literature are from United States. The aim is to compare the reaction to the announcements from Germany and from America. We study the influence of macro releases on intraday returns and volatility of the French and the German stock markets' indices within the two years period time. The flexible Fourier form framework is used to model intraday series of CAC40 and DAX. Two approaches are undertaken, with standardized surprises for announcements as well as dummy variables standing for announcements. We find that both markets, the French and the German, are characterized by similar intraday periodical pattern, both indices react very similarly to macro news and that the American announcements increase volatility slightly stronger than the German releases. By examining the impact of macro announcements over different time horizons, we are able to define how strong is the effect of macro information in the capital markets.

C716: **Intraday periodicity and intraday Levy-type jump detection**
Presenter: **Deniz Erdemlioglu**, FUNDP and K.U.Leuven, Belgium
Co-authors: Sebastien Laurent, Christopher Neely

Financial market volatility displays cyclical patterns particularly at the intraday level. We investigate the impact of intraday periodicity in the detection of big Levy-type jumps. To do so, we first design a Monte Carlo simulation to estimate periodicity in a robust way. We then propose a filtered big Levy jump test that accounts for periodic volatility. The results show that truncated maximum likelihood (TML) periodicity estimator has high relative efficiency and low bias in the presence of Levy jumps. Neglecting the periodic volatility decreases the power for Levy-type jump detection. Accounting for periodicity indeed improves the accuracy of jump identification.

C746: **Wavelet decomposition of stock market correlation using high-frequency data**
Presenter: **Lukas Vacha**, Academy of Sciences of the CR, Czech Republic
Co-authors: Jozef Barunik, Miloslav Vosvrda

A contribution on international stock market co-movement by studying its dynamics in the time-frequency domain is made. The novelty of our approach lies in the application of wavelet tools to high-frequency financial market data. A major part of economic time series analysis is done in the time or frequency domain separately. Wavelet analysis combines these two fundamental approaches allowing study of the time series 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 connect our approach to the dynamic conditional correlation approach as well. Using wavelet coherence, we uncover interesting dynamics of correlations between Central European and Western European stock markets using high-frequency data in the time-frequency space.

C920: **A functional filtering and neighborhood truncation approach to integrated quarticity estimation**
Presenter: **Dobrislav Dobrev**, Federal Reserve Board, United States of America
Co-authors: Torben Andersen, Ernst Schaumburg

A first in-depth look at robust estimation of integrated quarticity (IQ) based on high frequency data is provided. IQ is the key ingredient enabling inference about volatility and the presence of jumps in financial time series and is thus of considerable interest in applications. We document the significant empirical challenges for IQ estimation posed by commonly encountered data imperfections and set forth three complementary approaches for improving IQ based inference. First, we show that many common deviations from the jump diffusive null can be dealt with by a novel filtering scheme that generalizes truncation of individual returns to truncation of arbitrary functionals on return blocks. Second, we propose a new family of efficient robust neighborhood truncation (RNT) estimators for integrated power variation based on order statistics of a set of unbiased local power variation estimators on a block of returns. Third, we find that ratio-based inference, originally proposed in this context by Barndorff-Nielsen and Shephard (2002), has desirable robustness properties in the face of regularly occurring data imperfections and thus is well suited for our empirical applications. We confirm that the proposed filtering scheme and the RNT estimators perform well in our extensive simulation designs and in an application to the individual Dow Jones 30 stocks.

CS15 Room S261 CONTRIBUTIONS IN MODELLING AND FORECASTING FINANCIAL RISK Chair: Michele La Rocca

C232: **Forecasting volatility and jumps based on OHLC-data**
Presenter: **Janine Balter**, Saarland University, Germany

Building on newly developed consistent asymmetric jump estimators constructed from the Open, Highs and Lows of a price process, we make use of the well-known approximated long-memory HAR-model in order to forecast volatility. Applied on the S&P 500 index on the one hand we will show that jumps have forecasting power, whereby only short term negative jumps have a distinctive significant impact. On the other hand we will see that including a long range Leverage effect will improve the forecasting power by reducing the influence of negative jumps simultaneously. Furthermore, it turns out that both positive and negative jumps are sufficiently predictable contrary to the widley documented cases in literature. This new finding is due to the usage of the new OHLC-based (asymmetric) jump estimators.

C585: **Conditional heteroskedasticity and dependence structure in crude oil and US dollar markets**
Presenter: **Chih-Chiang Wu**, Yuan Ze University, Taiwan
Co-authors: Wei-Peng Chen

A flexible range-based volatility model is constructed to explore the volatility and dependence structures between the oil price and the US dollar exchange rate. An asset-allocation strategy is implemented to evaluate the economic value and confirm the efficiency of this model. The empirical results indicate that the use of price range information can not only enhance the explanatory power of volatility structures but also benefit investors by producing extra benefits of between 112 and 559 annualized basis points in an asset allocation strategy; less risk-averse investors can generate higher benefits. Moreover, an additional economic gain of between 11 and 33 annualized basis points can be achieved by taking the asymmetric dependence structure between crude oil and USDX markets into consideration.

C621: **The role of the information set for forecasting - with applications to risk management**
Presenter: **Hajo Holzmann**, Marburg University, Germany
Co-authors: Matthias Eulert

Point forecasts are issued on the basis of certain information. A larger amount of available information should lead to better forecasts. We show how the effect of increasing the information set on the forecast can be quantified by using strictly consistent scoring functions. Further, a method is proposed to test whether an increase in a sequence of information sets leads to distinct, improved h-step point forecasts. For the value at risk (VaR), we show that increasing the information set will result in VaR forecasts which lead to smaller expected shortfalls, unless an increase in the information set does not result in any change of the VaR forecast. We also discuss the role of the information set for evaluating probabilistic forecasts by using strictly proper scoring rules.

C677: **Financial risk management using high-dimensional vine copulas**
Presenter: **Eike Brechmann**, Technische Universitaet Muenchen, Germany
Co-authors: Claudia Czado

In light of recent financial market turmoils, a diligent financial risk management is crucial, in particular with regard to dependencies among different assets. In this talk we discuss the class of copula-GARCH models with particular focus on vine copulas, which are flexible multivariate copulas built up of arbitrary bivariate copulas. Vine copulas thus can capture a wide range of dependence and recently developed innovative model selection techniques allow for their use in high-dimensional applications. To investigate the Euro Stoxx 50 and its members we develop a vine copula based factor model, the Regular Vine Market Sector (RVMS) model, which takes into account market and sectorial dependencies and allows to separate systematic and idiosyncratic risk of financial assets. We further treat passive and active portfolio management using vine copula-GARCH models. In particular, we show how to forecast the Value-at-Risk of individual assets and larger portfolios. Our models are evaluated based on the forecasting accuracy and compared to the corresponding performance of relevant benchmark models. Diversification benefits and portfolio optimization are also discussed.

C822: **Optimal superposition policies for futures investments**
Presenter: **Rainer Schuessler**, Westfalische Wilhelms-Universitat Munster, Germany

The framework of dynamic asset allocation is transferred to the management of downside risk for trading strategies in futures markets. A general method is introduced to assess the portfolio-wide effects of simultaneously applying superposition rules (varying stop loss limits and different levels of leverage) to mechanical dynamic investment strategies. The key concept underlying our method is to systematically separate the management of portfolio-wide risk and the risk for individual assets. We show how to formulate the investment problem as a finite horizon Markov decision problem and how to solve it via dynamic stochastic optimization using backward recursion and Monte Carlo sampling. The portfolio-level view for risk management facilitates analysis and allows considering an arbitrarily large investment opportunity set. In our empirical study, we apply the method to investment strategies in commodity futures markets across 34 different commodity markets from 1995 until 2010. Results indicate substantial increase of utility for the considered investment strategies and risk preferences.

CS16 Room Bedford QUANTITATIVE RISK MANAGEMENT II Chair: Simon Broda

C459: **A new time-based quantitative model for risk management**
Presenter: **Duc Phamhi**, ECE Graduate School of Engineering, France

Most of Basel II and III financial regulations are based on data, treated in a statistical mindset. While regulators call for forward-looking risk management, they suggest no time-based modeling approach. The theoretical foundations of the application of filtering techniques, like sequential Monte Carlo, and interactive particles simulation, to risk management processes in banks are established. We first show how temporal dynamics in equations solves deficiencies of existing, risk non-sensitive, models. The new models are proposed as value-based, time varying, functions of rare catastrophic scenarios allowing arbitrage between risk mitigation decisions. Next, the affiliation is established from stochastic optimal control foundations, through reinforcement learning and temporal differences learning, to this approach. Some parts of this framework also borrow from hidden Markov model and the related Bayesian inference techniques that underlie interactive particle systems filters. Small models are presented to illustrate the usefulness of time-based process in modeling fraud risks. We conclude by showing how Basel II and Basel III regulations are still respected using this novel approach to risk capital.

C685: **Operational risk modelling: The impact of the Peaks-over-Threshold approach on risk measures**
Presenter: **Martin Kukuk**, University of Wuerzburg, Germany
Co-authors: Verena Bayer

The loss distribution approach using bank-internal data to estimate probability distribution functions for each business line is the most risk-sensitive methodology in order to measure a bank's capital requirement for operational risk. According to the disposals of the Basel Committee on Banking Supervision about operational risk, a bank has to demonstrate that its approach captures potentially severe, extreme loss events. For this purpose we apply the Extreme Value Theory's Peaks-over-Threshold (POT) approach for modelling operational loss data. The POT approach fits a univariate generalized Pareto distribution (GPD) to the upper tail of a given distribution function in a continuous manner and provides a better fit in the tails than traditional probability distributions. Since operational loss data show several "low frequency/high impact" events, the GPD is often suitable to model the heavy tails of the univariate loss distributions, whereas the center of the distribution is better modelled by, e.g., a lognormal distribution. We analyze operational losses of an external database which contains worldwide information on publicly reported operational losses. We apply the POT approach and demonstrate the differences to traditional probability distributions with the help of the risk measures value-at-risk and expected shortfall.

C826: **Outliers detection in credit risk multivariate data via rank aggregation**
Presenter: **Silvia Figini**, University of Pavia, Italy
Co-authors: Luisa Cutillo, Annamaria Carissimo

The problem of rank aggregation, the task of combining different rank orderings on the same set of units (preference lists) so as to achieve a single final ordering, is addressed. We explore the potentiality of Network Aggregation, a novel approach for rank aggregation inspired by graph theory. In particular, we underline the robustness of our proposal with respect to the presence of outliers. To this aim we introduce a new method for the detection of multivariate outliers which accounts for the data structure and sample size. The cut-off value for identifying outliers is defined by a measure of deviation of the empirical distribution function of the robust Mahalanobis distance from the theoretical distribution function. We investigate the performances of the overall method on a real financial data set composed of balance sheet information on Small and Medium Enterprises. The empirical evidence achieved shows that the Network Aggregation algorithm improves predictive performances in credit risk analysis.

C429: **Realized news impact curves**
Presenter: **Sven Christian Steude**, University of Zurich, Switzerland
Co-authors: Kerstin Kehrle, Marc Paolella

Stock price volatility is driven to a large proportion by the arrival of good or bad news. We demonstrate that harvesting preprocessed news sentiment data can improve volatility forecasting enormously. In particular, we promote the class of mixed normal conditional heteroskedastic models and show that it lends itself as an ideal candidate to model the asymmetric feedback between news and volatility. By generalizing the model such that market wide, sector wide and firm specific high frequency news are governing the mixing components and as such future volatility, large gains in terms of in–sample fit and out–of–sample forecasting performance can be realized highlighting the predictive power of news sentiment.

C421: **Model averaging for risk management in European stock markets**
Presenter: **Sorin Dumitrescu**, Academy of Economic Studies from Bucharest, Romania
Co-authors: Marius Acatrinei, Petre Caraiani, Radu Lupu

The current financial crisis raised the problem of the feasibility of risk models used for regulatory or business purposes in general. We analyze the main stochastic processes used for financial purposes in order to generate a better model for risk measurement. We use five-minute frequency for a set of European stock market indexes and build an out-of-sample test for both simple and complex jump processes, stochastic volatility processes and GARCH-family processes. We also build a comprehensive model in which all these models hold weights according to their forecast precision and compute its out-of-sample precision. We also approach the idea of co-movements for the stock market indexes and model the whole set using a principal component analysis recomposition of the data using wavelet multiresolution analysis with ortoghonal basis. The residuals of this fitting are then modeled with the above mentioned processes. The same analysis will be realized on different frequencies starting from five-minute to daily in pursuit of a characterization of the time scaling dynamics of the stock market returns. We expect to develop a framework for the selection of a better model for risk measurement.

Saturday 17.12.2011 16:35 - 18:40 Parallel Session F – ERCIM

ES10 Room S261 CLASSIFICATION AND DISCRIMINANT PROCEDURES FOR DEPENDENT DATA Chair: Andres M. Alonso

E547: **Spatial depth-based classification for functional data**
Presenter: **Carlo Sguera**, Carlos III de Madrid, Spain
Co-authors: Pedro Galeano, Rosa Lillo

We consider two functional generalizations of the notion of multivariate spatial depth: the functional spatial depth (FSD) and the kernelized functional spatial depth (KFSD). The first generalization is linked with the notion of functional spatial quantile (FSQ), in such a way that, under some weak assumptions, the FSD of a curve x is given by $1 - \|Q^{-1}(x)\|$, where $Q^{-1}(x)$ represents the inverse of the FSQ function at x. The second generalization is a kernelized version of the first one, and it enables to take into account the local structure of a functional data set, which in some cases might br more important than the global view provided by the first version. Furthermore, we evaluate how the considered functional spatial depths may help in solving an important statistical problem such as classification for functional data. In order to do that, we present a simulation study in which the performances of the functional spatial depths are compared with the performances of some existing functional depths. Finally, we apply spatial depth-based classification to real functional data sets.

E338: **Time series classification via the combination of functional data projections**
Presenter: **Javier Gonzalez**, University of Groningen, Netherlands
Co-authors: Alberto Munoz

The problem of time series discrimination in the Functional Data Analysis context is addressed. To this end we first introduce a general regularization methodology to obtain the projections of a set of time series onto a battery of different Reproducing Kernel Hilbert spaces. Such projections allow us to obtain several finite-dimensional representations of the series. We consider the Euclidean metrics induced by the previous representations and we combine them to obtain a single fusion metric. Finally we test a collection of discrimination procedures on several time series data sets, where distances are calculated using the fusion metric. We show that the use of the proposed fusion metric is able to consistently improve the performance of the considered discrimination procedures for a diversity of simulated and real time series data sets.

E521: **Comparison of several supervised classification methods for functional data**
Presenter: **Manuel Febrero-Bande**, Univ. of Santiago de Compostela, Spain

The supervised classification problem can be stated as follows: Assume we have a training sample $\{\mathcal{X}_1,\ldots,\mathcal{X}_n\},\{y_1,\ldots,y_n\}$ being $y_i \in \{1,\ldots,K\}$ the class label of $\mathcal{X}_i$. The aim of classification procedures is to assign a new coming observation $\mathcal{X}_{new}$ to a group using the information provided by the training sample. Supervised classification is a relevant problem, sometimes referred to as "discrimination" in the statistical community and also, a topic of leading current interest in the fields of Machine Learning (in computing science) and Pattern Recognition (in engineering). Several methods for supervised classification with functional data are revised or proposed. The comparison includes some well known methods based on k-nearest neighbors or kernel classifiers as well with others procedures relying on logistic regression or more sophiticasted regression models and also, others based on depth measures. Some applications of these methods to simulated and real data will be shown.

E328: **Robust functional classification for time series**
Presenter: **Andres M. Alonso**, Universidad Carlos III de Madrid, Spain
Co-authors: David Casado, Sara Lopez-Pintado, Juan J. Romo

The integrated periodogram is proposed to be used to classify time series. The method assigns a new time series to the group that minimizes the distance between the time series integrated periodogram and the group mean of integrated periodograms. Local computation of these periodograms allows us to apply this approach to nonstationary time series. Since the integrated periodograms are curves, we apply functional data depth-based techniques to make the classification robust. The method provides small error rates with both simulated and real geological data, improving on existing approaches, and presents good computational behavior.

ES14 Room Bloomsbury NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT I Chair: Guadalupe Gomez

E138: **Experimental design for modeling benign positional vertigo**
Presenter: **Santiago Campos**, University of Castilla-La Mancha, Spain
Co-authors: Jesus Lopez-Fidalgo

In practice the behavior of many stochastic processes or particular systems is usually modeled by using non-linear models involving differential equations. The models studied here are related to vestibular disorders, benign positional vertigo, to be precise. This is the most common cause of vertigo. D-optimal designs for non-linear models with two unknown parameters have been analytically calculated although the differential equations appearing are in a simple form. This analytical derivation to find D-optimal designs is not always feasible in practice. In these cases, we must resort to numerical algorithms.

E211: **A Bayesian multi-state model for estimating the progression of stage IV non-small cells lung cancer**
Presenter: **Carmen Armero**, Universitat de Valencia, Spain
Co-authors: Silvia Perra, Alicia Quiros, Stefano Cabras, M. Eugenia Castellanos, Mauro Javier Oruezabal, Javier Sanchez-Rubio

Bayesian reasoning and multi-state models are used to assess the progression of stage IV non-small cells lung cancer (NSCLC) through the disability model with three states: the initial one, which is determined by the stage IV patient's diagnosis time, the tumour progression and the patient's death.Transition probabilities between states are expressed in terms of the hazard rate functions for times between transitions, which we analyze through Weibull accelerated failure time regression models. Uncertainty about the parameters of the model is expressed in terms of its posterior distribution and it has been propagated to the hazard rate functions of the times between transitions. We can thus obtain the posterior predictive distribution for the transition probabilities, given the time and the covariates, which offers a satisfactory description of the dynamics of the system. Data for the study comes from the Infanta Cristina Hospital of Madrid, Spain, and consist of survival times for stage IV NSCLC patients and measures of several covariates that may be related to the disease, observed from January 2008 to December 2010.

E245: **FARMS: A new strategy for model selection**
Presenter: **Susana Perez-Alvarez**, Fundacio IrsiCaixa, Spain
Co-authors: Guadalupe Gomez, Christian Brander

Selecting a regression model, when based on large datasets with a big number of covariates, needs efficient methods to pick up the variables to be included in the final model. A variable selection method should find a subset of variables having the optimal prediction performance. Sometimes, this prediction is not optimized during the process of variable selection, and testing for all potential subsets of variables is not possible.

Consequently, suboptimal methods for variable selection are applied and the prediction performance of regression models is estimated separately. We propose a new method that combines Forward variable selection and All subsets Regression for Model Selection: FARMS. In order to explore its properties, we fit different models using common methods and FARMS. We tested also its robustness. We performed these comparisons on a dataset with host genetic and immunological information of over 800 individuals from Lima (Peru) and Durban (South Africa) with HIV infection. This dataset includes around 500 variables with information on HIV immune reactivity (around 400 predictive variables) and individual genetic characteristics (around 100 predictive variables). We compared all the models obtained and confirmed that the application of FARMS proved to be more time-efficient and have better statistical properties.

E205: **Global optimization strategies for non-linear dynamical models of cell metabolism based on recasting into power-law models**
Presenter: **Albert Sorribas**, University of Lleida, Spain
Co-authors: Carlos Pozo, Gonzalo Guillen-Gosalbez, Laureano Jimenez, Rui Alves, Alberto Marin-Sanguino

Successful design of newly engineered microbial strains for biotechnological purposes requires the use of mathematical models. With appropriate optimization techniques, one could identify the necessary modifications needed in order to achieve a given biotechnological goal. As appropriate models to perform such an analysis are necessarily non-linear and typically non-convex, finding their global optimum is a challenging task. Canonical modeling techniques, such as Generalized Mass Action (GMA), offer a possible solution because they have a mathematical structure that enables developing specific algorithms for global optimization. Based on the GMA canonical representation, in previous works we have defined a highly efficient optimization algorithm and a set of related strategies for understanding the evolution of adaptive responses in cellular metabolism. Here, we explore the possibility of recasting kinetic non-linear models into an equivalent GMA model, so that global optimization on the recast GMA model can be performed. With this technique, optimization is greatly facilitated and the results are transposable to the original non-linear problem. Our results show that recasting non-linear kinetic models into GMA models is indeed an appropriate strategy that helps overcoming some of the numerical difficulties that arise during the global optimization task.

E293: **Structured additive regression modeling of time between menarche and menopause in breast cancer screening women**
Presenter: **Bruno de Sousa**, Institute of Hygiene and Tropical Medicine , Portugal
Co-authors: Elisa Duarte, Thomas Kneib, Carmen Cadarso-Suarez, Vitor Rodrigues

A database containing approximately 260,000 data records of women that entered the Screening Program for the first time in the central region of Portugal is analyzed. It is believed that the period of time between the age of menarche and the age of menopause has been increasing over time. Therefore, a new variable called Window was defined as the difference between the age of menopause and the age of menarche, which represents a woman's years of fertility. The evolution in time and space of the variables Window and the Age of Menarche will be analyzed, exploring the possible associations with other variables, such as Contraceptive Pills, Pregnancy Status, Nursing Status and Purchasing Power Index. The results show that early menarche seems to occur in younger women, with decreasing fertility periods for women born after 1933. From an economical perspective, we conclude that counties with higher purchasing power show early menarche ages, and larger periods of women's fertility. The spatial analysis leads to the conclusion that interior counties have lower ages of menarche and larger periods of fertility.

ES17 Room Senate OUTLIERS AND CHANGE-POINTS IN TIME SERIES I Chair: Roland Fried

E086: **Monitoring the intraday volatility pattern**
Presenter: **Siegfried Hoermann**, Universite libre de Bruxelles, Belgium
Co-authors: Piotr Kokoszka, Robertas Gabrys

A functional time series consists of curves, typically one curve per day. The most important parameter of such a series is the mean curve. We propose two methods of detecting a change in the mean function of a functional time series. The change is detected on line, as new functional observations arrive. The general methodology is motivated by and applied to the detection of a change in the average intraday volatility pattern. The methodology is asymptotically justified by applying a new notion of weak dependence for functional time series. It is calibrated and validated by simulations based on real intraday volatility curves.

E263: **Robust nonlinear filtering of state-space models with applications**
Presenter: **Bernhard Spangl**, University of Natural Resources and Life Sciences, Vienna, Austria
Co-authors: Peter Ruckdeschel, Rudolf Fruehwirth

The problem of recursive filtering in state-space models is considered. The classically optimal Kalman filter for linear state-space models and its nonlinear extensions, the extended Kalman filter and the unscented Kalman filter, are well known to be sensitive to outliers, so robustness is an issue. Approaches to robustify both recursive nonlinear filtering algorithms, the extended Kalman filter as well as the unscented Kalman filter are discussed. For an implementation in R, the first two authors have been working on an R package robKalman, where a general infrastructure is provided for robust recursive filters and smoothers. The proposals are backed by corresponding implementations in R within the robKalman package and their performance is tested by simulation studies. Moreover, a special practical application focuses on track reconstruction of particles in high energy physics.

E168: **Robustification of Elliott's HMM-based online filter**
Presenter: **Peter Ruckdeschel**, Fraunhofer ITWM, Germany
Co-authors: Christina Erlwein

Robust statistics in the context of Hidden Markov Models (HMMs) is considered. More specifically we aim at a robustification of an Expectation-Maximization-type algorithm due to Elliott. Within this algorithm, recursive filters are derived to obtain information of the latent Markov chain underlying the HMM. The derivation of these recursive filters involves a change of probability measure technique, which we show to be prone to substitutive outliers. The corresponding M-step produces adaptive estimates for the model parameters, which make use of the recursive filters for processes of the underlying Markov chain. A robustification is proposed which in the filtering step extends results on optimal robust filtering and in the M-step uses strategies known to be useful in the context of Gaussian location-scale models. This model is implemented to R in a new package robHMM which is used to apply this algorithm in an asset allocation problem in discrete time and develop trading strategies to optimally invest in growth or value stocks.

E409: **Effects of outliers on the identification and estimation of asymmetric GARCH-type models**
Presenter: **Ana Perez-Espartero**, University of Valladolid, Spain
Co-authors: Esther Ruiz, Angeles Carnero

It is well known that outliers affect the identification and estimation of the parameters and the underlying volatilities in symmetric GARCH models. However, nothing is known about the effect of outliers on asymmetric conditional heteroscedastic models when the volatility response is different depending on whether past returns are positive or negative. We first analyze the effect of additive outliers on the identification of asymmetries by deriving the analytical properties of the sample cross-correlations between returns and future squared returns, in the presence of outliers. Second, we analyze whether positive and negative outliers of equal size have different impacts on the estimated model parameters and volatilities in the

context of the Threshold GARCH model (TGARCH). Different methods of estimation are considered: maximum likelihood, Gaussian quasi-maximum likelihood (QML) and QML based on maximizing the Student likelihood (QML-t). Given that there are no closed-form expressions for these estimators, this analysis is carried out by simulation. The results are illustrated with an empirical application.

ES18 Room Gordon EXTREME VALUE THEORY AND APPLICATIONS Chair: Michael Falk

E066: **Asymptotically unbiased estimation of the coefficient of tail dependence**
Presenter: **Armelle Guillou**, Universite de Strasbourg et CNRS, France
Co-authors: Yuri Goegebeur

A class of weighted functional estimators for the coefficient of tail dependence in bivariate extreme value statistics is introduced and studied. Asymptotic normality of these estimators is established under a second order condition on the joint tail behavior, some conditions on the weight function and for appropriately chosen sequences of intermediate order statistics. Asymptotically unbiased estimators are constructed by judiciously chosen linear combinations of weighted functional estimators, and variance optimality within this class of asymptotically unbiased estimators is discussed. The finite sample performance of some specific examples from our class of estimators and some alternatives from the recent literature are evaluated with a small simulation experiment.

E174: **A simple Bayesian combinatorial model for bivariate extreme shocks**
Presenter: **Pasquale Cirillo**, University of Bern, Switzerland

An intuitive construction for the Bayesian modeling of bivariate extreme shocks is given. We take into account N defaultable systems, assuming them to be exchangeable. Every system is made of two elements subject to random shocks of random magnitude. A system may fail because of the collapse of any (or both) of the two components. Now, let X_n and Y_n be the default times of the two parts constituting system $n = 1,...,N$. The couple (X_n, Y_n) gives us important information about the life of the system. Let Z_n, V_n and W_n be three mutually independent reinforced urn processes (RUP). Now set, for every $n = 1,...,N$, $X_n = Z_n + V_n$ and $Y_n = Z_n + W_n$. In this way, the dependence between X_n and Y_n is simply given by the common element Z_n. Exploiting the combinatorial properties of RUP, we propose a simple but effective Bayesian nonparametric approach to bivariate extreme shock models, also showing how the same approach can be extended to EVT in general.

E239: **Conditioning exceedances on covariate processes**
Presenter: **Rolf-Dieter Reiss**, University Siegen, Germany
Co-authors: Ulf Cormann

In several applied fields such as hydrology, climatology or finance one is concerned with the estimation and prediction of functional parameters of the upper tail of the conditional distribution $P(Y \in \cdot | X = x)$ of a response variable Y given a covariate $X = x$. The statistical inference is carried out within a certain Smith-Shively model of a point process of exceedances of the Y-value above a given threshold whereby the process also includes the exceedance covariates. The inference is proposed to be based on the conditional distribution of the previous point process given the point process of covariates. It is of importance that the conditional process merely depends on our target distribution, namely, $P(Y \in \cdot | X = x)$. In the special case of Poisson processes the conditional process may be found in the point process book by Reiss. Our results are valid within the broader model where the response variables are conditionally independent given the covariates, which also includes the Smith-Shively model. Densities are calculated so that likelihood-based estimators are feasible. It is numerically exemplified that the maximum likelihood principle leads to more accurate estimators within the conditional approach than in the unconditional one.

E403: **On extremes of Gaussian processes in a random environment**
Presenter: **Juerg Huesler**, University of Bern, Switzerland

A Gaussian process which crosses a random boundary as in a random environment is considered. The random environment may be of different types. For example, we may consider a random trend or a random variance function. We present results on the extremes of such a process $\xi(t)\eta(t)$, where $\xi(t)$ is a Gaussian process and $\eta(t)$ another process, being independent of $\xi(t)$. We assume that $\xi(t)$ is a standard locally stationary Gaussian process with covariance function $1 - r(t,t+s) \sim C(t)|s|^{\alpha}$ as $s \to 0$, with $0 < \alpha \leq 2$ and $C(t)$ a positive bounded continuous function. We consider the exceedance probabilities of $\xi(t)\eta(t)$ under certain restrictions of the random process $\eta(t)$. Some particular examples will be discussed where $\eta(t)$ is a random trend or random deviation process. Asymptotic exact results can be derived for such exceedance probabilities.

E410: **Testing for a generalized Pareto process**
Presenter: **Stefan Aulbach**, University of Wuerzburg, Germany
Co-authors: Michael Falk

A parametric family of continuous processes $\mathbf{X}$ on $[0,1]$, indexed by $\vartheta \in \Theta \subset \mathbb{R}$ is considered, where $\mathbf{X}$ is a generalized Pareto process in case $\vartheta = 0$. Based on n independent copies $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ of $\mathbf{X}$, we establish local asymptotic normality (LAN) of the point process of exceedances among $\mathbf{X}^{(1)}, \ldots, \mathbf{X}^{(n)}$ above an increasing threshold line. The corresponding central sequence provides an asymptotically optimal sequence of tests for testing $H_0 : \vartheta = 0$ against a sequence of alternatives $H_n : \vartheta = \vartheta_n$ converging to zero as n increases. These test statistics depend, however, on some additional and usually unknown model parameter. We, therefore, consider an omnibus test statistic sequence as well and compute its asymptotic relative efficiency with respect to the optimal test sequence.

ES25 Room Jessel MODEL VALIDATION Chair: M. Dolores Jimenez-Gamero

E020: **A nonparametric test for risk-return relationships**
Presenter: **Ingrid Van Keilegom**, Universite catholique de Louvain, Belgium
Co-authors: Juan Carlos Escanciano, Juan Carlos Pardo Fernandez

Nonparametric tests for testing parametric risk-return relationships are proposed. That is, tests for parametric restrictions between the conditional mean and the conditional variance of excess returns given a set of unobservable parametric factors are considered. A distinctive feature of our tests is that they do not require a parametric model for the conditional mean and variance, while allowing for flexible parametric risk-return relationships. Tests are based on the difference of the estimated restricted and unrestricted pricing errors' distributions. A suitable transformation of this difference renders the tests asymptotic distribution-free, with limits that are functionals of a standard normal variable. Hence, the tests are straightforward to implement. A simulation study compares the finite sample performance of the proposed tests. Finally, an application to excess returns of the S&P500 highlights the merits of our approach.

E106: **An affine invariant multiple test procedure for assessing multivariate normality**
Presenter: **Carlos Tenreiro**, University of Coimbra, Portugal

A multiple test procedure for assessing multivariate normality (MVN) will be presented. The new test combines a finite set of affine invariant test statistics for MVN through an improved Bonferroni method. The usefulness of such an approach is illustrated by a multiple test including Mardia's

and BHEP (Baringhaus-Henze-Epps-Pulley) tests that are among the most commonly recommended procedures for testing MVN. The results of a simulation study carried out for a wide range of alternative distributions indicate that the proposed multiple test procedure presents an overall good performance against the most recommended MVN tests in the literature.

E121: **Bootstrap model validation under bias selected data**
Presenter: **Jorge Luis Ojeda Cabrera**, Universidad de Zaragoza, Spain

The problem of model validation for selection biased data is considered. Selection biased observations are common in those studies where available data cannot be collected directly from the phenomena of interest. As a consequence, usual statistical tools lead to biased estimation and misleading inference for this type of data. While bias correction may be useful to provide unbiased estimators, it is not clear how to use it to perform model validation for general types of selection biased observations. In particular, residuals, which are the most widely used tool to address model validation, also suffer from the bias present in selection biased data. The aim is to introduce the problem of model validation for Selection Biased data through regression estimation, tackling some of the drawbacks aforementioned from the perspective of resampling. In order to do so, the "compensation" or "mean preservation" framework is introduced as a tool that provides both: consistent estimation, and appropriate residuals to perform model validation for selection biased data. It is shown that these residuals are particularly suitable to be used jointly with resampling methods to address the problem of model validation in this context.

E186: **Tests of fit via phi-divergence measures for biostatistics and reliability data**
Presenter: **Alex Karagrigoriou**, University of Cyprus, Cyprus
Co-authors: Ilia Vonta

Measures of divergence or discrepancy are used extensively in statistics in various fields. Divergence measures that are based on the Csiszar's class of measures, which is generated by a collection of functions phi, are considered. The class includes the Kullback-Leibler measure, the minimum discrimination measure, the Cressie-Read measure, the Pearson's measure and the Matusita measure, among others. More specifically, we propose a class of goodness of fit tests based on Csiszar's class of measures designed for grouped censored survival or reliability data. We derive the asymptotic distribution of the test statistic under the null hypothesis and under contiguous alternative hypotheses. We examine the case where the distribution under the null hypothesis is known, as well as the case where unknown parameters are involved in the assumed distribution. Simulations are presented to show the performance of the proposed test.

E198: **Testing for the symmetric component in skew-symmetric distributions**
Presenter: **M. Dolores Jimenez-Gamero**, Universidad de Sevilla, Spain

Skew-symmetric distributions have recently received much attention in the statistical literature. Most models of asymmetric distributions proposed allow for a continuous variation from symmetry to asymmetry, obtained by transforming an arbitrary symmetric distribution by means of a skewing mechanism. The resulting skewed distribution often shares some of the properties of its symmetric antecedent. So, for certain inferential purposes, because of this inheritance, it would be interesting to test if the symmetric generator belongs to a certain parametric family, that is to say, testing goodness-of-fit for the symmetric component in skew-symmetric multivariate distributions. With this aim, a characterization of the characteristic function of the symmetric component is first given. Based on this characterization, a test for testing goodness-of-fit for the symmetric component is proposed. The approach is quite general, and it can be applied to any skew-symmetric multivariate distribution. Finally, taking into account that the normal law is perhaps the most studied distribution, as a particular case of unquestionable interest, the generalized skew-normal family is considered, since the symmetric component of the distributions in this family is normal.

ES41 Room Court STATISTICS IN FUNCTIONAL AND HILBERT SPACES II Chair: Gil Gonzalez-Rodriguez

E170: **On the estimation of the L1-median of a functional variable with complex surveys: application to the electricity load curves**
Presenter: **Camelia Goga**, Universite de Bourgogne, France
Co-authors: Mohamed Chaouch

Mean profiles are widely used as indicators of the electricity consumption habits of customers. Unfortunately, it is well known that the mean is highly sensitive to the presence of outliers, for instance consumers with high level of consumption. An alternative to the mean profile is proposed: the L1-median profile, which is more robust. When dealing with large datasets of functional data (load curves for example), survey sampling approaches are useful for estimating the median profile avoiding storing the whole data. We propose here estimators of the L1 median trajectory using several sampling strategies. A comparison between them is illustrated by means of a test population of N = 18902 electricity meters for which the individual electricity consumption is recorded every 30 minutes during two weeks. We develop a stratification based on the linearized variable which substantially improves the accuracy of the estimator compared to simple random sampling without replacement. Some potential areas for future research are also highlighted.

E204: **Confidence bands for unequal probability and model assisted Horvitz Thompson estimators for functional data**
Presenter: **Herve Cardot**, Universite de Bourgogne, France

When collections of functional data are too large to be exhaustively observed, survey sampling techniques provide an effective way to estimate global quantities such as the population mean function. Assuming functional data are collected from a finite population according to a probabilistic sampling scheme, we estimate the mean function with a Horvitz-Thompson estimator. Under mild conditions on the population size, observation times, regularity of the trajectories, sampling scheme, and smoothing bandwidth, we prove a Central Limit Theorem in the space of continuous functions. We also establish the uniform consistency of a covariance function estimator and apply the former results to build global confidence bands for the mean function. The bands attain nominal coverage and are obtained through Gaussian process simulations conditional on the estimated covariance function. These results are illustrated on a real example of electricity consumption curves and various sampling strategies (stratified sampling, proportional to size sampling and model assisted estimation) are compared both in terms of accuracy of the estimations and width of the confidence bands.

E292: **Partial differential smoothing for surface estimation**
Presenter: **Laura M. Sangalli**, Politecnico di Milano, Italy
Co-authors: Laura Azzimonti, Piercesare Secchi, James O. Ramsay

In this work we propose a novel functional data analysis technique for surface estimation, based on a bivariate generalization of smoothing splines. The surface estimate is obtained via minimization of a penalized sum-of-square-error functional where the roughness penalty consists in the L2 norm of a second order partial differential operator. The minimization problem defining this surface estimator is shown to have a unique solution, which is approximated by means of the finite element method. The surface estimator is linear in the observations and some classical inferential results are derived. The described spatial smoother is especially well suited for applications where some knowledge of the problem suggests the choice of a partial differential operator modeling to some extent the phenomenon under study. With respect to classic spatial smoothing techniques, such as thin-plate splines, tensor product splines, kernel smoothing and kriging, the proposed method has also the advantage of efficiently dealing

with data distributed over irregularly shaped regions; moreover, it accounts for covariate information and has the capacity to comply with general conditions at the boundary of the domain.

E350: **A one-dimensional dispersion measure for multidimensional and functional spaces**
Presenter: **Alicia Nieto-Reyes**, Universidad de Cantabria, Spain
Co-authors: David Estebanez-Gallo

The variance is the dispersion measure par excellence in one-dimensional spaces. However, in multidimensional spaces we should use the variance-covariance matrix, which is quite a challenge when dealing with high dimensional spaces, atop the difficulty of interpretation. We will first propose two procedures for substituting by a number the variance-covariance matrix. The first one is based on the random projections method and the second one on distances. Then, we will see an application to test the covariance structure for the expression levels for sets of genes. Finally, we apply the procedure based on distances to functional spaces as a way to substitute by a number the variance function.

E711: **Interpretable support vector machines for functional data**
Presenter: **Belen Martin-Barragan**, Universidad Carlos III de Madrid, Spain
Co-authors: Rosa E. Lillo, Juan Romo

In many applications, getting a classifier with high correct classification rate is not the only issue. Practitioners are often concerned with the interpretability of the classifier, since they are reluctant to apply a classifier that they cannot interpret. In this paper we propose a new method, which we call *Interpretable Support Vector Machine for Functional Data* (ISVMFD) that produces SVM-based classifiers for functional data which has high classification accuracy and provides an interpretable coefficient function. The concept of interpretability might be widely discussed. The proposed method is extremely flexible to allow different choices for this interpretability.

ES43 Room Torrington ORDER-RESTRICTED INFERENCE AND APPLICATIONS Chair: Cristina Rueda

E044: **Order restricted inference for multivariate binary data with applications**
Presenter: **Ori Davidov**, University of Haifa, Israel
Co-authors: Shyamal Peddada

In many applications researchers collect multivariate binary response data under two, or more, naturally ordered experimental conditions. In such situations one is often interested in using all binary outcomes simultaneously to detect an ordering among the experimental conditions. To make such comparisons a general methodology for testing for the multivariate stochastic order between $K > 2$ multivariate binary distributions is developed. The proposed test uses order restricted estimators which, according to our simulation study, are more efficient than the unrestricted estimators in terms of mean squared error. The power of the proposed test was compared with several alternative tests. These included procedures that combine individual univariate tests for monotonic order, such as the union intersection tests and a Bonferroni based test. We also compared the proposed test with unrestricted Hotelling's T^2 type test. Our simulations suggest that the proposed method competes well with these alternatives. The gain in power is often substantial. The proposed methodology is illustrated by applying it to a two–year rodent cancer bioassay data obtained from the US National Toxicology Program (NTP).

E127: **Testing against linear inequality constraints in parametric regression**
Presenter: **Mary Meyer**, Colorado State University, United States of America
Co-authors: Jianqiang Wang

It is considered hypothesis testing in the ordinary least-squares regression model, where the null hypothesis is that a set of linear constraints holds for the parameters, and the alternative is that the constraints do not hold. Examples include testing whether order restrictions hold for a set of population means, whether a polynomial function is increasing over a range of the data, and whether a regression surface is increasing in both predictor variables. A novel yet simple test is proposed and it is shown that it has good large-sample properties. In particular, we show that the test size approaches the target along the boundary between the null and alternative sets, not just at the "least favourable" point. Simulations using some common regression models demonstrate that this approach can be rapid so that there is virtually no bias for some models and reasonable sample sizes, and the power of the test compares favourably to that of the standard likelihood ratio test and the modified "conditional" test.

E298: **Isotonized linear discriminant rules with application to medical studies**
Presenter: **Miguel Fernandez**, Universidad de Valladolid, Spain
Co-authors: David Conde, Cristina Rueda, Bonifacio Salvador

The problem of classifying observations into one of several populations when additional information on these populations is available is considered. The availability of this additional information is not unusual in statistical practice. For example, the severity of a disease (and therefore the population into which the individual is to be classified) may be known to be related with increased levels of a protein or group of proteins. Several approaches for the incorporation of this sort of additional information on the populations of interest in the definition of linear discriminant rules are described. The new rules have interesting properties both from the theoretical and practical point of view. Their good behaviour using simulation studies is shown. We apply them to real data dealing with bladder cancer obtaining much improved results. Moreover, it is shown that the true classification error of these rules is not properly estimated by the usual methods and, therefore, we propose new methods, specifically designed for these rules, for evaluating their true classification error.

E319: **Identification of tightly regulated temporally conserved cell-cycle genes in budding yeast, fission yeast and humans**
Presenter: **Shyamal Peddada**, National Institute of Environmental Health Sciences (NIH), United States of America
Co-authors: Miguel Fernandez, Cristina Rueda

A Cell division cycle is a well coordinated process in eukaryotes with cell-cycle genes exhibiting a periodic expression over time. There is considerable interest among cell-biologists to determine genes that are periodic in multiple organisms and whether such genes are also evolutionarily conserved in their time to peak expression. Interestingly, periodicity is not well conserved evolutionarily. A conservative estimate of number of periodic genes common to fission yeast (*Schizosaccharomyces pombe*) and budding yeast (*Saccharomyces cerevisiae*)("core set FB") is 35, while those common to fission yeast and humans (*Homo sapiens*)("core set FH") is 24. Since the cell-cycle is a carefully orchestrated process, we hypothesize that the temporal order of peak expression among the core set of genes is conserved between organisms. Using a novel statistical methodology we discover that *ace2, cdc18, mik1, histones (hhf1, hta2), klp5* are evolutionarily conserved in their temporal order in *S. pombe*, *S. cerevisiae* and *H. sapiens*. About 80% of FB genes and a third of FH genes are discovered to be temporally conserved between fission and budding yeasts and between fission yeast and humans, respectively. Methodology developed here could help biologists test other similar hypotheses.

ES37 Room Woburn BAYESIAN MODELLING AND COMPUTATION Chair: Antonio Lijoi

E269: **Prior information and dependence in regression**
Presenter: **Christopher Hans**, The Ohio State University, United States of America
Co-authors: Steven MacEachern, Agniva Som

We investigate prior distributions that are specifically designed to allow for the incorporation of prior information about the strength of the linear relationship between the response variable and the predictors in a regression model. The most commonly used standard prior distributions either assume a priori independence of regression coefficients or induce dependence solely via the empirical design matrix. While such priors may be good default priors in the absence of meaningful prior information, they are often incapable of encoding specific forms of prior knowledge available to researchers. In contrast to these approaches, the priors we consider inherit particular dependence structures that are generated through a variance decomposition, allowing for the direct incorporation of prior information about the strength of the regression relationship. We explore the structure of both the prior and posterior distributions, highlighting properties that are shared by "good" default priors. We compare the Bayesian variable selection and model uncertainty properties of our priors with those of standard priors. Computational methodology for working with this new class of regression models is highlighted throughout.

E270: **Efficient computation techniques for high dimensional Bayesian non parametric models**
Presenter: **Anjishnu Banerjee**, Duke University, United States of America
Co-authors: David Dunson, Surya Tokdar

Gaussian processes (GPs) are widely used in nonparametric regression, classification and spatio-temporal modeling, motivated in part by a rich literature on theoretical properties. However, a well known drawback of GPs that limits their use is the expensive computation, typically $O(n^3)$ in performing the necessary matrix inversions with n denoting the number of data points. In large data sets, data storage and processing also lead to computational bottlenecks and numerical stability of the estimates and predicted values degrades with n. To address these problems, a rich variety of methods have been proposed, with recent options including predictive processes in spatial data analysis and subset of regressors in machine learning. The underlying idea in these approaches is to use a subset of the data, leading to questions of sensitivity to the subset and limitations in estimating fine scale structure in regions that are not well covered by the subset. Motivated by the literature on compressive sensing, we propose an alternative random projection of all the data points onto a lower-dimensional subspace. We demonstrate the superiority of this approach from a theoretical perspective and through the use of simulated and real data examples. We then consider extensions of these approaches for dimensionality reduction in other non-parametric and probability models and connect them with a wide class of matrix approximations.

E450: **Flexible modelling of dependence in volatility processes**
Presenter: **Maria Kalli**, University of Kent, United Kingdom
Co-authors: Jim Griffin

It is proposed a novel volatility model that draws from existing literature in autoregressive stochastic volatility, aggregation of autoregressive processes, and recent work in Bayesian non-parametric modelling and sampling to create a dynamic SV model that explains long range dependence. By aggregating autoregressive log volatility processes and estimating the distribution of the autoregressive coefficient using a flexible Bayesian approach we provide more insight on the dynamic properties of such processes.

E683: **Variational approximations in geoadditive quantile regression**
Presenter: **Elisabeth Waldmann**, Goettingen University, Germany
Co-authors: Thomas Kneib

While ordinary regression analyses only the impact of covariates on the mean of a conditional distribution of a response variable, quantile regression analyses the impact on the quantiles. Estimation of parameters is commonly carried out by minimizing the sum of asymmetrically weighted errors (“check-function”). Treating this problem from a Bayesian viewpoint requires the specification of an error distribution leading to asymmetric weighting of the residuals as with the asymmetric Laplace distribution (ALD). Until now, inference in this context was mainly conducted by applying Markov chain Monte Carlo techniques (MCMC), which can be adapted to quantile regression by reparametrization of the ALD as a Gaussian distribution with weights and an offset. MCMC tends to be slow, especially for complex models, like for example geoadditive models. These are a very flexible class of models, which comprise linear as well as nonlinear covariates in terms of, for example, continuous, spatial or random effects. We propose the usage of variational approximations (VA) as an alternative to MCMC methods. The idea behind VA is to calculate the parameters iteratively using Kullback-Leibler distances, which is computationally less expensive but also leads to less accurate estimates. This talk compares VA and MCMC and their performance in geoadditive quantile regression.

E366: **A probability for classification based on the Dirichlet process mixture model**
Presenter: **Ruth Fuentes-Garcia**, UNAM, Mexico

We provide an explicit probability distribution for classification purposes when observations are viewed on the real line and classifications are to be based on numerical orderings. The classification model is derived from a Bayesian nonparametric mixture of Dirichlet process model with some modifications. The resulting approach then more closely resembles a classical hierarchical grouping rule in that it depends on sums of squares of neighbouring values. The proposed probability model for classification relies on a numerical procedure based on a reversible Markov chain Monte Carlo (MCMC) algorithm for determining the probabilities. Some numerical illustrations comparing with alternative ideas for classification are provided.

ES58 Room S264 STATISTICAL MONITORING AND ITS APPLICATIONS I Chair: Abdulkadir Hussien

E068: **Truncated sequential monitoring**
Presenter: **Edit Gombay**, University of Alberta , Canada

New sequential monitoring strategies are considered, and compared to open-ended strategies. Their advantages are demonstrated, and some examples from applications are presented. These include multi-armed longitudinal clinical trials, monitoring surgical performance, and time series type data. The methods are simple and relatively easy to apply. Their theoretical background relies on stochastic processes that can be approximated by Brownian motions. One can describe these monitoring procedures as the continuous time versions of the well-known Pocock and O’Brien-Fleming group sequential procedures.

E073: **Monitoring equality of two process capability indices**
Presenter: **Ejaz Ahmed**, University of Windsor, Canada

A Process Capability Index (PCI) is a numeric summary that compares the behavior of a product or process characteristics to engineering specifications. We propose a sequential procedure for testing whether two processes are equally capable by using the process capability index, Cpm.

We employ a nonsequential Wald-type statistic and provide its sequential version by Brownian motion approximations. We point out that, as a byproduct, the nonsequential Wald-type statistic used here provides an easily computable alternative to Boyels' approximate F-test. We give an algorithm for conducting the sequential test and we examine its performance by Monte Carlo simulations. Finally, we illustrate the method by testing capability improvement of an industrial process before and after calibration based on published data.

E076: **Efficient detection of multiple changepoints within wind energy time series**
Presenter: **Rebecca Killick**, Lancaster University, United Kingdom
Co-authors: Idris Eckley

The problem of detecting multiple changepoints in large wind energy data sets is considered. In this setting the amount of data being collected is continually increasing and consequently the number of changepoints will also increase with time. An efficient and accurate analysis of such data is of considerable interest to those working in the energy sector, since understanding the characteristics of wind energy, is central to reliable design and operation of wind farms. Detecting the presence of changepoints in wind energy time-series is of particular importance, since statistical and engineering modelling of the wind environment typically assumes stationarity of the environment (in time). Drawing on recent work on efficient search methods, we compare and contrast the effect of different approaches to this data, focusing in particular on computational and statistical aspects. Se conclude by highlighting the importance of such computationally efficient methods in a wind energy setting.

E078: **Some aspects of risk-adjusted monitoring charts**
Presenter: **Abdulkadir Hussein**, University of Windsor, Canada

A class of risk-adjusted charts for monitoring binary outcomes in the context of health care performance is reviewed. We point out some shortcomings of the currently used risk-adjusted charts and propose remedies in the form of truncated sequential charts. We illustrate the proposed methods by using an example on monitoring surgical performance.

ES69 Room Bedford IMPRECISE PROBABILISTIC MODELING TO SOLVE STATISTICAL PROBLEMS I Chair: Sebastien Destercke

E527: **Nonparametric predictive inference to assess three-group diagnostic tests**
Presenter: **Faiza Ali**, Durham University, United Kingdom
Co-authors: Tahani Coolen-Maturi, Frank Coolen

Assessing the performance of diagnostic tests is of importance in many application areas. We present nonparametric predictive inference (NPI) for assessment of accuracy of diagnostic tests involving three groups with real-valued or ordinal test results, and we consider the choice of cut-off points for such classifications. This is achieved by developing NPI lower and upper receiver operating characteristic (ROC) surfaces, which are bounds to the empirical ROC surface. NPI is a frequentist statistical approach which is based on relatively few modelling assumptions, in which lower and upper probabilities are used to quantify uncertainty. For inferences on diagnostic tests, interest tends to be explicitly in future use of such tests, so a predictive approach seems attractive for such assessments and choice of cut-off points.

E528: **The ordering of future observations from multiple sources**
Presenter: **Tahani Coolen-Maturi**, University of Kent, United Kingdom
Co-authors: Faiza Ali, Frank Coolen

Inferences based on information from multiple sources are common in practice. We present nonparametric predictive inference (NPI) for the ordering of real-valued future observations from multiple independent sources, where the data can include right-censored observations. We present an algorithm for calculating the NPI lower and upper probabilities for the event that the next future observations from these sources are ordered in a specific way. Several applications of these NPI lower and upper probabilities are explored, including multiple comparisons and ranked set sampling. They are also important for assessment of accuracy of diagnostic tests, and their use as predictive alternatives to several statistical tests is illustrated.

E572: **Generalised Bayesian inference with conjugate priors, and a link to g-priors for Bayesian model selection**
Presenter: **Gero Walter**, LMU Munich, Germany

In generalised Bayesian inference, sets of priors are considered instead of a single prior, allowing for partial probability specifications and a systematic analysis of sensitivity to the prior. Especially when substantial information is used to elicit the prior, prior-data conflict can occur, i.e., data that are very unlikely from the standpoint of the prior may be observed. This conflict should show up in posterior inferences, alerting the analyst and, e.g., lead to a revision of prior specifications. However, when conjugate priors are used, a reasonable reaction is not guaranteed. Mostly, prior-data conflict is just averaged out, and in Bayesian regression, conflict in one regressor leads only to a non-specific reaction across all regressors. Generalised Bayesian inference can amend this behaviour by encoding the precision of inferences via the magnitude of the posterior set. The simplified natural conjugate prior most suited for generalised Bayesian regression has a link to the so-called g-prior, which is used for model selection in classical Bayesian regression.

E573: **Elicitation and inference for the imprecise Dirichlet model with arbitrary sets of hyperparameters**
Presenter: **Matthias Troffaes**, Durham University, United Kingdom
Co-authors: Dana Kelly, Gero Walter

A common concern in Bayesian inference is sensitivity to the prior distribution, particularly when events of interest are very rare, in which case zero counts can have a pervading influence on any inference made. One way to analyse sensitivity to the prior is simply by using sets of priors. In this methodology, elicitation is much easier, as we can deal with partial probability specifications. In addition, it allows for robust decisions to be made, and enables policy makers to identify areas where additional information is required. We start out from the so-called imprecise Dirichlet model, a generalisation of the Dirichlet-Multinomial model, where one typically specifies a lower or upper probability (or both) for each category, along with a learning parameter, which determines how quickly we learn from data. In many applications, prior information is not near-vacuous, and handling prior-data conflict properly is non-trivial. In this case, the choice of sets of hyperparameters has received little attention so far. We focus on modelling and elicitation of sets of hyperparameters, and see that it is mandatory to ensure a range for the learning parameter, rather than a single value, for good performance of the posterior bounds.

E946: **Modelling the observational process using coherent lower previsions**
Presenter: **Gert de Cooman**, University of Ghent, Belgium
Co-authors: Arthur Van Camp

We deal with the general problem of learning about a random variable Y, given observations of another random variable X that is informative about Y. In doing so, we make a key distinction between X itself, which we interpret as latent or unobservable, and the observation that we actually make of it, which we identify with a new, manifest random variable O. Our beliefs about Y conditional on X are described by a coherent lower prevision $\underline{P}_s(\cdot|X)$ on $\mathcal{L}(\mathcal{Y}\times\mathcal{X})$, which is called the *inference model*. We will also assume that there is a *measurement model* about the relation

between O and X, which has the form $\underline{P}_{dm}(\cdot|O)$, defined on $\mathcal{L}(\mathcal{X} \times O)$. The central ancillarity assumption that we make is that O is epistemically irrelevant to Y, conditional on X, so $\underline{P}(g|x,o) = \underline{P}_s(g|x)$ for all g in $\mathcal{L}(\mathcal{Y})$, x in $\mathcal{X}$ and o in O. We investigate when it is possible to construct the least-committal conditional lower prevision $\underline{E}(f|O)$ on $\mathcal{L}(\mathcal{Y} \times \mathcal{X})$ that is coherent with both the sampling and the measurement model, and possibly with an additional prior assessment about X in terms of a lower prevision $\underline{P}_0$ on $\mathcal{L}(\mathcal{X})$. We show that our treatment encompasses and generalises a number of well-known approaches to important problems: dealing with missing or incomplete data, and vague or imperfect observations in the sense of Jeffreys, leading to a Generalised Jeffreys Rule.

Saturday 17.12.2011 16:35 - 18:40 Parallel Session F – CFE

CS97 Room B18 FINANCIAL ECONOMETRICS III **Chair: Willi Semmler**

C025: **Predicting housing prices according to expected future interest rate**
Presenter: **Nissim Ben David**, Emek Yezreel , Israel

A theoretical model is presented for predicting future housing prices as a function of the expected future interest rate, housing depreciation and rent rate. Focusing on the notion of arbitrage, where the returns on investing in housing is equal to the return on bonds, housing prices in the U.S. are forecast as a function of the expected future interest rate on corporate bonds graded at categories AAA or AA. It is got that a change in the expected future bond yield will lead to a change in future housing prices. Since we can use the present yield of bonds expiring at different points of time in the future to predict the expected future return of bonds, we can use the current information on bond yields in order to predict future housing prices.

C640: **Some issues with exponential STAR models for the modelling of exchange rate regimes**
Presenter: **Daniel Buncic**, University of St. Gallen, Switzerland

The exponential smooth rransition autoregressive (ESTAR) model has become one of the workhorse statistical models for the modelling of exchange rate regimes. We investigate identification issues in ESTAR models which are related to the size of the γ parameter that controls the shape and the magnitude of the exponential regime weighting function. A common empirical feature in many exchange rate series is that the inner regime thresholds are fairly wide apart. For the ESTAR model to be able to fit such wide thresholds, the γ parameter of the exponential function needs to be small. A consequence that follows from this is that for small values of the γ parameter, the exponential weighting function is well approximated by a simple quadratic function. This result leads to parameter identification issues and estimation difficulties in ESTAR models. Three well known empirical studies are used in a simulation study to show that the fitted ESTAR models cannot be estimated unless additional parameter restrictions are imposed on the model.

C670: **The origins of increasing trend in correlations among European stock markets: Evidence from smooth transition conditional correlation approach**
Presenter: **Mehmet Fatih Oztek**, Middle East Technical University, Turkey
Co-authors: Nadir Ocal

Multivariate generalized autoregressive conditional heteroscedasticity (MGARCH) models with smooth transition conditional correlation (STCC) are employed to evaluate stock market integration under a time varying framework. For the period January 1998-December 2010, we consider the correlation structure among new European Union (EU) member countries, (Poland, Czech Republic, Hungary, Bulgaria, and Romania) and EU core countries, namely Germany, France, Denmark, Sweden, UK and Greece as well as Turkey and EU core countries to examine whether increasing correlation levels and timing of increases in correlation is special to becoming a new member or special to regional dynamics. We also investigate the correlation structure between the US market and new member countries including Turkey to uncover whether the correlation structures between new members of EU and EU core countries are due to region-specific factors or governed by global trends. Empirical results imply increasing trend in correlations among all European countries including Turkey irrespective of being a new member country or not. The similarity of the correlation dynamics between Turkish and EU markets and Turkish and US markets suggests that the increasing correlation with EU cannot be attributed to accession strategy of EU or to regional specific developments but can be the result of global developments in financial markets.

C930: **The impact of scale effects on the prevailing Internet-based banking model in the US**
Presenter: **Alexandre Momparler**, Universitat de Valencia, Spain
Co-authors: Francisco J. Climent

Internet-based banks use a technology-intensive production process that may benefit from scale effects as they grow larger. It is analyzed whether the predominant Internet-primary bank in the USA generates technology-based economies of scale. There is evidence of both favorable and adverse technology-based scale effects. As the leading Internet-primary bank gets larger, the financial performance gap with traditional banks shrinks while some of its critical competitive advantages wear down. The results suggest that unless the prevailing Internet-primary bank preserves the distinctive advantages of the Internet-based business model as it improves financial performance, it might end up converging with its branching competitors.

C952: **Modelling dynamic dependencies between CDS and the equity market with regime switching copulas**
Presenter: **Fei Fei**, Cass Business School, United Kingdom
Co-authors: Ana-Maria Fuertes, Elena Kalotychou

The high trading volume of CDS index-related financial instruments has made the understanding of the relationship between CDS returns and stock market returns and volatilities important for financial analysts, traders and economic policy makers. This paper explores the extent to which default risk dependences with the equity market have changed during the recent financial crisis. To this end, we propose a regime switching (RS) dynamic copula that allows dependences to be time-varying with different patterns across regimes. Some attractive features of the model are, first, that the evolution of the copula parameter, which is determined by a time-varying regression function, makes the dynamic expression of correlations tractable. Second, the regime-switching setup is general enough to embed the non-regime-switching model as a special case. Third, the copulas are able to capture nonlinear correlations and dependences at tails. We find that the regime-switching dynamic copula model outperforms simpler static or time-varying copula models. Two dependence regimes emerge robustly over time in the sector indices, CDS Auto-Vstoxx or CDS SubFin-Stoxx, but appear shorter in length and more volatile for the aggregated CDS index. Higher tail dependences are observed during crisis periods than in normal periods.

CS26 Room B34 BAYESIAN EMPIRICAL MACROECONOMICS **Chair: Gary Koop**

C058: **An efficient approach to estimate and forecast in the presence of an unknown number of change-points**
Presenter: **Yong Song**, University of Toronto, Canada
Co-authors: John Maheu

A new efficient approach to model and forecast time series data with an unknown number of change-points is developed. Given a conjugate prior for the time-varying parameters which characterize each regime, the predictive density and the posterior of the change-points have closed forms. The prior is further modeled as hierarchical to exploit the information across regimes. This framework allows breaks in the variance, the regression coefficients or both. In addition to a time-invariant structural change probability, one extension assumes the regime duration has a Poisson distribution. A new Markov chain Monte Carlo sampler draws the parameters from the posterior distribution efficiently. This framework allows for the analysis of datasets of several thousand observations. Empirical examples illustrate the benefits of our approach to parameter estimation and forecasting.

C883: **Exchange rate risk premium, monetary policy and new Keynesian models**
Presenter: **Jorn Inge Halvorsen**, Norwegian Business School, Norway
Co-authors: Michal Zdenek

State-of-the-art versions of new Keynesian model employed at central banks typically show that modeling the exchange rate risk premium (adopting Bayesian estimation) as an exogenous stochastic process is of great help in order to fit aggregate data. We use higher order approximation techniques in order to make the exchange rate risk premium and portfolio holding endogenous in a baseline small open economy model and study its determinants and effect of changes in monetary policy. Our results demonstrate the risk premium is not invariant to policy changes, and that a welfare improving monetary policy of strict inflation targeting actually can bring about an increase in the risk premium. As such, our results demonstrate the invalidity of modeling practice of the risk premium employed at central banks and contributes to the ongoing policy discussion about the usefulness of new Keynesian models in forming quantitative policy advice.

C348: **Bayesian financial conditions indexes**
Presenter: **Dimitris Korobilis**, University of Glasgow, United Kingdom
Co-authors: Joshua C.C. Chan

Bayesian methods are used to construct global and regional financial conditions indexes. Bayesian model averaging and selection is used to determine in an automatic, data-based way which financial variables among hundreds of them should be used to extract a Global FCI. We also test thoroughly which econometric modelling assumptions are needed to maximize the informational content of FCIs for predicting recessions, including nonparametric factors or factor stochastic volatility. Our findings suggest the Bayesian FCIs have superior predictive ability to ad hoc methods based on principal components analysis and methods which ignore model selection/averaging.

C413: **Selecting predictors by Bayesian model averaging in bridge models**
Presenter: **Lorenzo Bencivelli**, Bank of Italy, Italy
Co-authors: Massimiliano Marcellino, Gianluca Moretti

The use of Bayesian model averaging (BMA) is proposed as a tool to select the predictors' set for bridge models. BMA is a computationally feasible method that allows us to explore the model space even in the presence of a large set of candidate predictors. We test the performance of BMA in now-casting by means of a recursive experiment for the euro area and the three largest countries. This method allows for a flexible choice of the information set month by month. We find that BMA based bridge models produce smaller forecast error than fixed composition bridges and perform at least as well as medium scale factor models.

C259: **Hierarchical shrinkage in time-varying parameter models**
Presenter: **Gary Koop**, University of Strathclyde, United Kingdom
Co-authors: Miguel Belmonte, Dimitris Korobilis

EU-area inflation is forecast with many predictors using time-varying parameter models. The facts that time-varying parameter models are parameter-rich and the time span of our data is relatively short motivate a desire for shrinkage. In constant coefficient regression models, the Bayesian Lasso is gaining increasing popularity as an effective tool for achieving such shrinkage. Econometric methods are developed for using the Bayesian Lasso with time-varying parameter models. Our approach allows for the coefficient on each predictor to be: i) time-varying, ii) constant over time or iii) shrunk to zero. The econometric methodology decides automatically which category each coefficient belongs to. Our empirical results indicate the benefits of such an approach.

CS28 Room B35 PROBABILISTIC FORECASTING Chair: Gael M. Martin

C243: **Improving asset price prediction when all models are false**
Presenter: **John Geweke**, University of Technology Sydney, Australia
Co-authors: Garland Durham

Three alternative sources of information about volatility potentially useful in predicting daily asset returns are considered: past daily returns, past intra-daily returns, and a volatility index based on observed options prices. For each source of information we begin with one or more state-of-the art models, and then work from the premise that all of these models are false to construct a single improved predictive distribution for daily S&P 500 index returns. The criterion for improvement is the log predictive score, equivalent to the average probability ascribed ex ante to observed returns. The first implication of the premise is that models within each class can be improved. This is accomplished by introducing flexibility in the conditional distribution of returns, in volatility dynamics, and in the relationship between observed and latent volatility. The second implication of the premise is that model pooling will provide prediction superior to the best of the improved models. This is accomplished by constructing ex ante optimal pools. All procedures are strictly out-of-sample, recapitulating one-step-ahead predictive distributions that could have been constructed for daily returns beginning January 2, 1992, and ending March 31, 2010. The prediction probabilities of the optimal pool exceed those of the best of the state-of-the-art models by 1.47% and the worst of these models by 7.75%.

C105: **Combining predictive distributions**
Presenter: **Tilmann Gneiting**, University of Heidelberg, Germany
Co-authors: Roopesh Ranjan

Predictive distributions need to be aggregated when probabilistic forecasts are merged, or when expert opinions expressed in terms of probability distributions are fused. We take a prediction space approach that applies to discrete, mixed discrete-continuous and continuous predictive distributions alike, and study combination formulas for cumulative distribution functions from the perspectives of coherence, probabilistic and conditional calibration, and dispersion. Both linear and non-linear aggregation methods are investigated, including generalized, spread-adjusted and beta-transformed linear pools. The effects and techniques are demonstrated theoretically, in simulation examples, and in case studies on density forecasts for S&P 500 returns and daily maximum temperature at Seattle-Tacoma Airport.

C251: **Properties of professional forecasters' probability forecasts**
Presenter: **Kenneth Wallis**, University of Warwick, United Kingdom
Co-authors: Gianna Boero, Jeremy Smith

The Bank of England Survey of External Forecasters, like the better-known US Survey of Professional Forecasters, is a quarterly survey of point forecasts and density forecasts of inflation and GDP growth. Survey average forecasts are published in the Bank's quarterly Inflation Report. The Bank follows the practice of the SPF in making available to researchers the individual responses to the survey, made suitably anonymous. We analyse the quarterly series of individual one-, two- and three-year ahead forecasts of inflation and growth over the period 2006-2010. It is seen that the published survey average forecasts mask considerable individual heterogeneity in both point and density forecasts. Persistent individual biases towards optimism or pessimism in point forecasts have been observed in studies of other forecast surveys, and we obtain the same finding. A new finding is that similar persistence appears in individual forecasters' relative levels of uncertainty, as expressed in their subjective probabilities.

Different forecasters use different models and methods, and their subjective probabilities reflect different views of forecast uncertainty. These individual characteristics need to be taken into account in panel data studies of forecasts.

C094: **Non-parametric estimation of forecast distributions in non-Gaussian, non-linear state space models**
Presenter: **Gael Martin**, Monash University, Australia
Co-authors: Jason Ng, Catherine Forbes, Brendan McCabe

The object of this paper is to produce non-parametric maximum likelihood estimates of forecast distributions in a general non-Gaussian, non-linear state space setting. The transition densities that define the evolution of the dynamic state process are represented in parametric form, but the conditional distribution of the non-Gaussian variable is estimated non-parametrically. The filtering and prediction distributions are estimated via a computationally efficient algorithm that exploits the functional relationship between the observed variable, the state variable and a measurement error with an invariant distribution. Simulation experiments are used to document the accuracy of the non-parametric method relative to both correctly and incorrectly specified parametric alternatives. In an empirical illustration, the method is used to produce sequential estimates of the forecast distribution of realized volatility on the S&P500 stock index during the recent financial crisis. A resampling technique for measuring sampling variation in the estimated forecast distributions is also demonstrated.

CS35 Room B36 MODELING AND INFERENCE ON ASSET PRICE BUBBLES Chair: Ivan Paya

C675: **Rational bubbles in US stock prices: A co-explosive vector autoregressive approach**
Presenter: **Tom Engsted**, CREATES, Aarhus University, Denmark
Co-authors: Bent Nielsen

We derive the parameter restrictions that a standard equity market model implies for a bivariate vector autoregression for stock prices and dividends, and we show how to test these restrictions using likelihood ratio tests. The restrictions, which imply that stock returns are unpredictable, are derived both for a model without bubbles and for a model with a rational bubble. In both cases the restrictions can be tested through standard chi-squared inference. The analysis for the no-bubble case is done within the traditional Johansen model for I(1) variables, while the bubble model is analysed using a co-explosive framework. In addition, we extract the bubble component as the explosive common trend from the Granger-Johansen representation of the co-explosive model. The methodology is illustrated using US stock prices and dividends for the period 1872-2000.

C284: **Identification of social interaction effects in financial data: inference of herd behavior via Monte Carlo simulations**
Presenter: **Tae-Seok Jang**, University of Kiel, Germany
Co-authors: Thomas Lux

It is pursued to estimate the parameters of a stochastic agent-based model (ABM) of an artificial financial market with social interactions among market participants. In our empirical evaluation of the ABM, we extend the method of indirect inference by using a non-overlapping block bootstrap method for weakly dependent data. This simulation -based estimator can identify the 'deep' parameters of the model by matching (as closely as possible) the moments of interest. We first explore the behavior of this estimator via a Monte Carlo study considering three cases: weak social interaction with a moderate level (standard deviation) of news about fundamentals, moderate social interaction with a weak level of news, strong interaction with a high level of news. Given our choice of moment conditions, the SMM estimator is able to consistently recover the true parameters. Our empirical application of the model to five major exchange rates suggests that the mimetic contagion of investors is an important factor to explain historical prices in various foreign exchange markets. According to our framework, historical volatility of returns in the financial markets under scrutiny can be decomposed into news innovations (65%) and social interaction effects (35%).

C553: **A possible speculative bubble in the price of gold**
Presenter: **Jedrzej Bialkowski**, University of Canterbury, New Zealand
Co-authors: Martin Bohl, Patrick Maurice Stephan, Tomasz Piotr Wisniewski

Motivated by the current gold price boom, we investigate whether the rapidly growing investment activities have triggered a new asset price bubble. We draw on the convenience yield model and use commodity dividends to derive gold's fundamental value. Based on the deviations of the actual gold price from its fundamental value, we apply a Markov regime-switching Augmented Dickey- Fuller test to detect ex post and identify ex ante speculative gold price bubbles. The empirical evidence is favorable for a fundamentally justified price level even during the current period of a drastically rising gold price.

C275: **Testing for asset price bubbles: the role of fat tails and endogeneity**
Presenter: **Ivan Paya**, Lancaster University Management School, United Kingdom
Co-authors: Efthymios Pavlidis, David Peel

A number of Monte Carlo experiments are undertaken in order to examine the efficiency of different estimation methods (OLS, FM-OLS, LAD, FM-LAD) in the asset price equation for different processes of the error term such as fat-tail distributions and different bubble processes. In a second stage, we investigate how the inaccuracy of the estimated coefficients and therefore the residuals of the asset price equation employing the regression methods mentioned above will impact on the bubble detection tests, typically cointegration and unit root tests. We analyze how likely it is to reject the non-existence of bubbles in those cases. We extend previous work in the literature by expanding the set of unit root tests to recently developed ones such as the quantile autoregression methods and time-varying unit root test. We apply all these bubble detection methods to a series of asset prices such as the Nasdaq, S&P500, house prices and use our results to make inferences about the presence of bubbles/explosive processes in those markets.

CS40 Room B33 DYNAMIC MODELLING OF REALIZED COVARIANCE MATRICES Chair: Giuseppe Storti

C095: **Measuring volatility transmission between the US and German stock markets**
Presenter: **Bastian Gribisch**, Christian-Albrechts-University Kiel, Germany
Co-authors: Vasyl Golosnoy, Roman Liesenfeld

Using a conditional autoregressive Wishart (CAW) framework for the realized (co)variances of the Dow Jones and DAX indices, we analyze intra-daily volatility spillovers between the U.S. and German stock markets. The proposed model specification accounts for 3 distinct intraday phases resulting from the partly overlapping opening hours in the U.S. and German markets. The impulse response analysis within the 3-phase CAW model reveals significant volatility spillover effects during a trading day, both from the U.S. to German market and vise versa, indicating that volatility is like a meteor shower, which rains down on earth.

C145: **A Kalman filter with EM approach for multivariate realized covariance estimation**
Presenter: **Fulvio Corsi**, University of Lugano, Switzerland
Co-authors: Francesco Audrino, Stefano Peluso

Motivated by the need of an unbiased and positive-semidefinite estimator of multivariate realized covariance matrices, noisy and asynchronous ultra-high-frequency asset prices are modeled in a state-space framework with missing data. Then the covariance matrix of the latent states is estimated through a Kalman smoother and Expectation Maximization (KEM) algorithm. In the expectation step, by means of the Kalman filter with missing data, we reconstruct the smoothed and synchronized series of the latent price processes. In the maximization step, we search for covariance matrices that maximize the expected likelihood obtained with the reconstructed price series. Iterating between the two EM steps we obtain a KEM-improved covariance matrix estimate which is robust to both asynchronicity and microstructure noise, and positive-semidefinite by construction. In a Monte Carlo simulation setting, we compare our KEM estimator with alternative covariance matrix estimates obtained with the Hayashi and Yoshida pairwise method and Multivariate Realized Kernel.

C177: **Forecasting covariance matrices: A mixed frequency approach**
Presenter: **Roxana Halbleib**, ECARES, Belgium
Co-authors: Valeri Voev

The potential of using mixed-frequency data for the purpose of covariance forecasting is explored theoretically and empirically. The idea is to use high-frequency (intraday) data to model daily realized volatilities combined with low-frequency (daily) data as input to the correlation model. The theoretical contribution is to derive conditions which ensure that a mixed-frequency forecast has a smaller mean squared forecast error than a similar pure low-frequency or pure high-frequency specification. The conditions are very general and do not rely on distributional assumptions of the forecasting errors or on a particular model specification. We show how mixed-frequency models can be implemented in practice and demonstrate the ease of empirically verifying the theoretical conditions. The analysis reveals that using high-frequency data is particularly useful for forecasting in periods of high volatility and correlation.

C954: **Econometric analysis of vast covariance matrices using composite realized kernels**
Presenter: **Kevin Sheppard**, University of Oxford, United Kingdom
Co-authors: Neil Shephard, Asger Lunde

We propose a composite realized kernel to estimate the ex-post covariation of asset prices. Composite realized kernels are a data efficient method where the covariance estimate is composed of univariate realized kernels to estimate variances and bivariate realized kernels to estimate correlations. We analyze the merits of our composite realized kernels in an ultra high dimensional environment, making economic decisions solely based on yesterday's data. The first application is a minimum variance portfolio exercise and this is followed by an investigation of portfolio tracking. The data set is tick-by-tick data comprising 473 US equities over the sample period 2006-2009. We show that our estimator is able to deliver a significantly lower portfolio variance than its competitors.

C171: **CAW-DCC: A dynamic model for vast realized covariance matrices**
Presenter: **Giuseppe Storti**, University of Salerno, Italy
Co-authors: Luc Bauwens

A dynamic model for realized covariance matrices is proposed, assuming a Wishart conditional distribution. The expected value of the realized covariance matrix is specified in two steps: a model for each realized variance, and a model for the realized correlation matrix. The realized variance model is taken within the menu of existing univariate models. The realized correlation model is a dynamic conditional correlation model. Estimation is organized in two steps as well, and a quasi-ML interpretation is given to each step. Moreover, the model is applicable to very large matrices since estimation can be done by the composite likelihood method.

CS55 Room G16 CONTRIBUTIONS IN FINANCIAL MARKET AND THE MACROECONOMY Chair: Ana-Maria Fuertes

C042: **Quantitative easing and bond yields: results from a macro-finance yield curve**
Presenter: **Alex Waters**, The University of Kent, United Kingdom
Co-authors: Jagjit Chadha

A macro-finance nominal and real forward curve is estimated for the UK over the period of inflation targeting, since 1993. The Kalman filter is used to extract the four dynamic latent factors of the term structure, as in accordance to a previous methodology to fit the term structure of interest rates. We condition these estimated factors on various macroeconomic variables in a dynamic seemingly unrelated regression and find that the interest rate forwards responds to a wide variety of macroeconomic data in a manner consistent with a simple NK model. We show using static impulse response analysis how one macroeconomic shock to each latent factor can simultaneously effect the forward curve across all of the maturities that we study. The model is used to estimate the supply effects from debt issuance on forward rates and also to gauge the impact of quantitative easing on forward rates. We find that the programme of QE from March 2009 to January 2010 is likely to have reduced the medium term forwards by an amount similar to some recent event studies.

C647: **Sovereign bond yield spreads: A time-varying coefficient approach**
Presenter: **Burcu Erdogan**, Universitat Trier, Germany
Co-authors: Kerstin Bernoth

The determinants of sovereign bond yield spreads across 10 EMU countries between Q1/1999 and Q1/2010 are studied. We apply a semiparametric time-varying coefficient model to identify, to what extent an observed change in the yield spread is due to a shift in macroeconomic fundamentals or due to altering risk pricing. We find that at the beginning of EMU, the government debt level and the general investors' risk aversion had a significant impact on interest differentials. In the subsequent years, however, financila markets paid less attention to the fiscal position of a country and the safe haven status of Germany diminished in importance. By the end of 2006, two years before the fall of Lehman Brothers, financial markets began to grant Germany safe haven status again. One year later, when financial turmoil began, the market reaction to fiscal loosening increased considerably. The altering in risk pricing over a time period confirms the need of time-varying coefficient models in this context.

C699: **Volatility at very high frequencies: New estimates, new interpretations**
Presenter: **Christian Mueller-Kademann**, Zurich University of Applied Sciences, Switzerland

A series of very high frequency volatility estimates across time, space, and markets is presented. All data sets feature a strong positive association between volatility and trading activity. It is suggested to interpret this finding as strong evidence against the standard objective approach to economic modelling. The implications of these results for macroeconomic policy and theory are discussed as are alternatives and suggestions for future research.

C871: **A risk-driven approach to exchange-rate modelling**
Presenter: **Piotr Keblowski**, University of Lodz, Poland
Co-authors: Aleksander Welfe

The CHEER hypothesis augmented with risk as perceived by financial investors making their decisions is discussed. We have found that in addition to price and interest rate differentials the sovereign credit default risk also significantly determines the exchange rate. Four long-run relationships have been identified: two connecting the term spreads with the inflation rates, one characterizing the behaviour of the exchange rate and one describing the inflation rate in Poland. With a system of nine variables we have shown that two of the five common trends are cumulated shocks to the exchange rate and sovereign credit default swap indices. These two common trends exert also a primary influence on the system's variables, as most loadings are related to them. The approach we propose seems to meet well the demands of the medium-run estimations of the equilibrium exchange rate.

C816: **Regulatory impact on price discovery in fragmented markets: the case of short selling constraints**
Presenter: **Thomas Katzschner**, Friedrich Schiller University Jena, Germany
Co-authors: Robert Jung

The establishment of multilateral trading facilities (MTFs) has increased the competition between trading platforms and has led to a fragmentation of European equity markets. As a consequence, a single security may be traded at several venues that are supervised by different regulatory authorities. Moreover, price discovery and market leadership of the traditional trading venues are under challenge. Within this framework, we seek to assess the impact of short selling constraints for stocks of financial institutions implemented in 2010 by the German federal financial supervisory authority on trading activity and price discovery. In our panel data setting, we observe a significant migration in trading volume, trading frequency and price discovery from the MTFs to Xetra. We conclude from this finding that the short selling ban moved high-frequency traders out of the market, affecting mainly MTFs with their low latency and low transaction cost business model.

CS76 Room B20 CONTRIBUTIONS IN TIME SERIES ECONOMETRICS I Chair: Alessandra Amendola

C620: **Wild bootstrap tests for autocorrelation in vector autoregressive models**
Presenter: **Paul Catani**, Hanken School of Economics, Finland
Co-authors: Niklas Ahlgren

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

C469: **Unit root Markov models**
Presenter: **Hans Arnfinn Karlsen**, University of Bergen, Norway

The classical unit root model can be written as a first order vector autorregressive model. This representation shows the Markov structure of this linear model. It is known that a nonlinear generalization of the unit root concept may be done by keeping the Markov property and translating the unit root property to null recurrence. However, by turning this around it is of interest to find conditions for null recurrence and, in particular, the so-called beta null recurrence for a linear vector autoregressive model. Results in this direction have recently been published. Such results are extended to more general parametric models of linear type. These extensions also avoid some of the technicalities in previous paper, and that is made possible by a more extensive use of the split chain.

C732: **The Fisher effect in the presence of time-varying coefficients**
Presenter: **Theologos Pantelidis**, University of Macedonia, Greece
Co-authors: Ekaterini Panopoulou

A resolution of the Fisher effect puzzle in terms of statistical inference is proposed. Our dataset of 15 OECD countries provides empirical evidence of time-varying coefficients in the data generating process of both the interest rates and inflation rates. We show that these time-varying dynamics crucially affect the behaviour of all the cointegration estimators considered, especially in small samples. When employing simulated critical values instead of the asymptotic ones, we provide ample evidence supporting the existence of a long-run Fisher effect in which interest rates move one-to-one with inflation rates in all countries under scrutiny except for Switzerland.

C799: **On testing for a bilinear unit root in financial time series**
Presenter: **Julio Angel Afonso Rodriguez**, University of La Laguna, Spain

A simple bilinear process with a unit root as a particular case of the stochastic unit root (STUR) family of non-stationary processes in first differences, and as an alternative to the linear unit root process is considered. Under a weak assumption on the bilinear parameter, we use the recently proposed stochastic limits for this process to show the consistency of some usual nonparametric tests of the null hypothesis of stationarity and to get the asymptotic distributions of some other tests of the linear unit root hypothesis. We found that the nonparametric variance-ratio test statistic leads us to incorrectly identify a stationary process, while the most commonly used tests against the alternative of a general STUR process show non-trivial power under a weak bilinear unit root process. Thus, as an alternative to the existing parametric test, we propose a new testing procedure based on a simple modification of the KPSS test with good size and power properties to consistently discriminate between a linear and a bilinear unit root. We derive the asymptotic null and alternative distributions and present an application to the series of daily returns of the IBEX 35 stock market index and the EURO/US Dollar exchange rate.

C788: **Polarization of forecast densities: A new approach to time series classification**
Presenter: **Shen Liu**, Monash University, Australia
Co-authors: Elizabeth Ann Maharaj

Time series classification has been extensively explored in many fields of study. Most methods proposed in the literature address this problem based on the raw data, the features of the data, or the models that generate the data. However, if the research interest is on a specific future time period (e.g. forecasting economic behaviours of European countries), the methods that directly relate to the properties of the forecasts are much more appropriate. A method based on the polarization of full forecast densities of the observed series is proposed. First, the bootstrap forecast

replicates incorporating bias-correction and stationarity-correction are obtained for each time series, and based on these replicates a non-parametric kernel density estimation technique is implemented to approximate the forecast densities. Then the dispersions of the forecast densities of pairs of series are estimated by a polarization measure, which captures the overlap between two distributions. Following the asymptotic distribution theory of the polarization measure, a hypothesis test is constructed to determine whether two forecast densities are significantly different. The proposed method is applied to both simulated and real data sets, and the results demonstrate desirable performance of this method, even when the sample size is small.

Sunday 18.12.2011 08:55 - 10:35 Parallel Session G – ERCIM

ES02 Room Gordon STATISTICAL ALGORITHMS AND SOFTWARE I Chair: Cristian Gatu

E482: **Fitting high-dimensional generalized linear models via generalized orthogonal-components regression**
Presenter: **Dabao Zhang**, Purdue University, United States of America
Co-authors: Yanzhu Lin, Min Zhang

Success of partial least squares in fitting multiple linear regression has motivated several extensions, on the basis of the iteratively reweighted least squares, to fitting generalized linear models. While these extensions may still be computational intensive, the regression coefficients can more or less diverge, especially for logistic regression. Nonetheless they pursue the convergence of regression coefficients and thus usually generate non-convergent component sequences. Following a supervised dimension reduction strategy, we propose to sequentially construct orthogonal components with each component the resultant of convergent construction, and simultaneously regress against these orthogonal components to fit a high-dimensional generalized linear model. Our proposed method is computationally privileged, and also produces both convergent regression coefficients and convergent orthogonal components. Both simulated and real data examples are presented to illustrate the competitive performance of the new approach in comparison with several existing methods.

E618: **Numerical solution of Levy-Ito type stochastic differential equations in Mathematica**
Presenter: **Florian Loecker**, WU Vienna, Austria

A Mathematica software package, able to numerically solve general stochastic differential equations of general Levy-Ito type in a lean, integrated and efficient manner, is presented. It was written with a special view towards the typical applications in mathematical finance and is able to accommodate multidimensional specifications, making the software ideally suited to explore dependence patterns in that framework. The package supports a wide array of approximation schemes, such as Euler- and Milstein-schemes, and offers the choice between equidistant or jump-adapted simulation. In the jump-adapted case, we employ a version of the classic inversion algorithm to sample jumps, which we modified in order to confront the challenges associated with the presence of possibly infinitely many small jumps. Jump times are sampled following the prescriptions of compound Poisson approximation. The actual solution is then achieved via compiled functions. The package also features a number of algorithms frequently used in finance, mainly related to change of measure, pricing and hedging.

E691: **Gaining inzight**
Presenter: **Chris Wild**, University of Auckland, New Zealand

iNZight ("insigh" with an NZ, New Zealand, pun) is gui-driven system for statistical data exploration written in R, though the user never has to interact with R. Our desire has been to provide a tool that will actively encourage the exploration of multivariate data sets and enable emphasis to be kept almost entirely on seeing what data is saying rather than learning how to drive software. We also want to keep the data students working with always "in their faces" to minimise abstraction. We have done these things using a drag and drop metaphor in which the software is driven by dragging the names of variables from the top of the data spreadsheet and dropping them in a small number of appropriate places. What is delivered instantly, and with almost no learning curve, is graphics (plots). To obtain numerical summaries and inferential information the user has explicitly to ask for them. Plots involving up to three variables require almost no system knowledge. Relationships involving up to six variables can be explored using only very basic plot types. We discuss pedagogical imperatives and describe system capabilities, design choices and the reasoning that led to them.

E705: **On a limitation of Markov chain Monte Carlo methods for categorical data**
Presenter: **Tamas Rudas**, Eotvos Lorand University, Hungary
Co-authors: Anna Klimova

A standard method of testing models specifying the canonical parameters of an exponential family for contingency tables (e.g., log-linear models) is to perform a walk among distributions having the same values of the mean value parameters (sufficient statistics) as the observed one, and compute the value of the Pearson statistic. Then the percentile of the Pearson statistic of the observed distribution is interpreted as an achieved p-value. This is conditional testing and one justification is that, under certain circumstances, conditional tests perform better than unconditional tests, although the literature on this issue is somewhat controversial. Another justification could be if the percentile of the Pearson statistic would not depend on the values of the mean value parameters, but only on the canonical parameters of interest, a situation called Neyman structure. One may expect this to be the case, because the canonical and the mean value parameters are variation independent. The talk illustrates that this is not true and explores the magnitude of the problem in simple cases. E.g., in a 2x2 table the model of independence assumes that the odds ratio is 1, but the Pearson statistic is not a monotone function of the deviation of the odds ratio from 1.

ES16 Room Bloomsbury STATISTICAL MACHINE LEARNING AND ROBUSTNESS Chair: Andreas Christmann

E063: **On support vector machines to estimate scale functions**
Presenter: **Robert Hable**, University of Bayreuth, Germany
Co-authors: Andreas Christmann

A main goal of regression is to derive statistical conclusions on the conditional distribution of the output variable Y given the input values x. Support vector machines (SVMs) are well established to estimate location functions like the conditional median or the conditional mean. The estimation of scale functions by use of SVMs has not attracted much attention so far, althouth modeling heteroscedasticity is important in many areas, e.g. to estimate the volatility in finance or in quality control. We investigate the estimation of scale functions by SVMs when the median function is unknown, too. We consider the median absolute deviation (MAD) and the interquantile range (IQR) as measures of scale. We will show that both approaches yield nonparametric consistent estimators.

E148: **Semiparametrically efficient inference based on signed ranks in symmetric independent component models**
Presenter: **Davy Paindaveine**, Universite libre de Bruxelles, Belgium
Co-authors: Pauliina Ilmonen

Semiparametric location-scatter models are considered for which the p-variate observation is obtained as $X = \Lambda Z + \mu$, where μ is a p-vector, Λ is a full-rank $p \times p$ matrix, and the (unobserved) random p-vector Z has marginals that are centered and mutually independent but are otherwise unspecified. As in blind source separation and independent component analysis (ICA), the parameter of interest is Λ. On the basis of n i.i.d. copies of X, we develop, under a symmetry assumption on Z, *signed-rank* one-sample testing and estimation procedures for Λ. We exploit the uniform local and asymptotic normality (ULAN) of the model to define signed-rank procedures that are semiparametrically efficient under correctly specified densities. Yet, as usual in rank-based inference, the proposed procedures remain valid (correct asymptotic size under the null, for hypothesis testing, and root-n consistency, for point estimation) under a very broad range of densities. We derive the asymptotic properties of the proposed procedures and investigate their finite-sample behavior through simulations.

E464: **Hypothesis testing and Bayesian inference: New applications of kernel methods**
Presenter: **Dino Sejdinovic**, University College London, United Kingdom
Co-authors: Kenji Fukumizu, Bernhard Schoelkopf, Alex Smola, Le Song, Bharath Sriperumbudur, Arthur Gretton

In the early days of kernel machines research, the “kernel trick" was considered a useful way of constructing nonlinear learning algorithms from linear ones, by applying the linear algorithms to feature space mappings of the original data. More recently, it has become clear that a potentially more far reaching use of kernels is as a linear way of dealing with higher order statistics, by mapping probabilities to a suitable reproducing kernel Hilbert space (i.e., the feature space is an RKHS). It will be described how probability distributions can be mapped to kernel feature spaces, and how to compute distances between these mappings. A measure of strength of dependence between two random variables follows naturally from this distance. Applications that make use of kernel probability embeddings include: a) Nonparametric two-sample testing and independence testing in high dimensional and non-Euclidean domains. For instance, we test whether text in English is translated from the French, as opposed to being random extracts on the same topic. b) Inference on graphical models, in cases where the variable interactions are modeled nonparametrically (i.e., when parametric models are impractical or unknown). In experiments, this approach outperforms state-of-the-art nonparametric techniques in 3-D depth reconstruction from 2-D images, and on a protein structure prediction task.

E753: **Some recent results for functional data analysis**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina

Functional data analysis provides modern analytical tools for data that are recoded as images or as a continuous phenomenon over a period of time. Because of the intrinsic nature of these data, they can be viewed as realizations of random functions often assumed to be in $L^2(I)$, with I a real interval or a finite dimensional Euclidean set. When working with more than one population, as in the finite dimensional case, a common assumption is to assume the equality of covariance operators. As in the multivariate setting, assuming equality of covariance operators is not satisfactory since the covariance operators may exhibit some common structure. We will briefly discuss the extension to the functional setting of the common principal component model that has been widely studied when dealing with multivariate observations. We present estimators of the unknown parameters as well as a robust approach for the common directions and their size related to the proposal given in the finite–dimensional setting. If possible, we will present a proposal for testing the hypothesis that the covariance operators of $k-$ populations of random objects are equal based on the norm of the difference among estimates of the operators.

ES19 Room Woburn ADVANCES IN ROBUST DATA ANALYSIS Chair: Luis A. Garcia-Escudero

E173: **Impact of contamination on the TCLUST procedure**
Presenter: **Christel Ruwet**, University of Liege, Belgium
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Agustin Mayo-Iscar

The TCLUST procedure is a robust clustering procedure that performs clustering with the aim of fitting clusters with different scatters and weights. As the corresponding objective function can be unbounded, a restriction is added on the eigenvalues-ratio of the scatter matrices. The robustness of the method is guaranteed by allowing the trimming of a given proportion of observations. The resistance to contamination of that procedure will be studied. Results concerning breakdown points and some new criteria in robust cluster analysis, such as the dissolution point and the isolation robustness, will be presented.

E411: **Similarity in k-sample problems**
Presenter: **Eustasio del Barrio**, Universidad de Valladolid, Spain

A k-sample problem , $k > 2$, where samples have been obtained from k (random) generators is considered. We are interested in identifying those samples, if any, that exhibit substantial deviations from a pattern constituted by most of the samples. This main pattern would be constituted by component samples which should exhibit some internal degree of similarity. This kind of problem, involving similarity, can be of interest in a variety of situations. As an example, consider an enterprise with a lot of factories with different suppliers for several components which interact in the final product. The interest focuses on analyzing if there are factories whose quality of production exhibits significant deviations from a generalized pattern. A null hypothesis of homogeneity is too strong to be considered as a realistic one, and similarity seems more appropriate. To detect deviations we need to use some pattern as reference, that in our setup is a hidden pattern. We develop a statistical procedure designed to search for a main pooling sample, detecting the samples that are significantly less similar with respect to (a pooled version of) the others. This is done through a probability metric, a bootstrap approach and a stepwise backward-search algorithm.

E415: **Robust analysis and data exploration with FSDA toolbox for MATLAB**
Presenter: **Francesca Torti**, University of Parma and University of Milano Bicocca, Italy
Co-authors: Marco Riani, Andrea Cerioli

The new Matlab software FSDA (Forward Search for Data Analysis, http://www.riani.it/MATLAB.htm) extends the MATLAB Statistics Toolbox to support a robust analysis of complex datasets which deviate from classical model assumptions, such as linearity or multivariate normality. The software is historically built around the Forward Search approach of Atkinson, Riani and Cerioli but it has been extended to the main traditional robust multivariate and regression techniques, including LMS, LTS, MCD, MVE, MM and S estimation. The main peculiarities of FSDA are (i) the presence of tools for dynamic interaction with a number of exploratory plots, (ii) the very detailed documentation system fully integrated with the MATLAB help, (iii) the particular care given to computational performances. The talk will show how to apply the main robust methods and interaction tools to a number of classical datasets that are part of FSDA. Outputs of special interest for their robust properties will be discussed.

E159: **The p-value line: some applications**
Presenter: **Alfonso Garcia-Perez**, UNED, Spain

Some parameters in Robust Statistical Analysis represent the usual trade-off between robustness and efficiency or power. This is the case of the tuning constant, in the Huber statistic, or the trimming fraction in the trimmed mean. It is proposed to use the p-value line to choose the values of these parameters because this function reflects both power and sensitivity.

ES21 Room S261 COPULA MODELLING AND COMPUTATIONAL ANALYSIS Chair: Richard Gerlach

E057: **Vine copulas with application to financial data**
Presenter: **Claudia Czado**, Technische Universitaet Muenchen, Germany

Using only bivariate copulas as building blocks, regular vine copulas constitute a flexible class of high-dimensional dependency models. They allow for different asymmetric dependence between different pairs of variables. Interest in these models has been growing steadily and they are finding successful applications to financial data. This multivariate copula class will be introduced and parameter estimation will be discussed. The flexibility of regular vines arises from the free choice of the regular vine tree sequence identifying the bivariate copula terms, the copula family for

each bivariate copula term and their corresponding parameter. Corresponding automated selection strategies for a given data set will be developed. The usefulness of this rich copula class in high dimensions will be demonstrated in a data set involving financial indices and stocks.

E195: **Likelihood inference for Archimedean copulas in high dimensions**
Presenter: **Marius Hofert**, ETH Zurich, Switzerland
Co-authors: Martin Maechler, Alexander McNeil

Explicit functional forms for the generator derivatives of well-known one-parameter Archimedean copulas are presented. They play an important role, e.g., for accessing conditional distributions, the Kendall distribution function, or the copula density. They also allow us to derive the density of asymmetric generalizations of Archimedean copulas such as Khoudraji transformed Archimedean copulas. After combining the theoretical results with a careful numerical implementation, maximum-likelihood estimation becomes feasible according to both precision and run time, even in large dimensions. In a large-scale simulation study, the performance of the maximum-likelihood estimator is compared to several alternatives proposed in the literature. A fast procedure for obtaining initial intervals for the likelihood optimization is constructed. Further, confidence intervals for the parameter vector can be derived based on the likelihood. Moreover, extensions to multi-parameter Archimedean families are feasible. All methods and examples presented are implemented in the open-source R package `nacopula` and can thus be easily accessed and studied.

E213: **Weighted scores method for regression models with dependent data**
Presenter: **Aristidis Nikoloulopoulos**, University of East Anglia, United Kingdom
Co-authors: Harry Joe, N. Rao Chaganty

There are copula-based statistical models in the literature for regression with dependent data such as clustered and longitudinal overdispersed counts, for which parameter estimation and inference are straightforward. For situations where the main interest is in the regression and other univariate parameters and not the dependence, we propose a "weighted scores method", which is based on weighting score functions of the univariate margins. The weight matrices are obtained initially fitting a multivariate normal copula, which admits a wide range of dependence. Asymptotic and small sample efficiency calculations show that our method is robust and nearly as efficient as maximum likelihood for fully specified copula models.

E860: **Invariant dependence structures under truncation**
Presenter: **Fabrizio Durante**, Free University of Bozen-Bolzano, Italy

In financial risk management, the construction of appropriate models for dependence is of obvious importance, due to the well recognized fact that neglecting dependence gives rise to a dramatic underestimation of the probability of occurrence of joint extreme events. Copulas provide a widely accepted tool for building such dependence models. In the recent literature, several authors have focused the attention on the characterization of a special class of copulas that describe the limit behavior of a random vector when some of the components are taking on extreme values and on the associated dependence properties. Following this line of research, we introduce a new family of bivariate copulas that can be interpreted as the limiting dependence structures when one component of a random vector is taking on extreme values. The introduced family can capture non-exchangeable dependence structures and can be easily simulated. Moreover, such a family presents strong probabilistic similarities with the class of Archimedean copulas from a theoretical and practical point of view. Related inference methods will be also discussed.

ES27 Room Jessel SEMIPARAMETRIC QUANTILE AND EXPECTILE REGRESSION Chair: Thomas Kneib

E029: **Nonparametric estimates of quantile curves for censored data**
Presenter: **Stanislav Volgushev**, Ruhr-Universitat Bochum, Germany
Co-authors: Holger Dette

Firstly, a new nonparametric estimate of conditional quantiles is proposed. It avoids the problem of crossing quantile curves (calculated for various $p \in (0,1)$). The method uses an initial estimate of the conditional distribution function in a first step and solves the problem of inversion and monotonization with respect to $p \in (0,1)$ simultaneously. The second part of the paper focuses on censored quantile regression, in particular twice censored data. It is demonstrated that the new estimates are asymptotically normal distributed and the performance of the new procedure is illustrated by means of a simulation study.

E203: **Fitting prediction intervals by boosting quantile regression**
Presenter: **Andreas Mayr**, FAU Erlangen-Nuremberg, Germany
Co-authors: Torsten Hothorn, Nora Fenske

Prediction intervals are a useful tool to express uncertainty in the prediction of future or unobserved realizations of the response variable in a regression setting. Standard approaches typically assume an underlying distribution function and use the variance of the estimation method to compute boundaries around the expected mean. An adaptation of gradient boosting algorithms is presented to compute intervals based on additive quantile regression, therefore avoiding distributional assumptions. The boundaries of prediction intervals are directly modelled by applying nonparametric quantile regression with linear as well as smooth effects, fitted by boosting, providing intrinsic variable selection and model choice. We apply this approach to construct prediction intervals for future body mass index (BMI) patterns of individual children based on a recent German birth cohort study with 2007 children. Since the BMI distribution in childhood is typically skewed depending on age, common parametric approaches are not appropriate here. The lengths of the estimated intervals are child-specific and increase, as expected, with the age of the child.

E161: **Semiparametric expectile regression**
Presenter: **Thomas Kneib**, Georg August Universitat Gottingen, Germany
Co-authors: Fabian Sobotka

Recent interest in modern regression modelling has focused on extending available model specifications beyond mean regression by describing the complete conditional distribution instead of only the mean. Quantile regression is the most common approach to achieve this goal even in the presence of heteroscedastic errors. Quantile regression estimates are obtained by minimising an asymmetrically weighted sum of absolute deviations from the regression line, a decision theoretic formulation of the estimation problem that avoids a full specification of the error term distribution. Expectile regression models that replace the asymmetrically weighted sum of absolute residuals with asymetrically weighted squared residuals are described. This allows us to use efficient iteratively weighted least squares approaches even in complex semiparametric models including for example nonlinear effects of continuous covariates, random effects or spatial effects. We discuss advantages and disadvantages of expectiles as compared to quantiles and illustrate their application along an example on childhood malnutrition in developing countries.

E083: **Additive models for quantile regression: Model selection and confidence bandaids**
Presenter: **Roger Koenker**, University of Illinois, United States of America

Additive models for conditional quantile functions provide an attractive framework for non-parametric regression applications focused on features of the response beyond its central tendency. Total variation roughness penalities can be used to control the smoothness of the additive components much as squared Sobelev penalties are used for classical L_2 smoothing splines. We describe a general approach to estimation and inference for additive models of this type. We focus attention primarily on selection of smoothing parameters and on the construction of confidence bands for

the nonparametric components. Both pointwise and uniform confidence bands are introduced; the uniform bands are based on the Hotelling tube approach. Some simulation evidence is presented to evaluate finite sample performance and the methods are also illustrated with an application to modeling childhood malnutrition in India.

ES52 Room Torrington MIXTURE MODELS IN R Chair: Bettina Gruen

E143: **Dependent mixture models with R**
Presenter: **Ingmar Visser**, University of Amsterdam, Netherlands
Co-authors: Maarten Speekenbrink

A broad class of dependent mixture models (or hidden Markov models) and the R-package depmixS4 that was written to estimate such models are presented. DepmixS4 includes standard Markov models, latent/hidden Markov models, and latent class and finite mixture distribution models. The models can be fitted on mixed multivariate data with distributions from the glm family, the (logistic) multinomial, or the multivariate normal distribution. Other distributions can be added easily, and an example is provided with the exgaus distribution. Parameters are estimated by the expectation-maximization (EM) algorithm or, when (linear) constraints are imposed on the parameters, by direct numerical optimization. We illustrate the model using varied data sets: univariate/multivariate models; single long time series with few change points; and multiple time series of regular switching behavior.

E149: **Clustering for multivariate continuous and discrete longitudinal data using R package mixAK**
Presenter: **Arnovst Komarek**, Charles University in Prague, Czech Republic

Multiple outcomes, both continuous and discrete are routinely gathered on subjects in longitudinal studies. We propose a model-based statistical method for clustering (classification) of subjects into a prespecified number of groups with a priori unknown characteristics on the basis of repeated measurements of all longitudinal outcomes. Its implementation in the R package mixAK is shown. We start by modeling the evolution of each outcome using the classical generalized linear mixed model (GLMM). Possible dependence between the values of different outcomes is captured by specifying a joint distribution of all random effects involved in the GLMM for each outcome. The basis for the subsequent clustering is provided by assuming a heteroscedastic mixture of multivariate normal distributions in the random effect distribution. Mainly for computational reasons we base the inference on a Bayesian specification of the model and simulation based MCMC methodology. We also discuss evaluation of uncertainty in the classification and adapt a recently proposed methodology for model comparison based on the posterior distribution of deviances to explore optimal number of clusters. Everything will be exemplified on the analysis of real data using the routines from the contributed R package mixAK.

E166: **Fitting finite mixtures of von Mises-Fisher distributions using the R package movMF**
Presenter: **Bettina Gruen**, JKU Linz, Austria
Co-authors: Kurt Hornik

In certain applications the direction implied by the observations is the only relevant information. If model-based clustering is applied to such data, a common approach is to assume that each component in the mixture distribution follows a von Mises-Fisher distribution. The R package movMF implements ML estimation of finite mixtures of von Mises-Fisher distributions with the EM algorithm. The different variants of the EM algorithm supported are "softmax" for the ordinary EM, "hardmax" for maximizing the classification likelihood and "stochmax" for the Stochastic EM. Estimation of the length of the parameter is only possible using numerical methods and different approximations and different methods are implemented. The functionality available in the package will be presented as well as the use illustrated.

E595: **Applying the R package gamlss.mx to investigate cognitive decline.**
Presenter: **Dimitrios Stasinopoulos**, London Metropolitan University, United Kingdom
Co-authors: Robert Rigby, Ardo Van den Hout, Gratiela Muniz

The use of the package gamlss.mx is described. The package has two main functions i) gamlssMX() for fitting finite mixtures regression models, and ii) gamlssNP() for fitting non-parametric random effect models. Both functions use the generalised additive models for location scale and shape (GAMLSS) as their engine. GAMLSS are semi-parametric regression models. They are parametric, in that they require a parametric distribution assumption for the response variable, and "semi" in the sense that the modelling of the parameters of the distribution, as functions of explanatory variables, may involve non-parametric smoothing functions. All the parameters of the distribution can be modelled as functions of the explanatory variables. The current R implementation of GAMLSS allows the fitting of more that 70 different distributions. This, in combination with the finite mixture feature of gamlssMX(), provides a very flexible tool for modelling highly skew and/or kurtotic response variables. The package gamlss.mx() will be used to investigate cognitive decline, where models are fitted to mini-mental state examination (MMSE) scores, the most widely used test to measure global cognition. MMSE scores take integer values in the 0-30 range, and its distribution has strong ceiling and floor effects. Results and diagnostics from the fitted model will be provided.

ES53 Room Senate FREQUENCY DOMAIN ANALYSIS Chair: Alessandra Luati

E224: **Stabilization of spectral matrix inversion and its uses**
Presenter: **Andrew Walden**, Imperial College London, United Kingdom

Multi-channel frequency-domain statistical methods such as multiple and partial coherence analysis and graphical modelling involve the inversion of spectral matrices over a range of frequencies. We discuss such inversion and show how the problem has some features which make it distinct from concentration matrix estimation in multivariate analysis. Techniques to stabilise spectral matrix inversion have been around and used for over 25 years; one simple approach is to create, at each frequency, a diagonal matrix with entries related to the estimated spectrum over frequencies, then multiply the diagonal matrix by a small weight parameter before adding it to the original spectral matrix before inversion. This is equivalent to adding white noise to the series and can also effectively suppress spectral leakage. The lasso-type estimator is an alternative approach and also involves a parameter choice. A real-data example shows how the parameters of the two methods can be chosen to produce almost the same Frobenius norms for the inverse spectral matrices. Another approach involves shrinkage-type ideas applied to the spectral matrix. It will be shown how the multitaper spectral estimation method is particularly suited to this scheme through estimation of the shrinkage weight parameter.

E236: **Prediction of stationary spatial processes with exponential expectrum**
Presenter: **Mohsen Pourahmadi**, Texas A&M University, United States of America

The predictors of a stationary time series are usually found either in the spectral- domain using the Kolmogorov-Weiner theory or in the time-domain by fitting finite-parameter autoregressive and moving average models to the data. Bloomfield's finite-parameter exponential spectral density has both of these features. Its extension to the stationary random fields setup is expected to play more prominent roles, since spatial time-domain models are not as flexible as those in the time series due to lack of a natural or unique definition of the past and future. The extension of finite-parameter exponential models to the spatial processes has been studied in the engineering literature in the contexts of texture and image analysis. However, the construction of predictors of stationary spatial processes with exponential spectrum has been lagging behind. A framework is developed for

forecasting such processes both in the time- and spectral- domain using recursive formulas, which express the predictor coefficients in terms of the espectrum or the Fourier coefficients of the logarithm of the spectrum. The framework is general enough to include long-memory or fractional exponential processes. We illustrate the procedure using both a simulation study and real data analysis.

E539: **Spectrum based inference for nonstationary processes**
Presenter: **Rainer Dahlhaus**, University of Heidelberg, Germany

For nonstationary processes there usually exists no unique spectral representation. For locally stationary processes it can be shown that the time varying spectral density exists uniquely in an asymptotic sense. Based on this result we use an estimate for the time varying spectrum to do inference on several aspects of the nonstationary process. We show that many statistics of interest can be written as a functional of this time varying spectral estimate. A key role in the theoretical treatment of these statistics is played by the empirical spectral process. We present a functional central limit theorem for such processes indexed by function spaces, a Bernstein type exponential inequality and a maximal-inequality. Furthermore, we indicate how the above results on the empirical spectral process can be used for the theoretical treatment of inference problems for locally stationary processes. In particular, we derive uniform rates of convergence for local parameter estimates and the time varying spectral density.

E605: **Rational spectral density models for lattice data**
Presenter: **Rajendra Bhansali**, Imperial College London & University of Liverpool, United Kingdom
Co-authors: Luigi Ippoliti, Richard Martin

Conditional autoregressive models and directly specified covariance models represent two widely-used classes of spatial models. We consider their generalization to models with rational spectra. Theoretical properties are discussed and fits to real data are compared. We show, in particular, that the models with rational spectra can arise from various filtering operations on conditional autoregressive models, such as addition, smoothing, sampling and coarser resolution, and examine the extent to which a low-order conditional autoregression may approximate a rational spectral density model by comparing the correlations and inverse correlations implied by both these classes of models. In addition, we also compare and contrast the fits provided by the rational spectral density models and the conditional autoregressive models for two texture image data sets.

ES81 Room S264 TIME SERIES MODELING AND COMPUTATION II Chair: Roland Fried

E155: **Sequential testing of changes in the drift of a stochastic process**
Presenter: **Josef G. Steinebach**, University of Cologne, Germany

Some recent results are presented about sequential monitoring procedures, which have been constructed for detecting an "abrupt" or "gradual" change in the drift parameter of a general stochastic process. We assume that the latter process, e.g., a renewal counting process, has been observed in discrete time and that it satisfies a certain (weak) invariance principle (with rate). It is shown that the tests can be constructed such that the "false alarm rate" attains a prescribed level (say) α and that the tests have "asymptotic power 1". A more precise analysis of the procedures under the alternative proves that the stopping times, suitably normalized, have a standard normal limiting distribution. A few results from a small simulation study are also presented in order to give an idea of the finite sample behaviour of the suggested procedures.

E501: **Subsampling inference for the autocovariances of long-memory time series**
Presenter: **Agnieszka Jach**, Universidad Carlos III de Madrid, Spain
Co-authors: Tucker McElroy

It is provided a self-normalization for the sample autocovariances (acvs) of a linear, long-memory (parameter d) time series with the innovations that have either finite 4th moment or are heavy-tailed with tail index $2 < \alpha < 4$. This circumvention is necessary because in the asymptotic distribution of the sample acv there are three rates of convergence that depend on the interplay between d and α. These three different rates (classical, tail-dominated, memory-dominated) lead to three different limit distributions. A self-normalized sample acv that eliminates that dependence is defined and the corresponding limit theorems formulated. The distribution of the self-normalized sample acv is approximated nonparametrically by subsampling, as its asymptotic distribution is still parameter-dependent. The subsampling-based confidence intervals for the process acvs are shown to have good empirical coverage rates in a simulation study. The impact of a data-driven block-size selection on the coverage in the subsampling procedure is assessed. The methodology is further applied to the log-squared returns of exchange rates.

E505: **Online signal extraction from data streams by robust moving window regression with automatic width adaption**
Presenter: **Matthias Borowski**, TU Dortmund University, Germany
Co-authors: Roland Fried

The online (also real-time or sequential) monitoring of data stream time series takes place in many fields. Such time series are often measured with high frequency, involving that the 'true' signal, which carries the relevant information, is corrupted by a varying amount of noise and outliers. Furthermore, the signal shows enduring and suddenly changing trends and level shifts. Online signal extraction is therefore a basic but challenging task. It has turned out that robust Repeated Median regression in a moving time window is an adequate approach for online signal extraction. However, the chosen window width n has large impact on the signal extraction. During periods of linear trends, a large n is requested to obtain smooth signal extractions. When a trend changes and/or a level shift occurs, n should be reduced to track the signal change with high accuracy. We present Repeated Median based procedures to detect such signal changes online. Filtering algorithms embedding these procedures then allow for the time-dependent and automatic selection of the window width. The filtering methods are compared w.r.t. the robustness and the power for signal change detection. We also present multivariate extensions of the filters to extract signals online from multivariate time series.

E651: **On testing stability in multivariate RCA models**
Presenter: **Zuzana Praskova**, Charles University in Prague, Czech Republic

Random coefficient autoregressive (RCA) models belong to a broad class of conditional heteroscedastic time series models because of their varying conditional variance and as such may be used in various applications. In the contribution we will consider a multivariate first-order RCA model and discuss a class of weighted estimators of matrices of unknown parameters that are computationally simpler than the conditionally weighted least-squares estimators, and, as numerical studies show, they behave well and better than the least-squares estimators. A test statistic for stability of parameters in the RCA models will be proposed and its asymptotic properties will be considered both under the null hypothesis (no change in parameters) and the alternative under minimal moment assumptions.

ES76 Room Bedford DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS II Chair: Sonja Kuhnt

E523: **Some kernels for Kriging models**
Presenter: **Olivier Roustant**, Ecole des Mines de St-Etienne, France

Kriging models, also called Gaussian processes, denote a class of spatial stochastic processes used to model time-consuming computer codes. Their stochastic nature is suited to uncertainty quantification, which is used for risk assessment as well as global optimization. Another advantage

of Kriging models is provided by their covariance kernel, that can be customized for a specific problem or objective. We will first present a class of particular ANOVA kernels that allow us to perform analytically a complete global sensitivity analysis, without recursion nor truncation issues. We will detail the two possible ways from which they are deduced: functional (with Reproducing Kernel Hilbert Spaces) or probabilistic. In a second part, we will illustrate the latest developments about kernels in the R package DiceKriging: addition, scaling, ... and creation of a new kernel.

E529: **Sequential screening with elementary effects**
Presenter: **Hugo Maruri Aguilar**, Queen Mary, University of London, United Kingdom
Co-authors: Alexis Boukouvalas, John Paul Gosling

The Elementary Effects (EE) method is a simple but effective screening strategy. Starting from a number of initial points, the method creates random trajectories to then estimate factor effects. In turn, those estimates are used for factor screening. Recent research advances have enhanced the performance of the elementary effects method and the projections of the resulting design. We concentrate on a new proposal which turns the elementary effects method into a sequential design strategy. After describing the methodology, some examples are given and compared against the traditional EE method.

E555: **Bayesian mixture of Gaussian processes for deterministic computer experiments**
Presenter: **Ofir Harari**, Tel Aviv University, Israel
Co-authors: David Steinberg

Use of Gaussian processes is a popular approach to analyzing data from computer experiments. Combining more than one Gaussian process in a surrogate model for computer simulation could prove useful when there is uncertainty regarding the family of correlation functions, or when one wishes to characterize both global trends and finer details, all in the same model. One such option is to fit a model of the form $Y(x) = \mu + pZ_1(x) + (1-p)Z_2(x)$, where both Z_1 and Z_2 are stationary Gaussian processes whose correlation functions differ in scale parameters. We suggest a fully Bayesian treatment of the problem, taking advantage of MCMC sampling methods and providing point estimates and Bayesian credible intervals with a high degree of success, according to simulation results.

E554: **Kernel interpolation**
Presenter: **Thomas Muehlenstaedt**, W.L. Gore & Associates GmbH, Germany

Surrogate interpolation models for time-consuming computer experiments are increasingly used in scientific and engineering problems. A new interpolation method, based on Delaunay triangulations and related to inverse distance weighting, is introduced. This method not only provides an interpolator but also uncertainty bands to judge the local fit, in contrast to methods such as radial basis functions. Compared to the classical Kriging approach, it shows a better performance in specific cases of small data sets and data with non-stationary behaviour.

ES65 Room Court SYMBOLIC DATA ANALYSIS Chair: Monique Noirhomme

E635: **A clustering algorithm for multiple relational data matrices**
Presenter: **Francisco de Assis Tenorio de Carvalho**, Universidade Federal de Pernambuco, Brazil
Co-authors: Yves Lechevallier

A clustering algorithm that is able to partition objects taking into account simultaneously their relational descriptions given by multiple dissimilarity matrices is presented. These matrices have been generated using different sets of variables and dissimilarity functions. These methods are designed to provide a partition and a prototype for each cluster as well as to learn a relevance weight for each dissimilarity matrix by optimizing an adequacy criterion that measures the fitting between the clusters and their prototypes. These relevance weights change at each iteration and are different from one cluster to another. Experiments with real-valued datasets from the UCI machine learning repository as well as symbolic data sets show the usefulness of the proposed clustering algorithm.

E377: **Conceptual clustering of symbolic data using a quantile representation: discrete and continuous approaches**
Presenter: **Paula Brito**, Universidade do Porto, Portugal
Co-authors: Manabu Ichino

Quantile representation provides a common framework to represent symbolic data described by variables of different types. It is based on the monotone property of symbolic objects, characterized by the nesting structure of the Cartesian join regions. On a discrete approach, the observed variable values are expressed by some predefined quantiles of the underlying distribution; alternatively, in a continuous setup, variable values are represented by the quantile function of the underlying distribution. Having a common representation model then allows for a unified analysis of the data, considering simultaneously variables of different types. A dissimilarity is used to compare data units: in the discrete approach this may be the Euclidean distance between standardized quantile vectors; in the continuous approach the Mallows distance between functions is appropriate. The proposed hierarchical clustering method proceeds bottom-up, merging at each step the two clusters with closest quantile representation. A discrete or continuous quantile representation for the new cluster is then determined from the (possibly weighted) mixture of the respective distributions. Even if Uniform distributions are assumed for the data, the formed clusters are generally not Uniform on each variable, allowing for different profiles to emerge. Examples allow illustrating and comparing the proposed approaches.

E387: **Distribution and symmetric distribution model - A linear regression model for histogram-valued variables**
Presenter: **Sonia Dias**, Instituto Politecnico de Viana do Castelo, Portugal
Co-authors: Paula Brito

About 30 years ago Schweizer advocated that “distributions are the numbers of the future". Following in his footsteps, Diday generalized the classical concept of variables in Multivariate Data Analysis and introduced Symbolic Data Analysis. In the new framework proposed, each individual or class of individuals may take a finite set of values/categories (multi-valued variables), an interval (interval-valued variable) or a frequency/probability distribution (modal-valued variable). Histogram-valued variables are a particular case of this latter kind of symbolic variables where for each variable and to each entity under analysis corresponds a finite number of intervals, with an associated weight. The observed distributions can be represented by a histogram or, alternatively, by its quantile function. In recent years, the development of statistical concepts and methods that use distributions or sets of numbers/categories instead of real numbers is expanding at a steady rate. The Distribution and Symmetric Distribution Model is a linear regression model for histogram-valued variables, where the observed distributions of the dependent and independent histogram-valued variables are represented by the quantile functions. It uses the Mallows Distance to evaluate the dissimilarity between the observed and estimated quantile functions. A goodness-of-fit measure is associated with the model. Some examples illustrate the proposed methodology.

E139: **A tree construction algorithm for complex data**
Presenter: **Monique Noirhomme**, University of Namur, Belgium
Co-authors: Teh Amouh

In supervised learning problems, we have a training set consisting of observations on a class variable Y for n entities described by means of p predictor variables $Y_1, \cdots, Y_p$. Variable Y is a dependent single-valued variable that takes its values from a set $\mathcal{C} = \{1, 2, \cdots, K\}$ of K classes. The goal is to find a model for predicting the value $Y(e)$ of Y for any new entity e, when given the values $Y_1(e), \cdots, Y_p(e)$ of variables $Y_1, \cdots, Y_p$ describing the new entity e. Many decision tree solution approaches are available in the literature. Basically they all allow only two possible cases regarding the values of the predictor variables : ordered values or nominal categories. Consequently they do not apply in cases where the values of predictor variable are neither ordered values nor nominal categories. Such cases are encountered in complex data analysis where the observed values on a given variable can be intervals or distributions or any other complex type data. Just as nominal categories, complex type data are unordered. But unlike nominal categories, measures of resemblance (similarity and dissimilarity) are possible with complex type data. We propose a decision tree learning algorithm that uses resemblance measures in its splitting strategy.

Sunday 18.12.2011 11:05 - 12:45 Parallel Session I – CFE

CSI02 Room Beveridge BAYESIAN ECONOMETRICS Chair: Gary Koop

C219: **Bayesian regularization in latent variable models through shrinkage priors**
Presenter: **Sylvia Fruehwirth-Schnatter**, Vienna University of Economics and Business, Austria

Bayesian methods are nowadays widely accepted among statisticians and econometricians, mostly because Bayesian inference in combination with MCMC methods allows great flexibility in modelling complex data. However, many researchers feel uneasy about choosing priors, because they want to be as objective as possible. In the hope to let the data speak themselves they try to identify non-informative priors. The present talk gives some reasons why it is often advantageous to use carefully selected informative priors instead of non-informative ones, in particular in cases where the likelihood function is highly non-regular. Examples include Bayesian inference for finite mixture models when the number of components is unknown and variable selection problems in latent variable models.

C227: **Mixtures of $g-$priors for Bayesian model averaging with economic applications**
Presenter: **Mark Steel**, University of Warwick, United Kingdom
Co-authors: Eduardo Ley

The issue of variable selection in linear regression modeling, where we have a potentially large amount of possible covariates and economic theory offers insufficient guidance on how to select the appropriate subset is examined. In this context, Bayesian model averaging presents a formal Bayesian solution to dealing with model uncertainty. Our main interest here is the effect of the prior on the results, such as posterior inclusion probabilities of regressors and predictive performance. We combine a Binomial-Beta prior on model size with a g-prior on the coefficients of each model. In addition, we assign a hyperprior to g, as the choice of g has been found to have a large impact on the results. For the prior on g, we examine the Zellner-Siow prior and a class of Beta shrinkage priors, which covers most choices in the recent literature. We propose a benchmark Beta prior, inspired by earlier findings with fixed g, and show it leads to consistent model selection. Inference is conducted through a Markov chain Monte Carlo sampler over model space and g. We examine the performance of the various priors in the context of simulated and real data. For the latter, we consider two important applications in economics, namely cross-country growth regression and returns to schooling. Recommendations to applied users are provided.

C427: **Massively parallel posterior simulation for Bayesian inference**
Presenter: **John Geweke**, University of Technology Sydney, Australia

Massively parallel desktop computing in the form of graphical processing units (GPUs) and complementary software is now well within the reach of individual academics. Motivated by this fact, it is tackled afresh the important problem of posterior simulation for Bayesian inference. Some important attributes of our method are (1) obviating model-specific posterior simulation algorithms, which impose substantial time costs on researchers; (2) posterior simulation that is one to two orders of magnitude faster than any existing serial algorithm; (3) systematically coping with irregular posterior distributions, including those that are multimodal; (4) very accurate approximations of posterior moments, and accurate approximations of their accuracy; (5) as a by-product, accurate approximations of marginal likelihood. These features are illustrated using several specific models and data sets.

CS03 Room B34 LARGE SWINGS IN MACROECONOMIC TIME SERIES Chair: Evi Pappa

C040: **Prior selection for vector autoregressions**
Presenter: **Domenico Giannone**, Free University of Brussels, Belgium
Co-authors: Michele Lenza, Giorgio Primicer

Vector autoregressions (VARs) are flexible time series models that can capture complex dynamic interrelationships among macroeconomic variables. Their generality, however, comes at the cost of being very densely parameterized. As a result, the estimation of VARs tends to deliver good in-sample fit, but unstable inference and inaccurate out-of-sample forecasts, particularly when the model includes many variables. A potential solution to this problem is combining the richly parameterized unrestricted model with parsimonious priors, which help controlling estimation uncertainty. Unfortunately, however, the issue of how to optimally set the weight of the prior relative to the likelihood information is largely unexplored. A simple and theoretically founded methodology for prior selection in Bayesian VARs (BVARs) is proposed. Our recommendation is to select priors using the marginal data density (i.e. the likelihood function integrated over the model parameters), which only depends on the hyperparameters that characterize the relative weight of the prior model and the information in the data. It is shown that the out-of-sample forecasting accuracy of our model not only is superior to that of VARs with flat priors, but is also comparable to that of factor models.

C062: **Do sticky information models match survey data of inflation expectations?**
Presenter: **Matthias Paustian**, Bank of England, United Kingdom
Co-authors: Alina Barnett

We test whether sticky information models match UK data of inflation expectations. The first approch is via a pervasive sticky information DSGE model estimated on UK macro data and on the mean of the survey of inflation expectations. The second approach is via reduced form Bayesian vector autregression. In each case, we first estimate the model to match the means of the survey data and the macro data. In a second step we then look that the model implied measures of cross sectional dispersion of inflation expectations and ask whether it matches the dispersion seen in the data on surveys. We conclude that while the models estimate substantial stickines in information updating, they do not replicate the salient pattern of dispersion that we see in the data.

C074: **Riots, battles and cycles**
Presenter: **Stephane Auray**, CREST-ENSAI and ULCO, France
Co-authors: Aurelien Eyquem, Frederic Jouneau-Sion

A theoretical framework is proposed to investigate the impact of military conflicts on business cycles, as well as defense policies through enrolment mechanisms. The framework is a variation of a real business cycle model that admits explicit solutions. We extend the initial model to account for specific (potentially large) shocks that destroy the stock of capital. We consider two types of dynamics on the depreciation rate of capital: short-term shocks, that may be interpreted as riots and captured by a moving average specification, and mid-term shocks, that may be interpreted as wars and captured by a Markov switching process. Using a panel data with 14 countries from 1900 onwards we show that our model is able to reproduce the large variations of consumption to product ratios observed during major crises. Moreover the different regimes of the Markov process are shown to be consistent with observed episodes of conflicts. Our model reproduces usual business cycle facts. We also characterize the macroeconomic dynamics after shocks on the depreciation rate of capital.

C024: **For an olive wreath: Olympic Games and anticipation effects in macroeconomics**
Presenter: **Evi Pappa**, EUI, UAB and CEPR, Italy
Co-authors: Markus Bruckner

The effects that hosting and bidding for the Olympic Games have on macroeconomic outcomes in a panel of 184 countries spanning the period 1950-2006 are examnined. Actual hosting of the Games generates positive investment, consumption, and output responses before, during, and after hosting. We detect anticipation effects: (i) bidding for the Olympic Games generates positive investment, consumption, and output responses at the time of the bidding; (ii) bidding for the Games has a transitory level effect. We confirm the presence of legacy effects: hosting the Games has a permanent level effect.

CS14 Room B33 EMPIRICAL MODELLING OF FINANCIAL FRAGILITY Chair: Andrea Cipolini

C031: **Quantiles of the realized stock-bond correlation**
Presenter: **Nektarios Aslanidis**, Universitat Rovira i Virgili and CREIP , Spain
Co-authors: Charlotte Christiansen

The realized stock-bond correlation based upon high frequency returns is scrutinized. Quantile regressions are used to pin down the systematic variation of the extreme tails over their economic determinants. The correlation dependence behaves differently when the correlation is large negative and large positive. The important explanatory variables at the extreme high quantile are the lagged correlation itself, the short rate, and the yield spread. At the extreme low quantile the VIX volatility index is also significant. The empirical findings are only partially robust to using less precise measures of the stock-bond correlation.

C129: **Wavelet analysis of financial contagion**
Presenter: **Iolanda lo Cascio**, ESADE Business School, Spain

The aim is to estimate a factor model fitted to financial returns to disentagle the role played by common shock and idiosincratic shocks in shaping the comovement between asset returns during periods of calm and financial turbulence. For this purpose, we use wavelet analysis and, in particular, the Maximum Overlapping Discrete Wavelet Transform, to decompose the covariance matrix of the asset returns on a scale by scale basis, where each scale is associated to a given frequency range. This decomposition will give enough moment conditions to identify the role played by common and idiosincratic shocks. A Montecarlo simulation experiment shows that our testing methodology has good size and power properties to test for the null of no contagion (that is, of absence of an increasing role of idiosincratic shocks during turmoil). Finally, using Full Information Maximum Likelihood, we fit our model to test first for the presence of contagion whithin the East Asian region stock msarkets during the 1997-1998 period of financial turbulence, and, then, whether there is contagion from an index of financial distress in the US to East Asia during the recent sub-prime crisis.

C128: **Variance swaps, corridor variance swaps and the variance risk premium: evidence from the Italian market**
Presenter: **Silvia Muzzioli**, University of Modena and Reggio Emilia, Italy

The aim is to investigate the information content of option-based variance measures (model free implied variance and corridor implied variance), in the Italian index options market, in order to shed light on the different parts of the risk neutral distribution of the stock price and explore the nature of the variance risk premium. The variance risk premium is computed as the difference between the ex-post realised variance and the variance swap rate, proxied by model free implied variance. Up and down corridor variance measures and corridor measures with symmetric and asymmetric cuts of the risk neutral distribution are exploited in order to break down the variance risk premium into different parts. The analysis is pursued by using intra-daily synchronous prices between the options and the underlying asset.

C396: **Wavelet analysis of asset price misalignments**
Presenter: **Andrea Cipollini**, University of Modena and Reggio Emilia, Italy
Co-authors: Iolanda lo Cascio

Asset price misalignments are analyzed through wavelet decomposition. The analysis, carried within the time-frequency domain, allows us to detect how far, in a given time period, financial time series, such as house or stock prices, are from their fundamental value. The latter is associated with the low frequency component of a given time series. Moreover, using wavelet analysis, we explore whether monetary policy can contribute to asset price misalignments.

CS45 Room G16 UNIVARIATE AND MULTIVARIATE VOLATILITY MODELS Chair: Christos Savva

C052: **A multiple threshold conditional correlation GARCH model**
Presenter: **Oscar Martinez**, Universitat Rovira i Virgili, Spain
Co-authors: Nektarios Aslanidis

A multiple threshold conditional correlation (MTCC) GARCH process is proposed to analyze the behavior of a portfolio of 22 assets of international equities, government bonds and currencies. The model is easy to estimate even when the number of regimes is very large. We discuss representation and estimation based on grid search maximum likelihood techniques. The model can be seen as an intermediate case between the constant conditional correlation (CCC) process of Bollerslev and the dynamic conditional correlation (DCC) model of Engle. As threshold variables we use the principal components extracted from the data and interpret these as common factors that describe the state of the macroeconomy. Our results show that three common factors and therefore eight regimes are needed to model the whole correlation matrix of our portfolio. Interestingly, often these factors summarize common information in the three financials market separately (e.g., first factor on stocks). We also perform a portfolio evaluation exercise to compare the model with the DCC process.

C093: **Dynamic GARCH models**
Presenter: **Nicos Pavlidis**, Lancaster University, United Kingdom
Co-authors: Efthymios Pavlidis

A technique for monitoring the performance of GARCH models in the presence of structural change of unknown form is presented. The new methodology makes use of the Kalman filter and bootstrapping to dynamically update GARCH estimates as new information arrives. Monte Carlo simulations and an application to high frequency stock market data illustrate the forecasting performance of the dynamic GARCH model with respect to standard models of conditional heteroskedasticity.

C102: **Co-movements between carbon, energy and financial markets: A multivariate GARCH approach**
Presenter: **Nicolas Koch**, University of Hamburg, Germany

Liberalization of energy sector and the ensuing creation of new markets for carbon emissions have increasingly triggered a need for understanding the volatility and correlation structure between carbon, energy and financial markets. The existence of structural changes in correlation patterns

among these markets and links of the changes to market conditions is documented. A smooth transition conditional correlation (STCC) GARCH model is applied to a set of EU emission allowance futures, energy futures and financial asset returns to investigate the conditional correlation dynamics during the period 2005-2011. The model allows correlation to vary smoothly between extreme states and the dynamics are driven by a transition variable. We use, on the one hand, time as transition variable which allows the correlation level to change over time and the VSTOXX index, on the other hand, to account for uncertainty and volatility in markets. The results point to a considerable increase of correlations between carbon and energy markets in recent years. We further observe higher and more variable correlations between carbon futures and financial asset returns during the current financial crisis, predicted by higher expected stock market volatility. The effects of news on the covariances are examined through news impact surfaces.

C414: **Skewness and the relationship between risk and return**
Presenter: **Christos Savva**, Cyprus University of Technology, Cyprus
Co-authors: Panayiotis Theodossiou

We provide further evidence of the effects of the conditional skewness on the relationship between the excess returns and conditional volatility with our model being able to endogenously measure the impact of the skewness. More specifically we propose a model that utilizes the non-centered skewed generalized t (SGT) distribution with heteroskedastic error term. SGT distribution provides a great flexibility in modelling the dataset and an excellent fit to the empirical distribution of the data.

CS46 Room B35 FINANCIAL MARKET AND THE MACROECONOMY Chair: Willi Semmler

C384: **On the forward-looking content of equity and bond markets for aggregate investments: a wavelet analysis**
Presenter: **Marco Gallegati**, Universita Politecnica delle Marche, Italy
Co-authors: James Ramsey

The purpose is to investigate the role of equity and credit markets for aggregate investment by exploring the link among Tobin's Q, corporate bond spreads and investment for the US. Specifically, we study the different forward-looking information content of the equity and credit markets for aggregate investments by examining their relationship on a scale-by-scale basis. Our results show that both equity and credit markets display a useful role for aggregate investment and that their forward-looking information content is frequency dependent. Indeed, from time scale regression analysis we find that Tobin's Q has larger explanatory power for aggregate investment than credit spreads over longer time horizons, while the opposite is true for the short- to medium-term time horizons. In addition, when we consider the instability of the q-relationship over time we find that corporate bond spreads lose their significance for aggregate investments by the early 1990s.

C813: **Employment projections with a matching-model Phillips curve**
Presenter: **Ekkehard Ernst**, International Labour Organization, Switzerland

A generalised search model with matching on financial, labour and product markets is analysed, taking the impact of market frictions on the pass-through of shocks onto employment dynamics into account. The model includes endogenous demand for money that arises from matching frictions on the financial market. Moreover, price posting on retail markets allows us to establish a forward-looking Search Phillips Curve. Labour markets are characterized by employment protection and wage negotiations. Unemployed workers benefit from a tax-financed insurance system. The calibrated version of the model is then used to derive medium-term projections for employment developments. In particular, the capacity of the model to help understand the role of market frictions and the financial accelerator in shaping employment recoveries is assessed against a traditional employment-elasticities approach.

C544: **Information in the yield curve: A macro-finance approach**
Presenter: **Hans Dewachter**, NBB, National Bank of Belgium , Belgium
Co-authors: Leonardo Iania, Marco Lyrio

An affine term structure model that incorporates macroeconomic and financial factors is used to study the term premium in the U.S. bond market. The results corroborate the known rejection of the expectation hypothesis and indicate that one factor, closely related to the Cochrane and Piazzesi (2005) factor (the CP factor), is responsible for most of the variation in bond premia. Furthermore, the model-implied bond premia are able to explain around 32% and 40% of the variability of one- and two-year excess returns and their out-of-sample performance is comparable to the one obtained with the CP factor. The model is also used to decompose yield spreads into an expectations and a term premium component in order to forecast GDP growth and inflation. Although this decomposition does not seem important to forecast GDP growth, it is crucial to forecast inflation for most forecasting horizons. Also, the inclusion of control variables such as the short-term interest rate and lagged variables does not drive out the predictive power of the yield spread decomposition.

C541: **Estimating a banking - macro model for the EU using a multi-regime VAR**
Presenter: **Willi Semmler**, New School, United States of America
Co-authors: Stefan Mittnik

A Banking-Macro Model is introduces and the linkages are estimated through a Multi-Regime VAR (MRVAR). We first introduce a dynamic model which is akin to the Bunnermeier and Sannikov model. The banking sector borrows and issues liabilities, accumulates assets, pays bonuses, and its equity can undergo large fluctuations. It faces collateral constrained borrowing which affects borrowing cost on the credit market. The instability of the banking sector may arise not only from loan losses but also from adverse movements of asset prices and its deteriorating effects on the banks' balance sheets. We solve different variants of the model numerically. We then estimate the model by using an MRVAR for an EU data set on real and financial variables. In contrast to previous studies on the financial-real interaction, for example of the financial accelerator type, our model shows local instability and globally multiple regimes. Whereas the common model leads, in terms of econometrics, to a one-regime VAR, we demonstrate the usefulness of an MRVAR. We estimate our model for the EU with am MRVAR using a constructed financial stress index and industrial production for the EU. We 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 and the size of the shocks. Small financial stress shocks may not matter, but large shocks are likely to have magnifying effects.

CS53 Room B36 BAYESIAN METHODS IN ECONOMETRIC AND FINANCIAL APPLICATIONS Chair: Ioannis Vrontos

C606: **A quantile regression approach to out-of sample equity premium prediction in the presence of model uncertainty**
Presenter: **Loukia Meligkotsidou**, University of Athens, Greece
Co-authors: Ekaterini Panopoulou, Ioannis Vrontos, Spyridon Vrontos

The objective is to apply the quantile regression approach to equity premium prediction. Quantile regression models can be used to predict a set of conditional quantiles of returns, offering a flexible way of understanding how the effects of various predictors change across the distribution of conditional returns. Furthermore, we account for model uncertainty, regarding the set of predictors that are more useful at predicting the

equity premium. To this end we employ the Bayesian model comparison approach within the quantile regression setting, in order to produce combined forecasts of the equity premium based on model averaging. For reasons of comparison, we apply the standard linear regression models that are currently used for equity premium prediction, either including a single predictor or all the available ones, as well as existing methods of combining individual forecasts from single predictor linear models. We compare the different modeling approaches, i.e. standard linear and quantile regression, and various different methods of producing combined forecasts, including model averaging, in terms of their predictive ability. All different forecasts are evaluated against the benchmark constant equity premium using various evaluation tests.

C625: **Performance evaluation of pension funds: The impact of non-normality and time-varying volatility**
Presenter: **Spyridon Vrontos**, University of Piraeus, Greece
Co-authors: John Vrontos, Loukia Meligkotsidou

The idea of modeling the conditional distribution of pension fund returns using a fat tailed density and a time-varying conditional variance is introduced. This approach takes into account the stylized facts of pension fund return series, that is heteroscedasticity and deviations from normality. We evaluate pension fund performance using multifactor asset pricing models, with the relevant risk factors being identified through standard model selection techniques. We explore potential impacts of our approach by analyzing individual pension funds and show that it can be economically important.

C644: **Multivariate regressions: An alternative modelling approach**
Presenter: **Dimitrios Giannikis**, Athens University of Economics and Business , Greece
Co-authors: Loukia Meligkotsidou, Ioannis Vrontos

A new class of multivariate predictive regression models for the analysis and modelling of financial time series is developed. This class of models allows different dependent variables to be affected by different risk factors/predictors, i.e. provides a way of understanding how the relationship between dependent variables and risk factors changes across different return series. In addition, we allow for time varying variances/covariances by modelling the errors using a multivariate GARCH model. We consider the problem of estimation and inference for this class of multivariate predictive regression models. We propose a full Bayesian approach to inference for automatic determination of the (different) risk factor/predictive variables that explain financial returns. A Markov chain Monte Carlo stochastic search algorithm has been designed which provides posterior model probabilities and takes into account model uncertainty by using Bayesian Model Averaging and produces robust inferences for the quantities of interest. This algorithm identifies valuable predictors which can be used to predict future returns. The proposed multivariate predictive model is particularly useful for the analysis of hedge fund strategies where the return series is explained by different risk factors, due to the different strategies advocated by fund managers.

C908: **Bayesian analysis of autoregressive models with multiple structural breaks**
Presenter: **Ioannis Vrontos**, Athens University of Economics and Business, Greece
Co-authors: Loukia Meligkotsidou, Elias Tzavalis

We suggest a Bayesian approach for inferring stationary autoregressive models allowing for possible structural changes (known as breaks) in both the mean and the error variance of economic series occurring at unknown times. Efficient Bayesian inference for the unknown number and positions of the structural breaks is performed by using filtering recursions similar to those of the forward-backward algorithm. A Bayesian approach to unit root testing is also proposed, based on the comparison of stationary autoregressive models with multiple breaks to their counterpart unit root models. Simulation experiments are conducted with the aim to assess the performance of the suggested inferential procedure, as well as to investigate if the Bayesian model comparison approach can distinguish unit root models from stationary autoregressive models with multiple structural breaks in the parameters. The proposed method is applied to key economic series with the aim to investigate whether they are subject to shifts in the mean and/or the error variance. The latter has recently received an economic policy interest as improved monetary policies have also as a target to reduce the volatility of economic series.

CS61 Room B18 LONG MEMORY TIME SERIES MODELS Chair: Anne Philippe

C472: **Detection of non constant long-memory parameter**
Presenter: **Frederic Lavancier**, University of Nantes, France
Co-authors: Remigijus Leipus, Anne Philippe, Donatas Surgailis

Detection of non-constant long memory parameter in time series is considered. The null hypothesis includes stationary or non-stationary time series with constant long memory parameter, in particular $I(d)$ series, $d > -.5$. The alternative corresponds to a change in the long memory parameter and gathers in particular an abrupt or gradual change from $I(d_1)$ to $I(d_2)$, $-.5 < d_1 < d_2$. Various test statistics are considered. They are all based on the ratio of forward and backward empirical variances of the partial sums. The consistency of the tests is proved under a very general setting. Moreover, the behavior of the test statistics is studied for some models with changing memory parameter. A simulation study shows that our testing procedures have good finite sample properties and turn out to be more powerful than the KPSS-based tests considered in previous works.

C561: **Seasonal modeling by SARFIMA and near unit root models**
Presenter: **Liudas Giraitis**, Queen Mary, University of London, United Kingdom
Co-authors: Karim Abadir, Walter Distaso

The objective of the paper is to introduce a new approach for a seasonal modelling that allows for long memory and does not suffer from drawbacks of GARMA model. The suggested SARFIMA model has tractable analytical expressions for the spectral density and autocovariance function and enables practical estimation. Near unit root alternatives are also discussed.

C085: **Semiparametric estimation of the volatility in long memory in stochastic volatility models**
Presenter: **Josu Arteche**, University of the Basque Country, Spain

Long Memory in Stochastic Volatility (LMSV) models are flexible tools for the modelling of a persistent dynamic volatility, which is typical of financial time series. Empirical researchers however have usually preferred more rigid ARCH based models because, among other reasons, the conditional variances are exact functions of past observations, which makes it very easy to obtain a series of estimated volatilities. In SV models the conditional variance is a random variable driven by innovations different from those driving the levels, and its estimation is not a simple task, especially if the volatility component shows long memory is what makes more traditional techniques, such as the Kalman filter, difficult to use. A feasible semiparametric Wiener-Kolmogorov based signal extraction strategy in the frequency domain is here proposed and compared with the time domain version of Harvey (1998). For that an estimator of the spectral density function is required. Two versions are considered and their consistency under stationary and non-stationary (pseudo spectral density in this case) but mean reverting long memory is proved. A fully nonparametric volatility extraction strategy is also introduced based on Singular Spectrum Analysis. Finally, an application to the daily Dow Jones Industrial index is included.

C682: **Linear processes with space varying memory**
Presenter: **Alfredas Rackauskas**, Vilnius University, Lithuania

Let H be a separable Hilbert space with the norm $||x||, x \in H$. We consider a linear process (X_k) defined by $X_k = \sum_{j=0}^{\infty} u_j e_{k-j}, \quad k \geq 1$, where $u_0 = I$ is the identity map, $(u_j, j \geq 1)$ is a sequence of linear bounded operators on H such that $\sum_{j=1}^{\infty} ||u_j||^2 < \infty$ and (e_j) are iid random elements in H with mean zero and finite second moment $E||e_1||^2 < \infty$. We shall discuss memory phenomenon of (X_k) via limit behavior of partial sums $\sum_{k=1}^{[nt]} X_k, t \in [0,1]$. In particular, we show that in the case where $u_j = j^{-D}$ with $(1/2)I < D < I$, an operator fractional Brownian motion with values in H defined via operator-valued Hurst exponent appears in a limit of partial sums.

CS64 Room B20 EFFICIENT MCMC ALGORITHMS FOR BAYESIAN FINANCIAL ECONOMETRIC MODELS Chair: Roberto Casarin

C500: **A Bayesian estimator of the multivariate covariance of noisy and asynchronous returns**
Presenter: **Stefano Peluso**, Swiss Finance Institute. Universita della Svizzera Italiana, Switzerland
Co-authors: Fulvio Corsi, Antonietta Mira

A multivariate positive semi-definite estimator of the covariance matrix of noisy and asynchronously observed asset returns is proposed. The dynamic linear model is the setting we use to deal with presence of noise in the data, and we treat the asynchronous time series as synchronous time series with missing observations. Our approach is Bayesian and we implement an augmented Gibbs sampler algorithm to sample the covariance matrix, the observational error variance matrix, the latent process and the missing observations of the noisy process. Our MCMC estimator is positive semi-definite by construction, and a simulation study compares it with other existing methodologies. Then we study the convergence properties and the robustness of our algorithm.

C624: **Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models**
Presenter: **Gregor Kastner**, Vienna University of Economics and Business, Austria, Austria
Co-authors: Sylvia Fruehwirth-Schnatter

Recent findings show that Bayesian inference for stochastic volatility (SV) models using MCMC methods highly depends on actual parameter values in terms of sampling efficiency. While draws from the posterior utilizing the standard centered parameterization break down when the volatility of volatility parameter in the latent state equation is small, non-centered versions of the model show deficiencies for highly persistent latent variable series. The novel approach of ancillarity-sufficiency interweaving has recently been shown to aid in overcoming these issues for a broad class of multilevel models. We demonstrate how such an interweaving strategy can be applied to stochastic volatility models in order to greatly improve sampling efficiency for all parameters and throughout the entire parameter range. Moreover, this method of "combining best of different worlds" allows inference for parameter constellations that have previously been unfeasible to estimate without the need to select a particular parameterization beforehand.

C903: ***SMC*^{2}: A sequential Monte Carlo algorithm with particle Markov chain Monte Carlo updates**
Presenter: **Nicolas Chopin**, CREST (ENSAE), France
Co-authors: Pierre Jacob, Omiros Papaspiliopoulos

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

Sunday 18.12.2011 11:05 - 12:45 Parallel Session I – ERCIM

ESI03 Room Woburn ADVANCES IN MULTIVARIATE ANALYSIS Chair: Ana Colubi

E202: **Multivariate linear L1 regression**
Presenter: **Hannu Oja**, University of Tampere, Finland

The univariate concepts of sign and rank are based on the ordering of the data. Unfortunately, in the multivariate case there are no natural orderings of the data points. An approach utilizing L1 criterion functions is therefore often used to extend the concepts of sign and rank to the multivariate case. The multivariate spatial sign and rank methods that are based on the use of Euclidean distance are discussed. We discuss the test and estimates based on spatial signs and ranks in the one sample location, several samples location and, finally, multivariate linear regression cases. Transformation-retransformation technique is used to obtain affine invariant tests and equivariant estimates. Limiting distribution of the tests and estimates are given and their asymptotical efficiencies are discussed and compared. The theory is illustrated via examples.

E688: **A new multiclass support vector machine based on L_p norms**
Presenter: **Patrick Groenen**, Erasmus University, Netherlands
Co-authors: Gertjan van den Burg

For the binary classification problem, support vector machines (SVMs) have become increasingly popular over the last ten years. Their out-of-sample prediction performance often compares favorably to alternative binary classification methods such as logistic regression and linear discriminant analysis. One of the reasons for this is that the SVM is robust against outliers, allows nonlinear prediction, and avoids overfitting by regularizing the loss function using a quadratic penalty. However, in the case that more than two classes need to be predicted often a series of binary SVMs are performed (one-versus-all or between all pairs of classes, one-versus-one). A disadvantage of such methods is that they are suboptimal. We propose a new multiclass SVM loss function based on assigning a direction for each class in a $K-1$ dimensional space, taking a linear combination of the observations in this $K-1$ dimensional space, and taking SVM hinge errors for each direction. Our contribution lies in how these errors are added: we propose to use the r-th power of an L_p norm of the hinge errors. This general loss function has some existing multiclass SVM loss functions as special cases. In addition, we propose a majorization algorithm for its minimization, thereby guaranteeing monotonic convergence.

E834: **Multivariate extremes: a directional approach**
Presenter: **Juan Romo**, Universidad Carlos III de Madrid, Spain
Co-authors: Henry Laniado, Rosa Lillo

A new approach to analyze multivariate extremes is proposed. It provides a multivariate order based on a concept that we will call “extremality". Given a unit vector, the extremality allows us to measure the “farness" of a point in $\mathbb{R}^n$ with respect to a data cloud or to a distribution in the vector direction. We establish the most relevant properties of this extremality measure and we give the theoretical basis for its nonparametric estimation. We include two applications in finance: a multivariate Value at Risk (VaR) with level sets built through extremality and a portfolio selection strategy based on the order induced by extremality.

ES01 Room Senate ROBUST ANALYSIS OF COMPLEX DATA SETS Chair: Stefan Van Aelst

E430: **Robust regression and model selection: sparse least trimmed squares**
Presenter: **Andreas Alfons**, K.U.Leuven, Belgium
Co-authors: Christophe Croux, Sarah Gelper

Since the availability of data sets with a large number of variables is increasing, sparse model estimation is a topic of high importance in modern data analysis. In addition, outliers are a common data problem in applied statistics. Therefore robust regression and sparse model estimation is addressed. Robust sparse regression allows for better prediction performance through variance reduction, while at the same time improving interpretability of the resulting models due to the smaller number of explanatory variables. Furthermore, computational problems in the case of high-dimensional data are avoided. By adding an L1 penalty on the coefficient estimates to the well-known least trimmed squares (LTS) estimator, a robust and sparse estimator is introduced. For the computation of this sparse LTS estimator, a C-step algorithm is presented. Last but not least, the performance of sparse LTS regression is assessed by means of a simulation study.

E449: **Robust analysis of periodicities in light curves using M-regression**
Presenter: **Anita M. Thieler**, TU Dortmund University, Germany
Co-authors: Roland Fried

In astroparticle physics one is often confronted with light curves, which are noisy time series of irregularly spaced observations with additional information about the observation accuracy for each time point. To search for periodicity in the signal underlying this time series, one can use the Deeming Periodogram, which is a variant of the Fourier Periodogram for irregularly sampled data. However, this method cannot cope with additional periodic structure in the sampling like periodic gaps. Methods to overcome this frequent problem mostly base on weighted least squares estimation. We compare such classical methods to robust analogues based on M-regression by simulation, considering the effect of dependent so-called red noise, which is often assumed for light curves. We propose a way to determine significant periods, taking into account that estimation at different periods cannot be considered as independent for irregularly spaced time series.

E451: **On asymptotic properties of weighted M-estimators**
Presenter: **Mohammed El Asri**, University of Avignon, France

The class of M-estimators (e.g. the spatial median) shows good properties in location problems. Some authors studied the asymptotic normality and some criteria of robustness, such as the breakdown point and the influence function. Here, our aim is to study a variant class of M-estimators, namely the class of weighted M-estimators. The weighted approach is particularly suitable for the analysis of data points having different levels of importance. We will present some theoretical properties of central limit type for these weighted M-estimators and discuss from simulations and examples the criteria to select such estimators based on their robustness and efficiency.

E532: **The robustness of the hyperbolic efficiency estimator**
Presenter: **Christopher Bruffaerts**, Universite libre de Bruxelles, Belgium
Co-authors: Bram De Rock, Catherine Dehon

The robustness properties of a specific type of orientation are examined in the context of efficiency measurement using partial frontiers. This so called unconditional hyperbolic α-quantile estimator of efficiency can be seen as an extension of the input/output methodology of partial frontiers. The influence function of this fully non-parametric and unconditional estimator is derived for a complete multivariate setup (multiple inputs and

outputs). Like for the input and output quantile estimators, the hyperbolic α-quantile estimator is *B*-robust. The asymptotic variance of this estimator is recovered from the influence function. Some examples are given to assess the relevance of this type of estimator and to show the differences with the input and output *alpha*-quantile estimators of efficiency from both a robustness and a statistical efficiency point of view. Finally, this methodology will be used to examine the efficiency of US-universities.

ES04 Room Bloomsbury TIME SERIES MODELING AND COMPUTATION I Chair: Andres M. Alonso

E023: **Forecasting with Markov-switching dynamic factor models**
Presenter: **Pilar Poncela**, Universidad Autonoma de Madrid, Spain
Co-authors: Maximo Camacho, Gabriel Perez-Quiros

The Markov-switching dynamic factor model for change point detection in multiple time series analysis is used. It is pursued to date the time of going into/getting out of a recession using macroeconomic time series. Within this model, the specificities of the day to day monitoring of economic developments such as ragged edges and mixed frequencies are accounted for. The consequences of increasing the number of series in the model are also analyzed. All these issues have been contemplated in the linear framework, but they have not been considered yet in the Markov switching context. It is also shown that extracting a linear univariate common factor and applying afterwards a Markov switching mechanism leads to wrong inference about the probability of being in a certain state. The dynamic common factor Markov switching model will be used to compute inferences of the percentage chance that a certain economy will face a recession in the short term.

E534: **Dynamic conditional score models**
Presenter: **Alessandra Luati**, University of Bologna, Italy
Co-authors: Andrew Harvey

We are concerned with modeling a time varying signal embedded in non-Gaussian noise. We consider a class of observation driven models where, conditional on past information, the observations are generated by a Student-t distribution. The dynamics of the time varying location (the signal) are governed by the score of the conditional distribution and the time varying signal is updated by a suitably defined filter with ARMA-type structure. This filter may be interpreted as an approximation to a computer intensive solution for the corresponding unobserved component models. We derive the moments and the dynamic properties of the observations and the asymptotic distribution of the maximum likelihood estimator.

E349: **Time series models for compass wind directions**
Presenter: **Michael Wiper**, Universidad Carlos III de Madrid, Spain

Models for time series of discrete, circular variables such as compass directions of winds are developed. In particular, we consider two different approaches. Firstly, we propose a hidden Markov model based on the use of the wrapped Poisson distribution and secondly, we consider a circularized INAR model. Variants of both models to allow for seasonality, etc. are also introduced. Inference for our models is carried out using Bayesian techniques and our results are illustrated with an application to a real-time series of daily wind directions at a weather site in northern Spain. Our models are also compared with some simpler alternatives.

E492: **Forecasting electricity prices and their volatilities using unobserved components**
Presenter: **Carolina Garcia-Martos**, Universidad Politecnica de Madrid, Spain
Co-authors: Julio Rodriguez, Maria Jesus Sanchez

The liberalization of electricity markets more than ten years ago in the vast majority of developed countries has introduced the need of modelling and forecasting electricity prices and volatilities, both in the short and long term. Thus, there is a need of providing methodology that is able to deal with the most important features of electricity price series, which are well known for presenting not only structure in conditional mean but also time-varying conditional variances. A new model, which allows us to extract conditionally heteroskedastic common factors from the vector of electricity prices is proposed. These common factors are jointly estimated as well as their relationship with the original vector of series, and the dynamics affecting both their conditional mean and variance. The estimation of the model is carried out under the state-space formulation. The new model proposed is applied to extract seasonal common dynamic factors as well as common volatility factors for electricity prices and the estimation results are used to forecast electricity prices and their volatilities in the Spanish zone of the Iberian Market. Several simplified/alternative models are also considered as benchmarks to illustrate that the proposed approach is superior to all of them in terms of explanatory and predictive power.

ES23 Room Gordon ADVANCES IN DISTANCE-BASED METHODS AND APPLICATIONS Chair: Aurea Grane

E395: **Influential subsets in distance-based (DB) prediction from genomic data**
Presenter: **Josep Fortiana**, Universitat de Barcelona, Spain
Co-authors: Anna Esteve, Eva Boj

In the clinical follow-up of HIV-infected patients one of the most challenging problems is interpreting patterns of resistance to High Activity Anti-Retroviral Therapy (HAART). This is due both to the high dimensionality of genomic data and to the huge number of available drug combinations. From a statistical perspective, this clinical issue translates into the problem of predicting drug susceptibility from a large number of genotypic predictors. For such data, typically with thousands of predictors, the concept of "influence of a single predictor" should be subsumed in the more general "influential predictors subset". Our procedure for detecting such subsets results from a clustering of explanatory variables which can be inserted as a natural stage within the DB prediction framework. To this effect, we endow the set of predictors with a joint metric, amalgamating informations from (1) mutual dependence, (2) influence on the response and (3) nearness within the genomic sequence. We illustrate the procedure on a collection of datasets, publicly available from the Stanford Data Repository, that has been widely used in papers studying HIV-1 resistance prediction from genomic data.

E402: **Applications of distance-based (DB) methods in actuarial science using R**
Presenter: **Eva Boj**, Universitat de Barcelona, Spain
Co-authors: Teresa Costa, Josep Fortiana

DB methods are applied to the problems of pricing and reserving in non-life insurance. These two problems exemplify the suitability of DB methods in actuarial practice. Using data from real portfolios we assess their advantages and drawbacks as compared with more conventional methods in such respects as goodness of fit and prediction, interpretability of results and computational cost. We use the dbstats R package, which has been developed by members of our Barcelona DB prediction project team and is available from the CRAN repository.

E508: **New trends in error correcting output codes**
Presenter: **Oriol Pujol**, University of Barcelona, Spain

Error correcting output codes (ECOC) is a distance based meta-learning framework that allows the extension of any binary classification system to the multi-class case. This framework is based on a coding step, where a set of binary problems are learnt by a set of base binary classifiers

and coded in a matrix, and a decoding step, where new samples are classified according to a distance-based decoding process. This framework has proven to be extremely effective when the base classifiers are discriminant boundary-based algorithms, i.e. adaboost or svm. Although ECOC have been traditionally just used for extending classifiers to the multi-class case, new challenges in machine learning require the adjustment of this technique to these new problems. Embedded computing with low resources, large scale learning with thousands of classes, online learning and knowledge discovery are some of the scenarios where ECOC can be successfully adapted. How such adaptation to these new challenges is done is the base of this contribution. Two main tools are considered, namely sub-linear and problem-based coding strategies. Additionally, new scenarios for ECOC are overviewed.

E531: **Profile identification via weighted related metric scaling: an application to Spanish dependent children**
Presenter: **Aurea Grane**, Universidad Carlos III de Madrid, Spain
Co-authors: Pablo Alonso, Irene Albarran

Disability and dependence affect health status and quality of life, therefore they are significant public health issues. The main purpose is to establish the existing relationship among different variables (continuous, categorical and binary) referred to children between 3 and 6 years old and their functional dependence in basic activities of daily living. We combine different types of information via weighted related metric scaling to obtain homogeneous profiles for dependent Spanish children. The redundant information between groups of variables is modelled with an interaction parameter that can be optimized according to several criteria. The goal is to obtain maximum explained variability in Euclidean configurations. Data comes from the Survey about Disabilities, Personal Autonomy and Dependence Situations -SDPS 2008- (Spanish National Institute of Statistics, 2008).

ES24 Room Torrington MIXTURE MODELS: APPLICATIONS AND EXTENSIONS Chair: John Hinde

E882: **Model-based clustering and classification of data with mixed type**
Presenter: **Ryan Browne**, University of Guelph, Canada
Co-authors: Paul McNicholas

We propose a mixture of latent variables model for the model-based clustering and classification of data comprising variables of mixed type. This approach is a generalization of latent variable analysis with model fitting carried out within the expectation-maximization framework. Our approach is outlined and a simulation study conducted to illustrate the effect of sample size and noise on the standard errors and the recovery probabilities for the number of groups. Our modelling methodology is then applied to two real data sets and their clustering and classification performance is discussed. We conclude with discussion and suggestions for future work.

E856: **Finite mixture model clustering of SNP data**
Presenter: **Norma Coffey**, National University of Ireland, Galway, Ireland
Co-authors: John Hinde, Augusto Franco Garcia

Sugarcane is polypoid and it is important to develop methods that identify the many different alleles and associated genotypes. Single nucleotide polymorphisms (SNPs) can give an indication as to the number of allele haplotypes present for a gene and such information could have implications in sugarcane breeding since high yield potential may be due to the presence of and/or different number of copies of, a specific allele(s) present at a gene locus. Clustering these data provides a means of identifying different genotypes and therefore it is necessary to develop a technique that can determine the number of clusters present, determine the angles between the clusters to identify different genotypes, and provide a probabilistic clustering to identify points that have high probability of belonging to a particular cluster (have a particular genotype) and those that are regarded as an unclear genotype. We use finite mixtures of orthogonal regression lines to cluster the data, since it is not clear which variable is the response and which is the explanatory variable and we require that a regression line can be fitted to the group parallel to the y-axis. We implement this technique in R and show its usefulness in clustering these data.

E823: **Classification of variables using the Watson distribution**
Presenter: **Adelaide Figueiredo**, Faculdade de Economia da Universidade do Porto, Portugal

For a fixed group of individuals we consider a sample of normalized variables, randomly selected from a population of variables, to which we associate a mixture of Watson distributions. At a first stage we identify this mixture through the EM algorithm to obtain the parameters of the mixture and homogeneous groups of variables. Then, a test of analysis of variance is applied to investigate if the groups are distinct. Finally, we define a classification rule to assign a new variable into one of the groups of variables. This approach aims to give a contribution for the problem of a priori selection of variables.

E801: **Mixture models for outliers**
Presenter: **John Hinde**, National University of Ireland, Galway, Ireland

Mixture models provide a highly flexible framework for making extensions to the basic linear model. The potential of mixture models for outlier detection/accommodation, to provide more general error distributions, and to allow for heterogeneous regression relationships, will be illustrated. For finite mixture models estimation can be easily implemented using the EM algorithm, however for outlier detection the procedure can be very sensitive to starting values. A related approach is to use a mixture of experts model, modelling the mixture proportions using the residuals from a standard model. This seems to perform well and the algorithm appears to be robust to the starting configuration. The methods are illustrated using the famous Brownlee stackloss data and the mixture of experts approach is shown to recover the standard outliers found in the literature, both for the raw data and after transformation of the response.

ES28 Room Bedford DESIGN AND ANALYSIS OF COMPUTER EXPERIMENTS I Chair: Sonja Kuhnt

E144: **Global sensitivity analysis of computer models by using orthogonal arrays and Fourier regressions**
Presenter: **Astrid Jourdan**, E.I.S.T.I., France

Global sensitivity analysis of computer models is usually performed by estimating the Sobol' indices with a Monte-Carlo method. When the simulations are time-consuming, a well-known method consists of replacing the computer code by a meta-model with a fast computation time. A new approach defined by two points is suggested. The first one is the use of a Fourier regression, also called complex linear model. This metamodel allows one to derive an analytical estimation of the Sobol' indices, and thereby avoids the Monte Carlo estimation. The second point concerns the design of the computer experiments. We study the interest of using an orthogonal array instead of a Latin Hypercube in the case of the estimation of the second-order sensitivity indices. One difficulty is the construction of an optimal orthogonal array for the Fourier regression. This approach is applied to analytical functions.

E222: **Dealing with asynchronicity in Kriging-based parallel global optimization**
Presenter: **David Ginsbourger**, University of Bern, Switzerland
Co-authors: Clement Chevalier, Janis Janusevskis, Rodolphe Le Riche

Kriging-based algorithms have become common place in the field of Computer Experiments. They rely on the iterative maximization of sampling criteria, like the Expected Improvement (EI). EI's explicit trade-off between promising regions and global exploration contributed to its popularity in sequential optimization. In such a framework, the next run starts after the previous one has finished. However, costly simulations are always more distributed over multiple machines. Recently, a multipoints EI criterion has been proposed for synchronous parallel optimization. Nevertheless, dealing with busy simulations remains an open problem. Extensions of the multipoints EI to the asynchronous case are proposed. We focus on a variant of EI, EEI, for the case where one or several new evaluations have to be done while the responses of some previous simulations are not known yet. Different issues regarding EEI's computation and maximization are addressed. EEI's estimation by Monte Carlo or by quadrature naturally raises the question of choosing the integration points, which may be addressed based on state-of-the-art results. More originally, we discuss the effect of EEI's estimation error on the approximate maximization of EEI. A theoretical discussion and some numerical experiments are proposed to better understand the ins and outs of this question.

E397: **Partial maximum entropy sampling criterion for computer experiments**
Presenter: **Noha Youssef**, London School of Economics, United Kingdom
Co-authors: Henry Wynn

In computer experiments, it is common to model the output as a Gaussian process (GP). The GP is defined by its covariance structure. The study aims at finding a model based optimal design for estimating the covariance parameters. The proposed method suggests approximating the process via the Karhunen-Loeve (K-L) expansion. As a result, the covariance function is also approximated using the same expansion. The covariance parameters now are embedded in the model as mean parameters. A new criterion, partial maximum entropy sampling, is used when the interest is in a subset of the model parameters whether derived from the covariance terms via the K-L method, the regression terms, or both. Spectral bounds, which are used for branch and bound algorithms for maximizing the entropy sampling are developed for the partial entropy case.

E455: **Total interaction indices for the decomposition of functions with high complexity**
Presenter: **Jana Fruth**, TU Dortmund, Germany
Co-authors: Sonja Kuhnt

The decomposition of a function by functional ANOVA provides a clear complexity reduction and enables the understanding of the function and its effects. It also provides a valuable aid, for example, for variable selection and optimization. To calculate the decomposition, terms for individual variables are computed to begin with, often by Sobol main indices, e.g. using the FAST method. Next variables with no obvious influence are removed from the decomposition. In order to consider interactions between variables, usually all second order indices are calculated and again unimportant interaction terms are removed. Due to high computational costs, calculation of higher order sensitivity indices is so far avoided. As a result decompositions are obtained which provide a good simplification, but are not satisfying for functions with higher order interactions. It is proposed a new kind of index, called total interaction index, which does not neglect the complex interaction structure of the function. We show connections with other types of indices, like total indices and closed indices and suggest two different ways of calculation including FAST and RBD-FAST. In addition, we apply the proposed index to test functions.

ES30 Room S264 BIOSTATISTICS I Chair: Gilbert Mackenzie

E315: **Advances in covariance modelling**
Presenter: **Gilbert MacKenzie**, University of Limerick, Ireland
Co-authors: Jing Xu

The history of the development of joint mean-covariance modelling over the last decade is traced. From Pourahmadi's seminal paper, recent advances in research are discussed, and the potential impact on longitudinal clinical trial design and analysis is remarked. The conventional approach to modelling longitudinal data places considerable emphasis on estimation of the mean structure and much less on the covariance structure, between repeated measurements on the same subject. Often, the covariance structure is thought to be a "nuisance parameter" or at least not to be of primary scientific interest. Consequently little effort is expended on modelling its structure. In particular, the idea that in longitudinal RCTs intervention might affect the covariance structure rather than, or as well as, the mean rarely intrudes. We shall argue that these ideas are rather passe and that from an inferential standpoint the problem is symmetrical in both parameters μ and Σ . Throughout, we will distinguish carefully between joint estimation, which is now relatively routine, and joint model selection, which is not.

E369: **Modelling covariance structure for incomplete multivariate longitudinial data**
Presenter: **Jing Xu**, Birkbeck, University of London, United Kingdom
Co-authors: Gilbert MacKenzie

Comparing with univariate case, two difficulties obviously arise from covariance modelling for incomplete multivariate longitudinal data: 1) more complicated covariance structures are caused by the possible correlations among multiple variables and repeated measures; 2) missingness may occur among multiple variables and repeated measures as well. Additionally, as in univariate case, the positive-definiteness constraint of covariance matrix is still the major obstacle for multivariate case. A data-based method to model covariance matrices is developed. It consists of block triangular factorisation, matrix logarithm and parameterisation. Using this method, the constrained and hard-to-model parameters of Σ_i are traded in for unconstrained and interpretable parameters. Within the marginal model framework, we show how to overcome the difficulty from unbalanced data (missing at random) by embedding the covariance matrix of the observed data for each subject to a larger covariance matrix of complete case and employing one of EM algorithms to find the maximum likelihood estimates of the parameters. Application to a set of bivariate visual data gives an illustration of our methodology.

E488: **Model selection in sparse high-dimensional contingency tables**
Presenter: **Susana Conde**, University College Cork, Ireland
Co-authors: Gilbert MacKenzie

The problem of searching sparse, high-dimensional contingency tables for important sets of interactions is addressed using log-linear models. Such data often arise in genetic and epidemiological applications. Typically, they lead to models in which many effects, especially those corresponding to the higher order interactions, are inestimable. In general, our goal is to select an optimum model containing important lower-order interactions, when they exist, as these are more likely to be interpretable. First we highlight the respective roles of binary and Yates' design matrix coding. Then we prove a new theorem which allows us to eliminate the majority of the inestimable effects, hierarchically. In this way, we can identify the maximal model, best-supported by the data. Finally, we check whether this model can be reduced further, by a backwards elimination algorithm, to yield a final minimal model. We also compare our stepwise algorithms with alternative methods based on regularisation, by considering a penalised likelihood approach, which incorporates: (a) the least absolute shrinkage and selection operator (LASSO) penalty and (b) the smooth LASSO

penalty. We considered a variety of scenarios including cases where $2^p \gg n$. Finally, we analyse some data, report our findings and discuss the key issues.

E443: **Optimal choice of reference subclass in categorical regression models**
Presenter: **Defen Peng**, University of Limerick, Ireland
Co-authors: Gilbert MacKenzie

Issues surrounding the choice of reference subclasses for categorical variables in parametric regression models are addressed. We focus on: (a) techniques which maximize the precision of the resulting estimators, (b) discrepancy from the ideal allocation and (c) multi-collinearity. First, we derive the optimal design allocation and provide a statistic, based on generalized variance and its distribution for measuring the discrepancy between the optimal allocation and the observed allocations occurring in observational studies. Next we focus on reference category subclass choice, obtaining general expressions for the variance-covariance matrix of the estimators in linear and other generalized linear models for a single categorical variable. Then we extend the investigation to multiple categorical variables by means of simulation studies, wherein we demonstrate the use of the techniques. Later we address the issue of multi-collinearity which is especially important when the regression model contains multiple categorical variables with sparse subclasses. In this case the model design matrix may become unstable or even less than full rank. To begin, we consider the case of linear model with a single categorical variable and rapidly extend our investigation into GLMs with multiple categorical variables and end by drawing some general conclusions about the various methods developed.

ES33 Room S261 ROBUST METHODS IN SMALL AREA ESTIMATION Chair: Isabel Molina

E116: **Controlling the bias of robust small area predictors**
Presenter: **David Haziza**, University of Montreal, Canada
Co-authors: Valery Dongmo Jiongo, Pierre Duchesne

The user demand for small area estimators has been growing in most countries. This led survey statisticians to develop theoretically sound and yet practical estimation procedures, providing reliable estimators for small areas. A popular estimation method is the so-called empirical best linear unbiased prediction (EBLUP). However, the EBLUP is sensitive to the presence of outliers. That is, including or excluding outlying units from its computation may have a large impact on its magnitude. In recent years, the problem of robust small area estimation has received considerable interest. We introduce two new robust small area estimators that are robust in the presence of outliers. Results of a simulation study that compares the performance of several robust estimators in terms of relative bias and relative efficiency will be presented.

E425: **On computing and tuning some simple and robust unit-level SAE estimators**
Presenter: **Tobias Schoch**, University of Applied Sciences Northwestern Switzerland, Switzerland
Co-authors: Beat Hulliger

The need for robust estimates of area-level characteristics is of increasing importance for production in Official Statistics (OS), since of EBLUP methods are vulnerable to outliers. Statisticians in OS often fit mixed linear models (MLM) and use EBLUP methods for Small Area Estimation (SAE). The robust EBLUP methods, e.g. M-estimator of an MLM's core parameters (called RML), as well as the robust prediction of random effects and area means, however, require more attention to the (non-trivial) issue of robustness-tuning and to computational aspects in general. In the context of an RML fitting exercise, users typically attempt to refer to the well-known robustness properties of simple location-scale models (such as the fact that the M-estimator of location with Huber psi-function obtains a relative asymptotic efficiency of approx. 95% w.r.t. maximum likelihood estimator at the true model if the tuning constant is 1.345). However, these properties do not always directly carry over to the RML method. As a consequence, the choice of how best to tune RML-estimators is much more involved. A simulation study of the computational aspects (incl. breakdown-point considerations) of robust estimators for unit-level SAE models is carried out and recommendations concerning the choice of robustness tuning constants are made.

E466: **Robustness analysis of unbalanced linear mixed modeling**
Presenter: **Roland Fried**, TU Dortmund University, Germany
Co-authors: Isabel Molina Peralta, Betsabe Perez Garrido, Anita Thieler

Robust approaches for fitting linear mixed models are investigated and compared. Starting from the simplest situation of fitting a constant mean μ in the presence of independent random group effects u_i and independent random errors e_{ij}, $Y_{ij} = \mu + u_i + e_{ij}$, with $i = 1,\ldots,I$ being the group index and $j = 1,\ldots,n_i$ denoting the individuals, we subsequently elaborate more complicated models with covariates. We start from the simple model given above, since methods which do not work there will hardly be useful in more complex situations. While balanced scenarios with equal group sizes $n_1 = \ldots = n_I$ have been analyzed frequently in the past, unbalanced designs have been given less attention so far. In small area estimation we are usually confronted with unbalanced designs and want to protect against the possibility of outlying individuals or groups. Our analysis aims at identifying which methods are useful under which scenarios and under which type of contamination.

E712: **Outlier robust domain estimation for business survey data**
Presenter: **Nikos Tzavidis**, University of Southampton, United Kingdom
Co-authors: Sabine Krieg, Marc Smeets, Chiara Bocci, Virginie Blaess

Outliers are a fact of life for any survey, and especially so for business surveys. If outliers are a concern for estimation of population quantities, it is safe to say that they are even more of a concern in small area estimation, where sample sizes are considerably smaller and model-dependent estimation is the norm. It is therefore of interest to see how outlier robust survey estimation can be adapted to this situation. The development of outlier robust small area methodologies has been the focus of recent small area literature. We review a range of outlier robust small area methodologies and apply these to business survey data from the Netherlands. We discuss both point and Mean Squared Error estimation (MSE) using analytic and bootstrap-type MSE estimators. Finally, we place some emphasis on providing practical guidelines to the survey practitioner for working with outlier robust small area methodologies.

ES36 Room Jessel PARAMETRIC AND SEMIPARAMETRIC HAZARDS MODELS AND ANALYSES Chair: M. Carmen Pardo

E069: **Inference for a semiparametric generalized logit-based proportional hazards model in survival analysis**
Presenter: **Martha Avendano**, Complutense University of Madrid, Spain
Co-authors: Maria del Carmen Pardo, Narayanaswamy Balakrishnan

A generalized logit-based proportional hazards model of survival analysis is introduced as an alternative to the popular Cox proportional-hazards model. Our model is based on a generalization of the logistic distribution and it is semiparametric, since it can be factored into a parametric part consisting of a regression parameter vector associated with the covariates and a non-parametric part that can be left completely unspecified. The partial loglikelihood approach can be applied to estimate the model parameters. The asymptotic properties of the parameter estimates are established. We compare our models with the Cox proportional hazards model by using one example.

E157: **EM and stochastic EM algorithms for duration mixture models under random censoring**
Presenter: **Laurent Bordes**, University of Pau, France
Co-authors: Didier Chauveau

Several iterative methods based on EM and Stochastic EM methodology are presented. They allow us to estimate parametric or semiparametric mixture models for randomly right censored lifetime data, provided they are identifiable. We consider different levels of completion for the (incomplete) observed data, and provide genuine or EM-like algorithms for several situations. In particular, we show that in censored semiparametric situations, a stochastic step is the only practical solution allowing computation of nonparametric estimates of the unknown survival function. The effectiveness of the new proposed algorithms is demonstrated in a simulation study and actual datasets.

E371: **Tools for the assessment of the linear regression model with an interval-censored covariate**
Presenter: **Klaus Langohr**, Polytechnical University of Catalonia, Spain
Co-authors: Guadalupe Gomez, M. Luz Calle

A linear regression model with one covariate which is interval-censored is considered. That is, for each individual the value of the covariate is only known to lie between two values that define the censoring interval. We found this situation in an AIDS clinical trial designed to compare different antiretroviral treatment regimens for HIV-infected persons, who had previously failed on a combination therapy involving the protease inhibitor Indinavir. A linear model relating the delay in initiating the new treatment and the log viral load level of the person at the time of initiating the new treatment is considered. The covariate in this model, the time from Indinavir failure to enrolment to the new treatment, is interval-censored. A likelihood approach to jointly estimate the regression coefficients of the model as well as the marginal distribution of the covariate has been proposed. We propose different tools for the assessment of the underlying assumptions of the linear model. We define new residuals for this model as the expected value of the theoretical errors conditional on the observed censoring interval. We compare the behaviour of the new residuals proposed with alternative residuals considered in previous works through simulation.

E724: **Conditional AIC for generalized linear and proportional hazards mixed models**
Presenter: **Florin Vaida**, University of California, San Diego, United States of America
Co-authors: Michael Donohue, Rosanna Haut, Ronghui Xu

Issues of model selection for generalized linear mixed models (GLMM) and proportional hazards mixed models (PHMM) are addressed. In earlier work we proposed the concept of conditional Akaike information for linear mixed effects models, and developed the corresponding criterion, cAIC, when the focus is on the clusters. Here we show how to extend the approach to GLMM and PHMM. While the derivation in the linear case used exact calculation, in the more general situation asymptotic approximations are necessary. The penalty term in the cAIC is related to the effective degrees of freedom for GLMM and PHMM developed recently in the literature. We apply the methods to the analysis of several medical studies.

ES59 Room Court IMPRECISE PROBABILISTIC MODELING TO SOLVE STATISTICAL PROBLEMS II Chair: Marco Cattaneo

E196: **Performing non-parametric homogeneity tests on interval-valued samples**
Presenter: **Olivier Strauss**, CIRAD, France
Co-authors: Sebastien Destercke

The problem of performing a non-parametric homogeneity test (Kolmogorov-Smirnov or Cramer-Von-Mises) on samples that are imprecisely observed and given as sets of intervals is considered. In this situation, test values become imprecise as well, and the test can result in an additional outcome corresponding to unknown outcome. As computing all possible values of the tests is complex and computationally expensive, we propose to use the notion of p-box to provide conservative bounds. We study some properties of these approximations, and perform various experiments to study the behaviour of these approximations. Finally, we apply our method to the comparison of medical images whose reconstruction leads to interval-valued images.

E420: **Imprecise preferences by means of probability boxes**
Presenter: **Enrique Miranda**, University of Oviedo, Spain
Co-authors: Ignacio Montes, Susana Diaz

One of the most popular preference measures between random variables is the notion of stochastic dominance, which is based on the comparison of their cumulative distribution functions. However, it is not uncommon to encounter situations where there is not enough information to fully elicitate the probability distributions of these variables, and it becomes necessary to look for robust alternatives. The notion of stochastic dominance is generalized to the case where there is uncertainty about the probability distribution of the random variables. The resulting notion entails the comparison of two sets of cumulative distribution functions, which can be equivalently be represented by means of probability boxes, or p-boxes. We propose a number of possible extensions of stochastic dominance, and investigate the relationships between them. In particular, we determine how are the preferences, derived from these extensions, affected by considering convex combinations, point-wise limits, and the difference between considering the finitely additive or countably additive subjacent probability distributions. The particular case of 0-1 valued p-boxes, which are the extreme points of the class of probability boxes, is also investigated.

E422: **Imprecise measurement error models and partial identification: towards a unified approach for non-idealized data**
Presenter: **Thomas Augustin**, University of Munich (LMU), Germany

Some first steps towards a generalized, unified handling of deficient, nay non-idealized, data are considered. The ideas are based on a more general understanding of measurement error models, relying on possibly imprecise error and sampling models. This modelling comprises common deficient data models, including classical and non-classical measurement error, coarsened and missing data, as well as neighbourhood models used in robust statistics. Estimation is based on an eclectic combination of concepts from Manski's theory of partial identification and from the theory of imprecise probabilities. Firstly, measurement error modelling with precise probabilities is discussed, with an emphasis on Nakamura's method of corrected score functions and some extensions. Secondly, error models based on imprecise probabilities are considered, relaxing the rather rigorous assumptions underlying all the common measurement error models. The concept of partial identification is generalized to estimating equations by considering sets of potentially unbiased estimating functions. Some properties of the corresponding set-valued parameter estimators are discussed, including their consistency (in an appropriately generalized sense). Finally, the relation to previous work in the literature on partial identification in linear models is made explicit.

E465: **On the implementation of likelihood-based imprecise regression**
Presenter: **Marco Cattaneo**, LMU Munich, Germany
Co-authors: Andrea Wiencierz

Likelihood-based imprecise regression (LIR) is a new approach to regression allowing the direct consideration of any kind of coarse data (including e.g. interval data, precise data, and missing data). LIR uses likelihood-based decision theory to obtain the regression estimates, which are in general

imprecise, reflecting the uncertainty in the coarse data. We address in particular the implementation of LIR, focusing on some important regression problems. From the computational point of view, the possible non-convexity of the estimated set of regression functions poses a considerable challenge.

Sunday 18.12.2011 14:15 - 16:20 Parallel Session J – CFE

CS07 Room B33 ADVANCES IN COMPUTATIONAL METHODS FOR DSGE MODELS Chair: Filippo Ferroni

C026: **Optimal disinflation under learning**
Presenter: **Christian Matthes**, Universitat Pompeu Fabra, Spain
Co-authors: Timothy Cogley, Argia Sbordone

Transitional dynamics that emerge after the adoption of a new monetary-policy rule are modeled. It is assumed that private agents learn about the new policy via Bayesian updating, and we study how learning affects the nature of the transition and choice of a new rule. In our model, uncertainty about the long-run inflation target matters only slightly, and the bank can always achieve low average inflation at relatively low cost. Uncertainty about policy-feedback parameters is more problematic. For some priors, the bank's optimal strategy is to adopt an incremental reform that limits the initial disagreement between actual and perceived feedback parameters. More ambitious reforms can succeed when priors permit agents to learn quickly enough. While fast learning is critical for the success of an ambitious reform, full credibility is not.

C087: **Interpreting the hours-technology time varying relationship**
Presenter: **Filippo Ferroni**, Banque de France, France
Co-authors: Cristiano Cantore, Miguel Leon-Ledesma

The time varying relation between hours and technology shocks is investigated. An RBC model with a Constant Elasticity of Substitution production function is proposed and estimated with Bayesian techniques. In the full sample, the specification fits the data reasonably well compared to the standard specification; we find (i) evidence in favor of an elasticity of substitution less than unity (in the Cobb-Douglas case such elasticity is one) and (ii) a sizable role for capital augmenting shock for business cycles fluctuations. In rolling sub-samples, we document that the transmission of technology shocks to hours worked has been varying over time. We argue that this change is due to the relative increase of the degree of factors substitution, that is labor and capital have become less complementary inputs in production as time went by.

C111: **Trend inflation, wage indexation and determinacy in the U.S.**
Presenter: **Efrem Castelnuovo**, University of Padova, Italy
Co-authors: Guido Ascari, Nicola Branzoli

An estimated monetary policy rule for the U.S. economy featuring time-varying trend inflation and stochastic coefficients is combined with a plausibly calibrated medium scale new-Keynesian framework embedding positive trend inflation. Contrary to some recent literature, we find that the impact of the decline in trend inflation on the likelihood of being in a determinate state is modest and limited to the second part of the 1970s. A change in the policy parameters is shown to be sufficient to drive the economy to a unique equilibrium regardless of the level of trend inflation. We identify wage indexation as the key-element supporting our "Taylor parameter only" result about the switch to determinacy.

C130: **Using survey data on inflation expectations in the estimation of learning and rational expectations models**
Presenter: **Arturo Ormeno**, University of Amsterdam, Netherlands

Do survey data on inflation expectations contain useful information for the estimation of macroeconomic models? This question is addressed by using data from the Survey of Professional Forecasters in the estimation of a New Keynesian model and compare its performance when solved under the Rational Expectations assumption and learning. Survey data on inflation expectations do not only serve as an additional moment restriction in the estimation of this model, but also helps us determining the forecasting model for inflation that agents most likely use in a learning setup. Our results reveal that the predictive power of the above-mentioned model is improved when resorting to available survey data and to an admissible learning rule for the formation of inflation expectations.

CS38 Room B36 SIGNAL EXTRACTION AND FORECASTING Chair: Pilar Poncela

C189: **Coincident and leading indicators using factor linear dynamic harmonic regression models**
Presenter: **Antonio Garcia-Ferrer**, Universidad Autonoma de Madrid, Spain
Co-authors: Marcos Bujosa, Aranzazu de Juan

One strand of growing interest within the multivariate literature lies in examining how the size and the composition of the data affect the forecasting performance. In several recent papers it was found that expanding the sample size simply by adding data that bear little information about the factor components does not necessarily improve forecasts and, when the data are too noisy, we can be better off by throwing away some data even though they are available. The goal is to obtain Composite Coincident (CCI) and Leading (CLI) indicators for the Spanish economy using monthly targeted predictors (cycle drivers) and dynamic factor models. The usefulness of the new indicators has been checked through both historical behavior and out-of-sample forecasting performance. When using the whole sample our CLI systematically anticipates the peaks and troughs of both the GDP growth cycles as well as the industrial production classical cycles with considerable anticipation. On the other hand, the use of our CLI provides considerable aid in forecasting annual and quarterly GDP growth rates during the recent recession. Using only real data available at the beginning of each forecast period our leading indicator one-step forecasts show considerable improvements over other benchmark alternatives including panels of consensus forecasts. Given the shortcomings of GDP in representing present economic conditions, our CCI takes into account the latest monthly information and provides a more accurate account of the general business conditions.

C190: **Specification and misspecification of models for measuring the output gap.**
Presenter: **Davide Delle Monache**, University of Rome. Tor Vergata, Italy
Co-authors: Andrew Harvey

The issues involved in specifying the trend component in an unobserved components (UC) model are first 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 MSEs of filtered and smoothed estimators of the trend using a recent algorithm. Cycles are then brought into the analysis and the efficiency, and robustness of various measures of the output gap are assessed; e.g. we compare a well-known orthogonal trend-cycle specification versus certain correlated components model. The correlation misspecification leads to a substantial loss in the smoother efficiency while the filters estimates retain high efficiency. Under model uncertainty it is interesting to compare the asymptotic values of the parameters of fitted model with the ones obtained in the literature. Moreover, we consider the estimation of the misspecified model based on the prediction error variance at different forecast horizons. The estimation based on the h-step ahead prediction error (with increasing h) increases the robustness of the trend component, because we concentrate on the low frequency of the power spectrum. The experience of recent years is examined.

C261: **Forecasting with multivariate models**
Presenter: **Carlos Cuerpo**, Universidad Autonoma de Madrid, Spain
Co-authors: Pilar Poncela

The absence of an agreed model to forecast the main economic aggregates at different time horizons remains one of the main challenges for the econometric analysis, especially in light of the recent subprime crisis. The main objective is to conduct a comparative analysis of the out-of-sample forecasting performance of both structural (Dynamic Stochastic General Equilibrium) and non-structural models (Dynamic Factor Models), based on Kalman filtering. We will check the robustness of our results by applying the exercise to Spanish, Euro area as well as US data. Our sample covers the period 1980Q1-2009Q4. The exercise will be performed over different stages of the business cycle and that will be reflected in the signals extracted by the different models, comparing also the forecasting performance through different periods.

C481: **Blind source separation for non-Gaussian time series using higher-order statistics**
Presenter: **Ester Gonzalez-Prieto**, MPIDR, Germany
Co-authors: Antonio Garcia-Ferrer, Daniel Pena

A new blind source separation approach that exploits both, the non-Gaussianity and the temporal structure, of the dataset is introduced. We propose a fourth-order temporal blind identification (FOTBI) algorithm, which identifies the set of underlying independent components by the joint diagonalization of several time-delayed fourth-order cumulant matrices. Some Monte Carlo experiments are carried out to investigate the performance of FOTBI. According to our results, FOTBI seems to be a good alternative for the separation of non-linear time series independent components.

C582: **Time series least angle regression for selecting predictive economic sentiment series**
Presenter: **Christophe Croux**, K.U.Leuven, Belgium
Co-authors: Sarah Gelper

A group-wise selection approach for forecasting time series using a large number of predictors is investigated. The proposed method identifies the most informative predictors in a linear dynamic time series model. In particular, we predict industrial production growth in Belgium from a large number of available economic sentiment indicators. The industrial confidence indicators in France and Belgium are identified as the most predictive indicators. The method applies group-wise least angle regression to time series models, and we call it TS-LARS. The time series LARS builds interpretable and parsimonious forecast models. It is a flexible method, fast to compute, and it ranks the predictors according to their predictive content. The time series LARS shows good performance, both in identifying the most relevant predictors and in attaining accurate forecasts.

CS42 Room G16 DYNAMIC CORRELATION MODELS Chair: Jeroen Rombouts

C364: **Locally constant modelling of multivariate volatilities via unbalanced Haar wavelets**
Presenter: **Piotr Fryzlewicz**, London School of Economics, United Kingdom
Co-authors: Rainer von Sachs

A technique for adaptively segmenting a volatility matrix into time intervals within which it is approximately stationary is proposed. This provides an appealing starting point for tasks such as volatility forecasting or portfolio selection in set-ups where there is evidence that change-points in market volatility are present, such as for example those due to the recent financial crisis. Our method is based on binary segmentation, via Unbalanced Haar wavelets, of the entire panel of squared returns simultaneously. To this end, we propose a new test statistic, which can be described as a “multivariate thresholded CUSUM" test for deciding whether a given change-point candidate is significant. The procedure leads to consistent estimation of the number and location of change-points, and performs very well in practice. Two essential ingredients of our procedure are: the variance-stabilising Fisz transform and the polarisation identity.

C462: **Volatility spillover in EU markets using DCC-MIDAS**
Presenter: **Farrukh Javed**, Lund University, Sweden

The role of the macroeconomic variables in determining the returns correlations between a number of European financial markets is examined. We first apply the GARCH-MIDAS (Mixed Data Sampling) model to examine the effect of macroeconomic variables (oil prices, interest rate and exchange rate) on the returns volatility. Next, we employ a DCC-GARCH model to analyze the time-varying correlations among returns, in which we can explore how the economic variables affect the conditional correlations. This approach enables us to decompose the correlation into its short-term and long-term components. In this model, we also investigate to what extent the correlations among economic variables from different countries affect the long term component of the return correlations. This is also the only study which discusses the effect of the economic variables on the correlations within the MIDAS context. The preliminary results for countries such as UK, Germany, Denmark and Sweden show that all the parameters are significant at a 5% level. There is an obvious long term and short term component in the correlations.

C257: **Multivariate Markov-Switching and change-point GARCH models**
Presenter: **Arnaud Dufays**, Universite Catholique de Louvain-la-Neuve, Belgium
Co-authors: Luc Bauwens, Jeroen Rombouts

Multivariate GARCH type volatility models with fixed parameters are too restrictive for long time series due to breaks in the volatility process. These features are ignored in standard GARCH models and result typically in an artificially highly persistent process when fitted. Since this may have severe consequences for volatility forecasting, more flexible parameterizations are needed. We consider multivariate Markov-switching GARCH and change-point GARCH models of the dynamic conditional correlations type. This allows us to first detect breaks in the individual time series, and second in the conditional correlations. We determine the optimal number of regimes or breaks by maximizing the marginal likelihood. We illustrate the models on simulated data, and we apply it to several index returns.

C609: **Jump robust daily covariance estimation by disentangling variance and correlation components**
Presenter: **Kris Boudt**, K.U.Leuven and Lessius University College, Belgium
Co-authors: Jonathan Cornelissen, Christophe Croux

A jump robust positive semidefinite rank-based estimator for the daily covariance matrix based on high-frequency intraday returns is proposed. It disentangles covariance estimation into variance and correlation components. This allows us to account for non-synchronous trading by estimating correlations over lower sampling frequencies. The efficiency gain of disentangling covariance estimation and the jump robustness of the estimator are illustrated in a simulation study. In an application to the Dow Jones Industrial Average constituents, it is shown that the proposed estimator leads to more stable portfolios.

C255: **Forecasting correlations during the late-2000s financial crisis: short-run component, long-run component, and structural breaks**
Presenter: **Francesco Audrino**, University of St. Gallen, Switzerland

The predictive power of the various components affecting correlations that have been recently introduced in the literature will be empirically investigated. We focus on models allowing for a flexible specification of the short-run component of correlations as well as the long-run component. Moreover, we also allow the correlation dynamics to be subjected to regime-shift caused by threshold-based structural breaks of a different nature. Our results indicate that in some cases there may be a superimposition of the long- and short-term movements in correlations. Therefore, care is called for in interpretations when estimating the two components. Testing the short-term forecasting accuracy of correlations during the late-2000s financial crisis yields mixed results. In general component models allowing for a richer correlation specification possess a (marginally) increased predictive accuracy. Economically speaking, no relevant gains are found by allowing for more flexibility in the correlation dynamics.

CS50 Room B34 BAYESIAN MODEL AVERAGING IN ECONOMETRICS Chair: Mark Steel

C097: **Fishing economic growth determinants using Bayesian elastic nets**
Presenter: **Jesus Crespo Cuaresma**, Vienna University of Economics and Business, Austria
Co-authors: Paul Hofmarcher, Bettina Gruen

A method is proposed to deal simultaneously with model uncertainty and correlated regressors in linear regression models by combining elastic net specifications with a spikes and slabs prior. The individual specifications nest ridge regression and the LASSO estimator and thus allow for a more flexible modelling framework than existing model averaging procedures. Our method has clear advantages when dealing with datasets of (potentially highly) correlated regressors, a pervasive characteristic of the model averaging datasets used hitherto in the econometric literature. We apply our method to a dataset of economic growth determinants and show that our procedure has superior out-of-sample predictive abilities when compared to the standard Bayesian model averaging methods currently used in the literature.

C115: **Dynamic panels with predetermined regressors: likelihood-based estimation and Bayesian averaging with an application to cross-country growth**
Presenter: **Enrique Moral-Benito**, Bank of Spain, Spain

Likelihood-based estimation of linear panel data models with general predetermined variables and individual-specific effects is discussed. The resulting (pseudo) maximum likelihood estimator is asymptotically equivalent to standard GMM but tends to have smaller finite-sample biases as illustrated in simulation experiments. Moreover, the availability of such a likelihood function allows applying the Bayesian apparatus to this class of panel data models. Combining the aforementioned estimator with Bayesian model averaging methods we estimate empirical growth models simultaneously considering endogenous regressors and model uncertainty. Empirical results indicate that only the investment ratio seems to robustly cause long-run economic growth. Moreover, the estimated rate of convergence is not significantly different from zero.

C209: **Assessing early warning indicators of economic crises**
Presenter: **Charis Christofides**, IMF, United States of America
Co-authors: Theo Eicher, Chris Papageorgiou

The recent economic crisis has given rise to a voluminous literature devoted to identifying leading indicators of economic crises. The hallmark of the literature has been the diversity of the approaches and crisis indicators that have been employed, so much so that a new literature has sprung up that summarizes insights in meta analyses. The fundamental difficulty is that the great number of alternative crisis theories motivate an usually large set of crisis indicators and the resulting model uncertainty has thus far not been addressed. As a result existing regressions overstate significance levels, since the uncertainty surrounding the validity of a particular theory has been largely ignored. We propose a methodology to resolve the model uncertainty inherent in this literature by employing Bayesian Model Averaging. Using the largest set of crisis indicators suggested to date, we reexamine previous results and probe their robustness.

C359: **Growth determinants, data revisions and supermodels**
Presenter: **Martin Feldkircher**, Oesterreichische Nationalbank, Austria
Co-authors: Stefan Zeugner

The problem of (agnostic) BMA results applied to growth empirics being highly senstitive to small perturbations of the dependent variable is addressed. In a cross country growth set-up using four different vintages of the Penn World Table (PWT) data we demonstrate that such instability is partly due to the overconfident 'default' $g-$prior framework employed in standard macroeconometric BMA literature. Instead, we propose to rely on the hyper$-g$ prior which adjusts to data quality and induces smaller shrinkage factors according to the data's considerable noise component. This in turn renders BMA results considerably more stable over different revisions of PWT growth data. Our results point to initial income, human capital and a set of regional dummy variables as robust growth determinants, both with respect to model uncertainty as well as PWT revisions.

C435: **Model averaging and variable selection in VAR-models**
Presenter: **Sune Karlsson**, Orebro University, Sweden
Co-authors: Shutong Ding

Bayesian model averaging and model selection is based on the marginal likelihoods of the competing models. This can, however, not be used directly in VAR models when one of the issues is which – and how many – variables to include in the model, since the likelihoods will be for different groups of variables and not directly comparable. One possible solution is to consider the marginal likelihood for a core subset of variables that are always included in the model. This is similar in spirit to a recent proposal for forecast combination based on the predictive likelihood. The two approaches are contrasted and their performance is evaluated in a simulation study and a forecasting exercise.

CS63 Room B35 COMPUTATIONAL AND ECONOMETRIC METHODS IN DERIVATIVES APPLICATIONS Chair: Panayiotis Andreou

C260: **Market efficiency, information flows and hedging performance in European and US carbon markets**
Presenter: **Jing-Ming Kuo**, University of Durham, United Kingdom
Co-authors: Yukun Shi

Carbon markets, which are designed to reduce global CO2 emissions, have experienced rapid ongoing developments even during the recent recession and have attracted considerable attention from policy makers and environmental investors. We provide an insight into the efficiency of carbon markets using weekly EU Emission Allowance (EUA) and Carbon Financial Instrument (CFI) futures data with a generalized quadratic ARCH-in-mean error correction model (GQARCH-M-ECM). The results indicate that the market efficiency is strongly violated in both European and US carbon markets. It is of interest that, in using a bivariate AR-GARCH model, we found that there is volatility spillover but no pricing transmission between the two markets. Finally, we employed naive hedge, OLS and three multivariate GARCH models to determine the optimal hedge ratio and to examine the performance of hedging strategies in carbon markets. The results show that carbon futures markets cannot effectively provide the risk management function for carbon market participants.

C408: **Empirical pricing kernel estimation using a functional gradient descent algorithm based on splines or trees**
Presenter: **Pirmin Meier**, University of St. Gallen, Switzerland
Co-authors: Francesco Audrino

A new methodology to estimate the empirical pricing kernel implied from option data is proposed. Contrary to most of the studies in the literature which use an indirect approach, i.e. estimating first the physical and risk-neutral densities and obtaining the pricing kernel in a second step, we follow a direct approach. Departing from an adequate parametric or non-parametric and possibly economically motivated pricing kernel, we apply a functional gradient descent (FGD) algorithm based on B-splines or regression trees. This approach allows us to locally modify the initial pricing kernel and hence to improve the final estimate. We illustrate empirically the estimation properties of the method and test its predictive power on S&P 500 option data, also in comparison with other recent approaches introduced in the empirical pricing kernel literature.

C416: **Investor sentiments, rational beliefs and option prices**
Presenter: **Anastasios Kagkadis**, Durham University, Greece
Co-authors: Panayiotis Andreou, Dennis Philip

The impact of the economic environment on the risk-neutral skewness extracted from S&P 500 index options and consequently on options prices is investigated. Previous research has reported that index option volatility smile is steeper (flatter) and the risk-neutral skewness of monthly index returns is more (less) negative when investors' sentiment about the stock market is more bearish (bullish). We re-investigate the time variation in the slope of volatility smile and the risk-neutral skewness as implied by the S&P 500 index options under a rational updating of investors' beliefs about the conditional density of index return. In particular, it is aimed to determine whether the relationship between the variation in the slope of the index volatility smile to the changes in aggregate error in investors' beliefs becomes weaker once a rich set of macroeconomic factors that drive the economy are taken into consideration. We do that by linking risk-neutral skewness to variables that proxy for investment opportunities and predict market returns. To construct such a rich set of rational predictors of index return, we use asymptotic principal component analysis on an extensive set of 133 monthly macroeconomic time-series.

C530: **A volatility smirk that defaults: The case of the S&P 500 index options**
Presenter: **Panayiotis Andreou**, Cyprus University of Technology, Cyprus

Modern financial engineering has dedicated significant effort in developing sophisticated option pricing models to replicate the implied volatility smirk anomaly. Nonetheless, there is limited empirical evidence to examine the causes of this anomaly implied by market options data. This study investigates how market default risk affects the shape of the S&P 500 index implied volatility functions. 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. Moreover, the results illuminate a set of economic determinants that are found to affect the shape of the S&P 500 index implied volatility functions. These factors are related to characteristics of the underlying asset and micro-structure variables characterizing the option market itself.

C617: **An empirical study of stock and American option prices**
Presenter: **Diego Ronchetti**, Swiss Finance Institute at the University of Lugano, Switzerland

An empirical study of the information content of daily share prices and American put and call option mid-quotes about their generating process is described. Considering stock return and its volatility as the risk factors and without parameterizing their historical joint dynamics, two results are empirically obtained. First, share prices and option mid-quotes are both necessary to identify at the same time the discount for uncertain stock return and return variance in the asset price formation. Second, constraining the nonparametric estimation procedure of the historical joint dynamics of the risk factors to satisfy an arbitrage-free pricing model is useful to get more stable estimates over time. As an illustration, time series of different estimates of historical conditional correlation of the risk factors, Sharpe ratio of an investment on the stock, return skewness and kurtosis are reported.

CS69 Room B20 MEASURING SYSTEMIC RISK Chair: Monica Billio

C075: **Wrapping it up: Risk exposures, spillovers, contagion and systemic risk**
Presenter: **David Veredas**, Universite Libre de Bruxelles, Belgium
Co-authors: Mardi Dungey, Matteo Luciani

Broadly speaking there are as many definitions and measures of systemic risk as articles on the area. Trichet reduced to three the ways in which systemic risk can happen: contagion, financial imbalances and negative aggregate shocks. Contagion is, in our view, the most natural one. Risk contagion represents the transmission channels of risk between firms having conditioned on common factors. In other words, it refers to the unexpected commonness and how new transmission channels are created in periods of turmoil. Following the macro-finance literature, we model contagion within dynamic factor models for realized risk measures. Contagion takes place if the correlations across the idiosyncratic shocks increase due to unexpected transmission mechanisms.

C122: **Systemic risk diagnostics: coincident indicators and early warning signals**
Presenter: **Bernd Schwaab**, European Central Bank, Germany
Co-authors: Siem Jan Koopman, Andre Lucas

A novel framework to assess financial system risk is proposed. Using a dynamic factor framework based on state-space methods, we construct coincident measures ('thermometers') and a forward looking indicator ('barometer') for the likelihood of simultaneous failure of a large number of financial intermediaries. The indicators are based on latent macro-financial and credit risk components for a large data set comprising the U.S., the EU-27 area, and the respective rest of the world. Credit risk conditions can significantly and persistently de-couple from macro-financial fundamentals. Such a decoupling can serve as an early warning signal for macro-prudential policy.

C320: **l^q-regularization of the Kalman filter for exogenous outlier removal: application to hedge funds analysis**
Presenter: **Serge Darolles**, CREST, France
Co-authors: Jay Emmanuelle, Duvaut Patrick

A simple and efficient exogenous outlier detection & estimation algorithm introduced in a regularized version of the Kalman filter is presented. Exogenous outliers that may occur in the observations are considered as an additional stochastic impulse process in the observation equation of the Kalman filter that requires a regularization of the innovation in the recursive equations of the Kalman filter. Regularizing with a l1 or l2-norm needs to determine the value of the regularization parameter. Since the innovation error of the KF is assumed to be Gaussian we propose to first detect the possible occurrence of a non-Gaussian spike and then to estimate its amplitude using an adapted value of the regularization parameter. The algorithm is first validated on synthetic data and then applied to a concrete financial case that deals with the analysis of hedge fund returns. We show that the proposed algorithm can detect anomalies frequently observed in hedge returns such as illiquidity issues.

C448: **CISS - A composite indicator of systemic stress in the financial system**
Presenter: **Manfred Kremer**, European Central Bank, Germany
Co-authors: Daniel Hollo, Marco Lo Duca

A new indicator of contemporaneous stress in the financial system is introduced. It is named Composite Indicator of Systemic Stress (CISS) and exemplified on the basis of euro area data. The CISS not only permits the real time monitoring and assessment of the stress level in the whole financial system and the potential impact of mitigating policy measures, but may also help better distinguishing and delineating past episodes of financial crises. The CISS arguably comprises the most important segments of an economy's financial system: financial intermediaries, money markets, equity and bond markets (non-financial issuers), as well as foreign exchange markets. Its main methodological innovation compared to alternative financial stress indicators is the application of standard portfolio theory to the aggregation of the underlying individual stress measures into the composite indicator. Precisely, the subindices are aggregated on the basis of weights which reflect their time-varying cross-correlation structure. As a result, the CISS puts relatively more weight on situations in which stress prevails in several market segments at the same time. We determine within a threshold regression framework a level of the CISS at which financial stress endangers real economic activity.

C489: **Network analysis: Contagion and systemic risk**
Presenter: **Monica Billio**, University of Venice, Italy
Co-authors: Lorenzo Frattarolo, Loriana Pelizzon

Network analysis is considered to capture the interconnectedness in the financial system. The network is defined using Granger causality tests among monthly returns of hedge funds, banks, brokers, and insurance companies. 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 of interconnectedness are also decomposed to distinguish between contagion effects and systemic risk.

CS92 Room B18 FINANCIAL ECONOMETRICS II Chair: Lorenzo Trapani

C067: **Using skewness to estimate the semi-strong GARCH(1,1) model**
Presenter: **Todd Prono**, Commodity Futures Trading Commission, United States of America

IV estimators with an instrument vector composed only of past squared residuals, while applicable to the semi-strong ARCH(1) model, do not extend to the semi-strong GARCH(1,1) case because of underidentification. Augmenting the instrument vector with past residuals, however, renders traditional IV estimation feasible, if the residuals are skewed. The proposed estimators are much simpler to implement than efficient IV estimators, yet they retain improved finite sample performance over QMLE. Jackknife versions of these estimators deal with the issues caused by many (potentially weak) instruments. A Monte Carlo study is included, as is an empirical application involving foreign currency spot returns.

C049: **Measures of financial risk**
Presenter: **S.Y. Novak**, Middlesex University Business School, London, United Kingdom

Modern approaches to financial risk measurement are overviewed. Popular measures of risk, e.g., Value-at-Risk and Expected Shortfall, appear static from a short-term investor point of view. It is argued that technical analysis offers tools for dynamic risk measurement. Strengths and weaknesses of the technical analysis approach to financial risk measurement are discussed. The arguments are illustrated by examples of real data.

C652: **Option pricing under Sign RCA-GARCH models - A comparative study**
Presenter: **Joanna Gorka**, Nicolaus Copernicus University, Poland

After Black and Scholes's groundbreaking work, the literature concerning pricing options has become a very important area of research. Numerous option valuation methods have been developed. The Black and Scholes formula assumes that the returns on the stock price follow a normal distribution with constant volatility. This assumption (of constant volatility) is often strongly violated. Therefore, the stochastic volatility models are used for option valuation. There are two types of stochastic volatility: in continuous time (more general stochastic processes, SV) and in discrete time (the class of GARCH models). It is shown how one can compute option prices using a Sign RCA-GARCH model for the dynamics of the volatility. The proposed method is compared to Black and Scholes's evaluation and other selected GARCH option pricing models.

C390: **A test of the efficiency of asset returns in the four-moment framework: An international study**
Presenter: **Minh Phuong Doan**, RMIT University, Australia
Co-authors: Heather Mitchell, Richard Heaney

The efficiency of the CAPM incorporating skewness and kurtosis factors in the four-moment framework is investigated in four major developed markets: the U.S, the U.K, Japan and Australia. In particular, we apply a well-known generalised multivariate method for efficiency tests in the two-moment and four-moment framework. As we find the returns are not normally distributed, we also apply the bootstrap method as a robustness check. We find that stock returns in the U.S, U.K and Australian markets are not mean-variance efficient but they are generally mean-variance-skewness-kurtosis efficient, whereas the Japanese stock returns are mean-variance efficient. The finding is important since it suggests that performance measurement and asset allocation should be based on the four-moment framework, rather than the conventional mean-variance framework, especially for the U.S, U.K and Australian markets.

C594: **Testing for (In)Finite Moments**
Presenter: **Lorenzo Trapani**, Cass Business School, United Kingdom

This paper develops a new class of tests for the null hypothesis that the k-th moment of a random sample does not exist, versus the alternative that the k-th moment exists and is finite. The tests are based on using the k-th sampe moment and the divergent part of the Law of Large Numbers. This entails that, under the null, the k-th sample moment diverges to infinity almost surely, whereas under the alternative it converges to a finite constant. This dichotomy is exploited by using a randomised testing procedure, based on generating a Gaussian random sample of size R and multiplying it by the sample k-th moment. The test is extended to be applied to residuals, and the issue of dependence is tackled by pre-whitening. It is shown that the test has power versus alternatives characterized by trending moments. Simulations show that the test has the correct size and that the power can be made very high upon choosing a suitable sample size R when applying the randomized procedure.

CP03 Room Chancellor's POSTERS SESSION III Chair: Cristian Gatu

C694: **An econometric analysis of selected economic indicators in Nigeria: A vector autoregressive (VAR) modeling approach**
Presenter: **Olushina Olawale Awe**, Obafemi Awolowo University,Ile-Ife,Nigeria, Nigeria

The analysis of some economic variables in Nigeria is addressed. The variables considered are: gross domestic product, money supply, investment, exchange rate, inflation rate, government expenditure, and interest rate on lending. Data on these variables collected over a period of 35 years were subjected to econometric analysis using vector autoregressive models. Traditionally, most economic variables are non-stationary, hence unit

root tests were performed on all the variables. All the variables were found to be non-stationary and integrated of either I(1) or I(2). Johansen's co-integration test reveals that we can have at least four co-integrating equations out of seven considered. The correlation matrix shows that three of the variables are highly related with GDP: money supply (0.965410), investment (0.955394) and exchange rate (0.9625980). Finally, the VAR model was used to forecast for GDP and the performance was compared with other models in literature by the use of root mean square error (RMSE).

C700: **Estimation and sensitivity analysis of business efficiency under free distribution methodology.**
Presenter: **Felipe Manuel Rosa-Gonzalez**, University of La Laguna, Spain
Co-authors: Enrique Gonzalez-Davila

Business competition involves comparisons with respect to the leading companies in the sector and is essential to achieve efficiency levels appropriate to the requirements. The calculation of efficiency or inefficiency of firms is a widely used measure while making business decisions. What is more, the verification and robustness of the estimation methods play a great role in assisting the decision-making process especially now, when the current global crisis increases the risk of a bad decision. We apply methods of free distribution on panel data to evaluate and classify firm inefficiencies in a textile business. Actual observed data will allow us to generate through simulation artificial populations which will be used to perform a sensitivity analysis. In particular, we analyze the influence of an inappropriate election in regression models and the impact of selecting the best company.

C756: **Bootstrap techniques for estimating the number of components in mixture analyses**
Presenter: **Athanase Polymenis**, University of Central Greece, Greece

Bootstrap techniques have been proved to be very powerful tools for the assessment of the number of components in finite mixtures. We use a new method, which is a combination of a bootstrap minimum information ratio (called modified MIR), with a bootstrap likelihood ratio. The proposed method was motivated by the fact that the bootstrap likelihood ratio includes the case where the data arise from a single normal distribution, whereas the modified MIR does not. Therefore, it would seem interesting to "complete" the modified MIR algorithm by inserting a bootstrap likelihood ratio algorithm only for the non-mixture case. This idea is also supported by the fact that the bootstrap likelihood ratio had an excellent performance in a previous simulation exercise, where the true normal distribution of the data was identified more than 94% of the times, for all degrees of components' separation. In order to implement the method, we note that both algorithms have the same number of corresponding steps, and, thus, can be easily combined into a single new algorithm. Simulation results obtained are encouraging. Finally, it is worth mentioning that the method is easily applied to multivariate data.

C782: **Information content of various realized volatility and jump estimators on the model-free implied volatility**
Presenter: **Michaela Barunikova**, Charles University Prague, Czech Republic
Co-authors: Jozef Barunik

Option prices are widely believed to carry information concerning expectations of the market participants about the future movement of underlying asset prices. To support this opinion, volatilities of the underlying asset are extracted from the option prices and their information content is evaluated. Until the model-free option implied volatility (MFIV) has been introduced, implied volatility was calculated from the inverted Black-Scholes formulae, which is known to suffer from moneyness bias. We study the information content of volatility forecasts from the MFIV as well as several realized volatility estimators. To capture the long memory feature of the volatility, we use the ARFIMA approach and we model call and put options at various dates of expiration separately. As jumps and noise in the high frequency data cause large bias to the volatility estimates, we decompose the realized volatility to its continuous part and jump part using recently proposed jump adjusted wavelet two scale realized volatility (JWTSRV). The study is performed on the unique datasets of options and intra-day futures data containing German, London and U.S. stock indices. The sample covers the 2008 financial crisis providing us with the opportunity to bring new results on the relation of MFIV with RV.

C797: **Missing observations in volatility contagion analysis. Bayesian approach using the MSV-MGARCH framework**
Presenter: **Krzysztof Osiewalski**, Cracow University of Economics and Sabre Holdings, Poland
Co-authors: Jacek Osiewalski

Usually observations of prices on different markets are not fully synchronous. The question is whether we should exclude from modelling the days with prices not available on all markets (thus loosing some information and implicitly modifying the time axis) or somehow complete the missing (non-existing) prices. In order to compare the effects of each of the two ways of dealing with partly available data, one should consider formal procedures of replacing the unavailable prices by their appropriate predictions. We propose a fully Bayesian approach, which amounts to obtaining the marginal predictive distribution for any particular day in question. This procedure takes into account uncertainty and can be used to check validity of informal ways of "completing" the data. We use simple hybrid MSV-MGARCH models, which can parsimoniously describe volatility of a large number of prices or indices. In order to conduct Bayesian inference, the conditional posterior distributions for all unknown quantities are derived and the Gibbs sampler (with Metropolis-Hastings steps) is designed. Our approach is applied to daily prices from different financial and commodity markets, including the period of the global financial crisis. We compare inferences about conditional correlation obtained in the cases of deleted or completed observations.

C888: **Advanced estimates of regional accounts: A mixed approach nesting spatial errors into State Space Models**
Presenter: **Riccardo Corradini**, ISTAT, Italy

The policies related to regional economic activity developed by European Union (EU) and the role played by regions as economic subject have determined a larger set of disaggregated statistics at a macro and regional level. A new technique to forecast regional accounts is illustrated. The aim is to show how spatial panel data models could be casted into a state-space formulation. The comparison is made with respect to a set of real Italian economic data. The regional levels of full time equivalents will be used to forecast the regional levels of real GDP subject to a spatial national constraint. From the data analysis we draw some conclusions on the comparative performances of the distinct models.

C918: **Monetary policy and inflation targeting in Egypt: An empirical study**
Presenter: **Ahmed Mabrouk**, Misr International University, Egypt
Co-authors: Marwa Elsherif

Since 1988, when the New Zealand Central Bank first developed the inflation targeting (IT) policy option, many economies have adopted, in one way or another, the IT regime. This regime-shift was justified to a certain extent by the difficulties posed by targeting the nominal exchange rate or money supply. In Egypt, following the move to a floating exchange rate regime in February 2003, inflation rates have increased considerably, reaching 17%. This rise in the inflation rate was mainly due to the pass-through effect of the Egyptian pound's depreciation. The Central Bank of Egypt (CBE) announced the shift to adopt price stability as an explicit monetary policy objective once the fundamental prerequisites are met. Using time series econometric techniques, namely the vector autoregressive (VAR) model and impulse response functions (IRF), we aim to analyze the monetary transmission mechanism in Egypt, and to assess how successful the CBE has been in satisfying the preconditions for, and making a smooth transition to IT. We review the recent literature on the transmission mechanism of monetary policy and inflation targeting in some

developing countries. The empirical methodology utilized in the analysis is introduced and empirical results are presented and interpreted. Finally, we conclude with policy implications.

C919: **The monetary transmission mechanism in South Africa: A VECM augmented with foreign variables**

Presenter: **Annari De Waal**, University of Pretoria, South Africa
Co-authors: Renee Van Eyden

A structural cointegrated vector autoregressive (VAR) model with weakly exogenous foreign variables suitable for a small open economy like South Africa is developed. This type of model is known as an augmented vector error correction model (VECM), referred to by VECX*. We compile the foreign variables using trade-weighted three-year moving average data for the 16 main trading partners of South Africa, to account for significant changes in trade shares over time. It is the first VECX* developed for South Africa and the first model that uses time-varying trade weights to create foreign variables for the country. We find three significant long-run economic relations: the augmented purchasing power parity, the uncovered interest parity and the Fisher parity. These long-run relations are imposed on the VECX* for an investigation of the monetary transmission mechanism in South Africa.

C958: **Dynamic correlations in exchange rate time series: a copula approach**

Presenter: **Krenar Avdulaj**, Charles University in Prague, Czech Republic
Co-authors: Jozef Barunik

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

Sunday 18.12.2011 14:15 - 16:20 Parallel Session J – ERCIM

ES08 Room Bloomsbury OPTIMAL DESIGN Chair: Steve Gilmour

E169: **Optimal designs for multinomial logistic regression**
Presenter: **Stefanie Biedermann**, University of Southampton, United Kingdom
Co-authors: Min Yang

Multinomial categorical responses are observed in areas as diverse as discrete choice experiments, toxicology or clinical trials. The complicated structure of the information matrix in such models makes the explicit determination of optimal designs difficult. A new algorithm for the numerical calculation of optimal experimental designs is proposed. Its use is not limited to multinomial models, but it can be used more generally. Its efficiency is demonstrated through application to a concrete example. Since the models considered are nonlinear in the parameters, the designs found are locally optimal. Therefore it is shown that the methodology can be modified to determine adaptive optimal designs, which are more robust to initial misspecification of parameter values.

E446: **Algorithmic choice designs for paired comparisons of partial profiles**
Presenter: **Heiko Grossmann**, Queen Mary University of London, United Kingdom

Choice and paired comparison experiments are widely used for measuring how individuals weigh the different characteristics of goods or services. To this end, a suitable experimental design is used to combine attribute levels into options or profiles and to further arrange these into pairs or larger choice sets. When the number of attributes is large, often incomplete descriptions of the options, which are known as partial profiles, are used to prevent respondents from using simplifying strategies which may have a detrimental effect on the results. An algorithm for generating efficient paired comparison designs is presented. It is based on the author's previous analytical work on optimal exact designs for partial profiles in paired comparison experiments. The algorithm combines combinatorial structures such as Hadamard and weighing matrices in order to obtain a design which has a block diagonal information matrix. This approach differs considerably from standard optimal design algorithms. As a practical advantage, the algorithm allows the construction of efficient designs with fewer pairs than needed by the optimal exact designs which can be constructed analytically.

E598: **Q_B-optimal saturated two-level main effects designs**
Presenter: **Pi-Wen Tsai**, National Taiwan Normal University, Taiwan
Co-authors: Steven Gilmour

We provide a general framework that incorporates experimenters' prior beliefs into the design selection process for the study of saturated two-level main effects designs, which are commonly used for screen experiments. We show that under the sets of priors with more weights on models of small size, p-efficient designs should be recommended; when models with more parameters are of interest, D-optimal designs would be better. Also, we present new classes designs which can be found between these two designs under different sets of priors. The way in which the choice of designs depends on experimenters' prior beliefs will be demonstrated for the cases when $N = 2 \bmod 4$.

E556: **Model-robust variance-component estimation and lack-of-fit test for split-plot and other multi-stratum response surface designs**
Presenter: **Peter Goos**, Universiteit Antwerpen, Belgium
Co-authors: Steven Gilmour, Heiko Grossmann

Estimation of the variance components is essential in the analysis of data from split-plot and other multi-stratum response surface designs. The common approach to do so is to use restricted maximum likelihood (REML) estimation for the variance components, starting from a response surface model. We recommend using REML estimation based on the full treatment model instead, and show the benefits of this approach. We also show how to carry out a lack-of-fit test for data from split-plot and other multi-stratum response surface designs, based on REML.

E847: **An improved algorithm for split-plot and multi-stratum designs**
Presenter: **Steven Gilmour**, University of Southampton, United Kingdom
Co-authors: Luzia Trinca

Many industrial experiments involve some factors whose levels are harder to set than others. The best way to deal with these is to plan the experiment carefully as a split-plot, or more generally a multi-stratum design. Several different approaches for constructing split-plot type response surface designs have been proposed in the literature in the last 10 years or so, which has allowed experimenters to make better use of their resources by using more efficient designs than the classical balanced ones. One of these approaches, the startum-by-stratum strategy, has been shown to produce designs that are less efficient than locally D-optimal designs. An improved stratum-by-stratum algorithm is given, which, though more computationally intensive than the old one, makes most use of the advantages of this approach, i.e. it can be used for any structure and does not depend on prior estimates of the variance components. This is shown to be almost as good as the locally optimal designs in terms of their own criteria and more robust across a range of criteria.

ES12 Room Jessel HIGH-DIMENSIONAL STATISTICS, SPARSITY AND APPLICATIONS Chair: Pierre Alquier

E089: **PAC-Bayesian and interacting MCMC techniques under the scope of sparse generalized additive model**
Presenter: **Benjamin Guedj**, UPMC, France
Co-authors: Gerard Biau, Eric Moulines, Pierre Alquier

Sharp risk bounds for exponentially weighted aggregates of Bayesian estimators and their explicit simulation in the framework of prediction are addressed. In more detail, the generalized additive model $Y = \sum_{j=1}^{p} f_j(X_j) + \varepsilon$ under a sparsity constraint is considered. Sharp PAC-Bayesian oracle inequalities will be provided for the proposed estimator, which consists in a version of the Gibbs aggregated estimator computed with respect to a prior distribution favoring sparse estimates and simulated using adaptive and interacting MCMC methods.

E108: **High dimensional instrumental regression and confidence sets**
Presenter: **Eric Gautier**, CREST - ENSAE ParisTech, France
Co-authors: Alexandre Tsybakov

We present an instrumental variables method for estimation in linear models with endogenous regressors in the high-dimensional setting where the sample size n can be smaller than the number of possible regressors K, and $L \geq K$ instruments. Heteroscedasticity is allowed and a prior knowledge of variances of the errors is not needed. The main results are upper bounds on the estimation error of the vector of coefficients in l_p-norms for $p \geq 1$ that hold with probability close to 1, as well as the corresponding confidence intervals. All results are non-asymptotic. These bounds are meaningful under the assumption that the true structural model is sparse or many coefficients are too small to matter. In our IV regression setting, the standard tools from the literature on sparsity, such as the restricted eigenvalue assumption are inapplicable. Therefore, we develop a new approach based on

data-driven sensitivity characteristics. We show that, under appropriate assumptions, a thresholded STIV estimator correctly selects the non-zero coefficients with probability close to 1. The price to pay for not knowing which coefficients are non-zero and which instruments to use is of the order $p\log(L)$ in the rate of convergence.

E133: **The group fused Lasso for multiple change-point detection**
Presenter: **Kevin Bleakley**, INRIA Saclay, France
Co-authors: Jean-Philippe Vert

The group fused Lasso for detection of multiple change-points shared by a set of co-occurring one-dimensional signals is presented. Change-points are detected by approximating the original signals with a constraint on the multidimensional total variation, leading to piecewise-constant approximations. Fast algorithms are proposed to solve the resulting optimization problems, either exactly or approximately. Conditions are given for consistency of these algorithms as the number of signals increases, and empirical evidence is provided to support the results on simulated and array comparative genomic hybridization data.

E463: **Subspace adaptiveness of compressive Fisher's linear discriminant classifier**
Presenter: **Ata Kaban**, University of Birmingham, United Kingdom
Co-authors: Bob Durrant

Dimensionality reduction by non-adaptive stable embeddings, such as random projections and compressed sensing have been gaining popularity for their computational advantages and theoretical guarantees. We study the use of such techniques for high dimensional learning problems, such as their effects on classifier performance. An analysis of Fisher's Linear Discriminant (FLD) classifier when some high dimensional data is only available in a randomly projected compressive form is presented. Without any sparsity requirement, provided subgaussian classes, the estimated generalisation error of compressive FLD is upper-bounded in terms of quantities in the original data space, and the number of measurements required for good generalisation grows with the log of the number of classes. Furthermore, we can show that when the data distribution lives in a linear subspace of the ambient space, the error is independent of the ambient dimension and depends only on the dimension of the relevant subspace. Hence, in such situations compressive FLD can be both more efficient and more accurate than FLD on the original data, which is unable to take advantage of this structure.

E030: **A second look at stability selection**
Presenter: **Richard Samworth**, University of Cambridge, United Kingdom
Co-authors: Rajen Shah

Stability selection was recently introduced as a very general technique designed to improve the performance of a variable selection algorithm. It is based on aggregating the results of applying a selection procedure to subsamples of the data. We introduce a variant, called Complementary Pairs Stability Selection (CPSS), and derive bounds both on the expected number of variables included by CPSS that have low selection probability under the original procedure, and on the expected number of high selection probability variables that are excluded. These results require no (e.g. exchangeability) assumptions on the underlying model or on the quality of the original selection procedure. Under reasonable shape restrictions, the bounds can be further tightened, yielding improved error control, and therefore increasing the applicability of the methodology.

ES22 Room S261 DEALING WITH RARE EVENTS: RESAMPLING-BASED METHODS Chair: M. Ivette Gomes

E306: **Computer-intensive methods in an adaptive estimation of parameters of rare events**
Presenter: **Manuela Neves**, Technical University of Lisbon, Portugal
Co-authors: Ivette Gomes, Fernanda Figueiredo, Dora Prata Gomes

In Extreme Value Analysis, there are a few primordial parameters, among which is the extreme value index, γ, measuring the right tail-weight. Another relevant parameter of extreme events, of real interest for dependent samples (the common situation in practice), is the extremal index, θ, which can roughly be defined as the reciprocal of the expectation of the extremes' duration. A great amount of investigation has been performed on the extreme value index estimation, but much less attention has been given to the estimation of the extremal index. Most of the semi-parametric estimators of these parameters show the same type of behaviour: nice asymptotic properties, but a high variance for small values of k, the number of upper order statistics used in the estimation, a high bias for large values of k, and a real need for the choice of k. After a brief introduction of some estimators of the aforementioned parameters and their asymptotic properties we shall propose, on the basis of bootstrap and jackknife computer-intensive methods, an algorithm for the choice of k and the adaptive estimation of γ and θ. A simulation study as well as some applications to real data will be provided.

E355: **Confidence intervals and hypothesis tests for order statistics of parameters**
Presenter: **Min-ge Xie**, Rutgers University, United States of America

The problem of constructing confidence intervals or hypothesis tests for extrema of parameters, for example of $\max\{\theta_1,\ldots,\theta_k\}$, is considered as one of "the existing problems where standard bootstrap estimators are not consistent and where alternative approaches also face significant challenges." Based on recent developments on confidence distributions, we propose a new resampling method to deal with the inference problem for the extrema of the parameters and also, more generally, for any order statistics of the parameters. This new resampling method can be viewed as an extension of the well-studied and widely-used bootstrap method, but it enjoys a more flexible interpretation and manipulation. We provide a large sample theoretical support for the proposed method. We also explore the theoretical performance of both the standard bootstrap and the proposed method, especially in the presence of ties or near ties among the θ_i's. Empirical performance of the proposed method is studied in numerical examples using both simulations and a real data set.

E486: **Resampling tail estimators and applications**
Presenter: **Margarida Brito**, CMUP & Univ. Porto, Portugal

Tail empirical and quantile processes are widely used in tail nonparametric estimation. The study of the corresponding bootstrapped versions is also of great interest. We provide some limit results for these processes, in analogy with the corresponding asymptotic properties of the tail processes. We shall further be concerned with the estimation of positive tail indices. As it is well known, bootstrap methods can provide accurate approximations to the distribution of pivotal quantities, such as the studentized mean. Since the introduction of the bootstrap, several modifications of the original method have been suggested in the literature for dealing with particular problems. We describe some resampling based procedures, adapted to this context of tail estimation, and investigate the accuracy of the approximation of the distribution of certain normalized tail estimators by means of Edgeworth expansions. One of the important applications of these methods is the construction of confidence bounds. This application will be illustrated with some particular examples, such as the estimation of the ruin probability in risk theory.

E059: **Resampling for endpoint**
Presenter: **Liang Peng**, Georgia Tech, United States of America

It is known that the full sample bootstrap method is inconsistent for approximating the distribution of the maximum, which is used to estimate the endpoint of a distribution function. We show that it works when the endpoint is estimated by Hall's estimator. Further, we propose an empirical likelihood method to construct a confidence interval for the endpoint.

E955: **Bootstrapping with fat-tailed asymmetry**
Presenter: **Adriana Cornea**, Imperial College London, United Kingdom
Co-authors: Karim M. Abadir

We use transformations to achieve approximate tail symmetry of the data before applying the bootstrap for inference on the parameters of interest. We start by illustrating the potential for large accuracy gains with a striking example where asymmetry leads to a failure of the bootstrap. We then propose a general methodology for such transformations.

ES26 Room Gordon MCMC FOR ESTIMATING DIFFUSIONS Chair: Frank van der Meulen

E099: **Irreducible MCMC schemes for diffusions using high frequency imputation**
Presenter: **Andrew Golightly**, Newcastle University, United Kingdom
Co-authors: Darren Wilkinson

Bayesian inference is considered for parameters governing nonlinear multivariate diffusion processes using discretely observed data that may be incomplete and subject to measurement error. Typically unavailable transition densities are replaced with an Euler-Maruyama approximation and a high frequency imputation approach is adopted to allow sufficient accuracy of the approximation. Two MCMC schemes that overcome well known problematic dependence between the parameters and missing data will be examined; one based on a reparameterisation and the other making use of a recently proposed particle MCMC scheme. Both approaches will be implemented using synthetic data generated from a stochastic differential equation description of a Lotka Volterra system.

E898: **Advanced MCMC methods for sampling on diffusion pathspace**
Presenter: **Konstantinos Kalogeropoulos**, London School of Economics and Political Sciences, United Kingdom
Co-authors: Alexandros Beskos, Erik Pazos

Markov chain Monte Carlo (MCMC) methods provide a flexible and powerful tool for estimating diffusion processes. Implementation is performed via data augmentation operating on a discrete skeleton of the diffusion path and the parameters. Most existing approaches implement the diffusion path updates via independence samplers. While this approach works well in some cases, its performance may become unacceptably poor in models with high nonlinearities, distant observations and challenging observation regimes. Moreover, performance may be improved through clever use of the target density gradient. We define and construct advanced MCMC methods, such as the Metropolis-adjusted Langevin algorithm (MALA) and the Hybrid Monte Carlo (HMC), on the Hilbert space of the diffusion paths. Our derivation relaxes previous strict continuity restrictions on the log-likelihood gradient and allows the use of MALA and HMC in non-standard observation regimes arising in real-world applications. We consider cases of observations with error, partial observations (stochastic volatility models, dynamical systems, hypoelliptic diffusions), and observations on diffusion functionals (hitting times in latent diffusion survival models). The algorithms are illustrated on simulated and real datasets where the results indicate that HMC outperforms other algorithms, by avoiding their random-walk-type behaviour, and constitutes a powerful method for tackling high-dimensional path-sampling problems.

E407: **Markov chain Monte Carlo for exact inference for diffusions**
Presenter: **Omiros Papaspiliopoulos**, Universitat Pompeu Fabra, Spain
Co-authors: Giorgos Sermaidis, Roberts Gareth, Alex Beskos, Fernhead Paul

Exact Markov chain Monte Carlo methods are developed for discretely-sampled, directly and indirectly observed diffusions. The qualification "exact" refers to the fact that the invariant and limiting distribution of the Markov chains is the exact posterior distribution of the parameters of interest. The class of processes to which our methods directly apply are those which can be simulated using the most general to date exact simulation algorithm. Various methods are introduced to boost the performance of the basic scheme, including reparametrizations and auxiliary Poisson sampling. We contrast both theoretically and empirically how this new approach compares to irreducible high frequency imputation, which is the state-of-the-art alternative for the class of processes we consider, and we uncover intriguing connections. All methods discussed are tested on typical examples.

E933: **Bayesian inference for a generalized class of Heston models**
Presenter: **Osnat Stramer**, University of Iowa, United States of America
Co-authors: Matthew Bognar

A more general parametric stochastic variance model for asset prices than the stochastic variance model of Heston is defined. Our model is based on a continuous-time version of the smooth transition autoregressive (STAR) models. Our models allow the volatility of the variance process and the leverage effect to increase with the level of the variance. We utilize two relatively new MCMC methods which allow for Bayesian inference in our class of models. The first method relies on straightforward re-parameterization techniques while the second method is based on the pseudo-marginal approach.

E281: **Nonparametric drift estimation for diffusions**
Presenter: **Frank van der Meulen**, Delft University of Technology, Netherlands
Co-authors: Harry van Zanten, Moritz Schauer

A diffusion with known diffusion function and unknown drift function is considered. A nonparametric Bayesian estimation procedure is proposed for the drift function in the setting of low frequency data and no parametric assumption on the drift. A prior for the drift is defined as a series expansion within a hierarchical basis, truncated at a random number of terms. We present a MCMC algorithm by which the posterior drift can be computed using the saturated space approach of Carlin and Chib. The hierarchical basis offers numerical advantages that simplify computations, especially for moves across models, for which we use reversible jump steps.

ES35 Room Woburn LONGITUDINAL DATA ANALYSIS Chair: M. Carmen Pardo

E047: **Antedependence models for longitudinal nonstationary data**
Presenter: **Vicente Nunez-Anton**, Universidad del Pais Vasco (UPV/EHU), Spain

Antedependence models are useful, albeit underutilized, generalizations of well-known stationary autoregressive models for longitudinal data.

Like stationary autoregressive models, antedependence models specify parsimonious parametric forms for the conditional mean and variance of each observation, given all the observations preceding it (as well as any observed covariates) from the same subject. However, antedependence models differ by allowing these parametric forms to change over the course of the longitudinal study. This makes them much more flexible than their stationary autoregressive counterparts and hence, as a result, they are often able to fit longitudinal data exhibiting nonstationary characteristics (e.g., increasing variances or same-lag correlations that change over time) quite well. We motivate these models and show where they sit in the broad spectrum of statistical models used for longitudinal data analysis. We begin by describing some important features common to many longitudinal data sets, especially the tendency for observations from the same subject to be correlated. We then briefly review how various classical methods for continuous longitudinal data analysis and a more modern parametric modeling approach deal with these correlations. Next, antedependence models, as well as other alternative and not commonly used models, are very briefly described. Finally, two real data sets are used to motivate their usefulness.

E137: **Autoregressive models for positive time series**
Presenter: **David Morina**, Universitat Autonoma de Barcelona, Spain
Co-authors: Pedro Puig, Jordi Valero

Let X_t be a stationary positive time series. In many applications it is usual to work with the time series of the logarithms, $Y_t = \log(X_t)$. Suppose that Y_t follows a simple AR(1) model, that is, it can be expressed as $Y_t = \alpha Y_{t-1} + W_t$, where W_t is a white noise with mean μ and variance σ^2 . There are many examples in practice where these assumptions hold very well. According to these assumptions, we shall show that the autocorrelation function (ACF) of X_t characterizes the distribution of W_t. Consequently, the knowledge of the empirical ACF of X_t can help to choose the distribution of the white noise W_t of the log time series Y_t. This result can also be used to construct a goodness of fit test for the classical AR(1) model, where the white noise W_t is normally distributed. Several examples of application will be also shown.

E231: **Modelling and analysis of multivariate ordinal categorical data in longitudinal setup**
Presenter: **Apratim Guha**, University of Birmingham, United Kingdom
Co-authors: Biswas Atanu

In recent years, there has been a tremendous continuing attention towards the modelling and analysis of longitudinal data, although not much attention is paid to the ordinal categorical data. The problem becomes more complicated if the observation at every longitudinal time point is multivariate, and every component of the response vector is ordinal categorical. We develop urn model-type joint probability models with different correlation structures when the data are of a longitudinal bivariate ordinal nature, and discuss the methods of their estimation. Some examples which motivate us to model and analyse such kind of data are also discussed.

E282: **Computational ease on marginalized models for multivariate longitudinal binary data via probit link**
Presenter: **Ozlem Ilk**, Middle East Technical University, Turkey
Co-authors: Ozgur Asar

Longitudinal data are composed of repeated measurements taken from the same subject over time. Generalized linear models with random effects and/or serial dependence are commonly used to analyze this type of data. However, interpretation and computation of marginal covariate effects can be difficult in such models. We propose a multi-level model for multivariate longitudinal binary data to have simplified and more intuitive interpretations. This model consists of a triple of regression models, which is permitting subject-specific inferences and dependence across time, while modeling the marginal mean responses. First proposed in 2007, this model with logit links is computationally intensive. This is leading to a limitation on its use. We propose the use of probit links to ease the computational burden. Model with both links are discussed. Markov Chain Monte Carlo Methods are used for the logit link, while maximum likelihood estimation is considered for the probit link. Although it requires the numerical integration of random effects, it is observed that this new computational algorithm is much faster. Simulation studies are carried out on the new algorithm. Moreover, the methods are illustrated on a real life dataset.

E496: **Issues in the analysis of longitudinal data with dependent observation schemes**
Presenter: **Richard Cook**, University of Waterloo, Canada
Co-authors: Meaghan Cuerden, Cecilia Cotton

There has been considerable attention given in recent years to the statistical issues related to incomplete longitudinal data due to drop-out. It is well known that when data are missing completely at random likelihood and estimating function approaches yield consistent estimators but approaches based on estimating functions yield inconsistent estimators when data are missing at random. In some settings, however, the amount of data available is dictated not by drop-outs but by a stochastic mechanism related to the response process. We consider such a problem in the context of platelet transfusion trials where the aim is to assess the effectiveness of different platelet transfusion products and responses are obtained at each transfusion. In this setting different patients will require a different number of transfusions and the need for a transfusion at any given time is influenced by the responses to previous transfusions. Likelihood and inverse probability weighted estimating function approaches will be explored in this setting, and an application to a recent platelet transfusion trial will be given for illustration.

ES40 Room Senate BAYESIAN NONPARAMETRICS MODELLING Chair: Jim Griffin

E134: **Nonparametric mixture modeling for Bayesian analysis of dose-response studies**
Presenter: **Athanasios Kottas**, University of California, Santa Cruz, United States of America
Co-authors: Kassandra Fronczyk

A Bayesian nonparametric modeling framework for replicated count responses in dose-response experiments is presented. The main focus will be on modeling and risk assessment in developmental toxicity studies, where the primary objective is to determine the relationship between the level of exposure to a toxic chemical and the probability of a physiological or biochemical response. Data from these experiments comprise clustered categorical responses and typically involve features that can not be captured by standard parametric approaches. To provide flexibility in the functional form of both the response distribution and the dose-response curves, the proposed mixture models are built from dependent Dirichlet process priors, with the dependence of the mixing distributions governed by the dose level. The practical utility of the methodology is illustrated with data from a toxicity experiment for which the dose-response curves for different endpoints have different shapes, including a non-monotonic, possibly hormetic, dose-response relationship. As time permits, we will discuss a more structured nonparametric mixture model for traditional bioassay experiments, including quantal responses, and the more challenging setting with an ordinal response classification.

E113: **Repulsive mixtures**
Presenter: **Francesca Petralia**, Duke University, United States of America
Co-authors: David Dunson

Mixture models have become extremely used for density estimation, clustering and as a component in flexible hierarchical models. In using mixture models for clustering, identifiability problems arise if mixture components are not sufficiently well separated and the data for the different sub-

populations contain substantial overlap. Insufficiently separated components also create problems in using mixture models for density estimation and robust modeling, as redundant components that are located close together can be introduced leading to an unnecessarily complex model as well as to various computational problems. Current practice in Bayesian mixture modeling generates the component-specific parameters from a common prior, which tends to favor components that are close together. As an alternative, we propose to generate mixture components from a repulsive process that favors placing components further apart. Efficient algorithms are proposed for fully Bayes and fast approximate posterior computation allowing uncertainty in the number of components. The methods are illustrated using simulated data and several applications.

E600: **On the stick-breaking representation for Gibbs-type priors**
Presenter: **Stefano Favaro**, University of Turin and Collegio Carlo Alberto, Italy
Co-authors: Antonio Lijoi, Igor Pruenster

Random probability measures are the main tool for Bayesian nonparametric inference, given their law acts as a prior distribution. Many well-known priors used in practice admit different, though (in distribution) equivalent, representations. Some of these are convenient if one wishes to thoroughly analyze the theoretical properties of the priors being used, others are more useful in terms of modeling and computation. As for the latter purpose, the so-called stick-breaking constructions certainly stand out. We focus on the recently introduced class of Gibbs-type priors and provide a stick-breaking representation for it.

E265: **Nonparametric stick breaking priors with simple weights**
Presenter: **Ramses Mena**, IIMAS-UNAM, Mexico
Co-authors: Ruth Fuentes-Garcia, Matteo Ruggiero, Stephen G. Walker

A relatively simple nonparametric prior is presented and explored within exchangeable and non-exchangeable settings. Some properties and estimations results within mixture, regression and measure-valued diffusion processes are discussed. If time allows some results based on the corresponding exchangeable partition probability functions will also be discussed.

E962: **A predictive study of Bayesian nonparametric regression models**
Presenter: **Sonia Petrone**, Bocconi University, Italy
Co-authors: Sara Wade, Stephen Walker

Dirichlet process (DP) mixture models are a popular tool for constructing Bayesian nonparametric regression models. They can allow for flexible modeling of the distribution of the response given the covariate by partitioning subjects into groups. Our aim is to examine the predictive performance of these models and ways to improve it. For prediction, the random partition plays a crucial role, and in regression settings, it is often reasonable to assume that this partition depends on the proximity of the covariates. We find that DP-based mixture models which do not incorporate this knowledge can preform quite poorly. And, while DP-based mixture models which encourage these desirable partitions perform remarkably better, the posterior mass of undesirable partitions may still be relatively high. Therefore, we propose to modify the prior distribution of the random partition by setting the probability of undesirable partitions to zero, while still maintaining certain properties of the DP. This allows the distribution of the partition to depend on the covariate in a simple manner and greatly reduces the total number of possible partitions, resulting in improved prediction. Numerical illustrations will be presented.

ES46 Room Bedford RECENT ADVANCES IN MULTI-STATE MODELS Chair: Jacobo de Una-Alvarez

E300: **Nonparametric methods for testing Markov condition in multi-state models**
Presenter: **Mar Rodriguez-Girondo**, University of Vigo, Spain
Co-authors: Jacobo de Una-Alvarez

Markov multi-state models are often used for studying medical processes that can be split in several transient states. A process is Markov if, conditionally to the present, its future evolution is independent of the past. It is important to test the validity of this assumption to avoid biased inferences. Traditionally, Markov condition is checked via a proportional hazards specification. However, this may not properly represent the dependence structure between the survival prognosis at the present time and the past. We provide new nonparametric methods for checking the Markov condition in the three-state progressive model. This model allows for three states (1, 2, 3) and two possible transitions ($1\rightarrow 2$ and $2\rightarrow 3$) in a progressive way. Two different tests are presented: (1) A local test based on future-past association at each present time t measured through the Kendall's Tau. (2) A global test based on the comparison between the Markovian and non-Markovian nonparametric estimators of the joint distribution of the successive transition times. In both approaches we take censoring into account. Bootstrap techniques are used to approximate the null distribution of the test statistics. Extensions to other multi-state models as the illness-death model are discussed.

E216: **Presmoothing the Aalen-Johansen estimator in an illness-death model**
Presenter: **Luis Machado**, University of Minho, Portugal
Co-authors: Ana Moreira, Jacobo de Una-Alvarez

In many medical studies, patients may experience several events through a follow-up period. The analysis in such studies is often performed using multi-state models. One important topic is the estimation of the transition probabilities, since it allows for long-term predictions of the process. The usual nonparametric estimator of the transition matrix for non-homogeneous Markov processes is the so-called Aalen-Johansen estimator. We propose a modification of the Aalen-Johansen estimator in the illness-death model based on presmoothing. Simulations show that the presmoothed estimators may be much more efficient than the Aalen-Johansen estimator. An illustration through real analysis is included.

E237: **Nonparametric regression for sojourn time distributions in a multistate model**
Presenter: **Somnath Datta**, University of Louisville , United States of America

Multistate models are generalizations of traditional survival data where an individual undergoes different types of events corresponding to transitions to various states of a system. Multistate event data that are right censored are considered. Under this setup, inferring on the state waiting (or sojourn) time distribution corresponding to a give transient state j is problematic, since neither the entry nor the exit times are fully observed. We introduce novel procedures to test the effect of a categorical covariate on the sojourn time distribution. We also introduce a Aalen type linear hazard model for the state waiting time distribution that can incorporate both discrete and continuous covariates. The methods are illustrated using a number of real data applications.

E238: **Nonparametric regression using partial least squares dimension reduction in multistate models**
Presenter: **Susmita Datta**, University of Louisville, United States of America

A method of constructing nonparametric regression estimators of state occupation probabilities in a mulltistate model is introduced. In order to tackle potentially a large number of predictors in modern genomic and proteomic data sets we use partial least squares to compute estimated latent factors from the transition times along with the covariates which are then used in an additive model in order to avoid the curse of dimensionality. We illustrate the methodology using a data set on non-small cell carcinoma and study the performance of the estimators through simulations.

E123: **Comparison of prediction models for competing risks with time-dependent covariates**
Presenter: **Per Kragh Andersen**, University of Copenhagen, Denmark
Co-authors: Giuliana Cortese, Thomas Gerds

Prediction models for competing risks cumulative incidence in the presence of an internal time-dependent covariate are discussed. Results from a bone marrow transplant study are presented. Cumulative risks of relapse and death over time are predicted, taking into account the effect of Graft versus Host Disease (GvHD), which patients may develop during their follow-up. Predictions are based on two regression approaches: a multi-state model with an intermediate state that represents the condition 'alive with GvHD' and the landmark approach which consists of a sequence of competing risks analyses at different landmark time points s, where GvHD(s) is included as a time-constant covariate. The landmark approach only uses the GvHD status recorded at each landmark point s, whereas its development after s is not taken into account, and cumulative risks are estimated either based on a Fine-Gray model or by plugging-in models for the cause-specific hazards. Prediction errors of cumulative risks over different sub-intervals $[s,t]$ of the follow-up period are compared. For computing prediction errors, we consider the expected Brier score and, under a covariate-independent censoring scheme, we estimate it by means of pseudovalues for subjects at risk at each time s.

ES49 Room Torrington NEW DEVELOPMENTS IN QUANTILE REGRESSION Chair: Stanislav Volgushev

E053: **Conditional quantile processes based on series or many regressors**
Presenter: **Ivan Fernandez-Val**, Boston University, United States of America
Co-authors: Alex Belloni, Victor Chernozhukov

The nonparametric quantile regression (QR) series framework is developed, covering many regressors as a special case, for performing inference on the entire conditional quantile function and its linear functionals. The entire conditional quantile function is approximated by a linear combination of series terms with quantile-specific coefficients and estimates of the function-valued coefficients from the data. Large-sample theory is developed for the empirical QR coefficient process, namely we obtain uniform strong approximations to this process by conditionally pivotal and Gaussian processes, as well as by gradient and weighted bootstrap processes. These results are applied to derive estimation and inference results for linear functionals of the conditional quantile function, such as the conditional quantile function itself, partial derivatives, average partial derivatives, and conditional average partial derivatives. Specifically, we obtain uniform rates of convergence, large sample distributions, and inference methods based on strong pivotal and Gaussian approximations and on gradient and weighted bootstraps. All of the above results are for function-valued parameters covering the pointwise results as a by-product. The results are illustrated with an empirical example on estimation of the price elasticity function of the individual demand for gasoline, as indexed by the individual unobserved propensity for gasoline consumption.

E065: **A quantile-based approach to spectral analysis of time series**
Presenter: **Tobias Kley**, Ruhr-Universitat Bochum, Germany
Co-authors: Holger Dette, Marc Hallin

An important tool in the analysis of hidden periodicities of time series is the periodogram. It is commonly used to estimate the spectral density of a weaksense stationary time series. Yet, there are drawbacks to this approach. Due to the utilization of the autocovariances, only linear dependencies are accounted for. Moreover, the existence of second moments has to be assumed and the estimators are nonrobust regarding impulsive noise and monotone transformations. Therefore, this approach can be considered quite restrictive. A quantile-based measure of autodependence to replace the autocovariances and avoid its drawbacks is introduced. Quantile regression is used to estimate the spectral representation of the generalized measure of dependence and the estimator's asymptotic properties are analyzed to legitimize it as a statistical measure in time series analysis. To study the finite sample properties, the estimator is then applied to example data sets.

E181: **Two-step regression quantiles: Advantages and applications**
Presenter: **Jana Jureckova**, Charles University in Prague, Czech Republic

Two-step α-regression quantile (RQ) in the linear regression model $Y_i = \beta_0 + \mathbf{x}_i^\top \beta + e_i$, $i = 1,\ldots,n$ first estimates the slope components β with a suitable (rank) R-estimate $\widehat{\beta}_R(\alpha)$ and then determines the $[n\alpha]$-quantile $\hat{e}_{[n\alpha]}$ of the residuals $\{e_i - \mathbf{x}_i^\top \widehat{\beta}_R(\alpha)\}$, $i = 1,\ldots,n$. It is asymptotically equivalent to the RQ of Koenker and Bassett, very close numerically even for small sample sizes, and the extreme regression quantiles of both types coincide exactly. Because $\widehat{\beta}_R(\alpha)$ is invariant to the intercept β_0, it plays only auxiliary role in estimating $F^{-1}(\alpha)$ and thus the quantile inference is based mainly on $\hat{e}_{[n\alpha]}$, $0 < \alpha < 1$. The possible inference based on two-step RQ will be described, including, for example, the confidence sets, testing, estimating the sparsity function and the Pareto index of the distribution.

E345: **Quantile regression in nonparametric location-scale models with censored data**
Presenter: **Cedric Heuchenne**, University of Liege, Belgium
Co-authors: Ingrid Van Keilegom

Consider the random vector (X,Y), where X is completely observed and Y is subject to random right censoring. It is well known that the completely nonparametric kernel estimator of the conditional distribution $F(\cdot|x)$ of Y given $X = x$ suffers from inconsistency problems in the right tail, and hence any quantile function $m(x) = F^{-1}(s|x) = \inf\{y : F(y|x) \geq s\}$ $(0 \leq s \leq 1)$ that involves the right tail of $F(\cdot|x)$ cannot be estimated consistently in a completely nonparametric way. An alternative estimator of $m(x)$ that, under certain conditions, does not share the above inconsistency problems is proposed. The estimator is constructed under the model $Y = m(X) + \sigma(X)\varepsilon$, where $\sigma(\cdot)$ is an unknown scale function and ε (with zero quantile and scale one) is independent of X. We obtain the asymptotic properties of the proposed estimator of $m(x)$, we compare it with the completely nonparametric estimator via simulations and apply it to a study of quasars in astronomy.

E392: **Multiple imputation in quantile regression**
Presenter: **Ying Wei**, Columbia University, United States of America
Co-authors: Yanyuan Ma, Raymond Carroll

A multiple imputation estimator for parameter estimation in a quantile regression model is proposed when some covariates are missing at random. The estimation procedure fully utilizes the entire data set to achieve increased efficiency, and the resulting coefficient estimators are root-n consistent and asymptotically normal. To protect against possible model misspecification, we further propose a shrinkage estimator, which automatically adjusts for possible bias. The finite sample performance of our estimator is investigated in a simulation study. Finally, we apply our methodology to part of the Eating at American's Table Study data, investigating the association between two measures of dietary intakes.

ES56 Room Court APPROACHES TO THE TREATMENT OF IMPRECISION OF STATISTICAL DATA Chair: Renato Coppi

E309: **The ontological and epistemic views of fuzzy data in the statistical reasoning process**
Presenter: **Renato Coppi**, Sapienza University of Rome, Italy

The problem concerning the nature of fuzzy data as used in statistical analysis is framed within the Informational Paradigm, according to which Empirical and Theoretical Information constitute the input of the Statistical Reasoning Process leading to Additional Information as a result of the analysis. Various forms of Uncertainty affect this process and propagate through it. Among them the imprecision of statistical data, commonly modeled by using fuzzy data. With reference to this source of Uncertainty, it is argued that the "ontological" or "epistemic" nature of fuzziness, in the above context, is itself a piece of theoretical information which should be explicitly considered when declaring the ingredients of the specific reasoning process under investigation. This implies, on one side, the utilization of the appropriate mathematical machinery (e.g. definition of the arithmetic operations and of the statistical notions of mean, variance, etc.) for handling the fuzziness of the data. On the other side, it involves the consideration of the impact of the specified nature of data imprecision (including total imprecision stemming from missing values) on the other pieces of theoretical information used in the reasoning process. Illustrations of the above concepts are provided with reference to classification and regression methods dealing with imprecise data.

E361: **Making sense of set-valued data: Ontic vs. epistemic representations**
Presenter: **Didier Dubois**, CNRS-IRIT, France
Co-authors: Eyke Huellermeier

Sets and likewise fuzzy sets may have a conjunctive or a disjunctive reading. In the conjunctive reading a (fuzzy) set represents an object of interest for which a (gradual) description in terms of a collection of simpler entities makes sense. In contrast disjunctive (fuzzy) sets refer to the representation of incomplete knowledge. They do not model real objects or intervals but partial information about an underlying precise object or quantity. In this case the set captures uncertainty, and its membership function is a possibility distribution. We call such sets epistemic. Epistemic uncertainty is the realm of possibility theory, with applications such as computing with fuzzy intervals, or imprecise regression and kriging. We try to explain the consequence of the distinction between ontic and epistemic representations on the handling of (fuzzy) set-valued data. In a probabilistic framework, this is also reflected by the distinction between random sets and belief functions. By fault of making this distinction, there is a risk of misunderstanding basic notions and tools, such as distance between fuzzy sets, variance of a fuzzy random variable, fuzzy regression, fuzzy models, fuzzy equations, conditioning fuzzy information, etc. We discuss several examples where the ontic and epistemic points of view yield different methods.

E468: **Ternary classification trees for imprecise data**
Presenter: **Roberta Siciliano**, Universita di Napoli Federico II, Italy
Co-authors: Massimo Aria, Valentina Cozza, Antonio D'Ambrosio

The framework is the statistical learning theory of Vapnik, i.e. learn from the experience (training sample) to generalize and provide useful answers (prediction, decision) in new cases. Classification trees will be considered as supervisor, consisting in a recursive partitioning of the predictor space (input) to induce a partitioning of the sample of cases into disjoint subgroups which are internally homogeneous and externally heterogeneous with respect to a categorical (often dummy) response variable (output). Predictors are usually of numerical or categorical type, with punctual measurements. It is provided a supervised classification tree-based methodology to deal with imprecise data; specifically predictors' measurements can be provided by a functional distribution or an interval of values. The proposed recursive ternary partitioning algorithm discriminates in a better way the ordering relationships and the data imprecision. Typical data structures of this type occur in many real life applications, where training data comes with intrinsic uncertainty that might be the result of imprecise measuring instruments such as in image recognition (in medicine, physics, robotics, etc.) or human judgments/observations in socio-economic fields. As a result, the proposed approach can be understood as a "subjectivist" view of imprecision formalizing the uncertainty concerning an underlying "crisp" phenomenon.

E477: **On the notion of disambiguation in learning from imprecise data**
Presenter: **Eyke Huellermeier**, University of Marburg, Germany
Co-authors: Didier Dubois

In recent years, an increasing number of publications has been devoted to the learning of models from imprecise data, such as interval data or, more generally, data modeled in terms of fuzzy subsets of an underlying reference space. Needless to say, this idea also requires the extension of corresponding learning algorithms. Unfortunately, this is often done without clarifying the actual meaning of an interval or fuzzy observation, and the interpretation of membership functions. We shall first argue that different interpretations call for different types of extensions of existing learning algorithms and methods for data analysis. In this regard, we specifically distinguish between an "ontic" and an "epistemic" interpretation of (fuzzy) set-valued data. Then, focusing on the epistemic view, we argue that, in model induction from imprecise data, one should not try to find a model that reproduces the data, but instead "disambiguates" the data. More specifically, this leads to a learning procedure that performs model identification and data disambiguation simultaneously. This idea is illustrated by means of two concrete problems, namely regression analysis with fuzzy data and classifier learning from ambiguously labeled instances.

E784: **Comparing Likert and fuzzy scales through some statistical tools**
Presenter: **Maria Angeles Gil**, University of Oviedo, Spain
Co-authors: Sara de la Rosa de Saa, Maria Teresa Lopez, Maria Asuncion Lubiano

Likert scales are widely employed in opinion, valuation, rating,... questionnaires which are usually associated with social sciences, medical diagnosis, etc. They often concern questionnaires with a pre-specified response format. For its statistical data analysis each categorical response is usually coded by an integer number, and statistical analysis is rather limited. The scale of fuzzy numbers can be used alternatively in opinion, valuation, rating,... questionnaires, to "express" the responses. Questionnaires could be designed with a free fuzzy response format and there are at present many available techniques for its statistical analysis. Intuitively, the fuzzy scale along with the free response format and available statistical developments allows us to exploit more information than Likert scales. An introductory comparative empirical analysis is to be presented, leading to conclude about the advantages of using the scale of fuzzy numbers versus Likert scales from a statistical point of view. For this purpose several indicators concerning diversity, variability and representativeness of certain summary measures have been considered.

ES72 Room S264 NETWORKING ON BIOSTATISTICS: THE BIOSTATNET PROJECT III Chair: Guadalupe Gomez

E323: **The use of the Youden index in diagnostic studies**
Presenter: **Elisa Maria Molanes Lopez**, Universidad Carlos III de Madrid, Spain
Co-authors: Emilio Leton

The effectiveness of a continuous diagnostic test or biomarker, for classifying a disease status, is usually characterized using the receiver operating

characteristic (ROC) curve. The ROC curve describes graphically the performance of the biomarker under several cut-off points. A key point in this methodology is to find an optimal threshold, in order to maximize the accuracy of the biomarker. There are several methods for doing this: the Northwest corner, the symmetry point, the Youden index, etc. We will concentrate on the latter one, studied recently by several authors. Firstly, it is shown that an empirical likelihood based approach for estimating the Youden index, J, and its associated threshold, c_J, is competitive with the parametric approach based on the first and second order delta method. Secondly, it is shown how to combine multiple biomarkers using an estimate of a reparametrization of their likelihood ratio function via copula function estimates. Using this new biomarker, we can use the approach above-described to estimate J and c_J. Thirdly, a real example, well known in the literature, is analyzed in order to illustrate the described techniques.

E339: **Designs for partial likelihood in survival analysis**
Presenter: **Maria Jesus Rivas-Lopez**, University of Salamanca, Spain
Co-authors: Jesus Lopez-Fidalgo

In survival analysis, optimal design appears, for example, in allocation of patients to one of two treatments under comparison. A proportional hazards model is often used to explore a possible effect of a covariate on the survival time. The main interest is usually to estimate the regression coefficients, leaving unspecified the baseline hazard function. We will study the case of an experiment to be carried out during the interval time [0,1], with all individuals starting the experiment at time 0 and time-to-event outcomes. Some observations may be censored, either from withdrawal censoring during the follow-up, or right censoring at the end of the follow-up. Withdrawal censoring and failure time are assumed to be independent and with known distributions. Conditional density functions of an observed time, given that it is a failure or a withdrawn time, are found, as well as probabilities of having a failure, a withdrawn and a right censored time. The information matrix for this model can be computed, but examples showing that it is not always adequate to work with it, lead us to the use of the partial information matrix, based on the partial likelihood widely used in the estimation of the Cox model. An illustrative example allows us to compare optimal designs for both types of likelihood.

E906: **D-optimal factorial designs for Poisson models in the context of toxicity studies**
Presenter: **Roberto Dorta Guerra**, Universidad de La Laguna, Spain
Co-authors: Enrique Gonzalez Davila, Josep Ginebra

D-optimal experiments for Poisson response data have been studied in recent years. On the other hand two-level factorials designs are much used as screening designs at the preliminary stages of an investigation when the outcome is continuous. We study the D-optimal factorial designs for a one-factor second-order model and two-factor interaction model. We compare the designs obtained with the D-optimal designs, studying its D-efficiency as well as the advantages and disadvantages that arise in the context of toxicity studies. It is shown that conventional wisdom about this kind of experiments does not apply when the response is Poisson.

E914: **Durability in building maintenance**
Presenter: **Carles Serrat**, Universitat Politecnica de Catalunya-BarcelonaTECH, Spain

Decisions about intervention in existing buildings are generally based on information coming from inspections, as a systematic tool for the identification of some injury in buildings. In this sense, in order to carry out an efficient preventive task and maintenance, knowledge of the evolution of injuries and their distribution are essential. However, this information, unfortunately, does not exist and there are few studies that describe the lifecycle of constructive elements in play; so we must use durability estimators based on inspections. The first goal of this paper is to introduce the time to event approach methodology for the estimation of the durability in buildings. The main problem of this methodology is the high variability of the resulting estimator, due to the censorship mechanism in the inspection process. So, the second goal of is to present a simulation study that aims to analyze the accuracy of the methodology and allows the design of an efficient inspection plan, in the sense of reducing the negative impact of censorship in the precision of the estimators.

E886: **Transformed Gaussian model for joint modelling of longitudinal measurements and time-to-event in R**
Presenter: **Ines Sousa**, Minho University - Portugal, Portugal

We propose a model for the joint distribution of a longitudinal variable and a single time-to-event. This model, which we will call joint transformed Gaussian model, assumes a multivariate Gaussian distribution for the vector of repeated measurements expanded with the natural logarithm of failure time. The marginal distribution of time-to-event is log-Normal distributed, as well as the conditional distribution of the repeated measurements. The variance covariance structure of this model is of main interest, so we consider different parametric structures for the covariance matrix of the proposed model. We will present an analysis of a data set as an example. This model is being proposed under a set of functions for R, to be submitted to CRAN. However, in this talk we will refer to better approaches to make the already existing functions more efficient.

EP03 Room Chancellor's POSTERS SESSION III Chair: Cristian Gatu

E842: **An optimal design to improve control charts for correlated observations**
Presenter: **Ioulia Papageorgiou**, Athens University of Economics and Business, Greece

Control charts are widely used in monitoring a manufacturing process. The design of the control charts involves the sample, the sampling interval and the control limits construction. The basic assumption that usually underpins all three stages is that the measurements sampled from the process are uncorrelated. Previous research work in the area lies in two axes. Either study how the presence of correlation affects the control chart limits or how one can test and avoid problems in reliability of the limits and the probability errors caused by the existence of correlation within the sample. Theoretical new limits when a fixed positive correlation is assumed are provided. We assume a positive decreasing type of correlation, a quite realistic model for correlation among the sampled units and make use of results on optimal sampling strategies when sampling from a correlated population. We propose the best unbiased estimate that is different than the sample mean and attempt to update the control chart limits and construct them in a sense that accommodates the specific type of correlation. A comparison between equal and variable sampling interval is conducted.

E862: **Adaptation to the survey ICT-H in Canary Islands of the dual frame methodology**
Presenter: **Enrique Gonzalez-Davila**, Universidad de La Laguna, Spain
Co-authors: Alberto Gonzalez-Yanes

The combination of information obtained in person and by telephone of a survey usually includes the definition of two scenarios, one associated with the census or population register, and another to phonebook. The application of dual frame methodology to surveys of availability or use of new technologies in households where there are target variables that can match, have different degrees of association, or be independent with the scenarios defined, are of particular interest in this work. This methodology is adapted to the particular case of the ICT-H survey conducted in the Canary Islands, where one of the scenarios is contained entirely in the other. Dual frame estimator used is the pseudo maximum likelihood estimator introduced by Skinner and Rao. The results are compared with those offered by the direct estimator, and the estimates given by the ICT-H survey conducted by the Spanish National Statistics Institute for the Canaries. Additionally, a simulation study of such survey on an artificial population similar to the actual population, allows us to evaluate the efficiency of its application.

E928: **Applying multivariate beta models and mixed-multivariate log-linear models to discrete test score distributions.**
Presenter: **Tomoya Okubo**, The National Center for University Entrance Examinations, Japan

Results of fitting mixed multivariate beta models and multivariate log-linear models to large-scale test score data are shown. Generally, the scores of test are distributed from 0 to full score. When we analyze the distribution of the scores, we often apply a probability distribution to the test scores to understand the distribution of the scores. Beta distribution is considered as an appropriate probability distribution for such a fitting because beta distribution ranges from 0 to 1. It is suitable to apply beta distribution to discrete test scored data because such data are usually distributed from 0 to full score. Dirichlet distribution is considered as a form of multi-dimensional beta distribution; however, it cannot calculate the correlation coefficient between two variables because it has an open standard (N-1)-simplex structure. Therefore, we will apply a multi-dimensional extension of the beta distribution that is different from the Dirichlet distribution. Moreover, multivariate log-linear models are applied for the discrete test score distributions to compare features of the models.

E869: **Pairwise likelihood based robust estimation of multivariate location and covariance**
Presenter: **Nicola Lunardon**, University of Padova, Italy
Co-authors: Luca Greco, Laura Ventura

Pairwise likelihood may be a useful tool for approximating likelihood based inference in the estimation of multivariate location and covariance matrices.The maximum pairwise likelihood estimator is sensitive to the occurrence of outliers, as well as its likelihood counterpart. As the minimum covariance determinant estimator is an effective tool in providing robust inference at general multivariate elliptically symmetric distributions, the employ of pairwise likelihood equations looks challenging. The validity of pairwise based MCD estimation procedures is investigated in the context of mixed linear models and first order autoregression.

E891: **The potentialities of Chinese airline market for Lisbon international airport: the empirical modelling analysis**
Presenter: **Jose Vicente**, University of Evora, Portugal
Co-authors: Andreia Dionisio, Manuela Oliveira

A comparative analysis between an empirical gravity model and a dynamic regression model with the objective to explain the potentialities of the Chinese airline market for Lisbon international airport (air passengers demand) is carried out. We confirm the viability to create some direct flights from Lisbon international airport to the Chinese airline market, with the strategy to attract the transfer passengers flow with origin in Latin America. Panel dada is used to determine the influence of the explanatory variables, on average number of passengers (air passengers demand). The results are creating by the Stata final outputs. We also demonstrate that the dynamic regression model used is more robust and better than the empirical gravity model, often considered as a reference method in the field of aviation. The most relevant variables on the dynamic regression model are PPP (gross national income (GNI) converted to international dollars using purchasing power parity rates), business and trade factor, and tourism and cultural factor. Furthermore, we find some possible explanations for the results.

E935: **On the linear combination of independent Gumbel random variables**
Presenter: **Filipe J. Marques**, The New University of Lisbon, Portugal
Co-authors: Carlos A. Coelho

The exact distribution of the linear combination of p independent Gumbel random variables is obtained as the sum of two independent random variables, the first one corresponding to the sum of a given number of independent log Gamma random variables multiplied by a parameter and the second one corresponding to a shifted Generalized Integer Gamma distribution. Given the complexity of this exact distribution, a near-exact distribution, with a flexible parameter, γ, is developed, based on the exact c.f. (characteristic function). This near-exact distribution corresponds to a shifted Generalized Near-Integer Gamma distribution. The expression of its density and cumulative distribution functions does not involve any unsolved integrals, thus allowing for an easy computation of values for the density and cumulative distribution functions with high precision. Numerical studies conducted to assess the quality of this approximation show its very good performance. Furthermore, the expressions obtained greatly surpass in manageability the existing ones.

E807: **Nonparametric bootstrap inference for quantiles and its application to the extremes**
Presenter: **Ahmet Sezer**, Anadolu University, Turkey
Co-authors: Betul Kan, Berna Yazici

In many fields of modern science, engineering and insurance, extreme value theory is well conducted. The essential question is that if things go wrong, how wrong can they go? To aswer this question we need to estimate and assess the statistical models for tail related measures. Examination of high level of quantiles allows us to measure those kind of riks. We emphasize using nonparametric bootstrap inference since it is able to estimate a measure of variability and bias. The nonparametric bootstrap is compared with the traditional parametric bootstrap methods in estimating the quantiles at different levels, especially for high quantiles We discuss and show the advantage of nonparametric bootstrap over the parametric bootstrap. Comparison of powers conducted by the simulation study for heavy tailed distribution such as generalized Pareto and Weibull distributions.

E948: **New copulas obtained by maximizing Rényi entropis and their use for multivariate data analysis**
Presenter: **Ali Mohammad-Djafari**, CNRS-Supelec-Univ Paris Sud, France
Co-authors: Doriano-Boris Pougaza

Determining a joint probability distribution function (pdf) from its marginals is an ill-posed inverse problem in the sense that the solution is not unique. Indeed, the expression of the joint pdf is equal to the product of its marginals and an extra terme which is a copula pdf. We propose to use entropy as a measure of selection. Looking for maximum entropy joint pdf given its marginals, we obtained a new way for constructing copulas, and using the Rényi or Tsallis-Havrda-Charvát entropies we obtained new families of copulas. We present this constructive way of creating new copula families and then we will see how we can use these copulas in multivariate data dimensionality reduction and components analysis.

E957: **Selection of interactions with Bayesian logic regression methods**
Presenter: **Magdalena Malina**, Wroclaw University, Poland
Co-authors: Sebastian Tworek, Malgorzata Bogdan

We consider a Bayesian version of logic regression method, based on Markov Chain Monte Carlo searching algorithm, (MCLR), in which a model selection is made by maximization of the likelihood function. We propose a similar approach, taking the prior probability for a number of leaves into account. As in MCLR the searching space is a space of possible models. We propose an intuitive importance measure for interactions, which is a posterior probability of appearing of the interaction in a model. A posterior probability of a model is delivered based on an asymptotic estimation provided by BIC and additionally taking into account the geometric prior for a number of leaves in a model. Basing on these probabilities we determine posterior probabilities for various interactions. We shall present results of a simulation study and a real data analysis for the modified method.

Sunday 18.12.2011 16:50 - 18:30 Parallel Session K – CFE

CS01 Room Bloomsbury MODELLING MULTIVARIATE FINANCIAL TIME SERIES Chair: Alessandra Amendola

C107: **Smooth filtering and likelihood inference in dynamic latent variables models**
Presenter: **Christian Brownlees**, UPF, Spain
Co-authors: Dennis Kristensen, Yongseok Shin

A smooth particle filter is proposed that resolves some of the issues that standard filters involve. In particular, the proposed filter is smooth in the parameter values and so leads to a simulated likelihood function which is smooth in the parameters. Furthermore, the filter does in general not degenerate and so is able to handle slowly varying state variables. We give rate results for the smooth filter and establish the asymptotic properties of the resulting simulated estimators - both frequentist and Bayesian ones. The usefulness of our proposed method is demonstrated with Monte Carlo studies, and application to estimation of continuous-time stochastic volatility models given first-step estimates of integrated volatility.

C164: **Financial density selection**
Presenter: **Genaro Sucarrat**, BI Norwegian School of Management, Norway
Co-authors: J. Miguel Marin

Simple but flexible methods are proposed for density selection of skewed versions of the two most popular density classes in finance, the exponential power (EP) distribution and the Student's t (ST) distribution. For the first type of method, which simply consists of selecting a density by means of an information criterion, the Schwarz criterion stands out, since it performs well across density categories, and in particular when the Data Generating Process (DGP) is normal. For the second type of method, General-to-Specific (GETS) density selection, the simulations suggest that it can improve the recovery rate in predictable ways by changing the significance level. This is useful because it enables us to increase (reduce) the recovery rate of nonnormal densities by increasing (reducing) the significance level, if one wishes to do so. Finally, the methods are illustrated in an empirical application.

C199: **Modelling predictive asymmetry in multivariate financial time series**
Presenter: **Nicola Loperfido**, Universite degli Studi di Urbino "Carlo Bo", Italy
Co-authors: Cinzia Franceschini

In many financial time series, present returns are negatively correlated with future volatility. This stylized fact, known as predictive asymmetry, may be generalized to the multivariate case by considering the covariances between vectors of present returns and products of future ones. Both graphical methods and testing procedures show that the multivariate SGARCH model satisfactorily describes their structure. Theoretical results will be illustrated by multivariate financial data from several European countries.

C540: **Fast indirect estimation of latent factor models with conditional heteroskedasticity**
Presenter: **Giorgio Calzolari**, Universita' di Firenze, Italy
Co-authors: Gian Piero Aielli, Gabriele Fiorentini

A large latent factor model, in which the volatilities of common and idiosyncratic factors are conditionally heteroskedastic, is considered. When the conditional variances of the latent variables are measurable functions of the econometrician's information set (i.e. the observed series), maximum likelihood estimation is straightforward. This is no longer true when conditional variances depend on past values of the common or specific unobservable factors. Some complex indirect inference procedures available in the literature calibrate the score of a Kalman filter approximation with inequality constraints on the auxiliary model parameters. We investigate the performance of computationally simple indirect estimators based on auxiliary models that do not require the Kalman filter implementation. In particular, we only need to deal sequentially with univariate linear regressions and univariate GARCH-type models. We fully discuss computational details of our method and we report the results of an extensive Monte Carlo experiment with empirically realistic designs to evaluate the performance of our proposal. Finally, we present an application on a moderately large number of weekly and daily returns from the stock market.

CS24 Room Woburn BAYESIAN EMPIRICAL MACROECONOMICS Chair: Gary Koop

C453: **Robust modeling of IPO market cycles using a regime switching model with an unknown number of regimes**
Presenter: **Markus Jochmann**, Newcastle University, United Kingdom
Co-authors: Fabrizio Casalin

IPO cycles in the US stock and bond markets are analyzed using an infinite hidden Markov model (IHMM). The IHMM is a Bayesian nonparametric approach to modeling structural breaks. It allows for an unknown number of breakpoints and is a flexible and attractive alternative to existing methods that fix the number of regimes a priori. Thus, we are able to expand the existing literature on IPO market cycles which normally assumes two regimes representing "hot" and "cold" IPO markets.

C507: **Density forecasts with opinion pools and dependent models**
Presenter: **Anthony Garratt**, Birkbeck, University of London, United Kingdom
Co-authors: James Mitchell, Shaun Vahey

Many studies of prediction pools in the economics literature ignore the dependence between the predictions of the models, by treating the predictions as orthogonal. Others, look for "optimal" combinations assuming that the predictions are linearly related. It is examined the case in which a copula opinion pool is used which nests linear-dependence between predictions as a special case. Copulae provide a general means of modeling dependence with non-linear predictive densities. We contrast the copula-based approach with alternative methods for handling dependence, including the Ensemble Model Output Statistics which allows for correlation (i.e. linear dependence) in the central mass of the predictive densities. To illustrate the effectiveness of the copula opinion pool in producing well-calibrated density forecasts we provide simulations which combine a GARCH process with an AR for a univariate analysis of a single variable with two experts. We then revisit an empirical application in the literaure, to produce density forecasts for S&P500 returns. Our results indicate that modeling this dependence using the copula opinion pool offers the prospect of sharper and better calibrated predictive densities. We conclude that forecasting performance is sensitive to the form of dependence between predictions from the models.

C587: **Efficiency measurement in a DSGE framework**
Presenter: **Camilla Mastromarco**, University of Salento - Lecce, Italy
Co-authors: Ulrich Woitek

Stochastic frontier analysis is a possibility to address the issue of cyclical factor utiliziation by estimating a single equation production frontier. The idea is to extend this approach and incorporate inefficiency in a dynamic stochastic general equilibrium (DSGE) model. It is straightforward

to replace the production function in the standard Ramsey model by a stochastic frontier function, e.g. $Y_t = Z_t X_t F(L_t, K_t)$, where Y_t is output, K_t is capital, L_t is labor, Z_t is a technology shock, and X_t measures efficiency, with X_t bigger than zero and less than one. Linearizing the first-order conditions of the model, we end up with a state space space representation, which allows to estimate the structural parameters either by maximum likelihood or by Bayesian methods using the Kalman filter . Assuming an AR(1) process for the log of efficiency, $logX_t = \rho_x \log X_{t-1} + \sigma_\eta \eta_t$ requires $\log X_t$ less or equal than zero so that X_t is bigger than zero and less than one. It is straightforward to adjust the prediction equations of the Kalman filter allowing η_t to follow a truncated normal distribution, if we assume that η_t follows a standard normal distribution in the untruncated case. By allowing the possibility of inefficient production in addition to the VAR component, we are able to be more specific about off-model dynamics than in previous works.

C766: **Time instability of the U.S. monetary system: Multiple break tests and reduced rank TVP VAR**
Presenter: **Dukpa Kim**, University of Virginia, United States of America
Co-authors: Yohei Yamamoto

Earlier attempts to find evidence of time varying coefficients in the U.S. monetary vector autoregression have been only partially successful. Structural break tests applied to typical data sets often fail to reject the null hypothesis of no break. Bayesian inferences using time varying parameter vector autoregressions provide posterior median values that capture some important movements over time, but the associated confidence intervals are often very wide and make the entire results less conclusive. We apply recently developed multiple structural break tests and find statistically significant evidence of time varying coefficients in the U.S. monetary vector autoregression. We also develop a reduced rank time varying parameter vector autoregression with multivariate stochastic volatility. Our model has a smaller number of free parameters and thereby yields tighter confidence intervals than previously employed unrestricted time varying parameter models. Consequently, we find further evidence of time varying natural rate of unemployment, core inflation, inflation persistence and monetary policy rule.

CS21 Room Jessel MACRO-FINANCE INTERFACE Chair: Herman Van Dijk

C586: **Direct and indirect Monte Carlo for simultaneous equations, instrumental variables and errors in variables models**
Presenter: **Nalan Basturk**, Erasmus University Rotterdam, Netherlands
Co-authors: Arnold Zellner, Tomohiro Ando, Lennart Hoogerheide, Herman van Dijk

In many areas of economics, sets of variables are jointly generated with instantaneous feedback effects. Typically, these instantaneous effects are captured by Simultaneous Equations Models (SEMs). We make a connection between SEMs, the basic instrumental variables (IV) model, and a simple errors-in-variables model (EV). They possess a common statistical structure and therefore create common problem for inference: possible strong correlation between a right-hand side variable in an equation and the disturbance. As workhorse model we take the IV model. We analyze the existence of the joint posterior density under flat priors. We further present weakly informative regularization priors and result in posteriors that are proper. In such a situation one still faces the appearance of highly non-elliptical shapes of the posterior and predictive distributions. Standard methods like Gibbs sampling or Direct Monte Carlo are then very inefficient sampling methods. As a contribution we present the MitISEM algorithm for efficient and robust sampling. We illustrate these model structures, properties of the posterior densities, and the proposed posterior sampler using simulated data, US and German income-education data and Fulton fish market data.

C636: **Bayesian factor model averaging and industry momentum strategies**
Presenter: **Herman van Dijk**, Erasmus University Rotterdam, Netherlands
Co-authors: Lennart Hoogerheide, Lukasz Gatarek

A novel approach to Bayesian estimation of the number of factors in factor models with fat tailed disturbances is developed. The Metropolis-Hastings (M-H) sampling algorithm is derived under a Jeffreys' and encompassing prior imposed on the covariance matrix. An eigenvalue decomposition is applied to extract information from the covariance matrix. Analytic expressions for acceptance-rejection weights in the M-H algorithm are derived and it is shown how these results might be applied to evaluate Bayes factors. This method is tested in the context of rank reduction testing. We present the application of these ideas to a financial speculative strategy known as momentum strategy. We implement Bayesian Model Averaging in order to investigate whether it has any impact on the profitability of this strategy. We find that the number of factors is time varying over our data.

C391: **A theoretical foundation for the Nelson and Siegel class of yield curve models**
Presenter: **Leo Krippner**, Reserve Bank of New Zealand, New Zealand

Yield curve models within the Nelson and Siegel (hereafter NS) class have proven very popular in finance and macrofinance, but they lack a theoretical foundation. It is shown how the level, slope, and curvature components common to all NS models arise explicitly from a low-order Taylor expansion around central measures of the eigenvalues for the standard generic Gaussian term structure model. That theoretical foundation provides an assurance that NS models correspond to a well-accepted framework for yield curve modeling. It further suggests that any yield curve from the GATSM class can be represented parsimoniously by a two-factor arbitrage-free NS model. Such a model is derived and applied to investigate changes in United States yield curve dynamics over the period from 1971 to 2010. The results provide evidence for material changes in the data-generating process for the yield curve between the periods 1971-1987, 1988-2002, and 2003-2010.

C741: **International financial transmission of the US monetary policy: An empirical assessment**
Presenter: **Nikola Mirkov**, University St.Gallen and Swiss Institute of Banking and Finance, Switzerland

This paper proposes a way to study the transmission mechanism of the US monetary policy to foreign yield curves. It elaborates a high-frequency identification strategy of monetary policy shocks in an international setting and uses a sample of 125 policy rate decisions of the Fed to extract realised policy shocks. The Fed decisions span from February 1994 to December 2008 and are divided according to the direction of the policy rate move and weather they were anticipated by the Fed funds futures market. A consistent, two-country term structure model is estimated on daily data and used to assess both instantaneous and lagged reaction of foreign interest rates and forward term premia to the Fed policy rate decisions. Empirical analysis of the US - UK model shows that the most of the movement in the UK yields around policy action days results from estimated term premia. A surprise policy action seems to produce a spike in the UK premia around the short- and mid-range maturities, independently from the direction of the policy rate move. The estimated lagged reaction of the UK yields to a policy decision of the Fed is also negative, after both hikes and cuts of the policy rate. The results hold for different market price of risk specifications, after two robustness checks and for both two-country and single-country model output.

CS44 Room Court VOLATILITY ESTIMATION AND FORECASTING Chair: Simona Sanfelici

C092: **On stochastic volatility models with long-memory in discrete and continuous time**
Presenter: **Frederi Viens**, Purdue University, United States of America
Co-authors: Alexandra Chronopoulou

It is commonly accepted that certain financial data exhibit long-range dependence. A continuous time stochastic volatility model is considered in which the stock price is geometric Brownian motion with volatility described by a fractional Ornstein-Uhlenbeck process. Two discrete time models are also studied: a discretization of the continuous model via an Euler scheme and a discrete model in which the returns are a zero mean iid sequence, where the volatility is a fractional ARIMA process. A particle filtering algorithm is implemented to estimate the empirical distribution of the unobserved volatility, which we then use in the construction of a multinomial recombining tree for option pricing. We also discuss appropriate parameter estimation techniques for each model. For the long-memory parameter, we compute an implied value by calibrating the model with real data. We compare the performance of the three models using simulated data and we price options on the S&P 500 index.

C305: **Inference for stochastic volatility models with jumps**
Presenter: **Jeannette H.C. Woerner**, TU Dortmund, Germany

In the framework of classical stochastic volatility models many different methods have been proposed to estimate volatility and detect jump components. They are mainly based on power variation, multipower variation, a combination of both, on a threshold approach or a comparison of the process on different time-scales. Some recent new approaches will be considered. One is the detection of jumps using extreme value theory. Two other methods, also working for non-semimartingale models are based on the correlation structure and on the fine structure of the processes respectively. We apply our results to the world stock index, which possesses long range dependence and jumps when expressed in currencies of unstable economies and to high frequency stock data which often exhibits short range dependence.

C331: **Limit theorems in the Fourier transform method for the estimation of volatility**
Presenter: **Arnaud Gloter**, Universite d'Evry Val d'Essonne, France
Co-authors: Emmanuelle Clement

Some limit theorems for a Fourier estimator of multivariate volatility proposed previously in the literature are proved. We first give a central limit theorem for the estimator of the integrated volatility assuming that we observe the whole path of the Ito process. Then the case of discrete time observations possibly non-synchronous is studied. In this framework we prove that the asymptotic variance of the estimator depends on the behavior of the ratio N/n where N is the number of Fourier coefficients and n the number of observations. We point out some optimal choices of N with respect to n to minimize this asymptotic variance. In the case of asynchronous data, we show that the estimator may be biased and correct this bias in a specific example.

C291: **Estimation of quarticity with high frequency data**
Presenter: **Maria Elvira Mancino**, University of Firenze, Italy
Co-authors: Simona Sanfelici

In recent years the availability of high frequency financial data has improved the capability to compute volatility in an efficient way. Nevertheless, the efficiency of all the methodologies proposed in accurately estimating the volatility builds on the observability of the true price process, while observed asset prices are contaminated by market microstructure effects. In order to produce feasible central limit theorems for all these estimators, and as a consequence feasible confidence intervals, it is necessary to obtain efficient estimators of the so called quarticity, which appears as conditional variance in the central limit theorems. Nevertheless, the studies about estimation of quarticity are still few. A new method is proposed to estimate quarticity which is based on a Fourier volatility estimation. The Fourier methodology allows to reconstruct the instantaneous volatility as a series expansion with coefficients gathered from the Fourier coefficients of the price variation. We prove that the Fourier estimator of quarticity is consistent in the absence of noise. We compute analytically the bias of the Fourier estimator of quarticity in the presence of microstructure noise and we propose a corrected Fourier estimator which is asymptotically unbiased under a suitable growth condition for the number of Fourier coefficients that have to be included in the estimation. The new methodology is tested in realistic Monte Carlo experiments. We make a comparative analysis of the performance of all the estimators existing in the literature. Our analysis shows that one of the most attractive features of the Fourier methodology is its ability to cope with microstructure effects.

CS59 Room Torrington VOLATILITY, HEAVY TAILS AND RISK Chair: Jean-Michel Zakoian

C096: **Parametric inference and forecasting in continuously invertible volatility models**
Presenter: **Olivier Wintenberger**, University Paris Dauphine, France
Co-authors: Sixiang Cai

The notion of continuously invertible volatility models that relies on some Lyapunov condition and some regularity condition is introduced. It is shown that it is almost equivalent to the volatilities forecasting efficiency of the parametric inference approach based on the Stochastic Recurrence Equation (SRE). Under very weak assumptions, the strong consistency and the asymptotic normality of an estimator based on the SRE is proved. From this parametric estimation, a natural forecast of the volatility that is strongly consistent is deduced. This approach is successfully applied to recover known results on univariate and multivariate GARCH type models where our estimator coincides with the QMLE. In the EGARCH(1,1) model, this approach is applied to find a strongly consistent forecast and to prove that our estimator is asymptotically normal when the limiting covariance matrix exists. Finally, some encouraging empirical results of our approach on simulations and real data are given.

C335: **Estimating the marginal distribution of heavy tailed time series**
Presenter: **Christian Francq**, CREST and University Lille 3, France
Co-authors: Jean-Michel Zakoian

A method for estimating the parametric marginal distribution of a stationary time series is proposed. The estimator is obtained by maximization of the "quasi marginal" likelihood, which is a likelihood written as if the observations were independent. The consistency and asymptotic normality of the estimator are established under mild assumptions on the dependence structure. Applications of the asymptotic results to the estimation of stable and Generalized Pareto distributions are proposed. The theoretical results are illustrated on financial index returns.

C382: **Maximum likelihood estimator for a conditional heteroscedastic model with alpha-stable innovation**
Presenter: **Guillaume Lepage**, CREST, France
Co-authors: Christian Francq, Jean-Michel Zakoian

A conditional heteroscedastic model with alpha-stable innovation is studied. We show that, under a few assumptions on the set of parameters, the model is stationary and the maximum likelihood estimator is consistent and asymptotically normal even if the true process of the innovations has infinite variance. Then, we look at a more general result using the fact that only the alpha-stable distributions have a domain of attraction.

C657: **Tracking illiquidities in intradaily and daily characteristics**
Presenter: **Gulten Mero**, University of Cergy-Pontoise and THEMA, France
Co-authors: Serge Darolles, Gaelle Le Fol

Two types of liquidity problems are distinguished; namely, liquidity frictions and illiquidity events. The first one is related to order imbalances that are resorbed within the trading day. It can be assimilated to "immediacy cost" and impacts the traded volume at the intraday and daily frequencies

while affecting the price increments only at the intraday periodicity. The second one is inherent to the long lasting liquidity problems and is responsible for the presence of stochastic volatility. We extend the mixture of distribution hypothesis with liquidity frictions (MDHL) framework to account for the presence of the illiquidity events. We then propose a two-step signal extraction formulation of the MDHL model in order to separate the two liquidity problem impacts on the daily returns and volume. We also provide, for a set of FTSE100 individual stocks, long lasting illiquidity indicators.

CS65 Room Gordon FINANCIAL MARKETS CONTAGION Chair: Gaelle Le Fol

C032: **Running for the exit: Distressed selling and endogenous correlation in financial markets**
Presenter: **Lakshithe Wagalath**, University Paris 6, France
Co-authors: Rama Cont

A simple multi-period model of price impact in a market with multiple assets is proposed, which illustrates how feedback effects due to distressed selling and short selling lead to endogenous correlations between asset classes. It is shown that distressed selling by investors exiting a fund and short selling of the fund's positions by traders may have non-negligible impact on the realized correlations between returns of assets held by the fund. These feedback effects may lead to positive realized correlations between fundamentally uncorrelated assets, as well as an increase in correlations across all asset classes and in the fund's volatility which is exacerbated in scenarios in which the fund undergoes large losses. By studying the diffusion limit of our discrete time model, we obtain analytical expressions for the realized covariance and show that the realized covariance may be decomposed as the sum of a fundamental covariance and a liquidity-dependent 'excess' covariance. Finally, we examine the impact of these feedback effects on the volatility of other funds. Our results provide insight into the nature of spikes in correlation associated with the failure or liquidation of large funds.

C316: **Liquidity contagion: A look at emerging markets**
Presenter: **Jeremy Dudek**, CREST Paris Dauphine, France
Co-authors: Gaelle Le Fol, Serge Darolles

Emerging economies have passed an important stress test during the period 2008-2009 and are now the key drivers for global growth of the world economy. The explosive expansion that emerging markets are showing during the last few years is mainly due to substantial returns that investors made. However, literature exposes that poor liquidity is one of the main reasons for foreign institutional investors not investing in emerging markets. The main contribution is to propose to analyze the liquidity of emerging markets and their effects on the contagion from one another in order to prevent illiquid events or systemic risk. Emerging market liquidity measures are the basis for the sovereign debt market, the deviations of the covered interest parity or the BIL for the foreign exchange market. One classical way to identify contagion is to observe an increase in correlation between asset returns during a crisis period. However, since crisis periods are typically characterized by an increase in volatility, it is crucial to discriminate between volatility and pure-correlation rises. One way to tackle this problem is to use a state-space model with a time-varying volatility specification and apply it to both returns and liquidity indicators.

C330: **Moment component analysis: An illustration with international stock markets**
Presenter: **Michael Rockinger**, HEC Lausanne, Switzerland
Co-authors: Eric Jondeau, Emmanuel Jurczenko

It is well known that non-normality plays an important role in asset and risk management. However, handling a large number of assets has long been a challenge. We present a statistical technique that extends Principal Component Analysis to higher moments such as skewness and kurtosis. This method allows us to identify factors that drive the co-skewness and co-kurtosis across assets. These factors have interesting interpretations, for instance as hedges against increases in volatility among certain assets. We illustrate this approach using 37 international stock indices sampled at weekly frequency, for a total of 763 observations. We assert that both the co-skewness and co-kurtosis structures can be summarized with a small number of factors. This method is both fast and able to handle large portfolios under non-normality. Estimations using a rolling window reveal interesting commonalities over the business cycle.

C246: **Survival of hedge funds: Frailty vs contagion**
Presenter: **Patrick Gagliardini**, Universita della Svizzera Italiana and Swiss Finance Institute, Switzerland
Co-authors: Serge Darolles, Christian Gourieroux

The rather short lifetimes of a majority of hedge funds and the reasons of their liquidation explain the interest of investors and academics in hedge fund survival analysis. The dependence between liquidation risks of individual hedge funds is considered. This dependence can either result from common exogenous shocks (frailty), or be due to contagion phenomena, which occur when an endogenous behaviour of a fund manager impacts the Net Asset Values of other funds. We introduce dynamic models able to distinguish between frailty and contagion phenomena, and to test for the presence and magnitude of such dependence effects, according to the age and management style of the fund.

CS70 Room Bedford EVALUATING FINANCIAL PERFORMANCES Chair: Bertrand Maillet

C352: **The dependence between performance measures and the construction of a composite performance index**
Presenter: **Michele Costola**, University of Padua, Italy
Co-authors: Massimiliano Caporin

A composite index (CI) to be used within stock picking or equity screening rules is considered. We first analyse at the single asset level the dependence that exists across a set of performance measures including, among others, the well-known Sharpe, Sortino, and Omega indices. We perform this explorative analyses by fitting under a restrictive iid assumption NIG and GHY distributions on financial returns time series. The estimated parametric densities are used to generate thousands of simulated paths for the assets. Those, in turn, are used to estimate the performance measure under study and to recover their densities, and their copula. On the basis of observed outcomes, we propose two approaches for the construction of composite indices: the first makes direct use of the copula functions, while the second is based on more restrictive assumptions on the performance measure dependence.

C577: **A survey on the four families of performance measures**
Presenter: **Gregory Jannin**, University of Paris-1 Pantheon-Sorbonne, France
Co-authors: Massimiliano Caporin, Francesco Lisi, Bertrand Maillet

Measuring the performance of funds is a key issue in the finance industry. Since the introduction of the original Sharpe ratio in 1966, a large number of measures of portfolio performance have been proposed in the academic as well as practitioner literature. After a review of the relevant literature, we have identified more than 50 performance measures that we classify into four families according to their common characteristics. We first present relative performance measures we have defined in terms of return per unit of risk. We start with the original Sharpe ratio and then group variations around this measure together in this first family. Secondly, we present the class of absolute performance measures beginning with the

most famous one, the Jensen alpha, before introducing other similar measures. Thirdly, we gather into a supplementary category general measures based on specific features of the portfolio return distribution, starting with the Gain/Loss ratio and the Omega measure. Finally, the fourth category groups together some performance measures that are directly grounded on explicit utility functions.

C576: **Towards a generalized performance measure**
Presenter: **Bertrand Maillet**, University of Venice, Italy
Co-authors: Monica Billio, Gregory Jannin, Loriana Pelizzon

Most of the performance measures proposed in the financial and academic literature are subject to be gamed in an active management framework. One of the main reasons of this drawback is due to an incomplete characterization by these measures of the studied return distributions. It is recalled how the most representative performance measures can be significantly influenced (managed or gamed) by fund managers. We also introduce a new flexible and robust generalized performance measure, characterizing the whole return distribution, and hardly gamable. More precisely, it takes into account the moments of the return distribution and incorporates a specific Probability-weighted Function (PwF), reflecting investors' preferences. The new performance measure is also well adapted for analyzing performance of hedge funds and more peculiarly in the presence of derivative instruments characterized by non-Gaussian return distributions.

C697: **Different mutual fund reward-to-risk performance measures**
Presenter: **Pilar Grau-Carles**, Universidad Rey Juan Carlos, Spain
Co-authors: Luis Miguel Doncel, Jorge Sainz

One of the most popular measures of a portfolio performance is the Sharpe Ratio. This reward-to-risk ratio is calculated as the expected return of a portfolio to its standard deviation. But the use of this measure has some drawbacks. The main stems from the use of the standard deviation as a measure of risk when returns do not follow a normal distribution. In this case, if returns do not follow a symmetrically elliptically distribution, alternative measures based on the concept of Value-at-Risk (VaR) such as the Reward-to-VaR, Reward-to-modified VaR, and Reward-to-Expected Shortfall can be used to measure the down-side risk. An alternative approach is the use of bootstrap methodology to take into account risk estimation by calculating the Double Sharpe ratio. The objective is to ascertain whether the choice of a particular performance measure has an impact on the ranking of alternative investments. We show, using data from UK mutual funds, that different measures produce different rankings: in the number of changes in the ten bottom/top performing funds, in the funds that maintain an equal order in the ratings and in the funds that out/under-perform the benchmark. Also, Spearman rank correlation coefficient, Kendall tau and Cohen Kappa show different rankings for different measures. Finally, we show that the disagreement between the rankings is greater when there are major differences in the higher moments of the distribution of returns.

Sunday 18.12.2011 16:50 - 18:30 Parallel Session K – ERCIM

ESI04 Room Senate SPACE-TIME MODELLING IN DISEASE MAPPING Chair: Lola Ugarte

E158: **Detecting space-time interactions in disease mapping when using CAR models**
Presenter: **Lola Ugarte**, Public University of Navarre, Spain
Co-authors: Tomas Goicoa, Jaione Etxeberria, Ana F. Militino

Data on disease incidence or mortality over a set of contiguous regions have been commonly used to describe geographic patterns of a disease helping epidemiologists and public health researchers to identify possible etiologic factors. Nowadays, the availability of historical mortality registers offers the possibility of going further describing the spatio-temporal distribution of risks. The literature on spatio-temporal modelling of risks is very rich and it is mainly focussed on the use of conditional autoregressive (CAR) models from a fully Bayesian perspective. The complexity of the estimation procedure makes the Empirical Bayes approach a plausible alternative. In this context, it is of interest to test for interaction between space and time, as an absence of space-time interactions simplifies modelling and interpretation. A score test is derived, as well as a bootstrap approximation of its null distribution. A parametric bootstrap test will be also provided for comparison purposes. Results will be illustrated using brain cancer mortality data from Spain in the period 1996-2005.

E250: **Latent clustering and grouping in Bayesian mixed effect spatio-temporal models for small area disease risk**
Presenter: **Andrew B. Lawson**, Medical University of South Carolina, United States of America
Co-authors: Jungsoon Choi

A flexible approach to contextual modeling of geo-referenced small area healthy data is considered. Our models examine contextual spatial effects and hidden grouping of covariate parameters in the fixed effect part. We examine different spatial correlation prior distributions for the group labeling of spatial units, including both Ising/Potts and threshold CAR and mixture prescriptions. We focus on two different applications: low birth weight variation within counties of Georgia, USA, and SEER cancer registry data for prostate cancer in the state of Louisiana.

E546: **Hierarchical Bayesian modelling to assess divergence in spatio-temporal disease mapping**
Presenter: **Annibale Biggeri**, University of Florence, Italy
Co-authors: Dolores Catelan

Spatio-temporal areal data such those arising in disease mapping are considered. We do not focus on relative risk estimation, which is the usual goal of disease mapping, i.e. to investigate the geographical and temporal distribution of the risk of diseases by area. Instead our inferential goal is to identify areas which have relative risk of a given disease in a given temporal period that diverges from a reference. We propose a hierarchical modelling approach to the problem and show how to specify a full range of informative null priors. Spatial or spatio-temporal structured priors were considered in all the proposed models to take into account the underlying variability of baseline risk among areas and time periods. The conflict between the null prior and the likelihood at a single node in the graphical model is explored by a cross-validation posterior predictive probability integral. Then, selective inference needs appropriate post-processing of cross-validation posterior predictive p-values for multiple testing. Previous use of cross-validation posterior predictive distributions to detect outlying observations in disease mapping failed to address the selection effect. We review this issue in the context of hierarchical Bayesian models and take advantage of a real example on lung cancer in Tuscany (Italy).

ES13 Room S264 ROBUST METHODS FOR FINANCIAL APPLICATIONS Chair: Kris Boudt

E118: **Long-term asset tail risks in developed and emerging markets**
Presenter: **Stefan Straetmans**, Maastricht University, Netherlands
Co-authors: Bertrand Candelon

The tail of financial returns is typically governed by a power law (i.e. “fat tails”). However, the constancy of the so-called tail index which dictates the tail decay has been hardly investigated. The finite sample properties of some recently proposed endogenous tests for structural change in the tail index are studied. Given that the finite sample critical values strongly depend on the tail parameters of the return distribution, a bootstrap-based version of the structural change test is proposed. Our empirical application spans a wide variety of long-term developed and emerging financial asset returns. Somewhat surprisingly, the tail behavior of emerging stock markets is not more strongly inclined to structural change than their developed counterparts. Emerging currencies, on the contrary, are more prone to shifts in the tail behavior than developed currencies. Our results suggest that extreme value theory (EVT) applications in risk management or in assessing the (changing) propensity to financial crises can safely assume stationary tail behavior over long time spans, provided one considers portfolios that solely consist of stocks or bonds. However, our break results also indicate it is advisable to use shorter estimation windows when applying EVT methods to emerging currency portfolios.

E125: **Fitting semiparametric Markov-switching models to electricity prices**
Presenter: **Dennis Tuerk**, Maastricht University, Netherlands
Co-authors: Michael Eichler

Lately Markov-switching models seem to have become the standard tool in electricity price modeling. We propose the use of a semiparametric Markov-switching model using robust estimation techniques as an alternative to commonly applied estimation approaches. The model in combination with the estimation framework is easier to estimate, needs less computation time and distributional assumptions. To show its advantages the proposed model is compared with a well established Markov-switching model in a simulation-study. Further, the model is applied to Australian log-prices. The results are in accordance with the results from the simulation-study, indicating that the proposed model might be advantageous whenever the distribution of the spike process is not sufficiently known. The results are thus encouraging and suggest the use of our approach when modeling electricity prices and pricing derivatives.

E301: **Sparse and robust factor modelling**
Presenter: **Peter Exterkate**, Aarhus University, Denmark
Co-authors: Christophe Croux

Factor construction methods are widely used to summarize a large panel of variables by means of a relatively small number of representative factors. A novel factor construction procedure is proposed that enjoys the properties of robustness to outliers and of sparsity; that is, having relatively few nonzero factor loadings. Compared to more traditional factor construction methods, it is found that this procedure leads to better interpretable factors and to a favorable forecasting performance, both in a Monte Carlo experiment and in an empirical applications to macroeconomic forecasting.

E440: **Nonparametric tests for intraday jumps: impact of periodicity and microstructure noise**
Presenter: **Jonathan Cornelissen**, K.U. Leuven, Belgium
Co-authors: Kris Boudt, Christophe Croux, Sebastien Laurent

Robust analysis of high frequency financial data requires accurate tests to detect intraday price jumps. These tests need to take into account the stylized fact that opening, lunch and closing of financial markets induce a periodic component in the volatility of high frequency returns. Accounting for this periodicity in the implementation of nonparametric intraday jump tests is important to avoid a size distortion. We review this topic and extend it by explicitly taking microstructure noise into account in the analysis of the size and power of intraday jump test statistics.

ES32 Room S261 DIAGNOSTIC TESTS FOR INDEPENDENT AND TIME-SERIES DATA Chair: Simos Meintanis

E262: **Modelling stationary time series by continuous time processes**
Presenter: **Alejandra Cabana**, Universitat Autonoma de Barcelona, Spain
Co-authors: Argimiro Arratia, Enrique Cabana

Denote by $O\mathcal{U}_p$ the family of stationary processes obtained by the iterated application of p operators $\xi(t) \mapsto \int_{-\infty}^{t} \exp^{-\lambda(t-s)} d\xi(s)$ with parameters $\lambda_1, \lambda_2, \ldots, \lambda_p$ to a Wiener process w. The parameters λ_j are assumed to have positive real parts. The $O\mathcal{U}_1$ processes are the well known Ornstein Uhlenbeck processes. These processes can be used as models for stationary processes of a continuous parameter, and also to represent the discrete time processes obtained by sampling an $O\mathcal{U}_p$ process at a given rate. When $p = 1$ these are autorregressive processes of order one. For larger p the $O\mathcal{U}$ and AR processes differ. The $O\mathcal{U}_p$ processes are linear combinations of p OU processes with exponential decay λ. When $x \in O\mathcal{U}p$ is observed at instants $0, 1, \ldots, n$ the estimation of the parameters $\phi_1, \ldots, \phi_p$ related to $\lambda_1, \lambda_2, \ldots, \lambda_p$ by means of the polynomial identity $\prod_{j=1}^{p}(1+\lambda_j s) = 1 - \sum_{j=1}^{p} \phi_j s^j$ is performed by means of the solution of linear equations similar to the normal equations for linear models. The estimation of parameters, their asymptotic behaviour for large n and consistent tests of the null hypothesis $x \in O\mathcal{U}_p$ are considered.

E312: **Sequential monitoring of stability of time series**
Presenter: **Marie Huskova**, Charles University in Prague, Czech Republic
Co-authors: Zdenek Hlavka, Claudia Kirch, Simos Meintanis

Procedures for monitoring changes in autoregressive time series while controlling the overall size of the sequential test will be presented. Changes both in the error distribution and autoregressive parameters will be discussed. The proposed procedures utilize the empirical characteristic function of properly estimated residuals. Since the asymptotic null distribution contains unknown parameters, a proper bootstrap procedure is proposed in order to actually perform the test. Results on the finite - sample performance of the new methods will be presented. The focus will be on developing a proper version of bootstrap, computational aspects, choice of tuning constants and simulations.

E342: **Diagnostic tests for the location-shift assumption**
Presenter: **Olivier Thas**, Ghent University, Belgium
Co-authors: J.C.W. Rayner

The Wilcoxon rank-sum test is often considered as the nonparametric version of the t-test for comparing means. However, the hypotheses of the Wilcoxon test can only be expressed in terms of means under restrictive distributional assumptions. The most common assumption demands that the distributions of the two populations belong to a location-shift family, i.e. the distributions agree in shape except for a location shift. We present tests for testing the location-shift assumption. A first test is based on the Wasserstein distance between the two empirical quantile functions. A second class of tests is based on the set of semiparametric efficient score statistics. In an empirical power study we compare the tests. Finally, we present some meaningful graphical diagnostic tools for assessing the location-shift assumption.

E524: **The probability weighted empirical characteristic function and goodness-of-fit testing**
Presenter: **Norbert Henze**, Karlsruhe Institute of Technology (KIT), Germany
Co-authors: Simos Meintanis

We introduce the notion of the probability weighted empirical characteristic function (PWECF), of which the empirical characteristic function is a special case. Some of its properties are studied, and the potential use of the PWECF for goodness-of-fit testing is examined.

ES44 Room G16 MODELLING THE EXTREMES Chair: Carl Scarrott

E308: **Conditional analysis for multivariate extremes in finance**
Presenter: **Ye Liu**, Lancaster University, United Kingdom
Co-authors: Jonathan Tawn, Anthony Ledford

The impact of extreme events can be devastating. A good evaluation of the extreme risk of an investment can be the thin line between success and failure. Our research focuses on the application of multivariate extreme value theory to financial data. Traditional multivariate analysis, including the popular copula-based models, tend to make implicit assumptions about joint behaviour of extreme financial returns. Our study shows that these assumptions are often falsely imposed which may cause a very biased inference related to rare events, especially those beyond observed levels. To overcome this we propose a conditional approach based on a given multivariate extreme value model. Our new model is well suited for use with financial data and can deal with large number of variables jointly without the risk of over-parameterization. Its key features include the flexibility to capture all known types of tail dependence via a semi-parametric method as well as to account for intuitive factors through hierarchical modelling. We start by introducing the key issues we are aiming to address and demonstrating the shortfalls of existing models. These are followed by a quick introduction to the proposed model including its theoretical foundation and inference methods. Finally, we look at some typical extreme value risk measures of a typical financial data set and compare the results between the new model and some popular existing models.

E399: **Modelling non-stationary extremal behaviour via mixture modelling**
Presenter: **Carl Scarrott**, University of Canterbury, New Zealand
Co-authors: Anna MacDonald, Dominic Lee

Extreme value models are typically used to describe the distribution of rare events. Generally, an asymptotically motivated extreme value model is used to approximate the tail of some population distribution. One of the key challenges with even the simplest application of extreme value models is to determine the "threshold" above which the asymptotically motivated models provide a reliable approximation to the tail of the population distribution. A plethora of recent articles have considered the development and usage of mixture models for threshold estimation and quantifying the corresponding uncertainty. One particularly flexible mixture model allows for the usual GPD/PP representation for the upper tail behaviour, with the threshold as a parameter and the bulk distribution below the threshold captured by a non-parametric kernel density estimator. This mixture model is extended to describe non-stationary features. Unlike other mixture models seen in the literature, the benefit of this particular mixture model is that the non-stationarity in the threshold and GPD/PP parameters can be modeled in the usual way(s). The kernel bandwidth is assumed to not vary or vary very slowly, which is a safe assumption for most applications. Results from simulations and an application using Bayesian inference will be presented.

E484: **A Dirichlet process mixture model in fitting peaks over threshold**
Presenter: **Xin Zhao**, University of Canterbury, New Zealand
Co-authors: Dominic Lee, Marco Reale, Les Oxley, Carl Scarrott

A Dirichlet process mixture extreme model for capturing peaks over threshold distributions is proposed. The model relaxes the assumptions of the classical generalized Pareto distribution for the excesses above the threshold by allowing the extremes to come from random distributions rather than be identically distributed. A simulation study investigates the power and small sample behaviour. We apply the mixture model on financial time series which exhibit well known stylized features such as heavy tails and non-stationary for estimation of tail risk.

E562: **Extreme value modeling of survival times**
Presenter: **Alberto Alvarez-Iglesias**, National University of Ireland Galway, Ireland
Co-authors: John Newell, Carl Scarrott, John Hinde

A key challenge in survival analysis is estimation of the mean residual life function, which at time t measures the expected remaining lifetime of an individual that survived up to time t. This function completely determines the distribution of the survival times, so can be used as an alternative to the survivor function. Often estimation of the mean residual life is complicated by the upper tail of the survival distribution not being unobserved as, for example, when patients are still alive at the end of the study period. Observations of this type are right censored, but differ from other types of right censoring (drop out or lost to follow up) in the sense that after the maximum observed time there is no information at all. Various approaches have been developed to estimate the mean residual life in the face of this severe right censoring. In this work, a novel semi-parametric method is demonstrated which combines existing nonparametric methods and an extreme value tail model is presented. where the limited sample information in the tail (up to censoring) is used to estimate the upper tail behavior. This approach will be demonstrated with simulated and real-life examples.

ES45 Room B36 MIXTURE MODELS: THEORY AND DATA ANALYSIS Chair: Marco Riani

E244: **Fitting regression mixtures to contaminated data**
Presenter: **Domenico Perrotta**, European Commission, Italy

Suppose we want to cluster data originated from an unknown number of populations, in the presence of outliers and with groups that overlap considerably, forming one or more dense areas. Under this scenario, we have demonstrated elsewhere that reasonable and rather stable multivariate groups can be obtained by clustering only a subset of points selected with an appropriate thinning process. Here, we show that the approach can be also beneficial when data are generated around a number of straight lines which are therefore fitted by a linear regression mixture. This is shown for different mixture estimation methods and a Forward Search approach.

E322: **Constraints in mixture modelling**
Presenter: **Agustin Mayo Iscar**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Alfonso Gordaliza, Carlos Matran

Mixture modelling is nowadays a very important tool for the statistical practitioner. Several procedures can be applied for fitting mixtures to data sets by applying EM algorithms. Moreover, if no proper constraints are posed on the component scatter parameters, the fitted mixtures may include non-interesting spurious components. With this problem in mind, we propose a mixture modelling approach which in a computationally efficient way imposes constraints controlling the appearance of spurious components.

E340: **Parsimonious linear Student-t cluster weighted model**
Presenter: **Salvatore Ingrassia**, Universita di Catania, Italy
Co-authors: Simona C. Minotti, Antonio Punzo

The linear Student-t Cluster Weighted Model (CWM) represents a robust and flexible mixture approach to modelize the joint density $p(y,x)$ of a random vector $(Y,X) : \Omega \to \mathbb{R}^{d+1}$, with $X \in \mathbb{R}^d$, and $Y \in \mathbb{R}$, which can be considered as the input and the output variable, respectively. Moreover, suppose that Ω can be partitioned into g groups, that is $\Omega = \Omega_1 \cup \cdots \cup \Omega_g$. We assume that $X|\Omega_j$ has a multivariate Student-t distribution $t_d(x;\mu_j,\Sigma_j,\nu_j)$ and $Y|x,\Omega_j$ has a Student-t distribution $t\left(y;\mu(x;\beta_j),\sigma_j^2,\zeta_j\right)$, for $j = 1,\ldots,g$. We propose a parsimonious approach to the modelization of the density $p(y,x)$, which consists of considering the linear Student-t CWM as the generalization of a family of models obtained by allowing: 1. the distributions of $X|j$ or $Y|X,j$ to be equal among clusters; 2. the degrees of freedom (ν and/or ζ) to tend to infinity (so obtaining the normal distribution). From a convenient combination of these two constraints, we obtain 12 parsimonious and easily interpreted models which are appropriate to describe various practical situations. A real dataset exemplifies the proposed methodology. Finally, the performance of various model selection criteria, in leaning toward the correct model and the true value of g, is analyzed via a wide simulation study.

E354: **A comparison of different multivariate clustering methods**
Presenter: **Aldo Corbellini**, University of Parma, Italy

The ability to successfully classify different data groups is a matter of great interest. The relative performance of different classification methods is disussed using different datasets contaminated with various degrees of outliers. The methods reviewed include MCLUST, Tclust, fix point clustering and those based on the forward search. Focus is put on the robustness and stability of the different approaches under different contamination settings.

ES47 Room B33 SEMIPARAMETRIC MODELS WITH INCOMPLETE DATA Chair: Ingrid Van Keilegom

E215: **Semiparametric kernel density estimation with doubly truncated data**
Presenter: **Jacobo de Una-Alvarez**, University of Vigo, Spain
Co-authors: Carla Moreira

Doubly truncated data appear in several fields, e.g. Survival Analysis, Epidemiology, Astronomy and Economy. Under random double truncation, only observations falling between two random magnitudes (the truncation times) are recruited, which results in a sampling bias. A semiparametric kernel density estimator under double truncation is presented. It is assumed that the joint distribution of the truncation times belong to a given parametric family. Asymptotic properties such as consistency and asymptotic normality are discussed. Finite sample performance is investigated through simulations. It is demonstrated that the semiparametric estimator outperforms the purely nonparametric kernel density estimator. A real data illustration is provided.

E225: **A goodness-of-fit procedure for semiparametric copula models under random censoring**
Presenter: **Olivier Lopez**, UPMC Paris VI, France
Co-authors: Svetlana Gribkova, Philippe Saint Pierre

A new goodness-of-fit procedure for semiparametric copula models under a bivariate censoring scheme is studied. The work is motivated by a data-set containing lifetimes of married couples who subscribed to a certain type of insurance contract. These bivariate data are subject to bivariate random censoring, with the particularity that the two components of the censoring vector are identical. We provide the asymptotic behavior of our test-statistic, along with a bootstrap procedure that allows us to compute the critical values. The extension of our approach to other kinds of bivariate censoring schemes is also discussed.

E321: **Complete case analysis revisited**
Presenter: **Ursula Mueller**, Texas A&M University, United States of America

A general method is presented for obtaining limiting distributions of complete case statistics for missing data models from those of the corresponding statistics when all data are observed. This provides a convenient tool to obtain the asymptotic behaviour of complete case versions of established methods without (reproducing) lengthy proofs. It is well known that a statistical analysis which ignores cases that are only partially observed does not always perform well and that an approach which imputes missing values often has better properties. However, there are situations where a complete case analysis is appropriate. The methodology is illustrated by analysing three inference procedures for partially linear regression models with responses missing at random. Firstly, asymptotically efficient estimators of the slope parameter are derived. Secondly, an asymptotically distribution free (ADF) test for fitting a normal distribution to the errors is derived. Finally, we obtain an ADF test for linearity, i.e. for testing that the nonparametric component of these models is a constant.

E584: **Flexibly extending the classical Koziol-Green model by a copula function.**
Presenter: **Roel Braekers**, Hasselt University, Belgium
Co-authors: Auguste Gaddah

In survival analysis, we are interested in the time until an event. Due to practical reasons, we do not fully observe this time. A second independent random variable, a censoring time, obscures the observation process and only the smallest of both times is observed together with an indicator variable. The classical approach to estimate the distribution function for the lifetime under informative censoring time is by the Koziol-Green model. In this model, it is assumed that the survival function of the censoring time is a power of the survival function of the time until an event. We generalize this model and assume a parametric function for the relationship between these distribution functions. Hereby we link this assumption to a slide of a parametric copula function between the observed variables. We propose a maximum likelihood estimator for the copula parameter and construct a semi-parametric estimator for the lifetime distribution. As results, we show the consistency and asymptotic normality of the copula parameter. We derive the weak convergence of the semi-parametric distribution estimator. In a simulation study, we investigate the finite sample performance of these estimators and apply this model to a practical data set.

ES51 Room B18 ADVANCES IN SOFTWARE FOR TREE MODELS Chair: Achim Zeileis

E329: **TINT R-package for advanced subgroup analysis**
Presenter: **Elise Dusseldorp**, TNO-Dutch Center for Applied Scientific Research, Netherlands
Co-authors: Iven Van Mechelen

When multiple treatments are available, it is often difficult to decide which treatment is best for an individual given his/her state and characteristics. Empirical evidence is lacking. The method TINT focuses on supplying such empirical evidence by means of advanced subgroup analysis. It answers the question: "Who reacts better to treatment A and who reacts better to treatment B?" TINT results in a binary tree that subdivides the patients into subgroups on the basis of their characteristics; these subgroups 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 optimal size of the tree is determined with a bias-corrected bootstrap procedure. The method is appropriate for data from randomized clinical trials with many patient characteristics that could interact with treatment in a complex way. We will show applications of the method using the R-package TINT.

E401: **evtree: Evolutionary learning of globally optimal classification and regression trees in R**
Presenter: **Thomas Grubinger**, Innsbruck Medical University, Austria
Co-authors: Achim Zeileis, Karl-Peter Pfeiffer

Commonly used classification and regression tree methods like the CART algorithm are recursive partitioning methods that build the model in a forward stepwise search. Although this approach is known to be an efficient heuristic, the results of recursive tree methods are only locally optimal, as splits are chosen to maximize homogeneity at the next step only. An alternative way to search over the parameter space of trees is to use global optimization methods like evolutionary algorithms. The evtree package implements an evolutionary algorithm for learning globally optimal classification and regression trees in R. CPU and memory-intensive tasks are fully computed in C++ while the partykit package is leveraged to represent the resulting trees in R, providing unified infrastructure for summaries, visualizations, and predictions. evtree is compared to rpart, the open-source CART implementation, and ctree (conditional inference trees). On several benchmark problems from the UCI machine learning repository, evtree models offered at least similar and most of the time increased predictive accuracy.

E424: **partykit: A toolkit for recursive partytioning**
Presenter: **Achim Zeileis**, Universitat Innsbruck, Austria
Co-authors: Torsten Hothorn

Recursive partitioning methods, or simply "trees", are simple yet powerful methods for capturing regression relationships. Hence, many different algorithms have been suggested in both the statistics and machine learning communities and many standard algorithms are available as R packages, e.g., in rpart, RWeka, party, and many others. However, no common infrastructure is available for representing trees fitted by different packages. Consequently, the capabilities for extraction of information - such as predictions, printed summaries, or visualizations - vary between packages and come with somewhat different user interfaces. Similarly, implementations of new tree models might also require new infrastructure, e.g., for multi-way splits or more complex models in the leafs. To overcome these difficulties, the partykit package offers a unified representation of tree objects along with predict(), print(), and plot() methods. Trees are represented through a new flexible class "party" which can, in principle, capture all trees mentioned above. The package is currently under development at R-Forge and already provides conversion methods for trees of classes rpart, J48, and pmmlTreeModel as well as a re-implementation of conditional inference trees. Furthermore, the new packages evtree and CHAID employ the partykit infrastructure.

E510: **Detecting threshold interactions in supervised classification and regression: STIMA**
Presenter: **Claudio Conversano**, University of Cagliari, Italy

Simultaneous Threshold Interaction Modeling Algorithm (STIMA) has been implemented in the R environment as a tool enabling us to automat-

ically 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 Model or on the Regression Trunk Model. 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. Different specifications of the generalized linear model with threshold interaction effects can be provided by STIMA on the basis of the nature of the response variable. Results on real and synthetic data are presented in order to compare the performance of STIMA with that of alternative methods.

ES55 Room B35 FUZZY SETS IN STATISTICS Chair: Gil Gonzalez-Rodriguez

E197: **A kernel framework for learning relations from paired-comparison data**
Presenter: **Willem Waegeman**, Ghent University, Belgium
Co-authors: Tapio Pahikkala, Antti Airola, Tapio Salakoski, Bernard De Baets

Recently the problem setting of inferring relations between pairs of data objects has been widely investigated in the machine learning community. To this end, current approaches typically consider datasets containing crisp relations, so that standard classification methods can be adopted. However, relations between objects like similarities and preferences are in many real-world applications often expressed in a graded manner. We present a general kernel framework for inferring such relations from paired-comparison data and for ranking data objects according to these relations. This framework extends existing approaches because both crisp and valued relations are considered, and it unifies existing approaches because different types of valued relations can be modeled, including symmetric and reciprocal relations. Important links between recent developments in fuzzy set theory and machine learning are in this way established and applications appear in areas like bioinformatics, information retrieval and social network analysis. Experimental results for different application domains will be presented.

E478: **On a fuzzy regression model with crisp input and trapezoidal fuzzy output data**
Presenter: **Concepcion Roldan**, University of Jaen, Spain
Co-authors: Antonio Roldan, Juan Martinez-Moreno

The problem of analyzing linear and non-linear regression relationships considering crisp input and trapezoidal fuzzy output data is considered.The fuzzy regression will be formulated based on a minimum criterion which uses different distances between fuzzy numbers. The different distances considered are not the usual ones for fuzzy sets, but can be useful in some scientific contexts. This is particularly interesting, since it could provide alternative estimates that can be calculated easily and offer different models from which the experts could choose according to their needs. The results are illustrated by means of a simulation experiment.

E479: **Output distributions of aggregation functions: a statistical study**
Presenter: **Luis J. Rodriguez-Muniz**, University of Oviedo, Spain
Co-authors: Luigi Troiano, Irene Diaz

Aggregation functions have been studied in depth in order to investigate their analytical properties. In particular, researchers have paid attention to conjunctive, disjunctive and compensatory functions in order to characterize them. But few studies have been developed about the behavior of the outputs. To study the distribution of output values and how it is related to input values becomes relevant in order to determine what choices of aggregation operators are significantly relevant. We focus on pairwise comparing aggregation functions on the basis of their output distributions in order to answer questions such as: What is the relevant output difference among different parameter values of parametric aggregation functions? Has the number of data points to aggregate any effect on the output distributions? How can the effect of input distribution on the behavior of outputs be measured? Examining output differences will be useful to practitioners in choosing between alternatives with substantial statistical differences.

E566: **Modelling decisors' subjective attitude towards negotiation in e-democracy and e-cognocracy problems**
Presenter: **Pedro Teran**, Universidad de Oviedo, Spain
Co-authors: Jose Maria Moreno

Citizens' participation in public decision-making brings a huge multiplication of the number of actors involved. In such a multicriteria, highly multiactor, group decision problem, even if all parts wish a consensus there is a risk that the problem's mathematical solution remains so removed from a minority's point of view that they feel their perspective and values have been 'left out' or 'effaced' from the decision process. We propose a method to incorporate the decisors' attitude towards negotiation. It allows for more flexibility and a posterior analysis of the decision process which explores the solution space taking into account their declared negotiation attitude. Precisely, we obtain a Monte Carlo approximation to the probability distribution on all possible rankings of the alternatives (the group preference structure). Each time, a multicriteria problem is solved using the Analytic Hierachy Process method with aggregation of individual judgements. The input of the AHP step is a matrix of numerical pairwise comparisons for each decisor. These are obtained by fusing their initial judgements with their attitude towards negotiation in a fuzzy set, then applying artificial sampling. The method was tested with 42 decisors; the output will be discussed and compared to that of conventional AHP and voting methods.

ES74 Room B34 OUTLIERS AND CHANGE-POINTS IN TIME SERIES II Chair: Christophe Croux

E460: **Interventions in INGARCH processes and their effects on the estimation of autocorrelation**
Presenter: **Tobias Liboschik**, TU Dortmund, Germany
Co-authors: Roland Fried

Different approaches to include interventions within the framework of integer-valued GARCH (INGARCH) models for count time series are studied. There is a model in the literature where an intervention affects the non-observable underlying mean process at the time point of its occurrence and additionally the whole process thereafter via its dynamics. As an alternative, we consider a model where an intervention directly affects the observation at the time point of its occurrence, but not the underlying mean, and then also enters the dynamics of the process. While the former definition describes an internal change, the latter one models an external effect on the observations. We adopt conditional likelihood estimation and tests for intervention effects to our alternative way of modelling interventions. Both models are compared analytically and using simulated and real data examples. We study the effect of misspecification on the fitted intervention model. Moreover, the literature often only treats the special case of INGARCH(1,1) processes. For selecting the model order of a general INGARCH(p,q) process, the autocorrelations and partial autocorrelations are insightful. We illustrate by simulations how interventions affect the estimation of the autocorrelation function considering classical and simple robust estimators.

E662: **Least squares estimation in models with multiple change-points**
Presenter: **Dietmar Ferger**, University of Dresden, Germany

We consider $n \in \mathbb{N}$ independent observations $X_i = X_{ni}, 1 \le i \le n$, with $q \in \mathbb{N}$ *change-points* $1 \le \tau_{n1} < \tau_{n2} < \ldots < \tau_{nq} < n$, i.e. for every $1 \le r \le q+1$ the random variables in the r-th section have a common distribution: $X_i \sim Q_r$ for all $\tau_{n,r-1} < i \le \tau_{nr}$ with $\tau_{n0} := 0$ and $\tau_{n,q+1} := n$. Here, the probability measures $Q_1, \ldots, Q_{q+1}$ possess unknown expectations $\alpha_1, \ldots, \alpha_{q+1}$. Our main issue is to estimate the vector $(\underline{\tau}_n, \underline{\alpha}) := (\tau_{n1}, \ldots, \tau_{nq}, \alpha_1, \ldots, \alpha_{q+1})$. We investigate the pertaining least squares estimator $(\hat{\underline{\tau}}_n, \hat{\underline{\alpha}}_n)$. Under the assumption that $\tau_{nr} - \tau_{n,r-1} \to \infty$ for all $1 \le r \le q+1$ and $\alpha_{r+1} \ne \alpha_r$ for all $1 \le r \le q$, we prove distributional convergence of $\hat{\underline{\tau}}_n - \underline{\tau}_n$ and identify the limit variable $\underline{T} = (T_1, \ldots, T_q)$. It turns out that each component T_r is equal to the minimizing point of a certain random walk Γ_r on $\mathbb{Z}$ with a positive drift function given by $(\alpha_{r+1} - \alpha_r)^2 |k|$, $k \in \mathbb{Z}$. Moreover, it will be shown that $\hat{\underline{\alpha}}_n$ is a weakly consistent estimator for $\underline{\alpha}$.

E533: **Recent results on volatility change point estimation for stochastic differential equations**
Presenter: **Stefano Maria Iacus**, University of Milan, Italy

The problem of change point has been considered initially in the framework of independent and identically distributed data by many authors. Recently, it naturally moved to context of time series analysis. Indeed, change point problems have originally arisen in the context of quality control, but the problem of abrupt changes in general arises in many contexts like epidemiology, rhythm analysis in electrocardiograms, seismic signal processing, study of archeological sites and financial markets. For discretely observed, one-dimensional ergodic diffusion processes, some authors considered a least squares approach. The problems of the change-point of drift for continuously observed ergodic diffusion processes have been treated previously. For general Itô processes some authors have considered quasi-maximum likelihood estimation. We review recent theoretical results on change point analysis for the volatility term in discretely observed stochastic differential equations and their software solutions for the R statistical environment.

E021: **Testing for structural changes in the dependence structure at an unknown point in time**
Presenter: **Dominik Wied**, TU Dortmund, Germany
Co-authors: Walter Kraemer, Herold Dehling

There are many empirical hints that the dependence structure of financial assets cannot be assumed to be constant over time. A formal statistical test for constant correlation between two time series is presented. The test only needs mild assumptions, e.g. the possible change point need not be known a priori. This test can be applied to detect "shift contagion" or to develop trading strategies. For the latter case, the results of the test can be used to detect time points in which a portfolio change might be necessary. Such a trading strategy works well in practice and can help to hedge against financial crises. As a second contribution the paper presents corresponding change point tests for copulas or copula-based dependence measures. The procedures are robust against outliers and work well without the assumption of finite moments. We present the extension of a previously suggested test for constant copula in a given point to the case of mixing random variables and present a new test for overall copula constancy. The latter is based on the integral over the empirical copula and simultaneously tests for constancy of Spearman's rho.

ES67 Room B20 HEURISTICS IN FINANCE Chair: Dietmar Maringer

E258: **Modeling and prognosis of European interest rate swaps using genetic programming**
Presenter: **Gabriel Kronberger**, University of Applied Science Upper Austria, Austria
Co-authors: Stefan Fink, Michael Affenzeller

Interest rate swaps are an important financial instrument for risk management for loans and liabilities but are also used for speculative purposes. The market for interest rate swaps is highly liquid and less affected by fluctuations caused by the fear of credit losses as for instance government bonds. Reliable prediction models for the interest rate swap rates are important decision support instruments for treasurers or hedgers. In this paper we describe an approach for modeling and prognosis of daily and monthly interest rate swap rates using symbolic regression. Symbolic regression is a heuristic modeling approach based on genetic programming which is capable of identifying non-linear and conditional models automatically. Another potential advantage of symbolic regression is that it produces white-box models (i.e. mathematical formulae), which can be inspected and analysed easily by domain experts. Interest rates are strongly affected by macro-economic conditions. Thus, a number of important macro-economic variables and indicators (e.g. CPI, unemployment rate, stock indices, ...) are used as additional input variables for the models. Finally, we present the results of our modeling approach and compare the accuracy of the symbolic regression models with the accuracy achieved by linear modeling approaches.

E268: **Continuous time Bayesian classifiers for intraday FX prediction**
Presenter: **Simone Villa**, Universita degli Studi di Milano-Bicocca, Italy
Co-authors: Fabio Stella

A set of continuous time Bayesian models for predicting intraday FX are developed, analyzed and compared. The models belong to a novel class of supervised classifiers that are named continuous time Bayesian network classifiers. This class specializes the framework of continuous time Bayesian networks to the case of supervised classification. The problems of Bayesian network classifier learning from high frequency data, structural and parametric, are addressed and two algorithms to solve them are presented. Inference algorithms, filtering and prediction, are presented in the case where FX data are fully observable and when latent variables are present. Numerical experiments concerning single FX, EUR/USD, and cross currency FX have been performed. Furthermore, a software environment, based on MATLAB, for learning continuous time Bayesian network classifier from high frequency data and to make inference is described.

E302: **Option pricing with grammatical evolution**
Presenter: **Christian Oesch**, University of Basel, Switzerland

A first approach in using Grammatical Evolution to value options is presented. Grammatical Evolution is a grammar-based form of Genetic Programing. Here it is used to price options in both an artificial Heston-World as well as on real world empirical Swiss Market Index options data. The form of the option pricing formula is evolved my means of an evolutionary process by training the algorithm on a sample set and the outcome is validated on an independent set of data. The results of the experiments show that Grammatical Evolution produces good results on the artificial data and can be evolved to beat the Black-Scholes pricing formula in the empirical data set.

E474: **Multivariate neural network estimation of bidirectional volatility spillover between U.S. and European government bond markets**
Presenter: **Gordon Dash**, University of Rhode Island, United States of America
Co-authors: Nina Kajiji

Prior modelling efforts are extended to engineer an efficient mapping of government bond volatility transmission between the U.S. and European government bond markets. The research introduces the K7 multivariate Bayesian-enhanced multiple-objective radial basis function artificial neural network (K7-MRANN). The method is used to simultaneously determine the relative impact of bidirectional volatility spillovers between the U.S.

government bond market and regional European counterparts. The research method deployed permits a re-examination of univariate modelling results while explicitly extending the model statement to uncover simultaneous bidirectional volatility spillover that is a result of the contemporary global liquidity and credit crisis. The performance of the simultaneous determinants model is calibrated against two alternative specifications using the univariate GARCH framework. Subsequently, MRANN and GARCH residuals are tested for latent effects. We find that the multivariate residuals are statistically devoid of economic covariance and conditional volatility effects. When compared to other recent studies, the policy inferences corroborate parametric findings of a weak US spillover effect into European government bond markets. The MRANN signed weights also provide new insight into the inverted volatility spillover effects attributed to Europe's evolving sovereign debt crisis.

Monday 19.12.2011 09:05 - 10:25 Parallel Session L – CFE

CS02 Room Jessel REALIZED VOLATILITY IN APPLICATIONS Chair: Francesco Audrino

C100: **Realized volatility models and alternative Value at Risk prediction strategies**
Presenter: **Spyros Xanthopoulos**, Athens University of Economics and Business and Bank of Greece, Greece
Co-authors: Dimitrios P. Louzis, Apostolos P. Refenes

The Value at Risk (VaR) prediction accuracy and efficiency of six ARCH-type models, six realized volatility models and two GARCH models, augmented with realized volatility regressors, is assessed. The a-th quantile of the innovation's distribution is estimated with the fully parametric method using either the normal or the skewed student distributions and also with the Filtered Historical Simulation (FHS), or the Extreme Value Theory (EVT) methods. Our analysis is based on two S&P 500 cash index out-of-sample forecasting periods, one of which covers exclusively the recent 2007-2009 financial crisis. Using an extensive array of statistical and regulatory risk management loss functions, we find that the realized volatility and the augmented GARCH models with the FHS or the EVT quantile estimation methods produce superior VaR forecasts and allow for more efficient regulatory capital allocations. The skewed student distribution is also an attractive alternative, especially during periods of high market volatility.

C117: **Realizing smiles: option pricing with realized volatility**
Presenter: **Nicola Fusari**, Kellogg School of Management, United States of America
Co-authors: Fulvio Corsi, Davide La Vecchia

A discrete-time stochastic volatility option pricing model is developed, which exploits the information contained in high-frequency data. The Realized Volatility (RV) is used as a proxy of the unobservable log-returns volatility. We model its dynamics by a simple but effective (pseudo) long-memory process, the Leverage Heterogeneous Auto-Regressive Gamma (HARGL) process. Both the discrete-time specification and the use of the RV allow us to easily estimate the model using observed historical data. Assuming a standard, exponentially affine stochastic discount factor, a fully analytic change of measure is obtained. An extensive empirical analysis of S&P 500 index options illustrates that our approach significantly outperforms competing time-varying (i.e. GARCH-type) and stochastic volatility pricing models. The pricing improvement can be ascribed to: (i) the direct use of the RV, which provides a precise and fast-adapting measure of the unobserved underlying volatility; and (ii) the specification of our model, which is able to accurately reproduce the implied volatility term structure.

C221: **Realized volatility: estimation, forecasting and option trading**
Presenter: **Gabriel Gonzalo Velo**, Univerista di Padova, Italy

The performance of alternative realized estimators of volatility and of different time series forecasting models is analyzed from an economic point of view. We estimate and forecast the volatility of the S&P500 and we define a trading strategy with S&P500 Index options. We focus on contracts with only one week to expiration, and speculate on the future level of volatility. The profit derives from a buy-and-hold trading strategy which is driven by different estimators, that are based on the intradaily returns and ranges. The volatility estimators we use are robust to microstructure noise and jumps. Differently, the considered time series models account for the stylized facts present in financial data such as long memory, non-normality and volatility of volatility. Finally, we identify the estimator and forecasting model that perform economically better in our set up, with strategies that result in positive annualized returns.

CS81 Room Senate COMPUTATIONAL ECONOMETRICS AND APPLICATIONS I Chair: Monica Billio

C560: **Time scales, wavelet realized volatility and jump variation: An empirical investigation for India**
Presenter: **Amlendu Dubey**, Indian Institute of Management Indore, India

A procedure for efficient volatility analysis of the intra-day data in a multi-scale framework is provided. First, we observe that realized volatility estimates are very sensitive to the time-scale of analysis and there is considerable heterogeneity in the shapes of the dependence of realized volatility on the time scale of analysis. Then we analyze further and find that intra-day squared returns – which are building blocks of the realized volatility estimates – itself shows multi-scaling behavior. These two observations show the importance of time-scale analysis of the realized volatility. Then we try to model the realized volatility, taking explicitly different time scales of analysis into account. We try empirical implementation of the concept of wavelet realized volatility with recent advances in wavelet shrinkage methods and try to provide robust time-scale dependent volatility estimates. At the end we observe that high intra-day volatility in the Indian equity markets is intimately related with the days of arrival of unexpected market related news. We document the news and events during the days when we observe relatively higher estimated integrated volatility and jump variation in our period of analysis.

C679: **The continuous wavelet transform: A primer**
Presenter: **Luis Aguiar-Conraria**, University of Minho, Portugal
Co-authors: Maria Joana Soares

Economists are already familiar with the discrete wavelet transform. However, a body of work using the continuous wavelet transform has also been growing. We provide a self-contained summary on the most used continuous wavelet tools. Furthermore, we generalize the concept of simple coherency to partial wavelet coherency and multiple wavelet coherency, akin to partial and multiple correlations, allowing the researcher to move beyond bivariate analysis. Finally, we describe the generalized Morse wavelets, a class of analytic wavelets recently proposed, and compare it to the more common Morlet wavelet. A user-friendly toolbox, with examples, is attached to this paper. We provide three applications with real data. In one of them, we show evidence that corroborates the arguments of Blanchard and Simon about the great moderation, who have argued that the great moderation started well before 1983. In the second application, with the help of cross wavelet analysis, we study synchronism in international stock market returns. Finally, we will illustrate the usefulness of higher order wavelet tools, such as the partial wavelet coherencies and the partial phase-difference, to study the linkages between oil prices and the stock markets after controlling for some other macroeconomic variables.

C776: **Measuring association between random vectors**
Presenter: **Oliver Grothe**, University of Cologne, Germany
Co-authors: Friedrich Schmid, Julius Schnieders, Johan Segers

Five measures of association between two random vectors $\mathbf{X} = (X_1, \ldots, X_p)$ and $\mathbf{Y} = (Y_1, \ldots, Y_q)$ are suggested. They are copula based and therefore invariant with respect to the marginal distributions of the components X_i and Y_j. The measures capture positive as well as negative association of $\mathbf{X}$ and $\mathbf{Y}$. In case $p = q = 1$, they reduce to Spearman's rho. Various properties of these new measures are investigated. Nonparametric estimators, based on ranks, for the measures are derived and their small sample behavior is investigated by simulation. The measures are applied to characterise strength and direction of association of bond and stock indices of five countries over time.

C713: **Analizing if Google helps to predict French youth unemployment**
Presenter: **Frederic Karame**, Universite of Evry Val d'Essonne, France
Co-authors: Yannick Fondeur

According to the rising "Google econometrics" literature, Google queries may help predict economic activity. The aim is to test if these data can enhance predictions for youth unemployment in France. As we have on the one hand weekly series on web search queries, and on the other hand monthly series on unemployment for the 15 to 24-year old group, we use the unobserved components approach in order to exploit all available information. Our models are estimated with a modified version of the Kalman filter taking into account the twofold issues of non-stationarity and multiple frequencies in our data. We find that including Google data improves unemployment predictions relatively to a competing model without search data queries.

CS98 Room S261 FINANCIAL ECONOMETRICS IV Chair: Michael Creel

C802: **Persistent stochastic betas and the statistical properties of stock returns**
Presenter: **Nikolaos Kourogenis**, University of Piraeus, Greece
Co-authors: Nikitas Pittis

The implications of persistent stochastic betas in factor models for the statistical properties of stock returns are analyzed. It is demonstrated that this assumption alone is sufficient to account for the most important stylized facts of stock returns, namely conditional heteroscedasticity, leptokurtosis, weak serial correlation over short horizons, asymptotic independence and aggregational Gaussianity. This means that the assumption of a persistent stochastic beta is sufficient to reproduce the observed non-linear dynamics of stock returns, even in the most extreme case in which both the factor and the non-systematic component follow Gaussian independent processes.

C827: **Nonparametric tests for selecting significant lags**
Presenter: **Mariano Matilla-Garcia**, UNED, Spain

We suggest a new nonparametric statistical test and procedures for selecting relevant lags in the model description of a general linear or nonlinear stationary time series. All the techniques are based on correlation integrals, and can be applied to conditional mean and conditional variance, being valid for heteroscedastic series. The test can be also used as a diagnostic tool.

C779: **Inventory investment and french business cycles**
Presenter: **Frederique Bec**, University of Cergy-Pontoise, France
Co-authors: Melika Ben Salem

Using recent developments in nonlinear time series econometric models, a growing number of empirical works find evidence of a bounce-back effect in real GDP growth rate data.To our knowledge, the origins of this bounce-back phenomenon have hardly been explored so far. Yet, a widely held belief points to the inventory investment behavior as a good candidate. We show the existence of a bounce-back effect in French inventory investment quarterly postwar data while this bounce-back effect is absent in final sales data. This supports the conventional wisdom that inventory investment exacerbates aggregate fluctuations and hence the recent theoretical DSGE models rely either on the production-costs smoothing or on the reduction of fixed order costs, which clearly predict a destabilizing role of inventory investment in the business cycle. By contrast, our empirical findings cast doubt on models based on the stockouts avoidance motive for holding inventories. Finally, our estimated bounce-back threshold autoregressions improve noticeably the short-run forecasting performance, when compared to linear or standard threshold autoregressions.

C629: **Indirect likelihood estimation: Specification testing and model selection**
Presenter: **Michael Creel**, Universitat Autonoma de Barcelona, Spain

Indirect likelihood is a simulation-based estimation method that uses an auxiliary statistic. Specification testing using the familiar J or Sargan test is explored, as well as model selection using the likelihood of the auxiliary statistic, computed using simulation and kernel density estimation. While the J or Sargan test has often been found to suffer from notable size distortions, we find that size is quite accurate when the covariance matrix of the statistic is estimated using simulations. Furthermore, the statistic has good power to detect a misspecified model. Model selection methods similar to information criteria, but based on the likelihood function of the auxiliary statistic, are also found to be useful for selecting the best model out of a set of candidates. The results are illustrated using nonlinear DSGE models, as well as models from genetics.

CS48 Room Bloomsbury CONTINUOUS TIME FINANCIAL MODELS Chair: Leopold Soegner

C307: **Parameter estimation of Heston type stochastic volatility models**
Presenter: **Justinas Pelenis**, Institute for Advanced Studies, Vienna, Austria
Co-authors: Leopold Soegner

The performance of Heston type stochastic volatility models in multivariate and multi-asset settings is examined when using joint stocks and options data. We perform (Bayesian) parameter estimation of a wide array of affine class models with various multi-factor, volatility and correlation specifications. This is done in order to evaluate both model performance and the computational ease and precision of parameter estimation in those specific settings. The contribution is to extend some of the current estimation techniques to the multi-asset setting and to determine the strengths and pitfalls of competing approaches in both simulated and real-world datasets especially when options data is incorporated in the estimation of stock price dynamics.

C353: **Continuous-time hidden Markov models: robust filters, estimation and portfolio optimization**
Presenter: **Joern Sass**, University of Kaiserslautern, Germany

A multi-dimensional continuous-time model in which the observation process is a diffusion with drift modeled as a continuous time, finite state Markov process is considered. This underlying drift process is not observable and can be filtered from the observations. An EMM algorithm for parameter estimation is based on this filter and on filters for related processes. Robust versions, in the sense that the filters depend continuously on the observations, are well known. We analyze the dependency of parameter estimation on the use of these robust versions and compare with other estimation methods like MCMC methods. In a market in which asset prices follow these dynamics and an additional riskless asset is available, we look at an investor maximizing expected utility of her portfolio at some terminal time. We study the impact on parameter estimation and optimization for robust versions compared to non-robust versions and discuss consequences for model choice.

C254: **Method of moments estimation and affine term structure models**
Presenter: **Leopold Soegner**, Institute for Advanced Studies, Austria

Parameter estimation of affine term structure models is investigated by means of the generalized method of moments. Recent literature derived a method to calculate the moments of affine stochastic processes. Here the moments are available in almost closed form. The calculation of the moments requires the evaluation of a matrix exponential. By matching the moments implied by the model with some sample moments we derive

parameter estimates. Two step GMM will be applied to increase the efficiency of the estimator. Asymptotics for the estimated parameters are provided. The estimation procedure is applied to simulated and empirical interest rate data.

CS57 Room S264 REAL-TIME MODELLING WITH MIXED FREQUENCIES Chair: Tommaso Proietti

C853: **Nowcasting euro-area GDP growth using a mixed frequency Global VAR model**
Presenter: **Silvia Lui**, National Institute of Economic and Social Research, United Kingdom
Co-authors: James Mitchell

Quarterly Euro-area GDP data are published at a lag of about 45 days. This hinders policymaking in real-time. Accordingly, the quality of GDP nowcasts produced at a shorter lag by exploiting available monthly indicator variables is evaluated. Given that Euro-area GDP is the aggregation of national GDP data, and since publication lags vary both by country and by indicator variable, we extend previous work by accommodating both cross-country and cross-variable dependencies when nowcasting. This is undertaken by extending the Global Vector-Autoregressive model both to accommodate mixed-frequency data and to reflect the GDP quarterly aggregation constraint. Nowcasts for the Euro-area are then produced as within-quarter data accumulates reflecting the differing publication lags and informational content of GDP, IP and qualitative survey data across the Euro-area countries. Using a novel real-time dataset for the Euro-area, and its constituent countries, in real-time simulations we find that the proposed multivariate approach, by efficiently conditioning on available information, delivers more accurate nowcasts than either nowcasting the Euro-area aggregate using aggregated indicator variables or aggregating national/disaggregate nowcasts. Our approach also has advantages over the "mixed-approach", which involves using disaggregate indicators, as well as aggregated ones, to nowcast Euro-area GDP growth.

C879: **A comparison of bottom-up approaches and direct forecasts of German GDP in a data-rich environment**
Presenter: **Rolf Scheufele**, Swiss National Bank (SNB), Switzerland
Co-authors: Katja Drechsel

A method for making efficient use of indicators to conduct early estimates of GDP growth in Germany is presented. We employ MIDAS regressions to circumvent the mixed frequency problem and use pooling techniques to effectively summarize the information content of the various indicators. More specifically, we investigate whether it is better to disaggregate GDP (either via total value added of each sector or by the expenditure side) or whether a direct approach is more appropriate when it comes to forecast GDP growth rates. Our approach combines a large set of monthly and quarterly coincident and leading indicators and takes into account the respective publication delay. In a simulated out-of-sample experiment we evaluate the different modeling strategies conditional on the exact state of information and depending on the model averaging technique. The proposed approach is computationally simple and can be easily implemented as a nowcasting tool. This method also allows us to retrace the driving forces of the forecast which enables the interpretability of the forecast outcome.

C666: **On the prediction of GDP revisions: Evidence for Switzerland**
Presenter: **Boriss Siliverstovs**, ETH Zurich, Switzerland

A small-scale mixed-frequency dynamic factor model is constructed using data for Switzerland. The factor model combines the quarterly GDP growth rate and the monthly survey indicators. We evaluate the forecasting performance of the model during the period of the recent financial crisis when accurate information on the current stance of the economy is especially in high demand. We demonstrate that this factor model produces more accurate forecasts than the alternative benchmark models such as a random-walk model and a first-order autoregressive model. More importantly, the factor model produces more accurate forecasts of the revised rather than first-published estimates of the GDP growth rate. We demonstrate that this remarkable finding could be explained by the fact that the factor model is useful for predicting not only directions of future GDP revisions but also their magnitude, at least during the period under scrutiny. We conclude that there seems to be a scope for improvement of how estimates of the GDP growth rate are produced in Switzerland: in particular, in the direction of reducing volatility of subsequent revisions.

C824: **Testing the functional restrictions on parameters in the MIDAS regressions**
Presenter: **Virmantas Kvedaras**, Vilnius University, Lithuania
Co-authors: Vaidotas Zemlys

Some issues of correct specification of the mixed data sampling regressions are addressed. The main contribution is the proposed test for the statistical acceptability of a functional constraint which is imposed in this kind of models on the infinite number of parameters. The asymptotic distribution of the test statistic is derived and the small sample performance is evaluated using Monte Carlo simulations.

CS91 Room Gordon COMPUTATIONAL ECONOMETRICS Chair: Alessandra Amendola

C648: **Efficient high-dimensional importance sampling in mixture frameworks**
Presenter: **Tore Selland Kleppe**, University of Bergen, Norway
Co-authors: Roman Liesenfeld

We provide high-dimensional and flexible importance sampling procedures for the likelihood evaluation of dynamic latent variable models involving finite or infinite mixtures leading to possibly heavy tailed and/or multi-modal target densities. Our approach is based upon the efficient importance sampling (EIS) and exploits the mixture structure of the model when constructing importance sampling distributions as mixture of distributions. The proposed mixture EIS procedures are illustrated with ML estimation of a student-t state space model for realized volatilities and a stochastic volatility model with leverage effects and jumps for asset returns.

C664: **Automated vine copula calibration using genetic algorithms**
Presenter: **Gregor Weiss**, TU Dortmund University, Germany
Co-authors: Miriam Padberg

During the last few years, research on vine copulas (pair-copula constructions) and their use in quantitative risk management has increased steadily. Though vine copulas allow for an extremely flexible modeling of the dependence structure of a high-dimensional random vector, their use is severely hampered by the problems of a) finding a suitable vine tree structure and b) selecting suitable parametric copula candidates as the vine's bivariate building blocks. Recent studies in this field have tried to tackle the second problem by truncating vines and selecting the pair-copulas sequentially. We propose two solutions for both mentioned problems: First, we employ evolutionary algorithms for minimizing the vine model's global AIC in order to select the pair copulas. In a second step, we then use a set of genetic algorithms in order to choose the vine's tree structure. Both steps are then combined yielding a fully automated calibration of a vine copula model. The performance of both the sequential heuristic as well as our proposed method is analyzed in a simulation study and exemplified in a risk management application for estimating portfolio VaRs and ES.

C671: **Variable selection in competing risks model for corporate exit**
Presenter: **Marialuisa Restaino**, University of Salerno, Italy
Co-authors: Alessandra Amendola, Luca Sensini

Firms may leave the market in several ways such as merger, voluntary liquidation or bankruptcy. Each form of exit can be caused by different factors. Selection of the best set of input variables influencing the exit is essential in predicting the default risk. Our aim is to investigate the performance of different variable selection methods, focusing on a statistical procedure suitable for the competing risks model. In this theoretical setting, the same variables might have different degrees of influence on the risks due to multiple causes and this effect has to be taken into account in the choice of the "best" subset. The proposed procedure, based on shrinkage techniques, has been evaluated by means of simulation study and empirical analysis on a data-set of financial indicators computed from a sample of industrial firms annual reports. The reached findings support the need to opportunely select the subset of explanatory variables also in terms of the predictive accuracy of the proposed default risk models.

C678: **Optimal and data driven smoothing for simuation-based inference**
Presenter: **Patrick Richard**, Universite de Sherbrooke, Canada

The problem of computing simulated P values based on a smooth estimate of the test statistics' null distribution rather than on the empirical density function is investigated. Such a procedure can be useful when the cost of performing a large number of simulations is high or when the number of simulations cannot be controlled by the econometrician. This issue was first explored for the bootstrap in a previous paper, where using extensive Monte Carlo simulations, it was found that Kernel smoothing allows one to smooth the null rejection probability of bootstrap tests with respect to the number of replications performed, and provides better power. That work is extended in three ways. First, we explore the impact of some characteristics of the statistics' distribution on the optimal smoothing parameter. We find that skewness and kurtosis are quite important quantities. Second, we propose a data driven smoothing parameter selection that takes into account the level of the test being performed. Finally, we provide a practical example where smoothed p-values may outperform regular ones.

CS05 Room Torrington CONTRIBUTIONS IN MULTIVARIATE FINANCIAL TIME SERIES Chair: Alain Hecq

C632: **Co-summability: From linear to non-linear co-integration**
Presenter: **Vanessa Berenguer-Rico**, Universidad Carlos III de Madrid, Spain
Co-authors: Jesus Gonzalo

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

C686: **Time varying dependence in high dimensional financial data sets**
Presenter: **Jakob Stoeber**, Technische Universitat Munchen, Germany
Co-authors: Claudia Czado

The modeling of dependencies between different financial assets has for a long time been restricted to correlations and the multivariate normal distribution. Arguably, misperceptions about extreme interdependencies have also been an important element of the recent financial crisis. We present techniques to study inhomogeneity in the dependence structure of high dimensional financial time series. The models are based on regular vine (R-vine) copulas, which are constructed entirely from bivariate copulas as building blocks. Doing so, the richness of the class of bivariate copulas is exploited to create an equally flexible high dimensional dependence model. Thus, we account for stylized properties of multidimensional financial data such as asymmetric dependencies and tail dependencies. R-vine copula distributions are then applied in a new copula-GARCH model of DCC type, allowing for the dependencies to vary over time. We compare this naive approach to a Markov Switching (MS) R-vine model and demonstrate that variations in dependence can be detected in various financial data sets. They are usually closely interrelated with periods of market stress. In such times the Value at Risk of an asset portfolio is significantly underestimated when changes in the dependence structure are ignored.

C795: **General hybrid MSV–MGARCH models of multivariate volatility - Bayesian analysis**
Presenter: **Jacek Osiewalski**, Cracow University of Economics, Poland
Co-authors: Krzysztof Osiewalski

A new class of volatility specifications, the hybrid multivariate stochastic variance – GARCH (MSV–MGARCH) models, was previously introduced. The conditional covariance matrix is a product of a univariate latent process and a matrix with a simple MGARCH structure (Engle's DCC, scalar BEKK). The aim was to parsimoniously describe volatility of a large group of assets. Their hybrid models, similarly as other specifications from the MSV class, require the Bayesian approach equipped with efficient MCMC simulation tools. The numerical effort to pay the hybrid MSF–SBEKK(1,1) type I model seems very useful due to its good fit and ability to jointly cope with as many as 50 assets. However, one latent process may be insufficient in the case of a highly heterogenous portfolio. We propose a general hybrid MSV–MGARCH structure that uses as many latent processes as there are relatively homogenous groups of assets. We present full Bayesian inference for such models and suggest MCMC simulation strategy. The proposed approach is used to jointly model volatility on very different markets. We formally compare joint modelling to individual modelling of volatility on each market.

C774: **Permanent transitory decompositions under short and long-run present value model restrictions**
Presenter: **Alain Hecq**, Maastricht University, Netherlands
Co-authors: Joao Victor Issler

We revisit present-value (PV) relationships in economics in the context of vector-autoregressive (VAR) models. Whereas the early literature on PV relationships has exploited the existence of long-run restrictions in VAR parameters (cointegration), our focus is the existence of short-run restrictions (common-cyclical features) in VAR parameters. Several novel contributions are presented, but two stand out. First, we show that PV relationships entail a weak-form common feature relationship as in a model proposed in a previous paper, and also a polynomial serial-correlation common feature relationship. That represents restrictions on dynamic models (VARs) which allow new tests for the existence of PV relationships to be developed. Because these relationships occur mostly with financial data, we propose in addition to usual canonical correlation procedures, tests based on generalized method of moment (GMM) estimates, where it is straightforward to propose robust tests in the presence of heteroskedasticity and serial-correlation in the data. Their good performance is confirmed in a Monte-carlo exercise. Second, in the context of asset pricing, we propose applying a permanent-transitory (PT) decomposition under both common-trend and common cycle restrictions.The techniques discussed

here were applied to historical annual long- and short-term interest rates as well as on stock prices and dividends for the US.

CS43 Room Bedford CONTRIBUTIONS IN DERIVATIVE PRICING Chair: Paolo Foschi

C191: **Semi-nonparametric estimation of the call price surface under no-arbitrage constraints**
Presenter: **Matthias Fengler**, St. Gallen (HSG), Switzerland
Co-authors: Lin-Yee Hin

When studying the economic content of cross sections of option price data, researchers either explicitly or implicitly view the discrete ensemble of observed option prices as a realization from a smooth surface defined across exercise prices and expiry dates. Yet despite adopting a surface perspective for estimation, it is common practice to infer the option pricing function, and consequently each state-price density or empirical pricing kernel, for each expiry date separately, slice by slice. A semi-nonparametric estimator for the entire call price surface which yields a parsimoniously parametrized representation of the data is suggested. The estimator is based on tensor-product B-splines of arbitrary degrees and respects no-arbitrage constraints. Consequently one directly obtains families of arbitrarily smooth state-price densities across all time horizons, and – combined with a suitable choice of the family of physical densities – surfaces of the pricing kernel.

C736: **Capturing skewness and kurtosis by fitting the QQ-plot: A simple approach with an application to option pricing**
Presenter: **Antoni Vaello-Sebastia**, Universitat de les Illes Balears, Spain
Co-authors: Unai Ansejo, Aitor Bergara

A new distribution based on a polynomial expansion of the Gaussian quantiles, which nests the well known Cornish-Fisher expansion, is presented. Using third-order polynomials we obtain analytical expressions for the distribution and density functions, also yielding a skweness-kurtosis region much wider than the usually considered Gram-Charlier expansion. We discuss three different estimation methodologies using stock index data. We also apply our density for option pricing and hedging obtaining closed-form formulae. Finally, we conduct an empirical application with S&P 500 options data, showing that our option pricing model outperforms other models based on semi-nonparametric densities of similar order.

C821: **Option pricing via nonparametric Esscher transform**
Presenter: **Camila Epprecht**, Puc-Rio and Paris 1-Pantheon-Sorbonne, France
Co-authors: Manoel Pereira, Alvaro Veiga

A nonparametric version of the Esscher transform for risk neutral option pricing is introduced. Traditional parametric methods require the formulation of an explicit risk-neutral model and are operational only for a few probability density functions for the returns of the underlying. We make only mild assumptions on the price kernel and there is no need for the formulation of the risk-neutral model for the returns. First, we simulate sample paths for the returns under the historical distribution P. Then, based on the Esscher transform, the sample is reweighted, giving rise to a risk-neutralized sample from which derivative prices can be obtained by a simple average of the pay-offs of the option to each path. We compare our proposal with some traditional pricing methods in several exercises with artificial and real data.

C861: **Pricing of American options in local volatility models**
Presenter: **Paolo Foschi**, University of Bologna, Italy

Recently, several authors have proposed different analytical approximations for pricing European options when the underlying model exhibits a non-constant volatility and/or jumps in its trajectory. Some of these approximations are used to efficiently solve the free boundary problem related to the pricing of American options. The price of an American option is formulated as the solution to an optimization problem given in terms of either Barrier or Lookback Option prices that can be computed by means of the above explicit approximations. A preliminary numerical comparison of the proposed method is presented.

CS66 Room Court CONTRIBUTIONS TO BAYESIAN ECONOMETRICS Chair: Rachida Ouysse

C516: **Benchmark priors revisited: On adaptive shrinkage and the supermodel Eect in Bayesian model averaging**
Presenter: **Stefan Zeugner**, Universite Libre de Bruxelles, Belgium
Co-authors: Martin Feldkircher

Predominant in the Bayesian Model Averaging literature, default prior frameworks such as fixing Zellner's g tend to concentrate posterior mass on a tiny set of models. This supermodel effect is demonstrated and it is proposed to address it by a hyper-prior, whose data-dependent shrinkage adapts posterior model distributions to data quality. Analytically, existing work on the hyper-g-prior is complemented by posterior expressions essential to full Bayesian analysis and to sound numerical implementation. A simulation experiment illustrates the implications for posterior inference. Finally, the merits of the hyper-g-prior both in terms of predictive performance and stability of posterior results are demonstrated using a prominent growth data set employing four different vintages of the Penn World Table income data.

C596: **Bayesian moving average and principal components forecasts for large dimensional factor models**
Presenter: **Rachida Ouysse**, University of New South Wales, Australia

To overcome the challenges of dimensionality, many forecast approaches proceed by somehow reducing the number of predictors. Principal component regression (PCR) approach proposes computing forecasts as projection on the first few principal components of the predictors. Bayesian model averaging (BMA) approach combines forecasts to extract information from different possible relationships between the predicted variable and the predictor variables. Recent findings suggest there are theoretical and practical reasons to connect these two strands of the literature. Empirical evidence for connecting these two seemingly different approaches to forecasting is provided. We study the performance of Bayesian model averaging as a forecasting method based on large panels of time series as an alternative to principal components. We show empirically that these forecasts are highly correlated implying similar out-of-sample mean-square forecast errors. Applied to forecasting Industrial production and inflation in the United States, we find that the set of variables deemed *informative* changes over time which suggests temporal instability. The results can also be driven by the nature of the macroeconomic data which is characterized by collinearity and that the variable selection is sensitive to minor perturbations of the data. The empirical results serve as a preliminary guide to understanding the behavior of BMA under double asymptotics, when the cross-section and the sample size become large.

C738: **Bayesian semiparametric dynamic Nelson-Siegel model**
Presenter: **Cem Cakmakli**, University of Amsterdam, Netherlands

A Bayesian semiparametric dynamic Nelson-Siegel model where the error distributions of the factors are modeled according to a Dirichlet process mixture is proposed. An efficient and computationally tractable algorithm is implemented to obtain Bayesian inference. The semiparametric structure of the factors enables us to estimate the density of the factors and to capture various forms of non-normalities such as fat tails, skewness and nonlinear dependence between factors using a unified approach. The potential of the proposed framework is examined using simulated data and US bond yields data. The results show that our framework can identify two different periods with distinct characteristics. The framework is also

able to capture the non-linear dependence structure between the factors. While the relatively stable years of late 1980s and 1990s are comprising the first period, the second period captures the years of severe recessions including the recessions of 1970s and 1980s and the recent recession of 2007-2008 together with highly volatile periods of Federal Reserve's monetary policy experiments in the first half of 1980s.

C912: **Interaction terms and restricted model spaces in Bayesian model averaging**
Presenter: **Mathias Moser**, Vienna University of Economics and Business Administration, Austria

When dealing with Bayesian Model Averaging in large model spaces, the selection of "reasonable" models is a difficult task. So-called dilution priors have gained importance and can be used to adjust for similar models. This is especially relevant for model spaces that contain interacted covariates. In such a setting models which only include interacted variables can catch the effect of the omitted parent variable. To compensate for such models the literature proposes the use of dilution priors according to the strong or weak heredity principle. Accordingly a strong heredity prior would exclude models with interaction but no parent variable completely from the model space (assign them a zero model probability). This very informative prior has been criticised as being too strict in recent literature. A weak heredity prior which penalizes such uncommon models, but does not exclude them from the model space entirely is proposed. Both strong and weak heredity priors are compared using simulated data and several benchmark models. Furthermore, the predictive performance of the weak heredity prior is evaluated.

CS75 Room Woburn CONTRIBUTIONS IN TIME SERIES AND PANEL DATA ECONOMETRICS Chair: Jean-Pierre Urbain

C695: **Individual size and time period effects on the unconditional fixed effects negative binomial regression estimator**
Presenter: **Wan Fairos Wan Yaacob**, Universiti Teknologi MARA, Malaysia
Co-authors: Mohamad Alias Lazim, Bee Wah Yap

The unconditional maximum likelihood method has been widely acknowledged as an alternative method to conditional estimator in estimating fixed-effects model for panel count data. Substantial literatures had discussed the ability of this method in estimating the true fixed effects with better sampling properties though it is subjected to incidental parameter problem. The effect of individual size and time dimension on the consistency of the unconditional fixed-effects Negative Binomial (FENB) model parameter estimates is assessed using Monte Carlo simulation study. Simulation results show that the performance of the unconditional fixed effects estimator deteriorates substantially for small individual size. For a fixed individual size, better estimates were obtained when the time period is longer.

C750: **How banks' funding costs affect interest margins**
Presenter: **Arvid Raknerud**, Statistics Norway, Norway
Co-authors: Bjorn Helge Vatne, Ketil Rakkestad

We use a dynamic factor model and a detailed panel data set with quarterly accounts data on all Norwegian banks to study the effects of banks' funding costs on their retail rates. Banks' funds are categorized into two groups: customer deposits and long-term wholesale funding (market funding from private and institutional investors including other banks). The cost of market funding is represented in the model by the three-month Norwegian Inter Bank Offered Rate (NIBOR) and the spread of unsecured senior bonds issued by Norwegian banks. Our estimates show clear evidence of incomplete pass-through: a unit increase in NIBOR leads to an approximately 0.8 increase in bank rates. On the other hand, the difference between banks' loan and deposit rates is independent of NIBOR. Our findings are consistent with the view that banks face a downward-sloping demand curve for loans and an upward-sloping supply curve for customer deposits.

C875: **A panel analysis of the Fisher effect with an unobserved I(1) world real interest rate**
Presenter: **Gerdie Everaert**, Ghent University, Belgium

The Fisher effect states that inflation expectations should be reflected in nominal interest rates in a one-for-one manner to compensate for changes in the purchasing power of money. Despite its wide acceptance in theory, much of the empirical work fails to find strong favorable evidence. The Fisher effect in a panel of 15 OECD countries over the period 1972-2007 is examined. Standard panel cointegration tests suggest that nominal interest rates and inflation cointegrate but with an estimated slope coefficient on inflation which is significantly bigger than 1. Using the PANIC approach we detect a small non-stationary common component in the error terms of this alleged cointegrating relation. This implies the regression results to be spurious. One possible interpretation for the detected non-stationary common factor is that it reflects permanent common shifts in the real interest rate induced by e.g. shifts in time preferences, risk aversion and the steady-state growth rate of technological change. We next control for an unobserved non-stationary common factor using both the Common Correlated Effects (CCE) estimation approach and the Continuously Updated bias-corrected (CupBC) estimator. The estimated slope coefficient on inflation is now found to be close to, and insignificantly different from, 1.

C812: **On the applicability of the sieve bootstrap in time series panels**
Presenter: **Jean-Pierre Urbain**, Maastricht University, Netherlands
Co-authors: Stephan Smeekes

The applicability and the validity of the univariate sieve bootstrap in cross-sectionally dependent time series panels with finite cross-sectional dimension (N) is considered. The multivariate sieve bootstrap is not feasible in a panel even with a moderate N due to the curse of dimensionality. On the other hand, the univariate sieve bootstrap has been applied successfully in panel data with contemporaneous cross-correlation. Using the final equations of a VARMA model, we show that the univariate sieve bootstrap can be applied in a much more general class of panel data that allows for various forms of cross-sectional dependence, including Granger causality relations and common factors. We focus in particular on time series panels with unit roots, including the special case where the members of the panel are cointegrated. We provide theoretical results on the validity of the sieve bootstrap for these models. A simulation study shows that the performance of the sieve bootstrap in finite samples is satisfactory in comparison to other bootstrap methods that appear to be better suited to deal with cross-sectional dependence.

Monday 19.12.2011 10:55 - 12:35 Parallel Session N – ERCIM

ES05 Room B34 BAYESIAN NONPARAMETRIC PRIORS **Chair: Igor Pruenster**

E314: **A Bayesian nonparametric model for ranking**
Presenter: **Francois Caron**, INRIA Bordeaux Sud Ouest, France
Co-authors: Yee Whye Teh

A Bayesian nonparametric model to model partial rankings of an infinite number of individuals is presented. Each individual is assigned a positive skill rating parameter that is inferred from partial rankings. The model is able to handle a potentially infinite number of individuals. Markov Chain Monte Carlo algorithms for inference and several applications are shown.

E937: **Bayesian nonparametric covariance regression**
Presenter: **Emily Fox**, University of Pennsylvania, United States of America
Co-authors: David Dunson

Although there is a rich literature on methods for allowing the variance in a univariate regression model to vary with predictors, time and other factors, very little has been done in the multivariate case. Our focus is on developing a class of nonparametric covariance regression models, which allow an unknown p x p covariance matrix to change flexibly with predictors. The proposed modeling framework induces a prior on a collection of covariance matrices indexed by predictors through priors for predictor- dependent loadings matrices in a factor model. In particular, the predictor-dependent loadings are characterized as a sparse combination of a collection of unknown dictionary functions (e.g, Gaussian process random functions). The induced covariance is then a (sparse) quadratic function of these dictionary elements. Our proposed framework leads to a highly-flexible, but computationally tractable formulation with simple conjugate posterior updates. Theoretical properties are discussed and the methods are illustrated through simulations studies and the Google Trends flu dataset.

E852: **Vectors of dependent random probabilities**
Presenter: **Antonio Lijoi**, University of Pavia, Italy
Co-authors: Bernardo Nipoti

A significant part of the recent literature in Bayesian nonparametric inference has been focusing on the proposal of models that can accommodate for more general forms of dependence than exchangeability. We will present examples of dependent completely random measures that can be characterized in terms of a sequence of canonical correlations. These, under suitable transformations, are used to define dependent random probabilities that identify the directing measure for sequences of partially exchangeable observations. A few applications to the determination of Bayesian inferences, with the aid of suitable MCMC sampling schemes, are discussed. An illustration with random hazard rate mixtures for the analysis of two-sample survival analysis data is finally provided.

E495: **Constructing exchangeable priors via restriction**
Presenter: **Sinead Williamson**, Carnegie Mellon, United States of America
Co-authors: Zoubin Ghahramani, Steven MacEachern

Exchangeable distributions such as the Chinese Restaurant Process and the Indian Buffet Process provide flexible and computationally appealing priors for nonparametric versions of finite mixture models and finite latent variable models. However, the form of the partition structure of the Chinese Restaurant Process, or the latent variable selection of the Indian Buffet Process, may be a poor match for our data. We show how we can develop a range of exchangeable priors over partitions and matrices that can be adapted to suit our prior beliefs about our data, and demonstrate some relationships with existing models.

ES07 Room B18 MIXTURE MODELS **Chair: Christian Hennig**

E676: **Model-based clustering via copulas**
Presenter: **Ioannis Kosmidis**, University College London, United Kingdom
Co-authors: Dimitris Karlis

The majority of model-based clustering methods is based on mixtures of normal models and generally on models that allow the construction of elliptical clusters. The restriction to elliptical cluster shapes usually leads to problems in real data applications because there might be the case that more than one mixture components may be required to represent a single group in the data. We propose the use of models that result as mixtures of distributions that are defined via copulas. This kind of mixtures have several advantages: i) one can construct clusters with shapes that are far from elliptical, ii) the use of copulas allows us to determine the marginal properties of the clusters separately from the within-cluster dependence structure, iii) one may cluster data according to the dependence structure between the available variables by selecting the appropriate copula from a dictionary of copulas, iv) clustering of non-continuous data and mixed-mode data is easy while remaining in the framework of model-based clustering. We apply model-based clustering via copulas in some real-data applications focusing on the aforementioned advantages. Furthermore, we discuss the challenges associated with estimation and model-selection within the new framework of mixtures of copula-based distributions.

E578: **Maximum likelihood estimation in the logistic regression model with a cure fraction**
Presenter: **Jean-Francois Dupuy**, National Institute of Applied Sciences of Rennes, France
Co-authors: Aba Diop, Aliou Diop

Logistic regression is widely used in medical studies to investigate the relationship between a binary response variable Y and a set of potential predictors X. The binary response may represent, for example, the occurrence of some outcome of interest ($Y = 1$ if the outcome occurred and $Y = 0$ otherwise). We consider the problem of estimating the logistic regression model with a cure fraction. A sample of observations is said to contain a cure fraction when a proportion of the study subjects (the so-called cured individuals, as opposed to the susceptibles) cannot experience the outcome of interest. One problem arising then is that it is usually unknown who are the cured and the susceptible subjects, unless the outcome of interest has been observed. In this setting, a logistic regression analysis of the relationship between X and Y among the susceptibles is no more straightforward. We develop a maximum likelihood estimation procedure for this problem, based on the joint modeling of the binary response of interest and the cure status. This procedure results in a mixture model for zero-inflated data. We investigate the identifiability of this model and the asymptotic properties of the proposed estimator. We describe a simulation study investigating the finite-sample behavior of this estimator.

E709: **Normal discriminant analysis via the 2-terms eigenvalue decomposition**
Presenter: **Luca Bagnato**, Universita di Catania, Italy
Co-authors: Francesca Greselin, Salvatore Ingrassia, Antonio Punzo

The classification problem in the case that groups are known and both labeled and unlabeled data are available is analyzed. The classification rule is

derived using Gaussian mixtures, with covariance matrices fixed according to a multiple testing procedure, which allows us to choose among four alternatives: heteroscedasticity, homometroscedasticity (common eigenvalue matrices between groups), homotroposcedasticity (common eigenvector matrices between groups), and homoscedasticity. The mixture models are then fitted using either only the labeled data or all available ones (labeled and unlabeled) adopting the EM and the CEM algorithms in the latter case. Applications on real data are provided in order to show the classification performance of the proposed procedure.

E770: **Using mixture of Gamma distributions for Bayesian analysis in an M/G/1 queue with optional second service**
Presenter: **Abdolreza Mohammadi**, University of Groningen, Netherlands
Co-authors: Mohammad Reza Salehi-Rad, Ernst Wit

We deal with Bayesian inference for an M-G-1 queuing system with optional second re-service. The semi-parametric model based on a mixture of Gamma distributions is considered to approximate both the general service and re-service time densities. Given observations of the system, the Bayesian procedure based on birth-death MCMC method is proposed to estimate system parameters, protective distributions and some performance measures related to this queuing system such as stationary system size and waiting time. The approach is illustrated with a simulated study.

ES34 Room B36 INVARIANT COORDINATE SELECTION AND DIMENSION REDUCTION Chair: Hannu Oja

E273: **Multivariate outliers detection using ICS**
Presenter: **Anne Ruiz-Gazen**, Universite Toulouse 1, France

Transformations of data sets in invariant coordinate systems using two scatter matrix estimators S_1 and S_2 are considered. Such transformations can be very powerful for multivariate outliers detection. In particular if S_1 is the usual scatter matrix estimator and S_2 is a robust scatter estimator, outliers are likely to appear on the coordinates associated with the largest eigenvalues of $S_1 S_2^{-1}$. We propose to compare the use of different robust estimators on several examples. The choice of the number of eigenvalues to be considered as relevant for detecting outliers will also be discussed in this context. Finally, we will compare this proposal with other multivariate outliers detection methods such as robust principal component analysis and discuss the advantages of ICS.

E358: **Supervised invariant coordinate selection**
Presenter: **Klaus Nordhausen**, University of Tampere, Finland
Co-authors: Eero Liski, Hannu Oja

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 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), allowing the second scatter matrix S_2 to be supervised; supervised ICS can then be used in supervised dimension reduction. Several families of supervised scatter matrices are discussed, and their use in supervised dimension reduction is illustrated with an example and simulations.

E376: **Invariant coordinate selection revisited: a symmetry group perspective**
Presenter: **Frank Critchley**, Open University, United Kingdom
Co-authors: Ana Pires, Conceicao Amado

The Invariant Coordinate Selection, or ICS, was introduced in a recent work as a general method for exploring multivariate data by comparing different estimates of multivariate shape. Examples of the method performing well for a wide range of problems, extending beyond the limits of existing theoretical support, are reported in the literature. Motivated by this, we provide complementary ICS theory based on the relevant symmetry group. A Fisher Symmetry condition is introduced for which elliptical symmetry is not required, yet under which a subset of the invariant coordinates is shown to correspond to the Fisher Linear Discriminant subspace, class identifications of data points remaining unknown. Again, a Symmetric Component Analysis (SCA) model is introduced in which independence is not required, yet under which the invariant coordinates are seen to correspond to the symmetric components. Links with corresponding results for Independent Component Analysis (ICA) are established, relevant symmetry groups again playing a key role. Illustrative examples are given and further developments briefly indicated.

E470: **Testing the equality of the roots in ICS**
Presenter: **David Tyler**, Rutgers, The State University of New Jersey, United States of America
Co-authors: Jue Wang

Invariant co-ordinate selection (ICS) is useful for detecting non-elliptical structures in multivariate data. The idea underlying ICS is that all scatter functionals are proportional to each other under elliptical distributions, but not under non-elliptical distributions. Consequently, ICS considers the eigenvalue-eigenvector decomposition of one estimate of multivariate scatter relative to another. Under particular non-elliptical models, such as mixtures of elliptical distributions and independent component models, ICS can be used to extract the important features of these models. An important first step in applying ICS is to judge whether or not all the differences in the eigenvalues in the ICS decomposition are due to chance or are due to the underlying model being non-elliptical. Tests for the equality of the eigenvalues or for a subset of the eigenvalues in an ICS decomposition are presented, and their asymptotic null distributions are derived. The tests depend upon which two estimates of scatter matrices one uses in the ICS method. To help ascertain the effect of the scatter matrices used in ICS, we present the asymptotic properties of ICS under local asymmetric alternatives to elliptically symmetric distributions. These local alternatives include mixtures of elliptical distributions and skewed elliptical distributions.

ES42 Room G16 POINT PROCESSES: MODELLING AND FORECASTING Chair: Paula Rodriguez Bouzas

E072: **Modelling the occurrence of extreme heat events using a bivariate Poisson process**
Presenter: **Ana C. Cebrian**, University of Zaragoza, Spain

It is shown how a multivariate Poisson process can be used in the modelling of the occurrence of extreme events in a stochastic bivariate framework. This approach is applied to characterize the occurrence of extreme heat events in daily maximum and minimum temperatures. The common Poisson shock model (CPSM) is a point process with univariate marginal Poisson processes that takes into account the dependence between them. The estimation of a CPSM can be highly simplified using an equivalent representation based on three independent Poisson processes. This representation also allows the model to be generalised for time varying parameters. In the analysis of the heat event generating process, the parameters are represented by parametric functions of temperature and other covariates to take into account the influence of global warming and seasonal effects. Some tools for the validation analysis are suggested, and the main contribution is the development of a test of independence between non-homogeneous Poisson processes, that is necessary to carry out this validation. The model also allows us to obtain projections using temperature series generated by general circulation models in different climate change scenarios.

E120: **A J-function for inhomogeneous point processes with applications**
Presenter: **Marie-Colette van Lieshout**, CWI, Netherlands

The analysis of data in the form of a map of (marked) points often starts with the computation of summary statistics. Some statistics are based on inter-point distances, others on the average number of points in sample regions, or on geometric information. In the exploratory stage, it is usually assumed that the data constitute a realisation of a stationary point process and deviations from a homogeneous Poisson process are studied to suggest a suitable model. Although stationarity is a convenient assumption, especially if - as is often the case - only a single map is available, in many areas of application, though, heterogeneity IS present. We describe summary statistics that are able to accommodate spatial and/or temporal inhomogeneity and illustrate the approach on data of Pakistani earthquakes.

E438: **Tests for outliers in power-law and inverse Gaussian models for event times**
Presenter: **Chrys Caroni**, National Technical University of Athens, Greece
Co-authors: Dimitris Stogiannis

It is important to have objective tests for detecting outliers in data. Unusual values may have strong influence on parameter estimates, and even on the choice of model, but their subjective removal is an unreliable procedure. We examine here outlier testing using likelihood ratio in two situations, with application chiefly to time-to-event data. In one case, the power law model $\alpha\beta t^{\beta-1}$ is used to model the intensity function of an inhomogeneous Poisson process for the occurrence of events in each of a set of units or systems. We develop a test for an outlying value of the growth parameter β. In the second case, we develop tests for outlying values of the parameters λ and μ of the inverse Gaussian distribution. This distribution can model a lifetime as the first hitting time on a boundary of an underlying Wiener process. Tests for these parameters are equivalent to testing for outliers in the drift and starting level of the underlying process In general, in order to apply the tests for β and λ it is necessary to transform exactly F distributed statistics to approximate standard normality. The procedures are shown to work well.

E346: **Behavior comparison of turning points of stock prices**
Presenter: **Nuria Ruiz-Fuentes**, University of Jaen, Spain
Co-authors: Paula R. Bouzas, George Atsalakis

The turning points of a stock price, the trend changing points in its daily closing values, are studied as a counting process. This process was well modelled as a Cox process. A Cox process goodness-of-fit test is used to compare the behavior between the turning points of different stock prices as well as of new collected data of a single stock price with its past history. The data available are the closing prices of ten stocks listed in Athens and New York stock exchanges. In a first application, choosing the counting process of the turning points of one stock price and taking it as the pattern, the goodness-of-fit test assesses if the turning points of any other stock follow the same model. Repeating the same procedure with all the available stocks, we can conclude that the turning points behave the same for all the stocks. In a second application, the pattern is the past of a stock price and the test assesses if the new data available in a future interval of time follow the same model. It came out that three of the stocks (ELL, EMP and HRAK) show differences from their past.

ES48 Room B33 MULTIVARIATE NONPARAMETRICS Chair: Daniel Vogel

E266: **Weighted-mean regions of a probability distribution**
Presenter: **Karl Mosler**, University of Cologne, Germany
Co-authors: Rainer Dyckerhoff

Central regions, based on different notions of data depth, have been originally introduced as affine equivariant, descriptive measures in multivariate data analysis. They are convex bodies that describe a data cloud regarding its location, dispersion, and shape in R^d. In an inferential setting, a central region serves as an estimator for the set-valued parameter of an underlying probability distribution, provided this parameter is well defined and a convergence result can be established. A new class of central regions for probability distributions on R^d is investigated, called weighted-mean regions. Their restrictions to an empirical distribution are the weighted-mean trimmed regions that have been recently introduced for d-variate data. The weighted-mean regions give rise to a new notion of stochastic orderings of variability, which is finally discussed.

E375: **Hettmansperger-Randles estimators for multivariate regression**
Presenter: **Sara Taskinen**, University of Jyvaskyla, Finland
Co-authors: Hannu Oja

Consider a multivariate linear regression model. It is well known that if the estimation of the unknown regression parameter is based on an L2 objective function, one obtains the least squares (LS) estimator, which is affine equivariant and optimal when the error distribution is multivariate normal. However, the LS estimator is sensitive to outliers and very inefficient when the error distribution has heavy tails. The use of the L1 estimation based on the so called spatial sign score function yields the multivariate least absolute deviation (LAD) estimator, which is robust and efficient under heavy-tailed distributions, but unfortunately not fully affine equivariant. We will illustrate how a fully equivariant LAD estimator is obtained using the transformation retransformation technique. Corresponding $k-$step estimators are also derived. The influence functions and limiting distributions of the proposed estimators are derived at the multivariate elliptical case. The theory is illustrated with simulation studies and real data examples.

E866: **The Fisher consistent transformed spatial sign covariance matrix and its properties**
Presenter: **Alexander Duerre**, TU Dortmund, Germany
Co-authors: Daniel Vogel, Roland Fried

The classical Pearson correlation estimator is strongly influenced by outliers. Many robust alternative estimators, like, e.g., Kendall's tau or Spearman's rho, are computationally expensive or require transformations dependent on the (generally unknown) population distribution to make them consistent for the moment correlation coefficient. A highly robust, easy to compute covariance estimator is the spatial sign covariance matrix. We propose a transformation of the estimator, which makes it Fisher consistent for the covariance matrix (up to scale) at elliptical distributions, thus allowing us to construct a Fisher consistent estimator for the moment correlation. We compute the asymptotic efficiency of this correlation estimator in two dimensions and study its finite sample efficiency by means of simulations. Furthermore, the estimator is compared in terms of robustness to many known correlation estimators. Our correlation estimator is highly robust, very fast to compute and hence particularly suited for high-dimensional, robust pairwise correlation estimation. Finally, we propose an approximation to the Fisher consistency transformation in dimensions larger than two and consider also the k-step version of the spatial sign covariance matrix.

E432: **A nonparametric test for change-points in correlation**
Presenter: **Daniel Vogel**, Ruhr-Universitat Bochum, Germany
Co-authors: Herold Dehling, Roland Fried

We propose an asymptotic change-point test for correlation based on Kendall's tau. Suppose we observe a bivariate time series $((X_i,Y_i))_{i=1,\ldots,n}$. The null hypothesis of the proposed test is that Kendall's rank correlation between X_i and Y_i stays constant for all $i = 1,\ldots,n$. We assume

$((X_i, Y_i))_{i=1,\ldots,n}$ to be a stationary functional of an absolutely regular process, but make no further assumption on its distribution. This large class of processes includes all common time series models as well as many chaotic dynamical systems. In the derivation of the asymptotic distribution of the test statistic the U-statistic representation of Kendall's tau is employed. Kendall's tau correlation coefficient possesses a high efficiency at the normal distribution, as compared to the normal MLE, Pearson's correlation measure. But contrary to Pearson's correlation coefficient it has excellent robustness properties and shows no loss in efficiency at heavy-tailed distributions. This combination of efficiency and robustness is the advantage of our test over previous proposals of tests for constant correlation. Furthermore, the asymptotic variance of Kendall's tau has a tractable analytic form (in contrast to Spearman's rho for instance), which facilitates the practical implementation of the test.

ES64 Room B35 FUZZY KNOWLEDGE EXTRACTION AND ENGINEERING Chair: Anne Laurent

E444: **Towards a fuzzy ontology learning from folksonomies**
Presenter: **Sadok Ben Yahia**, Faculty of Sciences of Tunis, Tunisia
Co-authors: Aicha Ben Jrad, Chiraz Trabelsi

Social bookmarking tools are rapidly grasped the interest of the web users as it can be flagged out by the overwhelming number of participants. In such spaces, users freely annotate resources by means of any keyword or tag, giving raise to flat organizational structures, aka, folksonomies. In this respect, needless to mention that ontologies, as an explicit shared conceptualization of human knowledge, can be of benefit for enhancing information retrieval metrics. However, the human knowledge is characterized by its uncertainty. In the real world, relations between entities often include subtleties that we cannot model it using crisp ontologies. As a solution, fuzzy ontologies have been proposed where fuzzy logic concepts are incorporated into ontologies. A novel approach for extracting fuzzy ontologies from folksonomies is introduced. The main thrust of the illustrated approach stands in putting the focus on the discovery of fuzzy relations (taxonomic and non-taxonomic ones) between concepts associated to the tags. The discovery process heavily relies on the structure of triadic concepts to discover implicit shared conceptualization. Interestingly enough, Wikipedia is used as an external knowledge source for the domains of the tags and extraction of relations. Empirical studies have confirmed that the illustrated approach can discover high quality fuzzy ontology which leads to significant improvement in search information retrieval performance of folksonomies.

E689: **Comparison of fuzzy clustering methods for interval-valued symbolic data**
Presenter: **Marcin Pelka**, Wroclaw Univeristy of Economics, Poland
Co-authors: Andrzej Dudek

Symbolic data, unlike classical data, can be described by many different symbolic variables. One kind of those variables are interval-valued variables – for instance height(w) = [3, 5], which means that the height of w varies form 3 to 5. There are many clustering methods suitable for interval-valued data. There are hierarchical clustering methods (agglomerative or divisive), partitional clustering methods (like k-means). Sometimes it is a very difficult task to cluster symbolic interval-valued data into one single cluster (separate clusters). This kind of data often tends to form not separate but fuzzy clusters. Due to this fact there are proposals of fuzzy clustering methods for symbolic interval-valued data in the literature. The article compares some of those methods - like: fuzzy c-means, adaptive fuzzy c-means, with methods whose results can be treated as fuzzy, like k-means for symbolic data and also pyramidal clustering when dealing with clusters of different shape or clusters with outliers and or noisy data.

E551: **Characterizing forest of fuzzy decision trees errors**
Presenter: **Christophe Marsala**, LIP6-UPMC, France
Co-authors: Maria Rifqi

Forest of fuzzy decision trees are known to be a powerful machine learning tool in terms of good classification rate. The idea of this paper is to focus on the misclassified examples of this method to improve it. To this aim, the preliminary step consists in understanding what are the sources of misclassifications. Concretely, we propose to characterize misclassified examples by finding different types of errors or different groups of errors. We modify J. Forest's algorithm of supervised segmentation to find these groups. Our general process is the following: first, we learn a forest of fuzzy decision trees on a learning dataset and we validate it on a validation dataset. We class the examples on two categories: those who are well classified by the majority of the trees and those who are badly classified by the majority of the trees. Then, this new dataset is characterized by supervised segmentation algorithm. For instance, we can obtain that the badly classified examples can be distinguished in two groups: those who can be interpreted as outliers and those who are close to the frontiers of decision. We apply our general process on several datasets of UCI machine learning repository to show its benefits.

E502: **Mixing multiple fuzzy modalities for fuzzy gradual pattern mining**
Presenter: **Anne Laurent**, University Montpellier 2, France
Co-authors: Sarah Ayouni, Lisa Di Jorio, Sadok Ben Yahia, Pascal Poncelet

Data mining has gained a lot of attention and is essential in many domains related to business intelligence, scientific data management and analysis (biology, health, environment, $\cdots$), etc. Many algorithms have been defined and studied, including pattern mining. Patterns can be simple itemsets, sequential patterns, of also gradual patterns. In the latter case, co-variations of attributes are considered, such as "the higher the age, the higher the salary". We focus here on the extension of these kinds of patterns to the fuzzy case. We aim at extracting fuzzy gradual patterns such as "the more the expression of gene A is almost 1.8, the more the expression of gene B is almost 0.2" from large databases. In this framework, the choice of the fuzzy subsets (e.g. "almost 1.8") is of great importance. We have recently studied how genetic algorithms can help to define one relevant fuzzy modality per attribute. We discuss how multiple fuzzy modalities can be defined on the attributes and the impact on the extracted patterns.

ES68 Room B20 MICROARRAY DATA ANALYSIS Chair: Taesung Park

E228: **Comparative evaluation of gene-set analysis methods in association with survival time**
Presenter: **Seungyeoun Lee**, Sejong University, Korea RoK
Co-authors: Jinheum Kim, Sunho Lee

Many gene-set analysis methods have been previously proposed and compared through simulation studies and analysis of real datasets for binary phenotypes. We focused on the survival phenotype and compared the performance of Gene Set Enrichment Analysis (GSEA), Global test (GT), Wald-type test (WT) and Global boost test (GBST) methods based on simulation datasets and a real example of ovarian cancer data. From the simulation datasets and a real example, GT , WT and GBST tests have high power in simulated datasets and seem to be too liberal in the real dataset, whereas GSEA1 and GSEA2 tests have low power in simulation datasets and tend to be too conservative in the real dataset. We also found that the power of the five tests is much higher under correlated genes than under independent genes, when survival is positively associated with genes. It seems that there is a synergistic effect in detecting significant gene sets when significant genes have within-correlation and the association between survival and genes is positive or negative (i.e., one-direction correlation).

E386: **Stochastic model based abnormality detection in a large-scale gene regulatory network**
Presenter: **Haseong Kim**, Imperial College London, United Kingdom
Co-authors: Erol Gelenbe

Thanks to the development of high-throughput and Omics technology, we can identify regulatory pathways by measuring the activities of biological molecules and estimating their causalities via statistical and mathematical algorithms. Modeling these pathways can provide insight of a system's dynamic behavior depending on its conditions. However, stochastic models describing the dynamics of a system are suffering from finding the steady-state of the system especially when a large number of genes are involved. The steady-state of a large-scale gene regulatory network is explained here by the closed-form solution of the G-network theory, which is a probabilistic queuing network model with special agents, such as positive and negative customers, signals and triggers. By assuming the microarray expression data of a case-control study obtained in a steady-state, we determine the abnormality of transition probabilities of packets (customers in a queuing system) which contains gene expression information and moves around genes in a network structure. Brain tumor mRNA expression data were collected from GEO database and we found abnormally activated/inactivated pathways in a KEGG based non-metabolic gene network structure consisting of 1305 genes.

E568: **NetworkProfiler: Uncovering cancer heterogeneity in transcriptome data**
Presenter: **Seiya Imoto**, University of Tokyo, Japan

Patient-specific analysis of molecular networks is a promising strategy for making individual risk predictions and treatment decisions in cancer therapy. Although systems biology allows the gene network of a cell to be reconstructed from clinical gene expression data, traditional methods, such as Bayesian networks, only provide an averaged network for all samples. Therefore, these methods cannot reveal patient-specific differences in molecular networks during cancer progression. In this study, we developed a novel statistical method, called NetworkProfiler, which infers patient-specific gene regulatory networks for a specific clinical characteristic, such as cancer progression, from gene expression data of cancer patients.

E574: **Gene set analysis for SNPs with low minor allele frequencies**
Presenter: **Taesung Park**, Seoul National University, Korea RoK
Co-authors: Jaehoon Lee, Soyeon Ahn

Although a number of genome-wide association studies (GWAS) have been successful in detecting common genetic variants associated with complex diseases and traits, missing heritability has not been completely explained by the common variants. SNPs with low minor allele frequencies (or rare variants) are expected to have plausibly moderate or large effect and to explain this missing heritability. However, it is not easy to perform association analysis using the SNPs with low minor allele frequencies. Collapsing multiple rare variants is one way of performing association analysis for the rare variants and has been shown to be powerful in detecting the effects of rare variants. In this research, taking advantage of collapsing multiple rare variants, we propose gene set analysis of rare variants in order to maximize the effect of rare variants. Our proposed gene set analysis considers the effect of multiple rare variants jointly and uses prior biological knowledge based on pathway information.

Monday 19.12.2011 10:55 - 12:35 Parallel Session N – CFE

CSI04 Room Beveridge FINANCIAL TIME SERIES MODELLING Chair: Christian Francq

C179: **Asymptotic equivalence of continuously and discretely sampled jump-diffusion models**
Presenter: **Feike C. Drost**, Tilburg University, Netherlands
Co-authors: Irene G. Becheri, Bas J.M. Werker

Local Asymptotic Normality (LAN) is established for continuous observations from jump-diffusion models with time-varying drift and jump intensity but known volatility. High-frequency discrete-time observations from this model contain, in an asymptotic and local sense, the same information about the parameters of interest. More precisely, we provide sufficient conditions on a jump identification mechanism that allows us to construct a central sequence for the continuous-time model, using discrete-time observations only.

C297: **Testing strict stationarity in GARCH models**
Presenter: **Jean-Michel Zakoian**, CREST and University Lille 3, France
Co-authors: Christian Francq

Testing strict stationarity of GARCH(1,1) models is the object. The asymptotic properties of the quasi-maximum likelihood estimator are studied without strict stationarity constraints. Except for the intercept, this estimator remains consistent and asymptotically normal in every situation. The asymptotic variance is different in the stationary and non-stationary situations, but is consistently estimated, with the same estimator, in both cases. Tests of strict stationarity and non-stationarity are proposed. The tests developed for the classical GARCH(1,1) model are able to detect non-stationarity in more general GARCH models. A numerical illustration based on stock indices and individual stock returns is proposed.

C563: **Reduced rank autoregression with volatility induced stationarity**
Presenter: **Anders Rahbek**, University of Copenhagen, Denmark

It is considered a multivariate autoregressive (VAR) model which allows for reduced rank in the conditional mean as is well known from cointegration analysis, while at the same time allows for lagged levels in the conditional covariance (ARCH). It is shown that not only the reduced rank relations (the "cointegrated relations"), but also the common trends driving the system may be strictly stationary despite the presence of unit-roots in the system. In particular, the common trends, while stationary, may have only fractional moments. Asymptotic theory is derived and an empirical illustration is given by term strcuture data.

CS09 Room Court HIGH FREQUENCY DATA MODELING Chair: Massimiliano Caporin

C206: **Market uncertainty and macroeconomic announcements: High-frequency evidence from the German DAX**
Presenter: **Klaus Wohlrabe**, Ifo Institute for Economic Research, Germany
Co-authors: Stefan Mittnik, Nikolay Robinzonov

Considering 64 macroeconomic announcements for the U.S., the Euro Area and Germany we analyze announcement effects on the German stock market using minute-by-minute DAX-index data. We employ a time series approach correcting for heteroscedasticity and find that a number of surprise announcements have an immediate and long-lived impact on the DAX. A separate analysis for high- and low-volatility periods, defined in terms of the VIX volatility index, reveals that the significant impacts found in the full-sample analysis are mostly due to influences taking place during states of high market-volatility, i.e. phases of high market uncertainty. Furthermore, the size of the impact is larger in these times. This indicates that market participants vary their attention as risk levels change. We link our results to some theoretical predictions in the literature.

C272: **Stylized facts and information asymmetry on high frequency precious metals spot prices**
Presenter: **Massimiliano Caporin**, University of Padova, Italy
Co-authors: Angelo Ranaldo, Gabriel Velo

Using nano-frequency data from a trading platform, we analyze the existence of high frequency financial data stylized facts on four precious metals: gold and silver (predominantly traded), platinum and palladium. We consider both prices (and therefore returns) as well as volumes and several liquidity measures. Those are extracted from a 100 millisecond ten-level order book matched with the complete list of executed trades on the platform. We focus on a two-year sample and, differently from previous studies in the area, we use spot prices and do not consider only gold. Our study shows that precious metals data have a behavior extremely similar to that of equities including the presence of returns asymmetry, volatility clustering, and intra-daily periodicities. Moreover, we are able to analyze the correlation across precious metals and between them and the US stock market (proxied by the S&P500 index). Finally, by combining the high frequency returns and the order flow (one of the liquidity measures we consider) we analyze the existence of information asymmetries in the precious metal market.

C347: **Indirect inference for long memory stochastic volatility model with high-frequency data**
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. An indirect inference estimation procedure for the long memory stochastic volatility model (LMSV) is proposed. The auxiliary model employs realized volatility, computed from high-frequency data. However, integrated volatility is measured by realized volatility with an error term, so that OLS provides downward biased estimates of the HAR-RV parameters. Thus HAR-RV is estimated correcting for the measurement error present in realized measures. A Monte Carlo simulation investigates the finite sample properties of the indirect inference estimates of the LMSV. We also address the question of the existence of true long memory versus the presence of level shifts in the volatility process. The encompassing model is the sum of the LMSV plus a level shift process and the indirect inference techniques allow us to estimate both. An application using high-frequency data on US stocks provides further evidence on the proposed methodology.

C160: **A dynamic multifactor model for high an low frequency volatility activity**
Presenter: **Paolo Santucci de Magistris**, Aarhus University, CREATES , Denmark
Co-authors: Stefano Grassi

The persistent nature of equity volatility as a mixture of processes at different frequencies is investigated by means of a dynamic multifactor model, which is a generalization of the heterogeneous autoregressive model for the realized volatility. The model is a time-varying additive cascade of different partial volatilities (generated by the actions of different types of market participants). It emerges that during financial crises the relative weight of the high frequency component dominates over the slowly moving ones. This result can be explained by the asymmetry in response to bad news (or negative shocks) than to good news (or positive shocks), due to the increase in the asymmetric propagation of volatility between long and short time horizons. An indirect inference estimation technique, based on the time varying HARV as auxiliary model, allows us to estimate

the dynamic parameters of various continuous-time stochastic volatility models. A long-horizon regression investigates the proportion of return predictability which is related to the high and low frequency volatility component.

CS18 Room S261 BAYESIAN FINANCIAL RISK MANAGEMENT Chair: Richard Gerlach

C172: **Stochastic volatility models and quantile regression using asymmetric Laplace error distribution via uniform scale mixtures**
Presenter: **Boris Choy**, University of Sydney, Australia
Co-authors: Nuttanan Wichitaksorn, Joanna Wang, Richard Gerlach

The Bayesian analysis of a stochastic volatility (SV) model with an asymmetric Laplace distribution (ALD) for the innovation term of the return is considered. We propose a uniform scale mixture representation for the probability density function of the ALD and show that the use of scale mixture will simplify the Gibbs sampler. The primary aim is to develop efficient computational methods for statistical inference of SV models with ALD. The second aim is to consider the Bayesian quantile SV model. Both SV model with ALD and Bayesian quantile SV model will be applied to stock market return data and the value-at-risk will be forecasted.

C220: **Evaluating value at risk and expected shortfall using generalised asymmetric volatility models**
Presenter: **Georgios Tsiotas**, University of Crete, Greece

Value at Risk (VaR) and Expected Shortfall (ES) are risk measures whose accurate estimation influences portfolio selection strategies. Their calculation involves the estimation of returns' volatility. Stochastic volatility (SV) models have gained considerable attention when dealing with time varying volatility. Volatility asymmetry, via the leverage effect hypothesis, is now expressed simultaneously with the skewness and the excess kurtosis hypothesis introducing generalised asymmetric SV models. By using real financial data series, the new models are compared to existing SV models for their forecasting performance in the accurate VaR and ES evaluation. Finally, model averaging strategies on generalised asymmetric SV models are also tested for their forecasting reliability in the VaR and ES evaluation in comparison to the single model alternatives. Results show some success in predicting the risk measures for various financial data series.

C389: **Calibration and filtering for multi factor commodity models with seasonality: incorporating panel data from futures contracts**
Presenter: **Gareth Peters**, University of New South Wales, Australia
Co-authors: Mark Briers, Pavel Shevchenko, Arnaud Doucet

A general multi-factor model for estimation and calibration of commodity spot prices and futures valuation is constructed. This extends the multi-factor long-short models previously developed in the literature in two important aspects: firstly we allow for both the long and short term dynamic factors to be mean reverting incorporating stochastic volatility factors and secondly we develop an additive structural seasonality model. In developing this non-linear continuous time stochastic model we maintain desirable model properties such as being arbitrage free and exponentially affine, thereby allowing us to derive closed form futures prices. In addition the models provide an improved capability to capture dynamics of the futures curve calibration in different commodities market conditions such as backwardation and contango. A Milstein scheme is used to provide an accurate discretized representation of the s.d.e. model. This results in a challenging non-linear non-Gaussian state-space model. To carry out inference, we develop an adaptive particle Markov chain Monte Carlo method. This methodology allows us to jointly calibrate and filter the latent processes for the long-short and volatility dynamics. We demonstrate the performance of our model and algorithm on both synthetic data and real data for futures contracts on crude oil.

C833: **Estimating portfolio value at risk using a skew-t copula-GARCH model**
Presenter: **Brent Hudson**, Essential Energy, Australia
Co-authors: Richard Gerlach

A new copula-GARCH model is proposed that simultaneously captures stock return asymmetry in both conditional mean and variance equations of its marginal distributions, as well as asymmetry in its dependence structure via a skew-t copula. Moreover, the marginal distributions are generalised to incorporate both skewness and excess kurtosis, thus enabling information on the first four moments of stock returns to be available in the one model. The performance of the proposed model is compared to a variety of existing univariate and multivariate GARCH models by estimating Value at Risk (VaR) over multiple time horizons on a stock portfolio consisting of five Dow Jones Industrial (DJI) stocks. While VaR back-testing has already been applied extensively in the literature, it has always been applied to compare models of the same dimension. Under the assumption that the portfolio weights are known, comparisons can be drawn between univariate and multivariate models together in the one study. In this example, the proposed skew-t copula-GARCH model outperforms the competing models in many cases, and also demonstrates significant asymmetries that are present in the data. These results clearly indicate the need to accurately capture the tail behaviour in stock returns for risk management purposes.

CS20 Room Senate NON LINEARITY AND BUSINESS CYCLES Chair: Dominique Guegan

C370: **Variable selection for prediction purpose of real economic activity**
Presenter: **Patrick Rakotomarolahy**, University of Paris 1, France
Co-authors: Dominique Guegan

Variable selection is a crucial step in the characterization of real economic activity dynamics. We consider some economic and financial variables which have been thought as the sources of different cycles in our sample history. The role of such variables in predicting real economic activity has been analyzed. But, such analyses were driven by the real economic dynamics which are usually taken to be linear. Therefore, it accounts for only linear relationship dynamic between the variables and the real economic activity. There are two main objectives. The first objective is on investigating possible linear or nonlinear relationship between the economic/financial variables and the real economic activity. This results in several types of relationships since each variable has one dynamic relationship to real economic activity. Then, we deal with mixture of linear and nonlinear dynamics. Analysis of this issue will be the second objective, which results in variable selection through predictive power of each variable which takes into account the relationship found during previous objetive.

C517: **A test for a new modelling: The univariate MT-STAR model**
Presenter: **Peter Martey Addo**, Universite Paris1 Pantheon-Sorbonne, France
Co-authors: Dominique Guegan

In ESTAR models it is usually quite difficult to obtain parameter estimates, as it is discussed in the literature.The problem of properly distinguishing the transition function in relation to extreme parameter combinations often leads to getting strongly biased estimators. This paper proposes a new procedure to test for the unit root in a nonlinear framework, and contributes to the existing literature in three separate directions. First, we propose a new alternative model -the MT-STAR model- which has similar properties as the ESTAR model but reduces the effects of the identification problem and can also account for cases where the adjustment mechanism towards equilibrium is not symmetric. Second, we develop a testing procedure to detect the presence of a nonlinear stationary process by establishing the limiting non-standard asymptotic distributions of the proposed test-

statistics. Finally, we perform Monte Carlo simulations to assess the small sample performance of the test and then to highlight its power gain over existing tests for a unit root.

C368: **A rank-based approach to cross-sectional analysis**
Presenter: **Ludovic Cales**, HEC - University of Lausanne, Switzerland
Co-authors: Monica Billio, Dominique Guegan

The aim is to study the cross-sectional effects present in the market using a new framework based on graph theory. 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. Practically, this modeling permits us to measure the marginal and joint effects of different cross-sectional 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.

C659: **Monetary policy, food inflation and the business cycle.**
Presenter: **Marc Pourroy**, University Paris 1 Pantheon-Sorbonne, France
Co-authors: Benjamin Carton, Dramane Coulibaly

The two recent episodes of food price surge, combined with the large widespread of inflation targeting as a framework for emerging economies' central banks, have raised the question of how monetary policy should react to these external shocks. We develop a small open economy dynamic general equilibrium model to address this issue. Our contributions to the literature are twofold. Firstly, we model an economy in which the core inflation index may be unknown by the central bank, and thus is not the first best index to target. Secondly, we stress out the role played by the particular form of food as a utility function. The latter is characterized by a low elasticity of substitution and Klein-Rubin form of minimum amount consumption.

CS23 Room S264 IDENTIFICATION-ROBUST INFERENCE AND LARGE MODELS Chair: Lynda Khalaf

C603: **Testing for GARCH effects: An exact procedure based on quasi-likelihood ratios**
Presenter: **Richard Luger**, Georgia State University, United States of America

A procedure is proposed to test for GARCH effects in possibly non-linear models with endogenous explanatory variables, including autoregressive models. The approach is based on the quasi-likelihood function, leaving the true distribution of model disturbances completely unspecified. A permutation principle is first established to test the null hypothesis of constant conditional variance in the basic location model. A pivotal bound to the null distribution of the quasi-likelihood ratio statistic is then derived to deal with the presence of nuisance parameters in richer conditional mean specifications. The bounding statistic is used with a Monte Carlo resampling technique which yields a level-exact and computationally inexpensive inference procedure. A simulation study illustrates the theoretical results and also shows that the test procedure has good power.

C604: **Efficient inference with time-varying identification strength**
Presenter: **Bertille Antoine**, Simon Fraser University, Canada
Co-authors: Otilia Boldea

Parameter instability and identification strength are two widespread concerns in applied economics. Moreover, parameter instability in a structural model may induce changes in the strength of identification. This issue, often neglected by researchers, may lead to incorrect, or inefficient inference. We develop break-point methods to detect both parameter instability and changes in the strength of identification, or strength of instrumental variables. If parameter instability is not linked to changes in instrument strength, we show the validity and efficiency of standard GMM inference methods, conditional on the break-points. On the other hand, if parameter instability is linked to changes in instrument strength, we develop methods to estimate the location and magnitude of such changes. We also derive the asymptotic distribution of 2SLS and GMM parameter estimates in different subsamples, and show which inference methods are valid and efficient, conditional on the instrument strength. Finally, we illustrate our methods via the new Keynesian Phillips curve for the US. We find that instrument strength changes twice between 1968 and 2005. We also show that our confidence intervals are more informative about the parameters of interest than the usual weak instrument robust confidence intervals.

C630: **Exact inference with time varying parameters in linear models**
Presenter: **Clement Yelou**, Statistics Canada, Canada
Co-authors: Jean-Thomas Bernard, Lynda Khalaf, Maral Kichian

Finite sample statistical inference methods in the univariate linear regression model with time varying parameters are proposed. The random-walk parameter model, where an LM test is used for testing the null hypothesis of constant parameters, represents a special case of the TVP process we consider. We show that the null distribution of the LM test statistic does not depend on any nuisance parameters, so exact p-values can be obtained using the Monte Carlo (MC) test technique. Since normality is not necessary for this MC test procedure, our test method will prove particularly useful in many empirical settings, e.g. modelling financial data. Our simulation study results show that the proposed MC test performs well, even in small samples. In addition, we propose constructing confidence sets for the variance of the TVP process by inverting the asymptotic LM test as well as our exact test. We provide an empirical application of these methods to an annual energy demand model with a random trend specified as a two-step stochastic process, using data for the province of Quebec (Canada) from 1962 to 2002.

C601: **Factor based identification-robust inference in IV regressions**
Presenter: **Lynda Khalaf**, Carleton, Canada
Co-authors: Georges Kapetanios, Massimiliano Marcellino

Robust methods for IV inference have received considerable attention recently. Their analysis has raised a variety of problematic issues such as size/power trade-offs resulting from omitted, weak or many instruments. The popular Anderson-Rubin test has the right size when the underlying first-stage model (that is, the model linking the structural equation's right-hand side endogenous variables to available instruments) is closed or is incomplete. Alternative methods are available that may outperform this statistic assuming a closed first-stage specification (that is, assuming that all instruments are accounted for). We show that information-reduction methods provide a useful and practical solution to this and related problems. Formally, we propose factor-based modifications to three popular weak-instruments-robust statistics, and illustrate their validity asymptotically and in finite samples. Results are derived using asymptotic settings that are commonly used in both the factor-analysis and weak-instruments literatures. For the Anderson-Rubin statistic, we also provide analytical finite sample results under usual assumptions. An illustrative Monte Carlo study reveals the following. Firstly, our factor-based corrections circumvent the size problems resulting from omitted or many instruments and improve the power of the Anderson-Rubin statistic. Secondly, once corrected through factor reduction, all considered statistics perform equally well. Results suggest that factor-reduction holds promise as a unifying solution to the missing instruments problem as well as a generic attractive approach to robust inference in the presence of many or weak instruments. An empirical study on New Keynesian Phillips Curves suggests that our factor-based methods can bridge the gap between structural and statistical macroeconomic models.

CS29 Room Woburn MODELLING AND FORECASTING FINANCIAL RISK Chair: Stefan Mittnik

C318: **Score-based range models**
Presenter: **Karl-Philipp Andres**, Cambridge Universiy, United Kingdom
Co-authors: Andrew Harvey

The range is employed as estimator for the volatility of a series. It is modelled by a dynamic equation driven by the the score of the conditional distribution. Monte Carlo studies for a wide range of distributions, including Gamma, Weibull, Burr and F, show the viability of this approach. The model is estimated for NASDAQ data and other series. Analytical standard errors are provided along with forecasts. We compare the results with some of the models in the existing literature.

C380: **Estimating the extended Heston stochastic volatility model with Jacobi stochastic leverage for S&P500 and VIX**
Presenter: **Isao Ishida**, Osaka University, Japan
Co-authors: Michael McAleer, Kosuke Oya

Leverage, or the negative correlation between return and its volatility, is a key feature in the time series analysis of equity price observations. We present new empirical evidence that the daily realized correlation, or realized leverage, between the S&P500 index and VIX, calculated using high-frequency intraday data, is time-varying. In popular one-factor stochastic volatility models, such as the Heston model, the leverage is a constant, representing the correlation between the two Brownian factors driving the price and volatility processes, respectively, and cannot accommodate the new empirical evidence. We propose a new GMM estimator for the extended Heston with stochastic leverage (EHSL) model that evolves according to the Jacobi diffusion process. This model has been discussed in the literature, but no satisfactory econometric procedure for estimating the model using time series observations seems to have been developed. We derive analytical expressions for the moments of the realized variances and covariation between the S&P500 index and VIX, and use them to construct moment conditions for a GMM estimator. We then apply the proposed estimator to high frequency S&P500 and VIX data and present the empirical results, thereby supporting the stochastic leverage extension to the original Heston model.

C431: **Optimal combination of risk forecasts under the Basel accord**
Presenter: **Paulo Araujo Santos**, Escola Superior de Gestao e Tecnologia de Santarem e CEAUL, Portugal
Co-authors: Juan-Angel Jimenez-Martin, Michael McAleer, Teodosio Perez Amaral

Under the Basel Accord, forecasting Value-at-Risk has become a daily task of banks and other Authorized Deposit-taking Institutions (ADIs), whereby daily forecasts are used to determine the capital requirements and associated capital costs. Recently, several risk management strategies based on combinations of risk forecasts from different risk models have been proposed. The empirical evidence has shown that these approaches can be profitable and safe for risk management, both in calm and turbulent financial periods. We propose a method for determining optimal combinations of risk forecasts under the Basel Accord and compare, in terms of daily capital requirements and violation penalties, the performance of the method with individual models and other approaches that are based on combinations of risk forecasts.

C790: **Solvency II calibrations: Where curiosity meets curiosity**
Presenter: **Stefan Mittnik**, University of Munich, Germany

The Solvency II regulatory framework specifies procedures and parameters for determining solvency capital requirements (SCR) for insurance companies in the EU. The proposed standard SCR calculations involve two steps. First, the risks of all individual business units, expressed in terms of Value-at-Risk (VaR), are measured and then, in a second step, aggregated to the company's overall SCR, using the so–called Standard Formula. The Standard Formula has two inputs: the VaRs of the individual risk units and their correlations. The appropriate calibration of these input parameters has been the purpose of various Quantitative Impact Studies that have been conducted during recent years. This paper demonstrates that the derivation of the calibration parameters for the equity-risk module, the most significant risk component of insurance companies' total SCR, is seriously flawed and gives rise to spurious and highly erratic parameters. As a consequence, an implementation of the Standard Formula with the currently proposed calibration settings is likely to produce inaccurate, unstable and biased capital requirements for equity-risk.

CS30 Room Jessel QUANTITATIVE ASSESSMENT OF FINANCIAL STABILITY AND MACRO-POLICIES Chair: Costas Milas

C498: **The risk of default and the term-structure of sovereign yield spreads**
Presenter: **Alena Audzeyeva**, Keele University, United Kingdom
Co-authors: Klaus Reiner Schenk-Hoppe

Information implied in observed sovereign rating transitions and defaults is used to decompose the term-structure of sovereign yield spreads of Mexico, Colombia and Brazil into the default and non-default components. The analysis suggests that the relative share of the default component within the yield spread varies over time and across maturity horizons. Default risk accounted for a rather small share (decreasing with maturity) of yield spreads for low investment grade Mexican and non-investment grade Colombian and Brazilian Eurobonds in 2003. This share increased significantly while yield spreads fell during 2003-2006 mainly due to non-default risk factors. Both of these trends reversed while the relative share of the default component became predominantly increasing with maturity in 2007-2008 for all three countries. Changes in the relative share of the default component vis-a-vis the non-default component along the maturity spectrum of the yield spread appear to be determined by both credit quality of a sovereign borrower and global market conditions.

C286: **A forecasting analysis of the inward cross-border mergers and acquisitions in the UK: A macroeconomic perspective**
Presenter: **Moshfique Uddin**, University of Pretoria, South Africa
Co-authors: Agyenim Boateng, Ruthira Naraidoo

Most of the growth in international production over the past decade has been carried out via cross-border mergers and acquisitions (CBM&As). Yet previous empirical work relating to CBM&As has been confined to firm-specific factors. This is against the backdrop that researchers have not been able to develop a coherent theory explaining the increasing trends of CBM&As activity. Furthermore, relatively little has been done in the field to address the forecasting performance of models explaining the wave behaviors of CBM&As. Building on prior studies, this study demarcates itself by analyzing the ability of linear and nonlinear parametric specifications as well as nonparametric and semiparametric models in forecasting the number of cross-border mergers and acquisitions inflows between 1987 and 2010 into the UK from a macroeconomic perspective. Our results indicate that macroeconomic factors cannot be neglected in accounting for a firm's decision to engage in mergers and acquisitions. More importantly, economic growth, country risk index, bond yields, the real exchange rate play a determining role in the forecasting models. We also note that forecasts constructed from nonparametric and semiparametric models perform particularly well based on rolling windows of estimation, and that there are gains from these models in forecasting CBM&As as the forecasting horizon lengthens.

C381: **Stabilization effects of the Euro area monetary policy**
Presenter: **Tatevik Sekhposyan**, Bank of Canada, Canada
Co-authors: Michael Owyang

The primary objective of the European Central Bank is price stability for the Euro Area. The current economic problems in the Euro Area countries suggest that it is possible for monetary policy stabilization effects in the aggregate and member economies to differ. In addition, the effects might vary across countries. To address this issue, we first estimate a Euro Area policy interest rate rule in a flavour of a Taylor rule. We then assess the importance of the country-specific differentials from the aggregate for the central bank interest rate setting. Further, we incorporate the empirical findings in a theoretical framework, where the country-specific stabilization effects of the aggregate monetary policy are formalized. We show that within typical parameter values the "Taylor principle" is satisfied for the aggregate, yet it is not satisfied for the currency union as a whole suggesting multiplicity of equilibria.

C084: **Debt sustainability and financial crises: Evidence from the GIIPS**
Presenter: **Costas Milas**, University of Liverpool, United Kingdom
Co-authors: Gabriella Legrenzi

The sustainability of the public finances of Greece, Ireland, Italy, Portugal and Spain (GIIPS), allowing for possible non-linearities in the form of threshold behaviour of the fiscal authorities is assessed. We provide some evidence of fiscal sustainability when debt gets "too high" relative to a threshold which is not necessarily fixed but varies with the level of debt relative to its recent history and/or the occurrence of a financial crisis. However, the Greek and Italian debt-to-GDP threshold levels (over which adjustment takes place) exceed 87% and rise further in periods of financial crises. This arguably adds to international investors' concerns, and as a result, raises the yields demanded for holding Greek and Italian debt. As debt is rolled over at high interest rates, fiscal prospects worsen making default more likely and adding to contagion effects from one Eurozone country to another.

CS33 Room Bloomsbury BAYESIAN ECONOMETRICS AND APPLICATIONS Chair: Teruo Nakatsuma

C140: **Bayesian analysis of latent threshold dynamic models**
Presenter: **Jouchi Nakajima**, Duke University, United States of America
Co-authors: Mike West

A general approach to dynamic sparsity modelling in time series and state-space models is described. Time-varying parameters are linked to latent processes that are thresholded to induce zero values adaptively, providing dynamic variable inclusion/selection. Bayesian model estimation and prediction in dynamic regressions, time-varying vector autoregressions and multivariate volatility models using latent thresholding are discussed. Substantive examples in macroeconomics and financial time series show the utility of this approach to dynamic parameter reduction and time-varying sparsity modelling in terms of statistical and economic interpretations as well as improved predictions.

C296: **GPGPU parallel computing for Bayesian portfolio selection with massive number of assets**
Presenter: **Kenichiro McAlinn**, Keio University Graduate School of Economics, Japan
Co-authors: Teruo Nakatsuma

Bayesian methods that utilize subjective views of investors have proven to be successful in solving a range of problems in finance, from portfolio selection to program trading. Regardless of the mounting supportive evidence for Bayesian methods, they have yet to suffice for practical use in everyday decision making by investors because of their time consuming nature. With the recent development of fast and inexpensive devices for parallel computing such as general purpose graphic processing units (GPGPU), however, formerly impractical computations that take hours or days can be completed in minutes or even seconds. With this new paradigm of parallel computing in mind, we investigate whether the use of GPGPU can ease the burden of computationally intensive Bayesian methods, in particular Bayesian portfolio selection with a massive number of assets. In our parallel computing approach, we simultaneously perform multiple runs of Gibbs sampling for the factor model of asset returns and compute posterior statistics necessary for Bayesian portfolio selection. In numerical experiments, we find that parallel computing by GPGPU substantively reduces computing time for portfolio selection, compared with the traditional CPU. We also demonstrate that introducing subjective views improves the performance of a portfolio.

C388: **Realized stochastic volatility with leverage and long memory**
Presenter: **Shinichiro Shirota**, University of Tokyo, Japan
Co-authors: Takayuki Hizu, Yasuhiro Omori

The daily return and the realized volatility are simultaneously modeled in the stochastic volatility model with leverage and long memory. In addition to the stochastic volatility model with leverage for the daily returns, the ARFIMA process is jointly considered for the realized volatilities. Using a state space representation of the model, we estimate parameters by Markov chain Monte Carlo methods. Model comparison with similar realized stochastic volatility models with short memory is conducted by computing marginal likelihood.

C727: **Transdimensional approximate Bayesian computation for model choice**
Presenter: **Genya Kobayashi**, Kobe University, Japan
Co-authors: Hideo Kozumi

The first attempts on an approximate Bayesian computation (ABC) algorithm for model selection with between-model moves are made. Although a few model selection algorithms have been proposed, none of them can move across the models with different parameter dimensions. The goal of this paper is to fill the gap. Although the celebrated reversible jump algorithm can be a natural basis for a new algorithm, constructing an efficient proposal distribution is expected to be quite difficult, since we cannot use the information of the likelihood. We, instead, rely on a flexible and efficient method, called the pseudo-marginal approach. Our algorithm produces an approximation to the marginal algorithm whose stationary distribution is the posterior model probabilities. It is shown that the model selection algorithm based on the pseudo marginal approach can be straightforwardly fit into the ABC framework and thus can be easily implemented. Using the simulated data, the efficiency of our method is demonstrated. We also present some numerical examples using real data.

CS41 Room Bedford DERIVATIVE PRICING Chair: Jeroen Rombouts

C022: **The equity premium and the maturity structure of uncertainty**
Presenter: **Abderrahim Taamouti**, Universidad Carlos III de Madrid, Spain
Co-authors: Bruno Feunou, Jean-Sbastien Fontaine, Romeo Tedongap

The linkages between the conditional distribution of returns and the equity premium across different investment horizons are studied. We focus on the broad class of affine Long Run Risk [LRR] economies and show that the term structures of risk-neutral moments is affine and can reveal the LRR

or stochastic volatility factors at the heart of these models. Empirically, the variance term structure predicts the equity premium and the variance premium. Analogously, the term structures of higher-order moments also predict the equity premium and the variance premium. Consistent with theory, we find the predictability of the equity premium based on the variance, skewness and kurtosis term structure can be summarized by two risk factors. Our empirical finding relies on reduced-rank regressions. It also allows us to recover the common risk factors underlying the risk-returns trade-off implicit in the equity premium, the variance premium and the risk-neutral distribution more generally.

C184: **Testing and evaluating dynamic correlations in terms of Dow Jones industrial average index options pricing**
Presenter: **Francesco Violante**, Maastricht University, Netherlands
Co-authors: Jeroen Rombouts, Lars Stentoft

The predictive performance of various multivariate volatility models is assessed in terms of index option pricing. The set of forecasting models for the conditional covariance of the portfolio of underlying assets, consists of a large range of specifications in the dynamic conditional correlation GARCH class. This class of models is particularly suited because it allows us to consistently estimate the risk neutral dynamics with a minimal computational effort in relatively large scale problems. We consider a large panel of options on the DJ30 index over the period 1996-2009. The accuracy of the pricing is measured by three objective functions which allow us to infer whether systematic over/under prediction occurs. Inference on the models' predictive ability is based on the model confidence set approach. We elaborate on the degree of model sophistication required to achieve an accurate pricing with respect to two dimensions, namely the moneyness and the time to maturity.

C419: **Closed form option pricing with asymetric heterokedastic normal mixture models**
Presenter: **Lars Stentoft**, HEC Montreal, Canada
Co-authors: Jeroen Rombouts

A closed-form option pricing formula for a spot asset whose dynamics fall within the class of heteroskedastic normal mixture models is developed. The model contains a closed form GARCH model as a special case when only one component is allowed. However, the more general model can accommodate skewness and leptokurtosis, and it is even possible to include components with jump like properties. In an empirical study using a large set of index options between 1996 and 2009 we find substantial improvements compared to several benchmark models. The results also show that including option data in the estimation is essential and models based on both historical returns and options outperform models based only on returns. Since closed-form formulas are needed for this, a modeling framework like the one derived here is essential.

CS51 Room Gordon FINANCIAL ECONOMETRICS MODELLING Chair: Elias Tzavalis

C538: **Noncausality and asset pricing**
Presenter: **Matthijs Lof**, University of Helsinki, Finland

Misspecification of agents' information sets or expectation formation mechanisms may lead to noncausal autoregressive representations of asset prices, in the sense that current observations seem to depend on both past and future realizations, rather than only on past realizations. Annual US stock prices are found to be noncausal, implying that agents' expectations are not revealed to an outside observer such as an econometrician observing only realized market data. A simulation study shows that noncausal processes can be generated by asset-pricing models featuring heterogeneous expectations.

C550: **Influence of market conditions on event-study: The case of merger and acquisition announcement effects**
Presenter: **Christodoulos Louca**, Cyprus University of Technology, Cyprus
Co-authors: Panayiotis Andreou, Christos Savva

Empirical research on mergers and acquisitions has attracted a lot of attention, especially regarding the economic significance of such events. Nevertheless, most of prior research fails to account for the impact of unrelated events during the estimation period of the event study. The existence of contaminating firm-specific events during the estimation window will most likely affect the estimation of the return-generating process and the variance of the parameters. Unlike prior literature that focuses solely on simulating contaminated estimation time-series of stock returns, this paper carries out specification analysis of event-study statistical tests using real market data comprised by merger and acquisition announcements in the period 1980 – 2009. Our results suggest that a two-state market model estimated using the Markov switching regression method sufficiently controls unrelated events occurring during the event study estimation window and leads to unbiased results and inferences.

C961: **Exponent of cross-sectional dependence: Estimation and inference**
Presenter: **George Kapetanios**, Queen Mary University of London, United Kingdom
Co-authors: Natalia Bailey, Hashem Pesaran

Cross-sectional dependence and the extent to which it occurs in large multivariate datasets is of great interest for a variety of economic, econometric and financial analyses. A common characteristic of such analyses is the need to quantify cross-sectional dependence especially when it is prevalent enough to materially affect the outcome of the analysis. We provide such a quantification and suggest a summary measure which we refer to as the exponent of cross-sectional dependence. We find that this measure can accommodate a wide variety of cross-sectional dependence manifestations while retaining its simple and tractable form. We, then, proceed to discuss estimation and inference of this measure. The inference problem is complex, involves handing a variety of bias terms and, from an econometric point of view, has noteworthy characteristics such as nonstandard rates of convergence. We provide a feasible and relatively straightforward estimation and inference implementation strategy.

C193: **Retrieving risk neutral moments and expected quadratic variation from option prices**
Presenter: **Elias Tzavalis**, Athens University of Economics and Business, Greece
Co-authors: Leonidas Rompolis

Exact formulas are derived for retrieving risk neutral moments of future payoffs of any order from generic European-style option prices. It also provides an exact formula for retrieving the expected quadratic variation of the stock market implied by European option prices which nowadays is used as an estimate of the implied volatility. To implement the above formulas to discrete sets of option prices, a numerical procedure is suggested and upper bounds of its approximation errors are provided. The performance of this procedure is evaluated through a Monte Carlo exercise. It also shows that ignoring the jump component of the underlying asset can lead to seriously biased estimates of the new volatility index suggested by the Chicago Board Options Exchange (CBOE). This is also confirmed by an empirical exercise based on market option prices written on the S&P 500 index.

CS60 Room Torrington ECONOMETRICS WITH R Chair: Achim Zeileis

C452: **Panel covariate augmented Dickey-Fuller tests with R**
Presenter: **Claudio Lupi**, University of Molise, Italy

The implementation of pCADF (panel covariate augmented Dickey-Fuller) tests in R is described. Furthermore, it is shown that pCADF tests have

interesting properties. In particular, compared to other popular panel unit root tests, pCADF tests have larger power and smaller size distortions even in the presence of cross-correlated time series. The R implementation, that includes other p values combination panel unit root tests, is part of the ongoing work on a new R package named punitroots. Development versions of punitroots are available from R-forge.

C454: **More on spatial models in R: spse**

Presenter: **Gianfranco Piras**, West Virginia University, United States of America

spse is a package that contains functions for fitting simultaneous systems of spatially interrelated cross-sectional equations. The model relates to a single cross-section of observations. Spatial dependence arises from two sources. On one hand, the error terms are assumed to be both spatially autocorrelated and correlated across equations. On the other hand, the value of the dependent variable in each equation depends on its spatial lag (and the lags of all other dependent variables). The flexible specification implemented allows also for additional (other than spatial) endogeneity. We consider the implementation of maximum likelihood and generalized moments estimators. Both the limited and full information instrumental variable estimators are implemented. The limited information estimation is a straightforward extension of the GS2SLS procedure in the context of a single equation model. In line with the changes and improvements made in spdep, the calculation of the Jacobian term in the ML estimation has been extended to various methods including the eigenvalues decomposition, sparse matrices, and the Chebyshev and Monte Carlo approximations.

C467: **punitroots: Infrastructure for panels with unit roots**

Presenter: **Christian Kleiber**, Universitaet Basel, Switzerland

Ongoing work on a new R package named punitroots is reported. It aims at providing infrastructure for panels with large time dimension, including several tests for unit roots in panels of time series as well as a number of datasets, old and new. An overview of some popular tests in the area and their implementation in R is provided, among them regression-type tests stemming from the classical LLC test and its variants and Fisher-type tests employing combinations of p-values. A comparison with other software packages (notably Stata) is included. Development versions of punitroots are available from R-forge.

C480: **ML estimation of spatially and serially correlated panels with random effects: an estimation framework and a software implementation**

Presenter: **Giovanni Millo**, Assicurazioni Generali SpA, Italy

It is described a procedure for maximum likelihood estimation of panel models incorporating: random effects and spatial dependence in the error terms; and/or a spatially lagged dependent variable; and possibly also a serial dependence structure in the remainder of the error term. We start by sketching a taxonomy of spatial panel models, beginning with the two basic random effects (RE) specifications used in the literature: the spatial autoregressive (SAR) RE model containing a spatially lagged dependent variable and a group-specific, time-invariant component in the error term, and the spatial error (SEM) RE model, with both a group-specific component and a spatial dependence structure in the error term. Extending the SEM specification, an encompassing model allowing for serial correlation in the residuals is considered. Restrictions of the full model give rise to 18 different specifications. It is discussed an efficient implementation of the estimation procedure in R, to be added to the splm package for estimation and testing of spatial panel models, illustrating it through some well-known examples from the literature.

Monday 19.12.2011 14:05 - 15:45 Parallel Session O – ERCIM

ESI01 Room Beveridge ROBUST METHODS Chair: Alastair Young

E208: **Robust correlations revisited**
Presenter: **Stefan Van Aelst**, Ghent University, Belgium

Several robust estimates of correlation are considered and their performance is compared when estimating correlations in a biotechnology application. In this data set the goal is to estimate the correlation between microRNA and mRNA expression levels. The expression levels are obtained by next generation sequencing which is a cheap technology, but also is error-prone, which creates the need for reliable robust estimates of correlation.

E525: **Robust cluster analysis based on trimming: Review and advances**
Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain
Co-authors: Luis A. Garcia-Escudero, Carlos Matran, Agustin Mayo-Iscar

The TCLUST methodology is aimed to perform robust cluster analysis to find out heteregeneous groups in data sets, in the presence of both background noise and any kind of outliers. TCLUST belongs to the class of model-based cluster methods and it assumes a normal model for the clusters. Some kind of constraints on the scatter matrices are needed in order that the maximization of the likelihood function can be solved. This talk is devoted to make a review of this methodology, including the main theoretical results, the robustness properties and computational aspects. Finaly, some recent developments of this methodology applied to the Common Principal Components problem and Fuzzy Clustering will also be presented and discussed.

E559: **Breakdown and efficiency in complex models**
Presenter: **Marco Riani**, University of Parma, Italy

Data are an overwhelming feature of modern life. As the amount of data increases, so do the challenges facing the statistician in trying to extract information from ever larger data sets. We argue that larger data sets are also more complex and require flexible multiple analyses in order to reveal their structure. Only then all information can be efficiently extracted. In this paper we consider complex regression and multivariate data which may contain different subpopulations and may be contaminated with the presence of multiple outliers. The goal of this work is to compare the information which can be extracted from the use of traditional robust estimators (M, MM, S, τ), which try to combine robustness and efficiency, with those which use a flexible level of trimming and are based on the forward search estimator.

ES03 Room B36 IMPRECISION IN STATISTICAL DATA ANALYSIS I Chair: Olivier Strauss

E811: **The mids/ldev/rdev characterization of a fuzzy number. Some statistical applications**
Presenter: **Beatriz Sinova**, Universidad de Oviedo, Spain
Co-authors: Ana Colubi, Maria Angeles Gil, Stefan Van Aelst

To model a lot of real life experiments, the scale of fuzzy numbers is used, since it integrates the mathematical manageability and variability of real numbers with the interpretability and expressivity to reflect the underlying imprecision (of valuations, ratings, judgements or perceptions) of the categorical scale. Examples and applications can be found in very different fields, from Engineering to Social Sciences. A new characterization of a fuzzy number, in terms of the mids and a kind of left and right deviations, is now introduced. Based on this characterization, an L^1 metric is proposed, and its properties are examined. The generalized Hausdorff metric defined for non-empty compact intervals is a particular case. Finally, as an application we will consider an extension of the notion of median for random fuzzy numbers, and its robustness with respect to the Aumann expected value is compared by means of some simulations.

E838: **Clustering imprecise data using the fuzzy EM algorithm**
Presenter: **Benjamin Quost**, University of Technology of Compiegne, France
Co-authors: Thierry Denoeux

The problem of clustering imprecise data using finite mixtures of multivariate Gaussians is addressed. Data imprecision is assumed to be represented by possibility distributions and mixture model parameters are estimated using the fuzzy EM algorithm, an extension of the EM algorithm making it possible to handle fuzzy data. Different shapes of possibility distributions are considered. When possibility distributions are multivariate Gaussians (with arbitrary shape and orientation), we provide closed forms for the parameter update. In other cases such as trapezoidal distributions, Monte-Carlo estimation techniques have to be used. A generic procedure is proposed, in which data are sampled according to the membership functions. Experiments with synthetic and real data, in both cases of Gaussian and trapezoidal possibility distributions, demonstrate the interest of our approach for clustering ill-known data.

E744: **Fuzzifying Likert scales with factor analysis techniques**
Presenter: **Maria Symeonaki**, Panteion University of Social and Political Sciences, Greece
Co-authors: Aggeliki Kazani, Kate Michalopoulou

A methodology for developing fuzzy Likert scales based on factor analysis techniques is presented. The proposed methodology enables the examination of operational definitions and relations between items-questions (manifest variables). The fuzzification of items-questions and intermediate (latent) variables provided by factor analysis results in constructing the necessary rules for the development of a fuzzy system to be used in analysis. The evaluation of the methodology presented is tested on a Likert scale that was used in a large-scale survey for measuring xenophobia in Northern Greece conducted by the National Centre of Social Research.

E343: **Incorporating imprecise prior knowledge in multiple output regression**
Presenter: **Jan Verwaeren**, Ghent Univeresity, Belgium
Co-authors: Willem Waegeman, Bernard De Baets

The incorporation of domain-specific knowledge can substantially boost the predictive performance of machine learning algorithms, in particular in situations where data are scarce, low in quality or not identically distributed. Although such knowledge is inherently available in real-world applications, human experts are often only capable of describing it in an imprecise manner. We consider the specific case of including prior knowledge in multiple output regression problems (a.k.a. multivariate regression problems). Here, domain knowledge often takes the form of input-output relations that are specified by human experts and rough assessments of output-output correlations. We present a multiple output support vector regression framework, in which a joint optimization problem is considered for all output variables and expert knowledge is encoded by adding soft constraints and a modified regularization term. The model fitting process takes in this way domain knowledge into account without

enforcing it explicitly. Additionally, we discuss an application in chemometrics where this type of imprecise expert knowledge can be easily obtained.

E156: **Messy data analysis of interval data**
Presenter: **Maria Ana Odejar**, University of Aberdeen, United Kingdom

A Bayesian Markov chain Monte Carlo (MCMC) algorithm using data augmentation is developed to provide separate estimates and inference about the willingness to pay (WTP) interval data of the resolved cases that is the true zeros, bid payers and bid refusers and the non-ignorable unresolved or missing cases which are the do not knows, protests and true-missing responses. The latent decision score and the corresponding WTP interval is modelled as a bivariate ordered probit model with truncated regimes. The missing decision scores for the unresolved cases are generated using a finite mixture model. A simple technique of restricting the error variances is introduced. Results show that inclusion of the non-ignorable missing responses is not detrimental to the quality of parameter estimates, as reflected in their standard deviations, sum of predictive ordinate goodness-of-fit criterion and predictive ability of the model in and out of the estimation sample, as well as in the mixing and convergence properties of the Markov chains generated by the MCMC Bayesian algorithm. Correlation of the latent decision score and the WTP errors is high even with inclusion of the unresolved cases.

ES82 Room B34 CLUSTERING AND CLASSIFICATION Chair: Luis A. Garcia-Escudero

E878: **Multi-label classification in the belief function framework using the evidential random k-labelsets method**
Presenter: **Sawsan Kanj**, University of Technology of Compiegne, France
Co-authors: Fahed Abdallah, Thierry Denoeux

Multi-label classification is used in many real-world applications, where each instance can belong simultaneously to different classes. An effective multi-label method, named RAKEL, randomly breaks the set of labels into a number of k labelsets smaller than that of the initial number of labels and trains a single-label classifier for each element of the powerset of this labelset. To decide on unseen instances, the decisions of the single-label classifiers are gathered and combined, usually by a voting process. In this way, we have uncertain information on some labels, especially when the initial number of labels is high. However, the choice of parameter k is critical, as a small value of k leads to ignoring the correlation between labels. To solve these problems, the RAKEL method is adapted under the belief function framework applied on set-valued variables. The novel method is able to output an ensemble of valued subsets of labels. These subsets are then aggregated using belief functions to obtain a final decision for unseen instance. Experiments on real world data sets, including gene, scene and emotions classification, show the effectiveness of the proposed approach as compared to the original one.

E892: **Spatial clustering through aggregation control**
Presenter: **Ayale Daher**, European University of Brittany, France
Co-authors: Thierry Dhorne

The problem of spatial units clustering when the number of clusters C and the number of spatially connected components R are fixed ($R \geq C$) is introduced. For classifying spatially localized sites, it may be suitable to take into account their relative positions or even to impose spatial constraints on the obtained clustering. When this is not imposed, it is common to get too fragmented partitions. A natural way to use spatial information is to consider a matrix of spatial distances in order to modify the distances or dissimilarities d_{ij} calculated with the non-geographic variables. As a result the clustering method uses a new dissimilarity matrix $D^* = \{d^*_{ij}\}$ combining geographic information and non-geographic, which tends to form connected clusters. The idea that guides our work is that, in general, a spatial clustering has no operational interest if it does not take into account the spatial structure of the sites. Instead of seeking a general solution to the optimization probleme initially set, we will use an algorithm to control the intra-class spatial aggregation for a fixed number of clusters in order to approach (possibly achieve) the intial constraint set on the number of regions. We present in detail the proposed method and show how the level of aggregation can be controlled by a parameter of the model. Finally we look for the parameter value that leads to the best solution. The results are illustrated on a real example.

E905: **Modelling the development of late-onset sepsis and corresponding length of stay within preterm neonatal care**
Presenter: **Kieran Payne**, Queen's University Belfast, United Kingdom
Co-authors: Adele Marshall, Karen Cairns, Emma McCall, Stanely Craig

Late-onset sepsis syndrome is a particular type of infection which occurs at 4-90 days of life and is acquired from the care giving environment. Leading to a significant increase in length of stay, the early detection or identification of an infant being at higher risk of developing late-onset sepsis can help prevent the development of the syndrome. This can then reduce what could be an extended period of time within neonatal care. The reduction in length of stay is of benefit to infant, parents and the health service, with the average cost per day in 2009/10 of neonatal care in Northern Ireland being approximately £1018. Using discrete conditional survival models this work illustrates how classification trees can be used to identify factors which increase the risk of new born infants developing late-onset sepsis. Conditioned on this decision the predicted length of time spent within neonatal care is modelled using a mixture of Coxian phase-type distributions and regression models. The resulting model can be used to predict which babies will more than likely acquire infection and their resulting length of stay. In particular, this approach will be beneficial for hospital management in advanced resource allocation.

E884: **Evolutionary algorithm to optimise a support vector machine for probability of default prediction**
Presenter: **Dedy Dwi Prastyo**, Humboldt University of Berlin, Germany
Co-authors: Wolfgang Haerdle, Russ A. Moro

Default risk prediction is very important for financial institutions. The prediction method must be precise and easily adaptable to false acceptances and false rejection, or Type I and Type II errors respectively, that usually si a serious problem in imbalanced data sets. Nonlinear classification technique, i.e. Support Vector Machines (SVMs), now is a very popular method to predict the probability of default (PD). One of the crucial issues in SVM is how to choose the value of its parameters. We explore Genetic Algorithms (GA) for the parameters optimisation to classify solvency status of Germany firms based on financial ratio information. GA is an evolutionary algorithm able to find a global optimum solution. We show that GA-SVM outperforms competing models such as discriminant analysis, logit model, and classification and regression trees (CART). GA-SVM is also more robust to the imbalanced classification problem applied in credit reform data set.

E810: **Identification of connected regions in density-based clustering methods: a new dimension-independent technique**
Presenter: **Giovanna Menardi**, University of Padova, Italy
Co-authors: Adelchi Azzalini

Density-based clustering methods hinge on the idea of identifying groups with the regions around the modes of the probability distribution underlying the data. Any section of such distribution, at a level k, identifies a region of the sample space having density greater than k. Clusters correspond to the maximum connected components of this region. Nonparametric methods are used to estimate the probability density function and allow for detecting the high density level sets. While they exhibit several advantages and generally better performance than traditional distance-based procedures, clustering methods based on density estimation involve a non-trivial computational effort, required for the identification of the connected

regions. The use of geometric tools such as the Delaunay triangulation, for instance, suitably generalizes the univariate procedure for detecting the connected components, because spatial tessellation defines a contiguity relationship in the space. However, the computational complexity grows exponentially with the data dimension, thus making the triangulation unfeasible for large dimensions. We discuss the use of an alternative procedure to the Delaunay triangulation for identifying the high-density connected regions. The proposed procedure claims the notable advantage of being independent from the dimensionality of data.

ES83 Room B33 CONTRIBUTIONS IN PARAMETRIC AND SEMIPARAMETRIC INFERENCE Chair: M. Dolores Jimenez-Gamero

E229: **Estimation of Cox proportional hazards models in the presence of a negatively correlated frailty**
Presenter: **Joseph Gastwirth**, George Washington University, United States of America
Co-authors: Wenjing Xu, Qing Pan

Survival analysis has been used to estimate the lost wages of individuals who have been unfairly denied an employment opportunity. In the context of discrimination in promotion, the lost opportunity may affect the individual's decision to leave. This situation can be modeled by two Cox PH models, one for promotion conditional on not leaving and one for the leaving process. To account for the negative relationship between the two outcomes, as being passed over for promotion would encourage an employee to look for alternative opportunities, in addition to the fixed covariates, a gamma frailty term which increases the risk of one process multiplicatively but decreases the risk of the other outcome is introduced. An MCEM algorithm for fitting the joint process is proposed. The asymptotic properties of the parameter estimates are derived and verified through simulation studies. The advantages of modeling the negative relationship over simply fitting two individual Cox PH models are explored. The method is applied to data from an actual legal case.

E798: **Interval estimation of parametric functions in partially non-regular log-exponential models**
Presenter: **Inmaculada Barranco-Chamorro**, University of Sevilla, Spain
Co-authors: M. Dolores Jimenez-Gamero

Two-parameter models that can be called partially non-regular log-exponential models are considered. The term log-exponential refers to the fact that they can be related to the exponential distribution via a logarithm transformation, whereas the term partially non-regular refers to one of the estimators, in which our inferences are based on, is regular and the other one is non-regular. Our aim is to illustrate the applicability of theoretical results about the limiting distribution of maximum likelihood estimator (MLE) of a given parametric function when sampling from these models. Specifically, we focus on the performance of Wald-type approximate confidence intervals. To reach this aim, we consider functions of interest such as Lorenz curve, Gini index and quartiles in the exponential and/or Pareto distribution. By using R, we have carried out numerical simulations and analyze the empirical probability coverage and the lengths of these intervals. Although we get good results, in certain particular cases some improvements can be proposed.

E951: **Closed likelihood-ratio testing procedures to assess similarity of covariance matrices**
Presenter: **Antonio Punzo**, Universita di Catania, Italy
Co-authors: Francesca Greselin

The study of similarity between k covariance matrices Σ_h, referred to as k groups, under the assumption of multivariate normality is extended. Our analysis is based on the reparameterization $\Sigma_h = \lambda_h \Gamma_h \Delta_h \Gamma_h'$ where λ_h, Δ_h, and Γ_h specify volume, shape, and orientation of the density contours in each group, respectively. By allowing each of these quantities to be equal or variable between groups, one obtains eight configurations – in which homoscedasticity and heteroscedasticity represent the limit cases – characterized by a different degree of similarity. Due to its desirable properties in the multiple testing framework, we introduce an easily implementable closed testing procedure allowing for a choice between the eight configurations. Likelihood-ratio tests are used as local tests in the procedure for their optimality properties. The proposed approach discloses a richer information on the data underlying structure than the classical existing methods, the most common one being only based on homo/heteroscedasticity. At the same time, it allows a more parsimonious parameterization, whenever the constrained model is appropriate to describe the real data. The new inferential methodology is finally applied to some well-known data sets, chosen in the multivariate literature, in order to exemplify its use. Our proposal has been also compared with some well-known likelihood-based information criteria.

E938: **Mixtures of regressions with Student-t errors**
Presenter: **Simona C. Minotti**, Universita di Milano-Bicocca, Italy

Mixtures of linear regression models are traditionally based on the normal assumption for the error terms and then the regression parameters are estimated by maximum likelihood using the EM algorithm. Under the normal assumption, the M step of the EM algorithm uses a weighted least squares estimate (LSE) for the regression parameters. It is well known that the LSE is sensitive to outliers or heavy tailed error distributions. Some authors have proposed replacing the least square criterion with some robust criteria, e.g. Huber's function, Tukey's biweight function. As an alternative we propose to model error terms by means of Student-t distribution, as it has been introduced in robust regression, to obtain Finite Mixture of Regressions with Student-t errors. ML estimation of parameters based on Student-t distribution is a form of M estimation, yielding robust estimates of location with a redescending influence function. With sufficient data, degrees of freedom may be estimated from the data by ML, yielding an adaptive robust procedure. Different methods will be compared on the basis of artificial and real datasets.

ES60 Room G16 COMPUTER-AIDED DATA ANALYSIS Chair: Dietmar Maringer

E733: **Forecasting financial time series with Twitter**
Presenter: **Ramon Xuriguera**, Universitat Politecnica de Catalunya, Spain
Co-authors: Marta Arias, Argimiro Arratia

The aim is to assess whether a sentiment index (SI) constructed from Twitter data can improve the forecasting of financial time series. To this end we have collected and processed a large amount of Twitter messages from March 2011 to the present date related to companies listed in NASDAQ, and have built a time series of real values that reflect the positive or negative mood of the public. We first tested for non-linearity and causality relationships between this SI and the stock time series, and based on the observed relations we trained a wide variety of forecasting models (linear regression, neural networks, support vector machines, and others) under many parameter settings, both with and without the SI; thus building a database on which we can verify the hypothesis of whether the inclusion of the SI improves forecasting. In order to cope with the hundreds of results obtained under the different experimental settings, we have developed a decision tree-based summarisation method of this information and have implemented it with the open source data mining toolkit Weka. We found that the Twitter-based sentiment index is especially helpful when paired with support vector machines.

E735: **Bayesian inequality constrained multidimensional scaling**
Presenter: **Kensuke Okada**, Senshu University, Japan

Parameter estimation and model selection in multidimensional scaling with inequality constraints on the distance parameters is considered. Traditionally, multidimensional scaling has been mostly applied in an exploratory context. However, there should also be many occasions in which researchers would like to evaluate their informative hypotheses on the distance parameters. A typical situation would be when a certain number of categories or groups exists between the objects, and it can be hypothesized that all within-group dissimilarities should be smaller than the dissimilarities between groups. To achieve this goal we employ the encompassing prior approach in which the inequality constrained model and the encompassing (unconstrained) model are compared. A Markov chain Monte Carlo estimation algorithm is used to simultaneously obtain the posterior estimates of both constrained and unconstrained model parameters. The post-processing is introduced to deal with the indeterminacy problem of the configuration matrix. The goodness of the constrained model is evaluated by calculating the Bayes factor in comparison with the encompassing model. The proposed method is illustrated by the analysis of artificial and real psychological data.

E845: **Multivariate Markov chain approximations**
Presenter: **Simon Scheuring**, University of Zurich, Switzerland

In order to solve equilibrium models numerically, it is necessary to discretise vector autoregressive processes (VAR) into a finite number of states. The univariate case for such Markov chain approximations is well studied; however, the multivariate case has been scarcely addressed in the literature. We argue that the common approach to apply multivariate Gaussian quadrature has difficulties for small numbers of states. To address this weakness, two alternatives are proposed: moment matching and direct bin estimation. Moment matching constructs the discrete Markov chain, such that the first moments are exactly the same as in the VAR. Direct bin estimation leaves out the first step of estimating the VAR and estimates the discrete Markov chain directly, by forming bins and then estimating transition probabilities with maximum likelihood. Finally, moment matching and bin estimation are compared to four different implementations of Tauchen's approach in a standard Lucas asset pricing framework.

E809: **Fuzzy GARCH Models**
Presenter: **Rui Jorge Almeida**, Erasmus University Rotterdam, Netherlands
Co-authors: Nalan Basturk, Uzay Kaymak, Joao M. C. Sousa

Time series data exhibits complex behavior including path-dependency and non-linearity. It is important to employ flexible methods, such as a mixture of GARCH models, to allow for possible changes in the nature of this complex behavior. This work proposes a novel flexible fuzzy GARCH model that attempts to capture complex data behavior. Issues of parameterization and estimation are discussed. Regarding the former, we extend a previously studied fuzzy GARCH model and propose a new model. This model can capture changes in the conditional variance over time and also more complex data features such as skewness and multimodality, making the proposed model more general than standard GARCH models and the previous fuzzy GARCH model. It is further shown that the proposed model can be related to advanced dynamic mixture models for conditional volatility of the data, as well as for data average. The dynamic mixture nature of the model is captured using fuzzy rules that establish relations between the defined variables. The model parameters are obtained using the maximum likelihood approach. The results are illustrated using simulated data with different specifications, and real stock market data.

E668: **A Bayesian autoregressive three-state HMM for switching monotonic regimes for microarray time course data**
Presenter: **Serena Arima**, Sapienza University of Rome, Italy
Co-authors: Alessio Farcomeni

The advent of microarray technology has enabled biomedical researchers to monitor changes in the expression levels of thousands of genes. The researchers are more and more interested in characterizing temporal patterns of gene expression. A general limitation of available approaches is that time series are usually assumed to be stationary, which is not justifiable biologically. Further, researchers are actually particularly interested in temporary trends in expression levels. A trajectory may change its regime because of the reaction to treatment or of natural developmental phase. We propose a Bayesian autoregressive hidden Markov model which accounts for switching monotonic regimes: we propose a three state latent Markov process that flexibly allows to model the trajectory as stationary or as non-stationary with increasing and decreasing trends. We illustrate the approach on identification of genes involved in embryo development of mice with the 22q11 deletion.

ES75 Room B35 CONTRIBUTIONS TO EXTREME VALUE THEORY AND APPLICATIONS Chair: Juerg Huesler

E602: **Extremal dependence and the ACE algorithm**
Presenter: **Keith Knight**, University of Toronto, Canada

The maximal correlation between two random variables X and Y was defined by Hans Gebelein in 1941 to be the supremum of $E[\psi(X)\phi(Y)]$ over transformations ψ and ϕ satisfying $E[\psi(X)] = E[\phi(Y)] = 0$ and $E[\psi^2(X)] = E[\phi^2(Y)] = 1$. Given a sample from a joint distribution, the maximal correlation (as well as the optimal transformations ψ and ϕ) can be estimated via the alternating conditional expectations (ACE) algorithm. In this talk, we will define a version of maximal correlation for bivariate extremal dependence, which depends on the random variables only via their copula, and discuss its estimation using an appropriate modification of the ACE algorithm.

E633: **A simple generalization of the Hill estimator**
Presenter: **M. Ivette Gomes**, University of Lisbon, Portugal
Co-authors: Dinis Pestana

We are interested in a very simple generalization of the classical Hill estimator of a positive extreme value index (EVI), the primary parameter of extreme events. Given a random sample of size n, $X_1, X_2, \ldots, X_n$, from a common model F, let us denote by $X_{1:n} \leq X_{2:n} \leq \cdots \leq X_{n:n}$ the associated ascending order statistics. The Hill estimator, the mean of the log-excesses $V_{ik} := \ln U_{ik}$, with $U_{ik} := (X_{n-i+1:n}/X_{n-k:n})$, can be regarded as the geometric mean, or equivalently the mean of order $p = 0$, of the set of fulcral statistics $U_{ik}, 1 \leq i \leq k$. Instead of this geometric mean, we shall now more generally consider the mean of order $p \geq 0$ of those fulcral statistics, and the class of associated EVI-estimators. Apart from the derivation of the asymptotic behaviour of the new class of EVI-estimators, we shall proceed with an asymptotic comparison, at optimal levels, of the members of such a class. Moreover, a small-scale simulation study will be developed as well as an application to real data in the fields of finance, insurance and environment.

E704: **Asymptotic behavior of max-type test statistic for detecting multiple change points**
Presenter: **Daniela Jaruskova**, Czech Technical University, Czech Republic

A sequence of independent random variables is observed sequentially in time. We test a null hypothesis claiming that all variables have the same mean (they are identically distributed) against the alternative that there are at least one and at most d shifts in their mean occurring at some unknown time points. We consider a modified max-type test statistic. Its exact distribution is very complex and that is why we look for a limit distribution. The limit variable is a maximum of a certain Gaussian non-homogeneous random field. Approximate critical values may be obtained from the tail

behaviour of the limit variable. We will present this approximation together with the comparison of approximate critical values with critical values obtained by simulations.

E843: **Bias reduction of the geometric-type estimator and high order quantiles**
Presenter: **Laura Cavalcante**, Universidade do Porto, Portugal
Co-authors: Margarida Brito, Ana Freitas

The estimation of the extreme-value index and of high order quantiles is a central topic in the extreme value analysis. Several estimators for the extreme-value index have been proposed in the literature. Motivated by the applications to risk theory, we consider here a geometric-type estimator for the tail index and, based on this estimator, we propose two bias corrected estimators and study the corresponding asymptotic behavior. We also consider the problem of estimating high order quantiles using the geometric-type estimator and we establish the asymptotic normality of the resulting estimator. Finally, we provide some simulations in order to illustrate the finite sample behavior of the proposed estimators.

E925: **Tail regression quantile process and its applications**
Presenter: **Jan Dienstbier**, Technical University of Liberec, Czech Republic
Co-authors: Jan Picek

In the contribution we examine properties of the tail regression quantile process, which we define for some intermediate sequence k as $\hat{\beta}_n(1-kt/n)$, $t \in [0,1]$, where $\hat{\beta}_n(\alpha)$ is the α-regression quantile calculated in the simple linear model $Y_i = x_i^\top \beta + E_i$ with $x_i \in R^p$, $E_i \sim F$ and F satisfying the second order condition. Using Bahadur representation for the regression quantile process $n^{1/2} f(F^{-1}(\alpha))(\hat{\beta}_n(\alpha) - \beta(\alpha))$ and the properties of the maximum regression quantile $\hat{\beta}_n(1)$ we get the approximation of the tail regression quantile process. This approximation enables a wide range of practical applications from which we are particularly interested in the estimation of the extreme value index. We introduce a class of location and scale invariant estimators being smooth functionals of the tail regression quantile process and we derive their consistency and asymptotic normality. We illustrate the performance of the estimators using simulated data as well as real data analysis.

ES84 Room B18 STATISTICAL ALGORITHMS AND COMPUTER EXPERIMENTS Chair: Cristian Gatu

E759: **A computational approach to nonparametric regression: Bootstrapping the CMARS method**
Presenter: **Ceyda Yazici**, Middle East Technical University, Turkey
Co-authors: Fatma Yerlikaya-Ozkurt, Inci Batmaz

Bootstrapping is a computer-intensive statistical method which treats the data set as a population and draws samples from it with replacement. This resampling method has wide application areas especially in mathematically intractable problems. In this study, it is used to obtain empirical distributions of the parameters to determine whether they are statistically significant or not in a special case of nonparametric regression, conic multivariate adaptive regression splines (CMARS). It is a successfully modified version of the well-known nonparametric regression model, multivariate adaptive regression splines (MARS), which uses conic quadratic optimization. Although CMARS performs better with respect to several criteria, it is at least as complex as MARS. To achieve a less complex CMARS model without degrading its performance, three different bootstrapping regression methods, namely, Random-X, Fixed-X and Wild Bootstrap are applied on four data sets with different size and scale. Then, performances of the models are compared using various criteria including accuracy, precision, complexity, stability, robustness and efficiency. The results imply that the Random-X method produces more precise, accurate and less complex models for medium size and medium scale data.

E881: **KrigInv, an R package for sequential inversion of expensive-to-evaluate black-box simulators**
Presenter: **Clement Chevalier**, University of Bern, Switzerland
Co-authors: David Ginsbourger, Victor Picheny, Yann Richet

The use of complex code modeling physical phenomena has become commonplace in many applications. In simulation-based reliability studies (e.g. in nuclear safety) one often pays a special attention to one scalar output of interest f(x), where x stands for a d-dimensional input parameter of the simulator. When a critical threshold T is fixed, retrieving the excursion set $\Gamma = \{x : f(x) > T\}$ in a limited number of evaluations of f is a challenging problem. We choose here to address this issue using a kriging metamodel. Starting from an initial set of evaluations at different x's the kriging metamodel delivers both a response surface (the kriging mean) and an associated measure of uncertainty (the kriging variance) that are very useful for designing sequential evaluation strategies. More specifically, we implemented in the KrigInv R package the latest developments on the so-called stepwise uncertainty reduction strategies for inversion and probability of failure estimation. As illustrated by different low-dimensional test cases (including a nuclear safety application), these strategies appear to successfully identify the set Γ by sequentially evaluating f at well-chosen points. We also discuss the proposed implementation of SUR strategies, both in terms of complexity reduction for the computation of the infill sampling criteria, and in terms of improved numerical integration based on ad hoc importance sampling techniques.

E758: **New software for predictive data mining: The conic multivariate adaptive regression splines (CMARS) algorithm**
Presenter: **Inci Batmaz**, Middle East Technical University, Turkey
Co-authors: Fatma Yerlikaya-Ozkurt, Ceyda Yazici

CMARS is an alternative approach to the well-known data mining tool Multivariate Adaptive Regression Splines (MARS). It is based on a penalized residual sum of squares (PRSS) for MARS expressed as a Tikhonov regularization (TR) problem. CMARS treats this problem by a continuous optimization technique, in particular, the framework of Conic Quadratic Programming (CQP). These convex optimization problems are very well-structured, herewith, resembling linear programs and, hence, permitting the use of interior point methods. For examining the efficiency of the developed method CMARS, we compare with other data mining tools by using several data sets, each with different features. According to the results, in general, CMARS produces more accurate, robust and stable models than the other estimation methods. CMARS, written in MATLAB, has been developed as a user-friendly computing environment. With this study, we intended to put the CMARS algorithm in use, and let its performance be evaluated by other modelers and data analysts in the fields of nonparametric regression/classification and data mining.

E583: **Stochastic version of the EM algorithm for analysis of generalized sample selection models**
Presenter: **Shahab Jolani**, Utrecht University, Netherlands

The multivariate generalization of the classical Heckman selection model is a possible candidate for analyzing nonignorable missing data in longitudinal studies. Estimation procedure in this model is often hampered by complex patterns of incompleteness, that make the implementation of the EM algorithm to be difficult. A stochastic version of the EM algorithm, the StEM algorithm, is proposed and developed to find estimates of the parameters in the case of nonmonotone missing data. A Monte Carlo method is also used to obtain the standard errors of the estimates. We evaluate the performances of the StEM algorithm using a simulation study. The proposed algorithm is also applied to the well known Mastitis data. We found that the StEM algorithm performs well for the analysis of sample selection models.

E830: **Flexible Bayesian modeling for emulation and calibration of stochastic simulators**
Presenter: **Marian Farah**, University of Cambridge/MRC, United Kingdom
Co-authors: Athanasios Kottas, Robin Morris

Stochastic computer simulators are being increasingly used in technology and science to model random systems, e.g., in population dynamics, biological processes, queuing networks in a communication system, and nuclear interactions. We propose a new framework for emulation and calibration of stochastic simulators, where the emulator is built using a flexible Bayesian nonparametric mixture model for the joint distribution of the simulator inputs and output. For calibration, we develop a semiparametric approach to link the simulator data with field observed data using a two-stage approach. In the first stage, the posterior distribution of the emulation model parameters is estimated solely based on simulator data, and then, in the second stage, the posterior distribution of the calibration parameters is approximated based on the field data as well as the posterior distribution of the emulation model parameters obtained in the first stage. The methods developed in this work are applied to a synthetic data example as well as PROPSET, a stochastic simulator that models the bombardment of a device with high-energy protons in order to determine the effect of radiation on spaceborne microelectronics.

ES50 Room B20 COMPUTATIONAL STATISTICS Chair: Klea Panayidou

E943: **Efficient algorithms for the re-estimation of the general linear and SUR models after deleting observations**
Presenter: **Stella Hadjiantoni**, University of Cyprus and Queen Mary, University of London, Cyprus
Co-authors: Erricos Kontoghiorghes

A new estimation method for the (downdating) problem of removing observations from the general linear model (GLM) after it has been estimated is proposed. It is verified that the solution of the downdated least squares method can be obtained from the estimation of an equivalent GLM where the original model is updated with the imaginary deleted observations. This updated GLM has a dispersion matrix which comprises complex covariance values. The model is formulated as a generalised linear least squares problem (GLLSP) and its solution is derived by employing the generalised QR decomposition (GQRD). Previous computations are efficiently utilized. The algorithm is based on hyperbolic reflections but no complex arithmetic is used in practice. Strategies for exploiting the special structure of the matrices are also developed. The downdating problem of Seemingly Unrelated Regressions (SUR) model is also considered.

E591: **A flexible instrumental variable approach to semiparametric expectile regression**
Presenter: **Fabian Sobotka**, University Oldenburg, Germany
Co-authors: Rosalba Radice, Giampiero Marra, Thomas Kneib

Classical regression model literature has generally assumed that measured and unmeasured or unobservable covariates are statistically independent. For many applications this assumption is clearly questionable. When unobservables are associated with included regressors and have an impact on the response, standard estimation methods will result in an inconsistent and biased estimate. Instrumental variable approaches are among the most used techniques for isolating the effect of a treatment variable in the presence of unobservable confounding. Another important issue is the possible presence of heteroscedasticity in real data sets. This means that traditional mean regression techniques may not be sufficient to describe a treatment-response relationship. Hence, model specifications beyond mean regression describing more general properties of the response distribution are required. Among the few available approaches, expectile regression represents an effective tool which has several preferable attributes. For one, expectile regression avoids the linear programming techniques that are needed to compute a quantile regression estimate. Moreover, the least squares estimation of expectile regression allows for flexible model structures. We provide a flexible instrumental variable estimation technique, based on a two-step approach, for expectiles in the presence of unmeasured confounding. Its quality is examined via an extensive simulation study. Further, corrected confidence intervals are defined which assure appropriate coverage probabilities. The methods are then used to assess the impact of education on the fertility of a sample of women in Botswana. The analyses are carried out with our R-package expectreg.

E722: **Smoothing survival densities in practise**
Presenter: **Maria Dolores Martinez-Miranda**, University of Granada, Spain
Co-authors: Maria Luz Gamiz Perez, Jens Perch Nielsen

A class of local linear density estimators based on filtered data and address the problem of bandwidth selection is considered. Filtered data is defined as survival data following the Aalen multiplicative hazard model and include i.i.d. observations and their possible independent left truncation or right censoring. Survival data is omnipresent in statistical applications and the density is one of the fundamental building blocks while constructing mathematical statistical models. In this general context we present first the classical crossvalidation and then we develop the recent independent identical distributed do-validation method for the density estimators in the class. Opposite to complicated plug-in procedures, do-validation shares the simplicity and the intuitive appeal of crossvalidation, but do-validation exhibits a better performance than its classical counterpart. Through simulation experiments we see that do-validation works well also in this more general context. Finally. we include an application with real data where crossvalidation and do-validation are used in developing the graphical test of parametric mixed hazard models proposed in a recent paper.

E690: **Variable selection of varying coefficient models in quantile regression**
Presenter: **Hohsuk Noh**, Universite Catholique de Louvain, Belgium
Co-authors: Kwanghun Chung, Ingrid Van Keilegom

Varying coefficient (VC) models are commonly used to study dynamic patterns in many scientific areas. Especially, VC models in quantile regression are known to provide a more complete description of the response distribution than in mean regression. In this paper, we develop a variable selection method for VC models in quantile regression using a shrinkage idea. The proposed method is based on the basis expansion of each varying coefficient and the regularization penalty on the l2-norm of the corresponding coefficient vector. We show that the proposed estimator is given as an optimal solution to the second order cone programming (SOCP) problem and that under suitable conditions the proposed procedure has consistency in variable selection. Furthermore, we show that the estimated relevant coefficients converge to the true functions at the univariate optimal rate. Finally, the method is illustrated with numerical simulations including the analysis of forced expiratory volume (FEV) data.

E829: **Tree learning and variable selection**
Presenter: **Klea Panayidou**, University of Frederick, Cyprus

The problem of tree learning in the case of binary variables is considered. We illustrate existing methods and show that in a Bayesian framework if we use a special type of priors we can easily perform tree selection. We introduce tree entropy as a measure of information in a set of variables and issues associated with the submodularity properties of this measure.

Monday 19.12.2011 14:05 - 15:45 Parallel Session O – CFE

CS77 Room S264 CONTRIBUTIONS IN TIME SERIES ECONOMETRICS II **Chair: Alessandra Luati**

C581: **Saddlepoint approximations to the distribution of the estimator of the parameter in a non-stationary AR(1) model**
Presenter: **Juan Carlos Abril**, Universidad Nacional de Tucuman, Argentina
Co-authors: Maria de las Mercedes Abril

We start assuming that we have n observations from AR(1) model with the coefficient which is less than or equal to one, where the errors are independent and identically distributed normal random variables with zero mean and constant variance. We study approximations to the distribution of the first order sample serial correlation coefficient as an estimator of the AR coefficient. Firstly, we develop the general case, that is when the coefficient is less than or equal to one, and then we particularize for the case of non-stationary unit root presence, that is when the coefficient is equal to one. The approximations are obtained using the saddlepoint expansions. Finally, we present an analysis of the results by comparing Monte Carlo simulations and the approximations obtained.

C608: **Nonlinearities in CDS-bond basis**
Presenter: **Kurmas Akdogan**, Central Bank of Turkey, Turkey
Co-authors: Meltem Gulenay Chadwick

Theoretically, the risk premium captured by credit default swap (CDS) and bond yield spreads should be equal. However, data reveals a significant difference between the two spreads. We explore the presence of a mean-reverting behavior in this difference (CDS-bond basis), for selected emerging markets, employing alternative threshold models (TAR, TAR-GARCH and ESTAR). Our results indicate a positive relationship between the speed of adjustment and the trading frequency of the sovereign CDS's and bonds. The TAR-GARCH model suggests that the adjustment of the CDS-bond basis is immediate for economies with more liquid CDS's and bonds, such as Argentina, Brazil and Mexico. The ESTAR model indicates that the adjustment displays a gradual pattern for the basis of the economies with less frequently traded bonds and CDS's.

C649: **Maximum likelihood estimation of a noninvertible ARMA model with autoregressive conditional heteroskedasticity**
Presenter: **Mika Meitz**, Koc University, Turkey
Co-authors: Pentti Saikkonen

Maximum likelihood estimation of a particular noninvertible ARMA model with autoregressive conditionally heteroskedastic (ARCH) errors is considered. The model can be seen as an extension to so-called all-pass models in that it allows for autocorrelation and for more flexible forms of conditional heteroskedasticity. These features may be attractive especially in economic and financial applications. Unlike in previous literature on maximum likelihood estimation of noncausal and/or noninvertible ARMA models and all-pass models, our estimation theory does allow for Gaussian innovations. We give conditions under which a strongly consistent and asymptotically normally distributed solution to the likelihood equations exists, and we also provide a consistent estimator of the limiting covariance matrix.

C742: **Nonlinear causality tests and multivariate conditional heteroskedasticity: A simulation study**
Presenter: **Efthymios Pavlidis**, Lancaster University, United Kingdom
Co-authors: Ivan Paya, David Peel

The performance of linear and nonlinear causality tests in the presence of multivariate conditional heteroskedasticity is assessed. Monte Carlo simulations show that tests based on the least squares covariance matrix estimator can frequently lead to finding spurious causality. The degree of oversizing tends to increase with the sample size and is substantially larger for the nonlinear test. On the other hand, heteroskedasticity-robust tests, which are based on a heteroskedasticity consistent covariance matrix estimator proposed by other authors and the fixed design wild bootstrap, perform adequately in terms of size and power. Consequently, causality in mean tests can be conducted without the need to specify the conditional variance process. As an empirical application, we re-examine the return-volume relationship.

C725: **A simple test for spurious regressions**
Presenter: **Antonio Noriega**, Bank of Mexico, Mexico
Co-authors: Daniel Ventosa-Santaularia

The literature on spurious regressions has found that the t-statistic for testing the null of no relationship between two independent variables diverges asymptotically under a wide variety of nonstationary data generating processes for the dependent and explanatory variables. We introduce a simple method which guarantees convergence of this t-statistic to a pivotal limit distribution, when there are drifts in the integrated processes generating the data, thus allowing asymptotic inference. We show that this method can be used to distinguish a genuine relationship from a spurious one among integrated (I(1) and I(2)) processes. Simulation experiments show that the test has good size and power properties in small samples. We apply the proposed procedure to several pairs of apparently independent integrated variables, and find that our procedure, in contrast to standard ordinary least squares regression, does not find (spurious) significant relationships between the variables.

CS80 Room Court ECONOMETRIC MODELLING AND APPLICATIONS II **Chair: Paolo Foschi**

C033: **Covariance and variance transform for unimodal distribution with applications to options**
Presenter: **Guoqing Liu**, Harbin Institute of Technology, China

For a unimodal random variable set S with fixed moments $ES^i = m_i$ $(i = 1,2,...,n)$ and mode m, a formula on covariance of any two functions of S is provided along the line of Khintchine transform. Upper bounds are derived on the covariance and variance of European call options and Gap options. The techniques are based on symmetrization, domination by quadratic functions of two variables and change of measures.

C041: **Predicting corporate financial distress based on PLS discriminant analysis and neural networks technique**
Presenter: **Mohamed Bannour**, University of Toulon Var, France
Co-authors: Youssef Fahmi, Maher Slouma, Sami Ben Jabeur

The aim is to compare two statistical methods to predict corporate financial distress. Partial least squares-discriminant analysis (PLS-DA) and artificial neural networks techniques (ANN) will be used. PLS discriminant analysis (PLS-DA) is a method connecting a qualitative variable dependent on a unit of quantitative or qualitative explanatory variables. ANN may be viewed as a non-parametric technique in which an output is non-linearly related to a collection of explanatory variables. It is proposed to use a French firm for which some financial ratios are calculated.

C055: **Regression analysis of multivariate fractional data**
Presenter: **Jose Murteira**, Universidade de Coimbra and CEMAPRE, Portugal
Co-authors: Esmeralda Ramalho, Joaquim Ramalho

Alternative regression models and estimation methods for dealing with multivariate fractional response variables are discussed. Both conditional mean models, estimable by NLS and quasi-maximum likelihood (QML), and fully parametric models (Dirichlet and Dirichlet-multinomial), estimable by ML, are considered. In contrast to previous papers but similarly to the univariate case, a new parameterization is proposed here for the parametric models, which allows the same specification of the conditional mean of interest to be used in all models, irrespective of the specific functional form adopted for it. It is also discussed at some length the specification analysis of fractional regression models. All the proposed types of tests of the conditional mean can be performed through artificial OLS regressions, with test statistics evaluated at QML or NLS estimates. An extensive Monte Carlo study is included to evaluate the finite sample properties of most of the estimators and tests considered.

C289: **Quantifying the impact of monetary policy operations on commercial bank rates**
Presenter: **Sven Wagner**, Saarland University, Germany
Co-authors: Stefan Kloessner

Against the background of the recent financial markets turbulences, our research question is whether there is a pass-through of money market rates to commercial bank rates. Based on a standard banking firm model, we use a Vector Autoregressive approach to model the links between bank rates and money market rates for multiple time series. But in contrast to standard pass-through studies, we use a spillover measure based on variance decomposition of the h-step-ahead forecast error, which we decompose into own variance shares and cross variance shares in order to derive a spillover index. Unlike other researchers, we use spillover measures which are robust against realignments of variables within the VAR model. In doing so, we are able to trace movements in commercial bank rates to movements in money market rates. Our empirical model uses data from the MFI interest rate statistics gathered by the European Central Bank. These are harmonized statistics of interest rates that credit institutions apply to euro-denominated deposits and loans held by the non-financial sector within the euro area. By using this large dataset, we are able to distinguish between pass-through effects concerning different maturities, different countries, and different time intervals.

C939: **A nonparametric ACD model**
Presenter: **Fausto Galli**, University of Salerno, Italy
Co-authors: Antonio Cosma

A nonparametric analysis of financial durations is carried out. We use an existing algorithm to describe nonparametrically the dynamics of the process in terms of its lagged realizations and of a latent variable, its conditional mean. The devices needed to effectively apply the algorithm to our dataset are presented. On simulated data, the nonparametric procedure yields better estimates than the ones delivered by an incorrectly specified parametric method. On a real dataset, the nonparametric analysis can convey information on the nature of the data generating process that may not be captured by the parametric specification. In this way, the nonparametric method proposed can be a valuable preliminary analysis able to suggest the choice of a "good" parametric specification, or a complement of a parametric estimation.

CS94 Room Gordon FINANCIAL APPLICATIONS Chair: Patrick Burns

C680: **Private equity benchmarks and portfolio optimization**
Presenter: **Lars Helge Hass**, Lancaster University Management School, United Kingdom
Co-authors: Denis Schweizer, Douglas Cumming

Portfolio optimization with private equity is based on one of three different indices: listed private equity indices, transaction-based private equity indices, and appraisal value based private equity indices. We show that none of these indices are appropriate for portfolio optimization. We introduce a new benchmark index for buyouts and venture capital. Our benchmark is updated monthly, adjusted for autocorrelation (desmoothing) and available contemporaneously. We show our benchmark enables superior quantitative portfolio optimization.

C717: **Stochastic surface models for commodity futures: A 2D Kalman filter approach**
Presenter: **Javier Fernandez-Macho**, University of the Basque Country, Spain

Commodity markets have increasingly attracted the interest of producers, investors and academics. Efficient markets have been developed for hard commodities like energy products and metals where long term maturities exist. It is assumed two-dimensional stochastic process where futures prices evolve both along the time axis in the real world and along a second dimension of maturities obtained in a risk-free world, thus forming a $T \times n$ grid generated by a surface model similar to those used in areas like geophysics, image restoration, neural networks, remote sensing, etc. In contrast, however, our model is neither isotropic nor invariant in both directions. The paper obtains the surface state-space form of a doubly stochastic Inhomogeneous Geometric Brownian Motion theoretical model and derives the corresponding two-dimensional Kalman filter that is recursive in both directions simultaneously. The proposed methodology uses the entire time-maturity dynamics of the full stochastic process, including links from all available maturities per period, as an alternative to standard vector Kalman filtering along the one-dimensional time line, which typically results in large states unfeasible to handle computationally. The technique is illustrated using a dataset of all daily observations of NYMEX coal futures contracts during a recent year.

C780: **Index tracking, cointegration and picking up stocks with genetic algorithms**
Presenter: **Fernando Fernandez-Rodriguez**, University of Las Palmas de Gran Canaria, Spain
Co-authors: Eduardo Acosta-Gonzalez, Reinaldo Armas-Herrera

An alternative procedure for constructing a tracking portfolio which follows the behaviour of the DJIA index employing only five stocks is provided. Our method is based on using long-term relationships between series provided by concept of cointegration. As the quality of the benchmark tracking highly depends on the stock selection procedure, the stocks are picked up using a heuristic optimization procedure, called genetic algorithm. The procedure is also extended to tracking the DJIA index with a plus of profitability of 5%, 10% and 15%. The portfolios obtained by our methodology track the index, with and without a plus of profitability. The evolution of the portfolios and the reference indexes present high correlation; on the contrary, the correlation between the tracking error and the DJIA index is practically zero. In terms of profitability, the selected portfolios have a return very close to the corresponding reference indexes.

C185: **A study of equity and housing bubbles spillover to REITs**
Presenter: **Ogonna Nneji**, University of Reading, United Kingdom
Co-authors: Chris Brooks, Charles Ward

A regime switching approach is used to determine whether prices in the stock, direct real estate and indirect real estate markets are driven by the presence of speculative bubbles. The results show significant evidence of the existence of periodically partially collapsing speculative bubbles in all three markets. We then develop and implement a multivariate bubble model to evaluate whether the stock and housing bubbles spill over into REITs. We find the underlying property market bubble to be a stronger influence on the securitized real estate market bubble than that of the stock

market. Furthermore, our findings suggest a multi-directional transmission of speculative bubbles between the direct real estate and stock markets, although this link is not present for the returns themselves.

C607: **Portfolio optimization inside out**
Presenter: **Patrick Burns**, Burns Statistics, United Kingdom

The usual approach to portfolio optimization is to focus on the utility and to consider the constraints as a minor addition. We start with the constraints, satisfy them, and only then inspect the utility. We get entirely different pictures coming from this direction. This gives us new insight into the optimization process. In particular, it suggests a framework for thinking about how to improve the selection of constraints. The computations involved include the generation of random portfolios – that is, sampling from the population of portfolios that obey all of the constraints. When real-world constraints are used, then efficiently sampling is a difficult problem. A custom-built method that includes a genetic algorithm and various greedy algorithms is used.

CS10 Room S261 CONTRIBUTIONS IN APPLIED FINANCIAL ECONOMETRICS Chair: Christopher Baum

C734: **Modeling dependence dynamics through copulas with regime switching**
Presenter: **Flavio Ziegelmann**, Universidade Federal do Rio Grande do Sul, Brazil
Co-authors: Osvaldo Silva Filho, Michael Dueker

Measuring dynamic dependence between international financial markets has recently attracted great interest in financial econometrics because the observed correlations rose dramatically during the 2008-09 global financial crisis. We propose a novel approach for measuring dependence dynamics. We include a hidden Markov chain (MC) in the equation describing dependence dynamics, allowing the unobserved time-varying dependence parameter to vary according to both a restricted ARMA process and an unobserved two-state MC. Estimation is carried out via the inference for the margins in conjunction with filtering/smoothing algorithms. We use block bootstrapping to estimate the covariance matrix of our estimators. Monte Carlo simulations compare the performance of regime switching and no switching models, supporting the regime-switching specification. Finally, the proposed approach is applied to empirical data, through the study of the S&P500 (USA), FTSE100 (UK) and BOVESPA (Brazil) stock market indexes.

C740: **Hedge fund replication: A Dynamic performance-adaptive local linear regression approach**
Presenter: **Donatien Tafin Djoko**, Universite de Neuchatel, Switzerland
Co-authors: Catalin Starica

The implementation of sequential investment theory for the purpose of hedge fund index return replication is considered. We construct a performance-adaptive, local linear regression smoother for predicting hedge fund returns. We focus on estimating different trajectories of the time-varying hedge fund returns exposures through various style specific market risk factors. The target replicated value is then derived by exponentially aggregating the set of elementary estimates based on their past performances in terms of cumulative square replication error. As an extension of the purely linear factor-based replication design, the proposed local linear implementation keeps the former methods intuitiveness and computational simplicity, while minimizing the mean square replication error. The approach is dynamic and able to uncover hidden path dependent patterns between hedge fund indexes and common market risk factors. Empirical results suggest that the model performs better - in the sense of Mean Square Replication Error, Cumulative Square Replication Error, Percentage of the direction change, and Correlation between the estimated series and the realized series - than 'static' and rolling-windows modeling. Additionally, the series constructed using the hybrid methodology share several statistical properties which are the ones exhibited by hedge fund indexes. The experiment highlights the power of performance-adaptive, factor-based rules in the context of time series hedge fund returns forecasting.

C264: **Copula-based Russian banking system capital adequacy modelling within Basel II IRB framework**
Presenter: **Irina Andrievskaya**, Higher School of Economics, Russia
Co-authors: Henry Penikas

The aim is to model the Russian banking system capital adequacy under the Basel II IRB approach and to test for Basel II procyclicality effect. The research is based on publicly available quarterly financial statements of all the Russian banks for the period 2004Q1-2010Q1. The methodology consists of three steps. Firstly, the copula structural shift for the joint banking system risk distribution is examined. The copula parameter value dynamics is investigated. Secondly, four copulas (Gaussian, Frank, Clayton, Gumbel) are compared to fit the joint risk distribution and consequently arrive at the risk-weighted assets (RWA) value as the value-at-risk of the aggregate (systemic) risk distribution. Finally, the capital adequacy ratio (CAR) is calculated for copula-based RWA estimate and for the one based on empirical joint risk distribution. The findings suggest that the copula shift date equals to 2005Q3 being associated with the drop in copula parameter value. The observed effect strongly corresponds to the moment Russia was granted investment grade credit rating. Basel II CAR was estimated at ca. 4-8%, while the Basel I actual value equalled to ca. 15-20%. However, the Basel II procyclicality effect proved to be non-significant.

C373: **The informative role of stock markets in firm investment decisions**
Presenter: **Houdou Basse Mama**, University of Hamburg, Germany

A large panel of European firms over the period from 1991 through 2010 to explore the analytical and empirical linkages between firm-level investment and stock price informativeness is used. The relevant theoretical and empirical literature alarmingly provides two contrasting explanations about the informative role of the stock market. A body of research posits a significant informative function of stock prices in guiding corporate capital expenditures. Another strand predicts, if any, only a marginal impact on corporate investment behavior of stock prices because the latter may convey an important element of irrationality. First, we examine whether investment spending at firm-level significantly responds to market valuations after controlling for potential endogeneity (absolute response test). After constructing a robust measure of stock price informativeness, we subsequently investigate how the informative function of the stock market measures up. Specifically, we hypothesize that greater price informativeness should trigger larger response sensitivity of firms' investment to their stock market valuations (relative response test). The answer to this issue is potentially important for practitioners, policy makers and academics, and particularly so in turbulent times where there is increasing political proclivity for damning apparent decoupling of capital markets from real economic activity.

C646: **Hidden decisions behind hidden orders: an empirical analysis of incoming iceberg orders in electronic trading**
Presenter: **Vahidin Jeleskovic**, University of Kassel, Germany
Co-authors: Nikolaus Hautsch, Klaus Belter

If traders want to hide their interest in an electronic market using iceberg orders, they have to make four decisions simultaneously: i) should an iceberg order be used, ii) if yes, how big the displayed, iii) and the hidden part, and iv) how aggressive the order should be. To account for the simultaneousness, we analyze the process of these trader's decisions via an unique simultaneous equations system that includes: a logit model (for the first traders' decision) and censored regressions (for the other ones) with explanatory variables. Our empirical analysis is based on data from the Copenhagen stock exchange that support the decomposition of total limit order size into its displayed and hidden parts. We find evidence of a clear trade-off between iceberg order aggressiveness and its quantity of the hidden part, on the one side, and the displayed size, on the other side. That

is, iceberg orders are more likely to be placed near to the opposite market side while simultaneously the ratio of the hidden size to the displayed one tends to become larger. Moreover, the bigger the hidden size the higher the probability for an iceberg order.

CS06 Room Bedford FILTERING Chair: Tommaso Proietti

C294: **Particle filter estimation of duration-type models**
Presenter: **Miguel Belmonte**, Strathclyde, United Kingdom
Co-authors: Omiros Papaspiliopoulos, Michael Pitt

Estimation of discrete-time parameter driven microstructure models for the analysis and prediction of financial duration times is considered. The models are written in a non-linear and non-Gaussian state-space form, comprising an unobserved Markov chain which intends to capture the predictable component of the data (e.g trading intensity) and conditionally independent observations. In this article we adopt filtering methods, rather than Markov chain Monte Carlo, for the estimation of parameters. Estimation of the unobserved state can be done very efficiently using particle filters, which as a by-product yield unbiased estimates of the likelihood function and density forecast model diagnostics. We investigate the performance and the relative merits of two approaches for parameter estimation, which are based on particle filters: the smooth particle filter (SPF) for maximum likelihood estimation, the particle marginal Metropolis-Hastings (PMMH) algorithm and the particle Gibbs sampler (PG). We analyse from an applied perspective the scaling properties of SPF, PMMH and PG with the number of observations T, since financial durations are characterised by thousands of observations.

C667: **An empirical analysis of some peripheral EU stock market indices: A wavelet correlation approach**
Presenter: **Josue Polanco-Martinez**, University of the Basque Country, Spain
Co-authors: Javier Fernandez-Macho

An empirical analysis of the correlation at various time scales among the stock markets of some of the European economies most affected by recent deficit/debt troubles is presented. We use daily stock market indices from January 3, 2000 until May 29, 2009 for five peripheral European countries (Portugal, Italy, Ireland, Greece and Spain). We decompose the daily stock market returns of these countries into different time scale components using the Maximal Overlap Wavelet Transform (MODWT), and then analyze the time-scale relationships among the peripheral EU countries as well as with the benchmark EU stock market index (the German DAX30). New graphical tools that may serve to facilitate the interpretation of wavelet analysis are provided. The results suggest that, despite an evident overall correlation among the stock markets, the degree of statistical correlation varies both in time and scale. The stock markets with the highest wavelet correlation between them and with DAX30 belong to Italy and Spain.

C763: **Extracting latent price and volatility processes through particle filtering**
Presenter: **Petr Zahradnik**, Charles University in Prague, Czech Republic

The aim is to employ advanced Bayesian filtering techniques to extract quadratic variation of asset returns under the assumption that observed asset prices are contaminated by the so-called microstructure noise. Current research turns around various parametric and nonparametric methods to extract volatility which, even though under different and usually quite general assumptions, always yield a trade-off between the sampling frequency and a bias caused by the noise. Surprisingly enough, not much focus was directed to filtering the noise and studying the returns thereafter. General particle filters are hence employed to extract the latent (true, efficient) asset returns and volatility. This approach adds another possibility to the already vast variety of methods to estimate volatility and offers substantially more generality as the filtered asset returns series may serve even for other portfolio management purposes.

C900: **Improving the reliability of real-time Hodrick-Prescott filtering using survey forecasts**
Presenter: **Jaqueson Kingeski Galimberti**, University of Manchester, United Kingdom
Co-authors: Marcelo L. Moura

Incorporating survey forecasts to a forecast-augmented Hodrick-Prescott filter, we evidence a considerable improvement to the reliability of US output-gap estimation in real-time. Odds of extracting wrong signals of output-gap estimates are found to reduce by almost a half, and the magnitude of revisions to these estimates accounts to only three fifths of the output-gap average size, usually an one-by-one ratio. We further analyze how this end-of-sample uncertainty evolves as time goes on and observations accumulate, showing that a 90% rate of correct assessments of the output-gap sign can be attained with five quarters of delay using survey forecasts. We also apply the same methodology for trend-gap decomposition of monthly time series of industrial production in a sample of 15 emerging economies (plus US). In this latter exercise we evidence that the gains from the incorporation of survey forecasts are strongly dependent on the quality of these forecasts.

C825: **A Structural model of dynamic market timing: Theory and estimation**
Presenter: **Marcel Rindisbacher**, Boston University, United States of America
Co-authors: Jerome Detemple

Dynamic timing strategies of a fund manager with private information are derived and analyzed. Endogenous timing strategies generated by various information structures and skills, and associated fund styles are identified. Endogenous fund returns are characterized in the public information of an uninformed observer. Econometric methods for style analysis are developed. New tests of timing skill are proposed and their detection ability is analyzed. An application to a universe of hedge fund indices shows significant timing ability in specific categories of hedge fund styles.

CS31 Room Woburn CONTRIBUTIONS IN BAYESIAN ECONOMETRICS AND APPLICATIONS Chair: Richard Gerlach

C642: **Forecasting carbon phase II moments using stochastic volatility models**
Presenter: **Per Bjarte Solibakke**, Molde University College, Norway

A general scientific model methodology implementing MCMC simulation methodologies is applied to build a multifactor stochastic volatility model for the NASDAQ OMX carbon front December forward contracts (phase II). Stochastic volatility is the main way time-varying volatility is modeled in financial markets. The main objective is therefore to structure a scientific model specifying carbon volatility as having its own stochastic process. Appropriate model descriptions broaden the applications into derivative pricing purposes, risk assessment and asset allocation. Risk and portfolio measures are reported, conditional one-step-ahead moments, particle filtering for one-step-ahead conditional volatility, conditional variance functions for evaluation of shocks, analysis of multi-step-ahead dynamics, and conditional persistence. The analysis adds market insight and enables forecasts to be made, thus building up methodologies for developing valid scientific models for carbon market applications.

C748: **Bayesian bandwidth estimation for local linear fitting in a nonparametric regression model**
Presenter: **Han Lin Shang**, Monash University, Australia
Co-authors: Xibin Zhang

The aim is to investigate a Bayesian sampling approach to bandwidth estimation for the local linear fitting in a nonparametric regression model.

The error density is approximated by a mixture of Gaussian densities with mean being individual residual and variance being a common bandwidth parameter. A Markov chain Monte Carlo algorithm is presented to sample bandwidths for local linear estimator of the regression function and error density simultaneously. Through simulation, we compare the performance of the proposed Bayesian approach with some benchmark methods, such as normal reference rule and cross-validation. We apply our approach to a nonparametric regression model used for state-price density estimation based on the S&P500 options data.

C820: **Bayesian treatment effects models for panel outcomes with stochastic variable selection**
Presenter: **Helga Wagner**, Johannes Kepler University Linz, Austria
Co-authors: Liana Jacobi, Sylvia Fruehwirth-Schnatter

A flexible Bayesian inferential framework for a multilevel-treatment effects model with continuous outcomes observed over subsequent time periods is proposed. We specify a joint model for a categorical treatment with level-specific and heterogeneous treatment effects on the longitudinal outcome data. To control for unobserved subject specific factors a random intercept is included in the panel model. The treatment variable and the panel outcomes are assumed to be correlated, due to unobserved or unmeasured confounders. For inference we use MCMC methods and incorporate Bayesian variable selection to determine which covariates should be included in the model and to test for the existence of common and level-specific effects of the treatment as well as covariates. The model is applied to determine effects of incentives for maternal leave. We exploit a recent change in the parental leave policy in Austria that extended maternal benefits from 18 months, since birth of the child. to 30 months but left the period of job protection unchanged at 24 months. For our analysis we use data from the Austrian Social Security Register which contains complete individual employment histories since 1972 and also the reports number of births and maternity and parental leave spells for all Austrian employees.

C841: **A Bayesian semiparametric multiplicative error model for realized volatility**
Presenter: **Reza Solgi**, Swiss Finance Institute at the University of Lugano, Switzerland
Co-authors: Antonietta Mira

A semi-parametric multiplicative error model (MEM) for daily realized volatility is proposed. In traditional MEM models, the innovations are typically assumed to be Gamma distributed (with one free parameter that ensures unit mean of the innovations and identifiability of the model), however empirical investigations unveil the inappropriateness of this choice. In the proposed model, the conditional mean of the time series is modeled parametrically, while we model its conditional distribution nonparametrically. In particular, in order to model the innovations nonparametrically, we employ a Dirichlet process mixture model with the Gamma distribution as its kernel. The inference is performed using Markov chain Monte Carlo simulation. This model is applied to the time series of daily realized volatility of some indices, and it has been compared to similar parametric models. A comparison in terms of predictive performance of the proposed model relative to competing parametric models available in the literature has been conducted, showing better predictive performance, flexibility and robustness to mis-specification of our Bayesian semi parametric approach.

C665: **Forecasting with the double adaptive elastic-net Lasso - A Bayesian approach**
Presenter: **Deborah Gefang**, Lancaster University, United Kingdom

Recently, various Lassos have become popular for shrinkage estimation and variable selection when the dimensions of the predictors are high. Among them, elastic-net Lasso (e-net Lasso) is more appealing to the econometricians as it can easily handle the multicollinearity problem in a data based manner, while the other Lassos generally cannot. However, the traditional e-net Lasso does not have the oracle property that ensures the optimal large sample performance. To solve this problem, an adaptive e-net Lasso was proposed. But that approach might cause new problems as it depends on a two-stage estimation where the results of first stage e-net Lasso are used as weights to adjust the l_1 norm of the same e-net Lasso at the second stage. A double adaptive elastic-net Lasso (DALasso) is proposed using the results of principal components regression as weights to simultaneously adjust the l_1 and l_2 norms of e-net Lasso. We show that DALasso can achieve the oracle property at almost no additional computing cost. In empirical work, we examine the forecasting performance of DALasso using US macroeconomic data. We find that DALasso tends to forecast better than the adaptive e-net Lasso and VARs.

CS36 Room Torrington FORECASTING VALUE-AT-RISK Chair: Rodney Wolff

C056: **The limits of granularity adjustments**
Presenter: **Jean-David Fermanian**, Crest-Ensae, France

A rigorous proof of a granularity adjustment formula is provided to calculate loss distributions and risk measures (VaR) in the case of heterogeneous portfolio and random recoveries. As a significant improvement with respect to the literature, the technical conditions of validity and an upper bound for the remainder term are detailed. Moreover, the case of discrete loss distributions is explicitly considered. For some simple portfolio models, it is empirically proved that the granularity adjustments do not generate relevant results when the regularity conditions are not fulfilled.

C645: **Optimally harnessing inter-day and intra-day information for daily value-at-risk prediction**
Presenter: **Ana-Maria Fuertes**, Centro Universitario de la Defensa de Zaragoza, Spain
Co-authors: Jose Olmo

A robust encompassing test for Value-at-Risk (VaR) forecasts that builds on quantile regression theory is proposed. Our test naturally accounts for model risk and estimation uncertainty and is valid for any quantile forecasts from (non)nested models. We deploy it for two daily VaR models from the location-scale family to assess the merit of pooling inter-day and intra-day information. The techniques are illustrated for equity, FOREX, fixed income and commodity trading desks. In various cases, the out-of-sample 95% VaR predictions from an ARFIMA realized volatility model that exploits five-minute prices outperform the GARCH-based VaRs that exploit daily-recorded closing, high and low prices. By contrast, there is little evidence of encompassing for the 99% tail risk predictions which warrants optimal quantile combination.

C876: **Market value in the estimation of equity Value-at-Risk**
Presenter: **Alexandra Dias**, University of Leicester, United Kingdom

It is investigated whether market capitalization should be taken into account when computing the VaR of equity portfolios. We compute VaR for equity portfolios with different market capitalizations using different methods and considering crisis periods and non-crisis periods separately. We find that market capitalization is important for VaR computation. The VaR methods provide better estimates for higher market capitalization than low market capitalization equities. The more striking result is that VaR methods provide better VaR estimates during crisis periods than during non-crisis periods. The main reason for this result is that the VaR methods do not successfully model clusters of losses during non-crisis periods. Looking at a different dimension we also find that different VaR methods perform better for small and for large market value portfolios.

C864: **Non-parametric estimation of copulae**
Presenter: **Rodney Wolff**, The University of Queensland, Australia
Co-authors: Kohei Marumo

We apply expansion methods to construct copulae for asset return distributions and implied volatility distributions, to show the efficacy of the method for complex portfolios comprising non-linear assets. We show the superiority of Laguerre expansions over the more commonly used Hermite expansion, and demonstrate how Value-at-Risk can be computed in this way. We also apply this method to the time aggregation problem which is concerned with Value-at-Risk over a longer time horizon than one day.

C441: **The mechanics of VAR forecast pooling: A DSGE model based Monte Carlo study**
Presenter: **Steffen Henzel**, Ifo Institute, Germany
Co-authors: Johannes Mayr

The mechanics of VAR forecast pooling are analyzed and the forecast performance under varying conditions is quantified. To fill the gap between empirical and purely theoretical research we run a Monte Carlo study and simulate the data from different New Keynesian DSGE models. We find that equally pooling VAR forecasts outperforms single predictions in general, and that the gains are substantial for sample sizes relevant in practice. In contrast, the estimation of theoretically optimal weights or model selection is advisable only for very large data sets hardly available in practice. Most remarkable is that equally pooling forecasts from small-scale VARs can even dominate forecasts from large VARs including all relevant variables. Thus, we advocate the use of equally pooled predictions from parsimonious VARs as an easy-to-implement and competitive forecast approach.

CS49 Room Jessel CONTRIBUTIONS IN VOLATILITY ESTIMATION AND FORECASTING Chair: Simona Sanfelici

C877: **Consistent estimation of integrated volatility using intraday absolute returns for SV jump diffusion processes**
Presenter: **Shuichi Nagata**, Kwansei Gakuin University, Japan

Volatility estimation using high-frequency data has attracted considerable interest in the literature. We consider consistent estimation of integrated volatility (IV) in the presence of jumps. In order to construct a consistent estimator of IV, we focus on realized absolute variation (RAV), which are calculated from intraday absolute returns. RAV is known for predicting volatility well because of its robustness for jumps. However, it should be noted that RAV (or its square) is an inconsistent estimator of IV, even though it is robust to jump effects. We introduce our estimator as a natural extension of RAV, and show its consistency and asymptotic normality. We also show our estimator is asymptotically more efficient than another jump-robust estimator, bi-power variation. We analyze a simulation to assess the finite-sample behavior of our proposed estimator and compare its performance against alternatives. The results of a simulation complement the asymptotic result.

C911: **Forecasting the realized covariance matrix: A comparative approach**
Presenter: **Moritz Heiden**, University of Augsburg, Germany

Several of the latest methods to combine models for the realized covariance matrix with mixed data sampling strategies are compared. We give an overview on favourable decompositions of the covariance matrix, as well as some of the most recent models and their extensions. The major objective of our work lies in the assessment of whether or not simple mixed frequency regression models are able to adequately capture the dynamics of the covariance matrix and provide superior predictive results over different horizons compared to other models including data sampled at mixed frequencies. For this purpose, several models are specified and studied on a set of high-frequency data. To evaluate the models performance, out-of-sample forecasts at horizons up to one month are performed and compared using a model confidence set approach based upon multiple loss functions. We find that simple specifications within the mixed frequency regression framework yield the best results in forecasting on all horizons for three out of four loss functions. A deeper analysis of the forecasting method and loss functions reveals that the comparison is sensitive to the way over and underpredictions are evaluated.

C458: **A Copula-DCC model with daily range**
Presenter: **Filippo Spazzini**, Edison Trading SpA, Italy
Co-authors: Eduardo Rossi, Paolo Santucci de Magistris

A new parameterization of the conditional covariance matrix of financial returns based on ex-post realized measures is provided. The set-up is the DCC model where the log-conditional variances are modeled assuming as a proxy for the log-volatility innovations the standardized log-daily range, as in the Range-based EGARCH. The joint distribution is obtained using a t-copula function. In this way, the conditional distribution of the returns is non-Gaussian, with possible extreme dependence. The model is estimated adopting the Inference For Margins method. We firstly maximize the part of the log-likelihood that depends on the log-conditional volatility parameters and secondly the part that depends on the copula and correlation matrix parameters. A set of simulations shows the properties of QML estimators. Moreover, we investigate the forecasting performances of the Copula-DCC-Range-based model using both statistical and financial forecasting loss functions. An empirical analysis using intraday data for Brent and WTI crude oil futures will be shown. We will focus our attention on VaR forecasting performances since, after the decoupling of the two oil benchmarks, the cross-Atlantic spread is rapidly growing in importance both from a trading and risk-management perspective.

C610: **Forecasting volatility: Continuous time vs discrete time**
Presenter: **Helena Veiga**, Universidad Carlos III de Madrid, Spain
Co-authors: Carles Breto

The forecast performance of continuous and discrete-time volatility models is compared. In discrete time, we consider more than ten GARCH-type models and an asymmetric autoregressive stochastic volatility model. In continuous-time, a stochastic volatility model with mean reversion, volatility feedback and leverage is considered. We estimate each model by maximum likelihood and evaluate their ability to forecast the two scales realized volatility, a nonparametric estimate of volatility based on high-frequency data that minimizes the biases present in realized volatility caused by microstructure errors. We find that volatility forecasts based on continuous-time models may outperform those of GARCH-type discrete-time models so that, besides other merits of continuous-time models, they may be used as a tool for generating reasonable volatility forecasts. However, within the stochastic volatility family, we do not find such evidence. We show that volatility feedback may have serious drawbacks in terms of forecasting and that an asymmetric disturbance distribution (possibly with heavy tails) might improve forecasting.

C945: **On the effect of crisis on stock market predictability: The case of the Spanish stock market**
Presenter: **Rajesh Mohnot**, Middlesex University Dubai, United Arab Emirates

The predictability of the Spanish stock market returns is examined. Earlier studies suggest that stock market returns in developed countries can be predicted with a noise term but this study has specifically covered two time horizons; one pre-crisis period and the other one current crisis period to evaluate the stock market returns predictability. Since mean returns cannot prove all the time to be efficient predictor, variance of such returns do, hence various autoregressive models have been used to test the existence of persisting volatility in the Spanish stock market. The empirical results show that higher order autoregressive models such as ARCH(5) and GARCH(2,2) can be used to predict future risk in Spanish stock market

both in pre-crisis and current crisis period. We also found a positive correlation between Spanish Stock Market returns and the conditional standard deviations as produced by ARCH(5) and GARCH(2,2), implying that the models have some success in predicting future risk on the Spanish stock market. The predictability of stock market returns during crisis period is not found to be affected though the degree of predictability may be.

CS47 Room Bloomsbury FINANCIAL TIME SERIES Chair: Michele La Rocca

C163: **Estimated quasi-maximum likelihood estimator for GARCH models based on non-parametric MLE**
Presenter: **Taewook Lee**, Hankuk University of Foreign Studies, Korea RoK
Co-authors: Byungtae Seo

An estimated quasi-maximum likelihood estimator (NP-QELE) is developed for GARCH models based on the continuous-type normal mixtures. For the estimation of the mixing distribution, we employ the non-parametric MLE (NPMLE) and a certain fast computation algorithm for NPMLE. By adopting this method, we can suitably avoid the order selection problem in the application of the finite normal mixtures. The validity of our method is demonstrated by verifying the consistency of NP-QELE. Through a simulation study and real data analysis, we compare the performance of NP-QELE with the existing ones including QMLE based on the standard normal density and finite normal mixture QMLE in terms of the relative efficiency. It is noteworthy that NP-QELE provides an adaptive estimation method for various innovation distributions in the simulation study.

C461: **The number of regimes accross asset returns: Identification and economic value**
Presenter: **Mathieu Gatumel**, CERAG - University of Grenoble, France
Co-authors: Florian Ielpo

A shared belief in the financial industry is that markets are driven by two types of regimes. Bull markets would be characterized by high returns and low volatility whereas bear markets would display low returns coupled with high volatility. Modelling the dynamics of different asset classes (stocks, bonds, commodities and currencies) with a Markov-Switching model and using a density-based test, we reject the hypothesis that two regimes are enough to capture asset returns' evolutions for many of the investigated assets. Once the accuracy of our test methodology has been assessed through Monte Carlo experiments, our empirical results point out that between two and five regimes are required to capture the features of each asset's distribution. Moreover, we show that only a part of the underlying number of regimes is explained by distributional characteristics of the returns such as the kurtosis. A thorough out-of-sample analysis provides additional evidence that there are more than just bulls and bears in financial markets. Finally, we highlight that taking into account the real number of regimes allows us both to improve the portfolio's performance and the density forecast.

C548: **Volatility graphics for financial time series and volatility modeling**
Presenter: **Anthony Lawrance**, University of Warwick, United Kingdom

Exploratory graphics for the progression of volatility and for displaying the dependence of volatility on past behaviour are considered; they are particularly suitable for identifying volatility structure prior to parametric modelling. The importance of prior de-autocorrelation in assessing volatility is emphasised. Illustrations are based on current financial time series and the methods are justified using simulations generated by volatile models, such ARCH, GARCH and SV, and with comparisons to non-volatile linear models. Theoretical discussion of volatile models concerns the discrete time stochastic volatility (SV) model as a linear autoregressive moving average model in transformed variables with linear structure. Parametric restraint regions necessary for the strict (distributional) stationarity of GARCH(1,1) models and for their finite volatility, variance and expected absolute values are also discussed.

C747: **Robust estimation of semiparametric multiplicative volatility models**
Presenter: **Bonsoo Koo**, Monash University, Australia
Co-authors: Oliver Linton

We investigate a model in which we connect slowly time varying macroeconomic volatility with short-run financial volatility whose representation is given as a semi-strong GARCH (1,1) with heavy tailed errors. We focus on robust estimation of both macroeconomic and financial volatilities. We propose a nonparametric least absolute deviation estimation (LADE) approach for the macroeconomic volatility and suggest a semiparametric LADE approach to the financial volatility. Our estimation is based on a two-step procedure where macroeconomic volatility is estimated in the first step and financial volatility is estimated in the second step based on the estimate of macroeconomic volatility. We establish the relevant asymptotic theory of the proposed estimators. Numerical results lend support to our theoretical results.

C887: **Adaptive filtering for algorithmic pairs trading**
Presenter: **Seungjin Han**, University of Sheffield, United Kingdom
Co-authors: Kostas Triantafyllopoulos

Statistical arbitrage, an important topic of financial econometrics, plays a critical role in algorithmic trading and decision making in modern investment boutiques and hedge funds. In particular, pairs trading and its generalizations, have received considerable attention and popularity, since their initial application by Tartaglia's quantitative group at Morgan Stanley in the 1980's. Pairs trading, a market neutral trading strategy, is based on the assumption that a spread of two assets is mean-reverted, or that any fluctuations (upwards or downwards) of the spread are temporary. Then, it takes advantage of such temporary fluctuations, suggesting going short or long at different time points and thus realizing profits when closing a trading position. We propose detection of mean reversion in real-time, which enables the trader to identify tradable time intervals. For this, we employ a time-varying autoregressive model, which estimation is achieved by a novel combination of Kalman filter type recursions and adaptive forgetting. We develop novel adaptive forgetting, based on recursive least squares (RLS), and new Bayesian approach. Using Monte Carlo experiments, we compare the proposed algorithms with standard RLS algorithms with adaptive memory. Two real data sets, consisting of ETF assets, are used to illustrate the proposed methodology.

CS67 Room Senate FINANCIAL MODELING Chair: Panayiotis Andreou

C619: **A simple property for estimators of diffusion models**
Presenter: **L. Ramprasath**, Institute for Financial Management and Research, India
Co-authors: T.M. Durairajan

The definition and calculation of scale equivariant estimators of diffusion models, based on discretely sampled data is considered. These estimators, for example, ensure that a single time series of asset prices, expressed in different units of currency, do not result in structurally different models. We present examples of some popular estimators, proposed for various financial and biological applications, which are not equivariant according to this criterion. In some cases, even standard estimators like the MLE are not scale equivariant and this in turn leads to easy manipulation of key parameter estimates. We finally present a class of estimators based on optimal estimating functions, and prove that members of this class are always equivariant. The results are illustrated using the Gompertzian diffusion model for tumor growth, which is same as the Black-Karasinski interest

rate model in finance. The examples mentioned here force us to rethink about the estimators we typically use, for estimating the parameters of a diffusion model.

C775: **Modelling the European Central Bank official rate: a stochastic approach**
Presenter: **Luisa Cutillo**, University Parthenope Napoli, Italy
Co-authors: Alba Orlando, Maria Francesca Carfora

Following its main task of price stability in the euro area, the European Central Bank (ECB) increases or decreases interest rates in order to cool inflation or respectively to support economic growth. Monetary policy shows delayed effects on inflation and thus the ECB modifies interest rates on the basis of forecasts about the state of the economy over the coming quarters. The aim is to propose a stochastic model for the ECB official rate taking into account the expectations on the future state of the economy. We propose a MSVAR model for the dynamics of economic cycle, a Poisson process to describe the intervention times of the ECB and a Markov process to account for the direction and width of such interventions. The modelled economic variables are chosen via variable selection with respect to the historical BCE rates. The working hypothesis of our proposal is to link both the Poisson and the Markov processes to the evolution of the economic cycle as described by the MSVAR model. We show an application on Italian and suitably aggregated European data.

C916: **Fundamentalists, chartists and asset pricing anomalies**
Presenter: **Sandrine Jacob Leal**, ICN Business School, France

It is investigated whether serial correlations and excess volatility of stock returns may be explained through the interactions between fundamentalists and chartists. Fundamentalists are defined as agents who forecast future prices cum dividend through an adaptive learning rule. In contrast, chartists are defined as agents who forecast future prices based on the observation of past price movements. Numerical simulations reveal that the presence of both fundamentalists and chartists in the market generates trends in prices over short horizons and oscillations in prices over long horizons, as well as excess volatility of returns. Moreover, we find that the memory of the learning mechanism plays a key role in explaining predictability of returns as well as excess volatility. In particular, in the presence of chartists, (i.) short-run dependencies in financial time series can be explained by long memory in the learning mechanism of fundamentalists; (ii.) excess volatility of returns can be explained by short memory in the learning mechanism of fundamentalists.

C036: **The role of time-varying price elasticities in accounting for volatility changes in the crude oil market**
Presenter: **Christiane Baumeister**, Bank of Canada, Canada
Co-authors: Gert Peersman

There has been a systematic increase in the volatility of the real price of crude oil since 1986, followed by a decline in the volatility of oil production since the early 1990s. We explore reasons for this evolution. We show that a likely explanation of this empirical fact is that both the short-run price elasticities of oil demand and of oil supply have declined considerably since the second half of the 1980s. This implies that small disturbances on either side of the oil market can generate large price responses without large quantity movements, which helps explain the latest run-up and subsequent collapse in the price of oil. Our analysis suggests that the variability of oil demand and supply shocks actually has decreased in the more recent past preventing even larger oil price fluctuations than observed in the data.

C223: **Investment timing under debt issuance constraint**
Presenter: **Takashi Shibata**, Tokyo Metropolitan University, Japan
Co-authors: Michi Nishihara

The optimal investment timing decision problem of a firm subject to a debt financing capacity constraint is examined. It is shown that the investment thresholds have a U-shaped relation with the debt capacity constraint, in that they are increasing with the constraint for the more severe debt issuance capacity while decreasing for the less severe capacity. Although the financing constraint distorts investment timing, it may encourage the constrained levered firm to overinvest compared with the nonconstrained levered firm. Our result fits well with the related problems involving the internal financing constraint.

Monday 19.12.2011 16:15 - 17:35 Parallel Session P – ERCIM

ESI02 Room Beveridge BAYESIAN NONPARAMETRICS Chair: Stephen Walker

E372: **A Bayesian nonparametric approach to species sampling problems**
Presenter: **Igor Pruenster**, University of Torino & Collegio Carlo Alberto, Italy
Co-authors: Stefano Favaro, Antonio Lijoi, Ramses H. Mena

Species sampling problems have a long history in ecological and biological studies. Data are recorded from a population which is made of different species and, conditionally on a sample of size n, interest lies in the evaluation of the species variety featured by an additional sample of size m. Some recent results obtained by adopting a Bayesian nonparametric approach are reviewed. In particular, two issues are considered: (i) measurement of species richness; (ii) assessment of rare species variety. The former is evaluated by means of estimators of the number of distinct species to be observed in the enlarged sample of size $n+m$ and of the discovery probability, namely the probability of discovering a new species at the $(n+m+1)-$th step. As for (ii), one can rely on the predicted number of species with frequency less than a given threshold and on the estimation of rare species discovery probability. Inference is based on Gibbs-type priors which include, as special case, the two-parameter Poisson-Dirichlet process. Under this specification, the estimators admit closed form expressions. Natural applications are related to sequencing of genomic libraries, e.g. Expressed Sequence Tags (EST) or the analysis of T-cells data in biology.

E571: **Nonparametric Bayesian modelling in systems biology**
Presenter: **Katja Ickstadt**, TU Dortmund University, Germany
Co-authors: Martin Schafer, Jakob Wieczorek

First, a short introduction to nonparametric Bayesian modelling, including generalized Dirichlet process mixture models as well as Poisson/gamma models, is made. We will then introduce two specific applications from systems biology in more detail. The first application centers around the problem of spatial modelling of protein structures on the cellular membrane. Spatial effects such as clustering are supposed to influence signal transmission. Here, a variation of the Dirichlet process mixture model with mixtures of multivariate normals will be employed in order to understand the cluster structure of Ras, a small protein adherent to the plasma membrane. The main goal of the second example is to understand how several components of a biological network are connected. In particular, we will study cell-matrix adhesion sites. Bayesian networks are a main model class for such problems, however, they have the drawback of making parametric assumptions. We will employ a nonparametric Bayesian network approach to overcome this limitation. In both examples, generalizations of the Dirichlet process prior, like the Pitman-Yor prior, for nonparametric Bayesian inference will be discussed. Also, biological prior knowledge will be incorporated into the nonparametric Bayesian models.

E814: **Bayesian feature selection for classification of metabolite NMR Spectra**
Presenter: **Maria De Iorio**, University College London, United Kingdom

Feature selection, identifying variables in metabolite profiles which discriminate between samples in different treatment groups, is an important problem in metabolomics. We have previously described a Bayesian statistical model for biofluid NMR spectral data, which can be used to quantify a collection of target metabolites and summarise other key features of a spectrum. We extend the model to incorporate a feature selector for supervised classification and regression. The extended algorithm seeks parsimonious subsets of spectral variables that distinguish between treatment groups or optimise the variance explained in a regresion model. The Bayesian approach naturally aggregates the uncertainty associated with the estimation of the spectral variables, the uncertainty associated with classification or regression and the uncertainty associated with feature selection. We demonstrate our model and an associated MCMC algorithm by reanalysing a dataset from a recent study of the toxic response of the earthworm L. rubellis on exposure to copper. We identify a number of metabolites that respond to copper dosing and further identify regions of spectra, using wavelets, where fluctuations improve the predictive power of the regression model.

ES61 Room Woburn CONTRIBUTIONS IN NONPARAMETRIC STATISTICS Chair: Juan Carlos Pardo-Fernandez

E090: **A new independence test for continuous variables**
Presenter: **Marcus Vollmer**, University of Greifswald, Germany
Co-authors: Christoph Bandt

Analyzing dependencies of two or more variables belongs to the everyday work of many scientists. Beside the Chi-squared, Fisher and Barnard tests, which use frequency data, there are correlation methods and several tests for continuous variables. It will become more relevant to have an efficient test with great power to widely alternatives of independence, especially non-linear dependencies. Thus, a new independence test for continuous variables, which analyses the geometric arrangement of data in a plane, is presented. If you count the surrounding points in the rank scatter-plot, you will notice, that a low score indicates a high level dependency. The distribution function of the scores of random permutations, which represents arrangements of data, is suitable to calculate a p-value. Therefore the name of the test is GRaP, that means Geometry of Random Permutations. Several non-linear dependencies have been tested and compared to some famous tests. Power analysis showed that the GRaP independence test is very powerful for typical non-linear alternatives.

E765: **Testing for one-sided alternatives in nonparametric censored regression**
Presenter: **Juan Carlos Pardo-Fernandez**, Universidade de Vigo, Spain
Co-authors: Cedric Heuchenne

Consider two bivariate populations. In each population, the first variable is identified with a covariate and the second variable is identified with the response of a general nonparametric and heteroscedastic regression model. We assume that the response variable is subject to right censoring. In this talk, we will propose a procedure to test the null hypothesis of equality of the regression functions versus one-sided alternatives, that is, the regression function in the first population is less than the regression function in the second population. We introduce two test statistics for which we obtain the asymptotic normality under the null and the alternative hypotheses. Although the tests are based on nonparametric techniques, they can detect any local alternative converging to the null hypothesis at the parametric rate. The practical performance of a bootstrap version of the tests is investigated in a simulation study. An application to a data set about unemployment duration times illustrates the proposed methodology.

E792: **Blood-flow velocity field estimation via spatial spline models with a PDE penalization**
Presenter: **Laura Azzimonti**, Politecnico di Milano, Italy
Co-authors: Laura M. Sangalli, Piercesare Secchi, Maurizio Domanin

The aim is to estimate the blood-flow velocity field in a section of a carotid artery, using data provided by eco-doppler images. The eco-doppler signal measures, at each time instant, the frequencies of blood cell velocities in a small volume within the artery lumen. In this application, physical and physiological knowledge of the phenomenon suggests the theoretical shape of the velocity field. This a priori knowledge is used to regularize the estimate; in particular, it is included in the model as a partial differential operator defining the roughness penalty of the surface smoother.

Accurate surface estimation is achieved resorting to Finite Elements, which provides a basis for piecewise-polynomial functions. The proposed methodology is especially well suited for all those applications where some a priori knowledge of the problem (physical, chemical, mechanical or morphological) suggests the choice of a partial differential operator modeling to some extent the phenomenon under study.

E778: **On projection-type estimators of multivariate isotonic functions**
Presenter: **Abdelaati Daouia**, Universite catholique de Louvain, Belgium
Co-authors: Byeong U. Park

Let M be an isotonic real-valued function on a compact subset of $\mathbb{R}^d$ and let $\hat{M}_n$ be an unconstrained estimator of M. A feasible monotonizing technique is to take the largest (smallest) monotone function that lies below (above) the estimator $\hat{M}_n$ or any convex combination of these two envelope estimators. When the process $r_n(\hat{M}_n - M)$ is asymptotically equicontinuous for some sequence $r_n > 0$, we show that these projected estimators are r_n-equivalent in probability to the original unrestricted estimator. We provide some applications including estimation of conditional distribution, econometric and biometric functions.

ES73 Room Jessel CONTRIBUTIONS ON HIGH-DIMENSIONAL DATA ANALYSIS Chair: Ejaz Ahmed

E141: **Multivariate density estimation using geometric methods**
Presenter: **Leonard Hearne**, University of Missouri at Columbia, United States of America

A geometric density estimator is considered. This estimator uses the Delaunay tessellation to partition the support for an estimator into tiles. The proportion of probabilistic mass from observations on each tile divided by the content of the tile is used to estimate the probability density on each tile. Under suitable regularity conditions, geometric density estimators can be shown to be unbiased and consistent multivariate estimators. The level of density specificity is proportional to the number of tiles in the tessellation. To refine a density estimate, re-sampling methods can be employed. This partitions the support for the estimator into successively smaller tiles. The resulting refined density estimate is biased. By allowing some probability mass to be allocated beyond the convex hull in a Delaunay tessellation, these refined density estimators can again be made unbiased and consistent. This work has application in a broad class of multivariate estimation settings like finance, genetics, and medicine where geometric methods can be employed.

E543: **Testing for a large dimensional covariance matrix using the semicircle law**
Presenter: **Konstantin Glombek**, University of Cologne, Germany

New tests for sphericity of the covariance matrix and the covariance matrix itself of a $d-$dimensional multinormal population $X \sim N_d(\mu,\Sigma)$ are presented. These tests are applicable if the sample size n and d both go to infinity while $d/n \to y \in (0,\infty)$ provided that the limits of $\text{trace}(\Sigma^k)/d, k = 1,\ldots,8$, exist. The main idea of these tests is to check if the empirical spectral distribution of a suitable standardized sample covariance matrix obeys the semicircle law or not. Due to similarities of the semicircle law to the normal distribution, the proposed test statistics are of the type of the Jarque-Bera test statistic. Simulation results show that the new tests outperform the tests from the literature in some cases if y is small. Further, the new tests are comparable to the existing ones if $y < 1$.

E385: **Large portfolio optimization using wavelet thresholding**
Presenter: **Daniel Koch**, Universite Catholique de Louvain, Belgium
Co-authors: Sebastien Van Bellegem

The mean-variance portfolio optimization setting takes the form of a quadratic programming problem which uses the mean return vector and the cross-covariance matrix of a N-dimensional stationary process as input. Due to the fact that the process cannot be observed directly the mean return vector and the covariance matrix need to be replaced by estimates. In high-dimensional settings, e.g. when the number of assets is large relative to the sample size, the empirical covariance matrix is badly conditioned. Inversion of the covariance matrix is therefore unstable and portfolio optimization behaves poorly. We show that, under realistic assumptions, wavelet bases are such that the relevant information is concentrated on a small number of wavelet coefficients. In other words, wavelets achieve some decorrelation of the stationary process. We exploit this property and introduce a new thresholding rule of the empirical covariance matrix in the wavelet domain. In contrast to standard wavelet thresholding approaches, we do not operate on each wavelet coefficient at a time but on groups of coefficients. In some situations we also provide a time-domain interpretation of the thresholding rule. Simulation studies show the good performance of these decision rules compared to benchmarks and other methods of regularization such as Tikhonov regularization.

E726: **Sparse estimation of time-frequency surface for sound signals using regularized Bayesian methods**
Presenter: **Jeong-Ran Lee**, Seoul National University, Korea RoK
Co-authors: Gwangsu Kim, Yongdai Kim, Hee-Seok Oh

In this paper, a simple Bayesian method for the time-frequency inverse modeling of nonstationary signals is proposed. The model is based on the Gabor expansion, which requires a careful regularization through appropriate choices of variable selection schemes and prior distributions to cope with the overcompleteness. We introduce prior specifications motivated by the stochastic search variable selection (SSVS) in combination with Indian buffet process (IBP) that are tailored to achieve sparse time-frequency representations, and provide an effective MCMC algorithm for the inference. Applications of the proposed method are explored in the examples of signals such as sounds from a musical instrument, bat and speech.

ES79 Room Court CONTRIBUTIONS TO ROBUST ANALYSIS OF COMPLEX DATA SETS Chair: Christophe Croux

E590: **Robust multivariate coefficients of variation**
Presenter: **Gentiane Haesbroeck**, University of Liege, Belgium

When comparing the variability of data expressed in different units, the coefficient of variation is often used. While its definition is simple and well established in the univariate setting, several extensions to the multivariate case have been introduced in the literature. These multivariate coefficients of variation all rely in one way or another on the variance-covariance matrix (via its determinant, its trace or its inverse). When there are outliers in the data, the estimated values of the multivariate coefficients of variation may become unreliable and robust alternatives should be used. Applying the plug-in principle allows us to derive robust multivariate coefficients of variation in a straightforward way. In this talk, the robustness of the different coefficients of variation are compared by means of influence functions. Moreover, a real example based on compositional data will illustrate the attractivity of these robust measures of variability.

E639: **Robust estimation of conditional heteroscedastic models and forecasting of value-at-risk**
Presenter: **Kanchan Mukherjee**, Lancaster University, United Kingdom

The class of M-estimators for the parameters of the symmetric GARCH model and asymmetric GJR model of heteroscedasticity is introduced. We show, among other things, that the widely-used QMLE can estimate the underlying parameters consistently if and only if the innovation variance is one. Consequently, we propose alternative estimators of the transformed GARCH and GJR parameters and derive their asymptotic normality

under very weak assumptions on innovation moments such as the finiteness of merely fractional moment. Such results have important practical implications since many financial data reveal lack of higher moments. We also discuss application of these estimators to the forecasting of the value-at-risk, various backtesting procedures and other analyses of the financial data extending some recent results of Mukherjee (2008, Econometric Theory) and Iqbal and Mukherjee (2010, Statistics and Computing and 2011, Journal of Forecasting).

E915: **Asymptotic theory for iterated one-step Huber-skip estimators**
Presenter: **Bent Nielsen**, University of Oxford, United Kingdom
Co-authors: Soren Johansen

In regression analysis it is often an important concern to be able to detect outliers or other unsuspected structures. A very simple algorithm addressing this is first to obtain an initial estimator of the parameters, use this to discard observations with large residuals, and then re-run the regression. This is the one-step Huber-skip estimator. It is a special case of the one-step M-estimator in which the criterion function is not convex. A number of algorithms have been proposed which iterate such one-step estimators. We give an asymptotic fix-point result for such iterations of one-step Huber-skip estimators when the model has no outliers. The result is based on a tightness argument and allows regressors which are fixed, stationary, and non-stationary. We also apply the result to analyse the Forward Search Algorithm.

E088: **On the robustness of two-stage estimators**
Presenter: **Mikhail Zhelonkin**, University of Geneva, Switzerland
Co-authors: Marc Genton, Elvezio Ronchetti

Many estimators in the statistics and econometrics literature are obtained following a two-stage (2S) procedure. These procedures are mostly based on 2S Maximum Likelihood (ML) or Least Squares (LS) estimation. It is well known that classical ML and LS estimators are very sensitive to deviations from the underlying stochastic assumptions of the model or to outliers in the data. We present a general framework of treating 2S models based on M-estimation. The class of M-estimators has the advantage to include most of the 2S estimators available in the literature, to suggest a general way to robustify 2S estimators, and to clarify the structure of their asymptotic variance. Moreover, we provide an extension of these results to multi-stage procedures. Finally, we provide two specific examples of applications, namely the 2S least squares estimator and Heckman's sample selection bias correction procedure.

ES54 Room Senate IMPRECISION IN STATISTICAL DATA ANALYSIS II Chair: Angela Blanco-Fernandez

E475: **Dynamic directed evidential networks with conditional belief functions: Application to system reliability**
Presenter: **Wafa Lamari**, ISG Tunis, Tunisia
Co-authors: Boutheina Ben Yaghlane, Christophe Simon

Available knowledge in many applications is characterized by uncertainty which is either aleatory or epistemic. Graphical models are very powerful tools for representing uncertainty in knowledge. In this context, evidential networks are suitable and efficient when using the modified binary join tree proposed in the literature. In addition to the problem of uncertainty, another problem relying on the time-varying knowledge has to be overcome. We propose to use the evidential formalism to model knowledge in an uncertainty context and to handle the temporal evolution of uncertain knowledge, we develop the concept of dynamic evidential network with conditional belief functions (DDEVN) which is the evidential counterpart of dynamic Bayesian network (DBN) and dynamic possibilistic network (DPN). We extend the modified binary join tree to handle this temporal dimension. For the sake of illustration, the DDEVN is applied to the reliability study of a well-known three valves system.

E687: **Joint prediction of observations and states in time-series: A partially supervised prognostics approach based on belief functions and K-nearest neighbours**
Presenter: **Emmanuel Ramasso**, FEMTO-ST, France
Co-authors: Michele Rombaut, Noureddine Zerhouni

Forecasting the future states of a complex system is a complicated challenge that is encountered in many industrial applications covered in the community of Prognostics and Health Management (PHM). Practically, states can be either continuous or discrete: Continuous states generally represent the value of a signal while discrete states generally depict functioning modes reflecting the current degradation. For each case, specific techniques exist. In this paper, we propose an approach based on case-based reasoning that jointly estimates the future values of the continuous signal and the future discrete modes. The main characteristics of the proposed approach are the following: - The method relies on the K-nearest neighbours algorithm based on belief functions theory; - Belief functions allow the user to represent his partial knowledge concerning the possible states in the training dataset, in particular concerning transitions between functioning modes which are imprecisely known; - Two distinct strategies are proposed for states prediction and the fusion of both strategies is also considered. Two real datasets were used in order to assess the performance in estimating future break-down of a real system. The prediction accuracy obtained by the fusion process was close to 80% in both applications.

E728: **Two-sample inference about mean, variance and proportion using imprecise data**
Presenter: **Shubhabrata Das**, Indian Institute of Management Bangalore, India

Consider imprecise or fuzzy data as recorded in a Likert scale. We study how the imprecision in measurement scale of the true underlining variables distorts inference drawn from it. In particular, we focus on three basic parameters of statistical inference, namely mean, variance and proportion. Of particular interest is the comparison of parameters between two populations or two similar/ related characteristics. For the comparison of mean and proportion we look at shift in location, while for variance we look at possible change in scale. The study is expected to provide direction in determining suitable measurement scale and ensuing statistical inference from resultant data.

E835: **Time series classification by imprecise hidden Markov models: Supporting continuous variables**
Presenter: **Alessandro Antonucci**, IDSIA, Switzerland
Co-authors: Rocco de Rosa

Hidden Markov models are dynamical probabilistic models widely used for the modelling of temporal data. An imprecise-probabilistic version of these models, where sets instead of single distributions are adopted for the quantification of the model parameters, has been recently proposed. In these works, the higher freedom in the quantification only concerns the transition probabilities (among the discrete hidden variables), while the emission model for the (continuous) observable variables should be still assumed to be precise because of a number of technical limitations. These issues are fixed in the present work. We explore the possibility of learning an imprecise emission model, and we also show how inferences on models of this kind can be easily reduced to standard inferences on discrete models. This allows for the practical application of these models to classification. Such a new class of imprecise hidden Markov models is in fact tested on a supervised classification task, namely the recognition of video sequences depicting human actions. Preliminary experiments show the advantages of a more cautious quantification of the emission parameters.

ES78 Room Bloomsbury CONTRIBUTIONS IN APPLIED STATISTICS Chair: Agustin Mayo-Iscar

E764: **Modelling population dynamics from repeated surveys**
Presenter: **P. Mpesmpeas**, Athens University of Economics and Business, Greece

Recording and monitoring wildlife is crucial for the conservation of wild species. The most common type of information from a monitoring scheme is a time series of population abundances but the potential of such data for monitoring population changes can be limited. Recent work has shown that repeating the population survey can considerably increase the information obtained from the monitoring. We introduce new approaches to incorporating information from repeated surveys in state-space models for population dynamics. The methods include an approach which employs the individual observations at each period and an approach based on the use of aggregated samples, which is simpler and lends itself readily to data derived by any sampling procedure.We evaluate the methods using real data and a simulation study and outline various extensions of the work.

E880: **Propensity scores in observational studies**
Presenter: **Cara Dooley**, National University of Ireland, Galway, Ireland
Co-authors: John Hinde, John Newell

The propensity score was introduced many years ago to provide for balance in observed covariates in observational studies, which generally have no design before data collection. The propensity score is the conditional probability of being in the group of interest given the observed covariates. Two ways to incorporate the propensity score are: (i) using the propensity score to match the cases of interest to the most similar controls using the observed covariates or (ii) using the propensity score as a covariate in a regression model. In datasets with large imbalance in the number of subjects in each group, such as the example that will be examined here with 170 subjects of interest and over 20,000 control subjects, the estimation of the propensity score is not without problems. Often some form of stabilisation technique is used for the propensity score, but for the example of interest these had little effect. The issue in this case is that, due to the large number of controls, the model fails to adequately assign any subjects group of interest, assigning them all propensity scores of close to zero. We will explore possible methods for addressing this issue.

E857: **Graphical model representations of multivariate time series for road traffic flow forecasting**
Presenter: **Osvaldo Anacleto-Junior**, Open University, United Kingdom
Co-authors: Catriona Queen, Casper Albers

Traffic data have some characteristics that can be quite challenging to deal with from a statistical modelling perspective. In addition, on-line forecasts of multivariate traffic flow data are required to support real-time traffic management decisions. Multiregression dynamic models, which combine a representation of the multivariate time series by a directed acyclic graph with a state space model, have been shown to be a useful class of models in this field. However, there are still some modelling issues that must be addressed, such as measurement errors due to possible malfunctions of the data collection devices, different levels of traffic variability depending on the period of the day and the inclusion of additional traffic variables which can improve the predictive performance of traffic flow models. It will be shown how the multiregression dynamic model can be extended in order to deal with these issues, thus improving their suitability for multivariate short-term flow forecasting. Data from the intersection of three motorways – the M60, M62 and M602 in the west of Manchester – will be used for an application of the proposed methodologies.

E873: **Semi-nonparametric approximations to the distribution of Portmanteau statistics**
Presenter: **Deepak Sanjel**, Minnesota State University, United States of America
Co-authors: Serge Provost

Accurate semi-nonparametric approximations to the distribution of certain Portmanteau statistics that are expressible as sums of ratios of quadratic forms have been presented. Two methodologies, namely the symbolic computational approach and a recursive formula expressing joint moments in terms of joint cumulants, are being utilized to determine the exact moments of the portmanteau statistics. The density functions of those statistics are then approximated on the basis of those moments in terms of gamma density functions and Laguerre polynomials. As verified by a simulation study, the proposed approximations prove more accurate than those that are based on the asymptotic chi-square distribution, especially in the case of time series of short or moderate length.

Monday 19.12.2011 16:15 - 17:35 Parallel Session P – CFE

CS78 Room S261 STOCHASTIC VOLATILITY **Chair: Alessandra Luati**

C503: **Short-term GDP forecasting with a mixed frequency dynamic factor model with stochastic volatility**
Presenter: **Fabrizio Venditti**, Bank of Italy, Italy
Co-authors: Massimiliano Marcellino, Mario Porqueddu

A mixed frequency data small scale dynamic factor model with stochastic volatility is developed. The model picks up two major shifts in euro-area business cycle volatility: the former is associated with the 2001 recession and peculiar to the manufacturing sector; the latter, common to all indicators, characterizes the latest recession. We first use the model to show some useful applications for nowcasting and short term forecasting that arise from the Bayesian framework in which the indicator is set. In particular we show how the model allows us to compute probability distributions of (1) GDP forecasts and (2) contributions of innovations to the monthly indicators included in the model. We then turn to the point and density forecast evaluation. In line with findings in the vector autoregressions (VAR) literature we find that stochastic volatility contributes to a significant increase in density forecast accuracy.

C681: **A timely analysis of unconventional monetary policy via dynamic Nelson Seigel models**
Presenter: **Paul Veerhuis**, Australian Prudential Regulation Authority, Australia
Co-authors: Gareth Peters, Richard Gerlach

We present an analysis of the Reserve Bank of Australia (RBA) interventions in term money markets at the time of the financial crisis with the objective of identifying the level of success that the RBA had in relieving liquidity pressures in these markets. We use an arbitrage free dynamic Nelson Siegel model with stochastic volatility to jointly model Australian government, credit and LIBOR yield curves. The typical Nelson Siegel framework is extended to six factors with the first five factors representing the risk free and credit risk contribution to the price of unsecured bank lending. The remaining factor is then considered a proxy for liquidity risk in unsecured bank lending and a comparison of its time series is made to the point in time and volume of RBA interventions. Our analysis incorporates stochastic volatility in the observation residuals included via a Wishart process for the covariance structure of the residual covariance for contracts with different maturities. We perform joint calibration and filtering for the latent process and the model based on Adaptive Particle Markov chain Monte Carlo (PMCMC).

C710: **Measuring the trading volume in heterogeneous markets with stochastic volatility**
Presenter: **Eva Maria Ortega**, UMH, Spain
Co-authors: Jose Alonso

The trading volume refers to the number of shares or contracts of a security traded during a defined time period. Its relationship with the price changes has played a prominent role. Classical studies reflect that the volume is positively related to the magnitude of the price change in equity markets, however, in other markets, there are situations with divergence between them. The market microstructure advocates that the price movements are caused by arrival of new information and the process to incorporate it. The Modified Mixture of Distributions Model (MMD) states that the trading volume is a mixture of a Poisson distribution, unconditioning on the intensity of the information arrivals. We introduce three indices to measure the volume of shares with a fixed market characteristic, based on the MMD model, having stochastic volatility. Their probability distributions are analyzed as mixtures having arbitrary mixing distributions. We apply our results in the literature and new ones on stochastic comparisons of random sums and mixture of parametric distributions, to compare analytically the variability and the magnitude of the (specific) trading volumes, under different financial scenarios, emphasizing the effect of variations and correlations of some parameters. Bounds in closed-form based on Poisson compounds are built for the risk assessment in financial markets.

C803: **Annualizing volatility under long memory in high frequency variance**
Presenter: **Antonios Antypas**, University of Piraeus, Greece
Co-authors: Nikolaos Kourogenis

The relationship between the estimated daily variance and the variance of longer horizon returns under the presence of long memory in the conditional volatility is analyzed. A parametric specification of the conditional volatility offers the link between the long memory parameter of the model and the correct multiplier that must be used in order to obtain the annualized volatility. Extensive Monte Carlo experiments provide strong evidence that in this case the annualizing multiplier is significantly different from the square root of the number of days in the year. Specifically, it is shown that long memory in the conditional volatility causes the multiplier to decrease. Tables are presented that provide the value of the multiplier for a wide range of values of the long memory parameter of the model. Further, Monte Carlo experiments are provided for the assessment of the robustness of this approach when the volatility process is misspecified. The proposed method can be used in real world data, with possible applications to long horizon value at risk and option pricing.

CS13 Room Torrington BAYESIAN QUANTILE REGRESSION **Chair: Boris Choy**

C104: **Bayesian methods in quantile regression: a review**
Presenter: **Keming Yu**, Brunel University, United Kingdom

Classical quantile regression is an important regression method and model for the inference of conditional quantiles of a response variable given a vector of covariates. It can be used to measure the effect of covariates not only in the center of a distribution, but also in the tails. Bayesian methods involve formal combination through the use of Bayes's theorem of a prior distribution or belief about the value of a quantity of interest (for example, experience and expert opinion) based on evidence not derived from the study under analysis and a summary of the information available from the data collected in the study (or likelihood) to produce an updated or posterior distribution of the quantity of interest. Bayesian inference quantile regression has emerged in recent years. A brief review and discussion of the methods will be made, including Bayesian parametric quantile regression, Bayesian nonparametric quantile regression, Bayesian spatial quantile regression, Bayesian censored quantile regression, prior selection, variable selection and some applications and comparison examples.

C188: **Bayesian estimation and forecasting for semi-parametric conditional expected shortfall models**
Presenter: **Richard Gerlach**, University of Sydney, Australia
Co-authors: Cathy Chen, Liou-Yan Lin

The asymmetric-Laplace distribution has allowed likelihood, and consequently Bayesian, approaches to estimation and forecasting in quantile regression and dynamic quantile CaViaR models. Despite its wide-spread use, the incoherency of Value-at-Risk (VaR) as a risk measure has allowed conditional VaR or expected shortfall (ES) to become a potential alternative. Until recently, approaches to ES were mainly parametric. However, recently it has been illustrated how to use asymmetric least squares (ALS) to estimate semi-parametric expectiles, via conditional autoregressive expectile (CARE) models, and to utilise these in estimating dynamic ES. First, we extend the class of CARE models to include a

fully nonlinear threshold family. Second, we develop a new distribution, using the ALS criterion to build a probability density function, that we call the asymmetric-Gaussian (AG). The AG allows a likelihood solution to minimise the required ALS function for expectiles. We employ the AG together with a flat prior in a Bayesian approach to estimation and inference for expectiles via CARE models, allowing semi-parametric conditional ES models. A simulation study shows the properties of our Bayesian estimators, while an empirical study in multiple financial market indices compares our Bayesian CARE models to a range of competitors for forecasting ES, including GARCH specifications.

C552: **Bayesian copula-based skewed-EWMA quantile forecasting for portfolios**
Presenter: **Zudi Lu**, The University of Adelaide, Australia

With the increasing complexity of risks, how to estimate the risk of portfolios with complex dependencies is challenging. Recently, a skewed-EWMA procedure has been proposed by the presenter and his co-authors to calculate conditional quantile, or value-at-risk (VaR), for individual financial assets, which is derived from an asymmetric Laplace distribution and takes into account both skewness and heavy tails of the return distribution that are adaptive to the time-varying nature in practice by adjusting the shape parameter in the distribution. We will report an extended skewed-EWMA procedure to estimate the risk of complex portfolios with dependencies modelled via copula and Bayesian approaches. Monte Carlo simulation procedure that combines copula techniques with skewed-EWMA forecasting is developed. The empirical backtesting evaluation of the conditional quantile forecasting will be demonstrated in estimating extreme risks of some complex portfolios.

CS79 Room B20 FINANCIAL MARKETS Chair: Ana-Maria Fuertes

C616: **Long-term mean reversion and predictability of the US stock market returns**
Presenter: **Valeriy Zakamulin**, University of Agder, Norway

Formal tests of the random walk hypothesis in the prices of some U.S. stock portfolios over increasing horizons up to 40 years are performed. First of all, the results of these tests allow rejecting, at the conventional statistical levels, the random walk hypothesis for virtually all of the portfolios under investigation. Second, the results of these tests suggest not only the presence of mean reversion, but also some degree of predictability of multiperiod returns. Following the performance of out-of-sample forecast of multiperiod returns is analyzed. These latter tests demonstrate that for some stock portfolios the forecast based on the mean reverting model is statistically significantly better than the forecast based on the random walk model. In particular, we provide the evidence of the out-of-sample predictability of 13-18 year returns on the Standard and Poor's composite stock price Index, 4-5 year returns on the small cap stocks, and 16-18 year returns on the large cap stocks.

C913: **Predicting systemic financial crises with genetically optimized neural networks**
Presenter: **Peter Sarlin**, Abo Akademi University, Finland

A neuro-genetic (NG) model for predicting systemic financial crises in advanced and emerging economies with common indicators for capturing the build-up of domestic and global vulnerabilities and imbalances is developed. By using a genetic algorithm, the NG model specifies the (1) combination of inputs, (2) network configuration and (3) "user-specified" training parameters for a standard back-propagation artificial neural network (ANN). The performance of the NG model is evaluated by comparing it with stand-alone logit and ANN models in terms of usefulness for a policymaker. We show that the NG model provides better in-sample and out-of-sample performance as well as decreases in expertise and labor needed for, and uncertainty caused by, manual calibration of a predictive ANN model. We create and evaluate the models objectively using only in-sample information as well as an early-stopping procedure to achieve sufficiently parsimonious, but still non-linear, models for generalized processing of out-of-sample data. The result is an easy-to-use flexible tool for policymaking as it does not need supervision of an expert analyst and performs well in predicting out-of-sample the global financial crisis that started in 2007.

C673: **Private equity placements, cash dividend and tunneling: Empirical evidences from listed companies in China**
Presenter: **Yufang Zhao**, Huazhong University of Science and Technology, China
Co-authors: Xinping Xia, Hao Xiao, Yixia Wang

The relationship between private equity placements and cash dividend payout in listed companies which have achieved private equity placements is examined. With a sample of listed companies which have achieved private equity placements during the period of May 8,2006 -Dec 31,2009 on Shanghai and Shenzhen stock exchanges, we find that the Chinese listed companies pay out more cash dividend after private equity placement achieved than those unachieved; the companies with large shareholders participating in the private equity placements pay out more cash dividend than the companies without the principal shareholder partaking, the more the large shareholder hold shares, the more the companies pay out cash dividend. Meanwhile, we find that the companies that achieved the private equity placement pay out a lot of cash dividend in the placement years. These results indicate that the companies controlled by the large shareholders have significant tunneling preferences on cash dividend policies.

CS22 Room Bedford DENSITY FORECASTING USING REALIZED MEASURES Chair: Florian Ielpo

C154: **Testing for jumps in GARCH models, a robust approach**
Presenter: **Sebastien Laurent**, Maastricht University, Netherlands
Co-authors: Christelle Lecourt, Franz Palm

Financial series occasionally exhibit large changes. Assuming that the observed return series consist of a standard normal ARMA-GARCH component plus an additive jump component, a new test for additive jumps is proposed in an ARMA-GARCH context. The test is based on standardised returns, where the first two conditional moments are estimated in a robust way. Simulation results indicate that the test has very good finite sample properties, i.e. correct size and much higher proportion of correct jump detection than other tests. We apply our test on the YEU-USD exchange rate and find twice as many jumps as other tests.

C167: **The contribution of jumps for forecasting the density of returns**
Presenter: **Benoit Sevi**, Aix-Marseille School of Economics, France
Co-authors: Julien Chevallier, Florian Ielpo

The estimation of the jump component in asset pricing has witnessed a considerably growing body of literature. Of particular interest is the decomposition of total volatility between its continuous and jump components. Recent contributions highlight the importance of the jump component in forecasting the volatility at different horizons. A recent methodology is extended to measure the information content of intraday data in forecasting the *density* of returns at horizons up to sixty days. We extract jumps sequentially to have a measure of the jumps in returns. Then, we estimate a bivariate model of returns and volatilities where the jump component is independently modelled. Our empirical results for S&P 500 futures, WTI crude oil futures, the USD/JPY exchange rate and the MacDonald's stock confirm the importance of considering the continuous/jump decomposition for density forecasting.

C178: **Multivariate high-frequency-based volatility (HEAVY) models**
Presenter: **Diaa Noureldin**, University of Oxford, United Kingdom
Co-authors: Neil Shephard, Kevin Sheppard

A new class of multivariate volatility models that utilizes high-frequency data is introduced. We discuss the models' dynamics and highlight their differences from multivariate generalized autoregressive conditional heteroskedasticity (GARCH) models. We also discuss their covariance targeting specification and provide closed-form formulas for multi-step forecasts. Estimation and inference strategies are outlined. Empirical results suggest that the HEAVY model outperforms the multivariate GARCH model out-of-sample, with the gains being particularly significant at short forecast horizons. Forecast gains are obtained for both forecast variances and correlations.

CS27 Room Gordon ANALYSIS OF LARGE-DIMENSIONAL DATASETS: RECENT ADVANCES Chair: Marco Lippi

C241: **Recent advances in factor analysis: from stationary to evolutionary**
Presenter: **Giovanni Motta**, Maastricht University, Netherlands
Co-authors: Matteo Barigozzi, Marco Lippi

The main results of factor analysis for stationary time series are reviewed, while comparing static versus dynamic representations. Then we introduce the recently developed Evolutionary Factor Analysis, that generalizes the (identification and estimation) theory to the non-stationary framework. The behavior of different estimators applied to different factor models is studied and compared. In particular, we compare stationary versus evolutionary representations and the corresponding estimators by means of simulated data. Finally, we apply the different methods to US macroeconomic data.

C242: **Do Euro area countries respond asymmetrically to the common monetary policy?**
Presenter: **Antonio Conti**, Bank of Italy, Italy
Co-authors: Matteo Barigozzi, Matteo Luciani

The possible existence of asymmetries among Euro Area countries reactions to the European Central Bank monetary policy is investigated. By estimating a Structural Dynamic Factor model on a large panel of quarterly variables including data on the aggregate Euro Area as well as country-specific key economic variables, we find that member states react asymmetrically in terms of prices and unemployment, while no difference appears in terms of output. These results seem to be the consequence of structural local socio-economic factors on which national fiscal policies are more effective, rather than of European Central Bank policies.

C504: **Generalized dynamic factor models and singular ARMA models**
Presenter: **Bernd Funovits**, Vienna Graduate School of Economics, Austria
Co-authors: Elisabeth Felsenstein, Brian D.O. Anderson, Manfred Deistler, Weitian Chen

Data-driven modeling for Generalized Dynamic Factor Models (GDFM's) is considered as a problem of estimating a singular ARMA model (i.e. an ARMA model with singular innovation variance) from noisy data (with the noise being weakly dependent). We discuss the structure of singular AR and ARMA models and of the corresponding processes. Special emphasis is given to questions of identifiability, canonical forms and realization (in the sense of getting a system from a transfer function) of ARMA and State Space representations.

CS82 Room S264 COMPUTATIONAL ECONOMETRICS AND APPLICATIONS II Chair: Roderick McCrorie

C101: **Validating the CPC model using parametric and non-parametric inference based empirical algorithmic methods**
Presenter: **Januj Juneja**, San Diego State University, United States of America

The literature on common factors suggests that employing the hypothesis of common axes to financial data can lead to inaccurate inference regarding the fit of the common principal components (CPC) model. This research investigates the validity of conclusions drawn from this model, which is assessed by constructing simulation-algorithms and then employing statistical tests on the characteristics of resultant factors. The data used is generated by combining cross-sectional properties of the observed interest rate with a factor model representation. It establishes a link between the hypothesis of CPC and the data. The results imply that the inaccurate inference associated with the hypothesis of CPC lies in the inability of financial data to conform to the classical factor model assumptions. Nevertheless, through bootstrap procedures, we demonstrate that the key to determining the appropriateness of the fit of the CPC model to financial data lies in using principal components analysis (PCA) to disentangle common from local influences. This disentanglement allows us to propose an algorithm, nesting the Flury and Gautschi algorithm, which can validate the fit of the CPC model. The application of this algorithm is potentially very powerful to those who are interested in managing risk in international settings.

C643: **Overconfidence and credit cards**
Presenter: **Susan Kriete-Dodds**, University of Basel, Switzerland
Co-authors: Dietmar Maringer

In behavioural economics, overconfidence has become a key concept for explaining anomalies such as security misevaluation or excessive trading volume. While agent-based modelling can offer great insight into overconfidence, many models that consider overconfidence are instead experimental, empirical or market-based. An agent-based model is presented that examines the effects of overconfidence on credit card usage. Overconfidence is used here to explain why people who never intend to borrow on their credit card(s) do so anyway. The model contains consumption, two means of payment (credit card and cash), and a distortion to agents' income expectations via overconfidence. It was found that overconfidence leads to slightly more "accidental" borrowing and to increased turnover and credit card usage.

C901: **Computational methods for pricing Asian options: An evaluation**
Presenter: **Roderick McCrorie**, University of St Andrews, United Kingdom
Co-authors: Cao Liang

Up-to-date comparison and contrast among various analytical and numerical methods of pricing Asian options are provided. Using Lie symmetry methods, we discuss the hidden symmetries in the two-dimensional PDE for arithmetic Asian options, thereby explaining the generic difficulty in pricing them. Various methods are then compared and contrasted both analytically and in a sampling experiment, including the PDE approach; methods based on numerically inverting the Geman-Yor Laplace transform; Laguerre series methods; spectral expansion; and constructive complex analysis. Essentially, some comparative analyses in previous papers are re-evaluated and brought up-to-date, and recently modified Abate-Whitt methods are now seen to perform relatively better than before. Especial care is taken to treat the low volatility case robustly.

C745: **ICA based asset allocation**
Presenter: **Maximilian Vermorken**, University College London, United Kingdom
Co-authors: Francesca Medda, Thomas Schroeder

Since its introduction, authors have questioned the adequacy of the Mean-Variance criterion for the allocation of wealth. It fails to capture the non-Gaussianity of financial data as it does not include higher moments. We propose a new method to address the asset allocation problem. Starting from the traditional assumption of investor utility maximisation, we first operate a rotation of the asset space through Independent Component Analysis (ICA). The independent components are identified using all moments of return distributions of the considered assets. Using this rotation an analytical solution can be found for the utility maximisation problem, as integrals have closed form solutions. The asset allocation based on our method therefore uses all moments of the distribution in allocating wealth without making the sometimes limiting assumptions of other higher-moment based asset allocation methods. In the present paper the method is explained and tested. Test results show that the ICA based method clearly includes more aspects of the financial data when compared with Mean-Variance based portfolios. The method is applied for the composition of infrastructure funds where higher moment based asset allocation systems could be beneficial for the funds' performance.

CS62 Room B35 FINANCIAL ECONOMETRICS FOR RISK MANAGEMENT Chair: Chung-Ming Kuan

C791: **Predicting defaults with regime switching intensity: Model and empirical evidence**
Presenter: **Hui-Ching Chuang**, National Taiwan University, Taiwan
Co-authors: Chung-Ming Kuan

A default prediction model with regime-switching effects in the intensity function is proposed. In particular, the level of the intensity function and the risk exposures to observable risk factors in the intensity function are specified as regime dependent. We provide an estimation algorithm when the state variable is Markovian and illustrate the proposed model using the default data of US-listed companies during 1990-2009. Our test indicates that the regime switching effect in the intensity function is significant. A well-specified regime-switching intensity model outperforms intensity model in both in-sample and out-of-sample default predictions.

C077: **Weighted-Nadaraya Watson estimation of conditional expected shortfall**
Presenter: **Kengo Kato**, Hiroshima University, Japan

The problem of nonparametric estimation of the conditional expected shortfall (CES), which has gained popularity in financial risk management, is addressed. A new nonparametric estimator of the CES is proposed. The estimator is defined as a conditional counterpart of the sample average estimator of the unconditional expected shortfall, where the empirical distribution function is replaced by the weighted Nadaraya-Watson estimator of the conditional distribution function. We establish asymptotic normality of the proposed estimator under an α-mixing condition. The asymptotic results reveal that the proposed estimator has a good bias property. Simulation results illustrate the usefulness of the proposed estimator.

C783: **Idiosyncratic risk-based commodity strategies**
Presenter: **Adrian Fernandez-Perez**, University of las Palmas de Gran Canaria, Spain
Co-authors: Ana-Maria Fuertes, Joelle Miffre

A novel trading strategy that exploits the idiosyncratic risk level of individual commodities as signal for tactical asset allocation is developed. Residual risk is measured relative to hedging pressure-based benchmarks that explicitly take into account the natural propensity of commodity futures markets to be either in backwardation or in contango. While the resulting long-short portfolios remain useful for diversification and inflation hedging, they are a source of outperformance relative to long-only positions. The idiosyncratic risk signal is shown to have very low overlap with momentum and term structure signals. Among other robustness tests, we demonstrate that the performance is not eroded by transaction costs and cannot be attributed to liquidity risk.

C613: **Identification of price jumps, cojumps and tail dependence in financial asset prices**
Presenter: **Jin-Huei Yeh**, National Central Univ., Taiwan
Co-authors: Mu-Shu Yun

Price jumps in financial markets have been a very important issue in Finance since it would distort the decision we make in asset allocation and risk management, not to mention the financial implications of co-jumps in hedging systematic risk. We propose a simple approach for identifying jumps and co-jumps in asset prices utilizing the recent realized measures of variation in high frequency finance. The framework is informative in testing the existence of jumps, dating jump days, and quantifying jump contributions to price variations. Our suggested approach improves the over-alarm detection of jumps from previous formal tests. Moreover, we also find evidence of predictable jumps, in particular with respect to proxy that contains forward looking information. Generalizing the idea to test the presence of cojumps in realized covariances, this simple approach allows us to examine the existence of tail dependence and thus is capable of serving as a pretest before fitting any copula models. For the empirically investigated S&P 500 spot and futures indices, we found asymmetries in both the frequencies and magnitudes among those downward or upward comovements.

CS71 Room B33 REGRESSION TREES AND STRUCTURAL BREAKS Chair: Marco Reale

C317: **When long memory meets the Kalman filter: A comparative study**
Presenter: **Stefano Grassi**, Aarhus University and Creates, Denmark
Co-authors: Paolo Santucci de Magistris

The finite sample properties of the state space methods applied to long memory time series are analyzed through Monte Carlo simulations. The state space setup allows us to introduce a novel modeling approach in the long memory framework, which directly tackles measurement errors and random level shifts. Missing values and several alternative sources of misspecification are also considered. It emerges that the state space methodology provides a valuable alternative for the estimation of the long memory models, under different data generating processes, which are common in financial and economic series. Two empirical applications highlight the practical usefulness of the proposed state space methods.

C491: **Estimators for long range dependence: a simulation study**
Presenter: **Marco Reale**, University of Canterbury, New Zealand
Co-authors: William Rea, Les Oxley, Jennifer Brown

Results of a simulation study into the properties of 11 different estimators of the Hurst parameter, H, or the fractional integration parameter, d, in long memory time series are presented. We compare and contrast their performance on simulated Fractional Gaussian Noises and fractionally integrated series with lengths between 100 and 10,000 data points and H values between 0.55 and 0.90 or d values between 0.05 and 0.40. An application will be presented in the context of testing long memory vs structural breaks.

C658: **Multifractal height cross-correlation analysis: A new method for analyzing long-range cross-correlations**
Presenter: **Ladislav Kristoufek**, Charles University in Prague, Czech Republic

The new method for an analysis of long-range cross-correlations and multifractality – the multifractal height cross-correlation analysis is introduced. We show that the scaling of covariances of the absolute values of the series gives additional information about dynamics of two simultaneously recorded series and can cause divergence of the bivariate Hurst exponent from the average of the separate univariate Hurst exponents. A utility of the method is shown on several artificial series as well as real-world time series. We argue that even though the majority of the analyzed series (volume and volatility series of S&P500 and NASDAQ Composite indices, and returns and volatility series of WTI Crude Oil spot and futures prices) are cross-persistent, such cross-persistence is mainly caused by persistence of the separate processes and the fact that the series are correlated. The scaling of covariances of the absolute values of the examined processes is in good agreement with this result.

CS58 Room G16 CONTRIBUTIONS IN ECONOMETRICS AND FINANCIAL MARKETS Chair: Massimiliano Caporin

C162: **Competition and cascades in the financial markets: An agent-based network model of endogenous mergers**
Presenter: **Camillia Zedan**, University of Southampton, United Kingdom

There is an increasing trend towards global financial consolidation. Empirical evidence has shown that though consolidation can increase market efficiency, it can also increase systemic risk. Therefore, understanding the effects of company mergers on the financial markets is an increasingly pertinent issue. An agent-based model of endogenous merger formation in a simulated financial market, based on empirical analyses of UK merger data, is presented. We identify the conditions under which market competition is sufficiently disrupted to prompt cascading merger waves. We also identify methods of regulating such markets in order to minimise such destabilising effects whilst promoting efficient behaviour.

C896: **Modelling spillovers and measuring their persistence: Application to credit default swap premia**
Presenter: **Paola Donati**, European Central Bank, Germany

Before 2007, investors treated Euro area government bonds as essentially free of credit risk, irrespective of the actual fiscal stance of the issuer. Three years later, the market lost confidence in the sustainability of the public finances of Greece, Ireland and Portugal. Four years later, Belgium, Italy and Spain became affected. By the autumn of 2011, the cost of purchasing protection against sovereign default, as measured by credit default swap premia, rose to historical records for all euro area jurisdictions. Meanwhile, the perceived creditworthiness of euro area banks deteriorated severely. We investigate the propagation of the crisis using a multivariate frequency decomposition approach cast in the unobserved component framework to obtain, in real-time, a day-to-day quantitative evaluation of the extent to which the worsening (or improvement) in a sovereign CDS premium spreads to the CDS fees for other sovereigns and banks. The spectral decompositions are performed using a dynamic filter that exploits the time and frequency domain interpretation of the eigenvalues of state transition matrices of state space models employed. In doing so, we identify the long-lasting spillovers whose effect gets entrenched in the dynamics of the CDS premium of the contaminated country, or bank, to the point of steering its underlying trend.

C831: **Bias reduction in GARCH panels, with an analysis of hedge fund volatility**
Presenter: **Cavit Pakel**, University of Oxford, United Kingdom

It is well known that hundreds of observations are required to fit GARCH models. Focusing on the case of clustered GARCH dynamics, we employ a panel data approach to successfully estimate GARCH parameters in samples of around 150-200 time-series observations. In this setting, parameters that govern volatility dynamics are common across series, although long-run variances are potentially heterogeneous. In small samples, estimation of the common parameters is subject to the well-known incidental parameter bias. We use the integrated likelihood method to characterise this bias and obtain bias-reduced estimators. Simulations show that, even with as small as 150 time-series observations, a substantial portion of the bias is removed while attaining lower standard errors. This is in sharp contrast to traditional GARCH modelling. In a novel contribution to the literature, we utilise this approach to analyse volatility characteristics of hedge funds. As these panels are characteristically short, such analysis has hitherto been impossible under the GARCH framework. Results suggest that the sample distributions of hedge fund volatility display asymmetry and have large right tails. A second empirical exercise using stock market data shows that the integrated likelihood method achieves better forecasting performance compared to its alternatives.

C628: **Heterogeneous markets effects for asymmetric dynamic conditional correlation model with stock return and range**
Presenter: **Manabu Asai**, Soka University, Japan

The return and range model with dynamic conditional correlations (DCC) is considered. The new specifications for the asymmetric effects on log-volatilities and dynamic correlations, combined with the heterogeneous market effects, are suggested. The new DCC model can be estimated by the quasi-maximum likelihood method. The empirical analysis on Nikkei 225, Hang Seng and Straits Times indices shows the daily, weekly and monthly pattern of the asymmetric effects. The new DCC model also provides plausible one-step-ahead forecasts of the VaR thresholds for the period including the global Financial crisis.

CS73 Room B34 SHORT-TERM MACROECONOMIC FORECASTING: LESSONS FROM THE CRISIS Chair: Laurent Ferrara

C200: **Forecasting by factors, by variables, by both, or neither**
Presenter: **Jennifer Castle**, Oxford University, United Kingdom
Co-authors: Michael Clements, David Hendry

Forecasting US GDP and inflation over 1-, 4- and 8-step horizons is considered using a large dataset, with factors, variables, both, or neither. Model selection using Autometrics handles perfect collinearity and more regressors than observations, enabling all principal components (PCs) and variables to be included for model selection jointly with using impulse-indicator saturation (IIS) for multiple breaks. Results suggest that factor models are more useful for 1-step ahead forecasting than as the horizon increases, when selecting over variables tends to be better. Accounting for in-sample breaks and outliers using IIS is useful. Recursive updating helps, but recursive selection leads to greater variability, and neither outperforms autoregressions.

C146: **The return of non-linearity: Macroeconomic forecasting during the Great Recession**
Presenter: **Matteo Mogliani**, Banque de France, France
Co-authors: Laurent Ferrara, Massimiliano Marcellino

The debate on the forecasting ability of non-linear models has a long history, and the Great Recession provides us with an interesting opportunity for a re-assessment of the forecasting performance of several classes of non-linear models, widely used in applied research. We conduct an extensive analysis over a large quarterly database consisting of major real, nominal and financial variables for a large panel of OECD member countries. Although on average there are no systematic gains from non-linear models, during the Great Recession in general they largely outperformed standard linear specifications. This is particularly evident for some variables, countries and non-linear classes.

C091: **The possible shapes of recoveries in Markov-switching models**
Presenter: **Laurent Ferrara**, Banque de France, France
Co-authors: Frederique Bec, Othman Bouabdallah

The various shapes the recoveries may exhibit within a Markov-switching model are explored. It relies on the bounce-back effects and extends the existing methodology by proposing (i) a more flexible bounce-back model, (ii) explicit tests to select the appropriate bounce-back function, if any, and (iii) a suitable measure of the permanent impact of recessions. This approach is then applied to post-WWII quarterly growth rates of US, UK and French real GDPs.

CS74 Room B36 COMPUTER INTENSIVE METHODS IN ECONOMETRICS Chair: Oliver Scaillet

C180: **Bayesian semiparametric multivariate GARCH modeling**
Presenter: **John Maheu**, University of Toronto, Canada
Co-authors: Mark Jensen

A Bayesian nonparametric modeling approach for the return distribution in multivariate GARCH models is proposed. In contrast to the parametric literature, the return distribution can display general forms of asymmetry and thick tails. An infinite mixture of multivariate normals is given a flexible Dirichlet process prior. The GARCH functional form enters into each of the components of this mixture. We discuss conjugate methods that allow for scale mixtures and nonconjugate methods which provide mixing over both the location and scale of the normal components. MCMC methods are introduced for posterior simulation and computation of the predictive density. Density forecasts with comparisons to GARCH models with Student-t innovations demonstrate the gains from our flexible modeling approach.

C194: **Time-varying risk premium in large cross-sectional equity datasets**
Presenter: **Olivier Scaillet**, University of Geneva and Swiss Finance Institute, Switzerland
Co-authors: Patrick Gagliardini, Elisa Ossola

An econometric methodology is developed to infer the path of risk premia from large unbalanced panel of individual stock returns. We estimate the time-varying risk premia implied by conditional linear asset pricing models through simple weighted two-pass cross-sectional regressions, and show consistency and asymptotic normality under increasing cross-sectional and time series dimensions. We address consistent estimation of the asymptotic variance, and testing for asset pricing restrictions induced by large economies. The empirical illustration on returns for over twenty thousands US stocks from July 1963 to December 2009 shows that conditional risk premia are large and volatile in crisis periods. They exhibit large positive and negative strays from standard unconditional estimates and follow the macroeconomic cycles. The asset pricing restrictions are rejected for the usual unconditional four-factor model capturing market, size, value and momentum effects but not for a conditional version using instruments common to all assets and asset specific instruments.

C252: **Neural network sieve bootstrap for nonlinear time series analysis**
Presenter: **Michele La Rocca**, University of Salerno, Italy
Co-authors: Francesco Giordano, Cira Perna

A sieve bootstrap scheme, suitable for nonlinear processes, which uses feedforward neural networks as sieve approximators is discussed. Usage of this class of models appears to be justified for several reasons. Firstly, neural networks are popular models in nonlinear time series analysis for their good forecasting accuracy. Secondly, neural networks do not suffer much from the so-called curse of dimensionality and extension to high-dimensional models is expected to be quite straightforward. Thirdly, estimation of smoothing parameters (basically the hidden layer size) seems to be less critical than estimation of the window size in local nonparametric approaches. Finally, under quite general conditions, this class of models provides an arbitrarily accurate approximation to an unknown target function of interest. Like other sieve bootstrap schemes, NN-Sieve bootstrap retains the simplicity of the classical residual bootstrap and it is robust to model misspecification (being nonparametric in its spirit). Monte Carlo studies show that the NN-Sieve bootstrap has comparable performances to the AR-sieve bootstrap, when the process is linear, but it clearly delivers better results when the process is nonlinear, where it outperforms the AR- sieve bootstrap, the moving block bootstrap (MBB) and the kernel based bootstrap.

Authors Index

Aastveit, K., 27
Ababneh, A., 12
Abadir, K., 56, 74
Abdallah, F., 112
Abellan, J., 18
Abril, J., 117
Abril, M., 117
Acatrinei, M., 31
Acedanski, J., 10
Acosta-Gonzalez, E., 118
Addo, P., 105
Affenzeller, M., 91
Afonso Rodriguez, J., 44
Agostinelli, C., 17
Aguiar-Conraria, L., 93
Aguilera, A., 18
Ahlgren, N., 44
Ahmed, E., 37
Ahn, S., 103
Aielli, G., 81
Airola, A., 90
Akdogan, K., 117
Al-Saleh, M., 12
Alba-Fernandez, V., 20
Albarran, I., 60
Albers, C., 128
Aldea Mendes, D., 2, 9
Alessi, L., 10
Alfons, A., 58
Ali, F., 38
Almeida, R., 114
Alonso, A., 32
Alonso, J., 129
Alonso, P., 60
Alotaibi, R., 20
Alquier, P., 15, 72
Alvarez-Iglesias, A., 88
Alves, R., 33
Amado, C., 100
Amendola, A., 96
Amisano, G., 3
Amouh, T., 51
Anacleto-Junior, O., 128
Anatolyev, S., 5
Andersen, P., 77
Andersen, T., 30
Anderson, B., 131
Ando, T., 82
Andreou, P., 68, 109
Andres, K., 107
Andrews, J., 15
Andrievskaya, I., 119
Ansari, M., 11
Ansejo, U., 97
Antoine, B., 106
Antonucci, A., 127
Antypas, A., 129
Anyfantaki, S., 9
Arago, V., 3
Araujo Santos, P., 107
Aria, M., 78
Arias, M., 113
Arima, S., 114
Armas-Herrera, R., 118
Armero, C., 32
Arostegui, I., 16
Arratia, A., 87, 113
Arteche, J., 56
Arvanitis, S., 29
Asai, M., 133
Asar, O., 75
Ascari, G., 65
Asgharian, H., 10
Ashton, J., 4
Aslanidis, N., 11, 54
Atanu, B., 75
Atsalakis, G., 101
Audrino, F., 43, 67, 68
Audzeyeva, A., 107
Augustin, T., 63
Aulbach, S., 34
Auray, S., 53
Avarucci, M., 9
Avdulaj, K., 71
Avendano, M., 62
Awe, O., 69
Ayouni, S., 102
Ayton, P., 5
Azzalini, A., 112
Azzimonti, L., 35, 125

Bagnato, L., 99
Baguelin, M., 14
Bailey, N., 109
Baillie, R., 24
Baker, R., 18
Balakrishnan, N., 62
Balasuriya, J., 5
Balter, J., 30
Bandt, C., 125
Banerjee, A., 28, 37
Bannour, M., 117
Barbiero, A., 20
Barigozzi, M., 6, 131
Barnett, A., 53
Barquinero, J., 15
Barranco-Chamorro, I., 20, 113
Barunik, J., 6, 29, 70, 71
Barunikova, M., 70
Basse Mama, H., 119
Basturk, N., 82, 114
Batmaz, I., 115
Baum, C., 4
Baumeister, C., 124
Bauwens, L., 43, 66
Bayer, V., 31
Bec, F., 94, 134
Becheri, I., 104
Bedowska-Sojka, B., 29
Bellini, T., 17
Belloni, A., 77
Belmonte, M., 41, 120
Belter, K., 119
Ben David, N., 40
Ben Jabeur, S., 117
Ben Jrad, A., 102
Ben Salem, M., 94
Ben Yaghlane, B., 127
Ben Yahia, S., 102
Bencivelli, L., 41
Berenguer-Rico, V., 96
Bergara, A., 97
Bernales, A., 24
Bernard, J., 106
Bernoth, K., 43
Beskos, A., 74
Beutner, E., 9
Bhansali, R., 50
Bialkowski, J., 42
Bianchi, D., 3
Biau, G., 72
Bickel, P., 15
Biedermann, S., 72
Biggeri, A., 86
Billio, M., 3, 69, 85, 106
Biswas, A., 12
Blaess, V., 62
Blanco-Fernandez, A., 20
Bleakley, K., 73
Boateng, A., 107
Bocci, C., 62
Boente, G., 47
Boero, G., 41
Bogdan, M., 80
Bognar, M., 74
Bohl, M., 42
Boj, E., 59
Boldea, O., 106
Bordes, L., 63
Borgy, V., 7
Borowski, M., 50
Borrotti, M., 13
Bouabdallah, O., 134
Bouaddi, M., 24
Boudt, K., 66, 86
Boukouvalas, A., 51
Bouzas, P., 101
Boyarchenko, S., 25
Braekers, R., 89
Brander, C., 32
Branzoli, N., 65
Brechmann, E., 30
Breto, C., 122
Briers, M., 105
Brito, M., 73, 115
Brito, P., 51
Broda, S., 5
Brooks, C., 118
Brown, J., 132
Browne, R., 60
Brownlees, C., 81
Bruckner, M., 54
Bruffaerts, C., 58
Buchner, B., 2
Buikis, M., 8
Bujosa, M., 65
Buncic, D., 40
Buono, D., 26
Burns, P., 119

Caballero-Aguila, R., 22
Cabana, A., 87
Cabana, E., 87
Cabras, S., 32
Cadarso-Suarez, C., 15, 16, 18, 33
Cai, S., 83
Cairns, K., 112
Cakmakli, C., 97
Cales, L., 106
Calle, M., 63
Calzolari, G., 81
Camacho, M., 59
Campos, S., 32
Candelon, B., 86
Cano, A., 19
Cantore, C., 65
Caporin, M., 84, 104
Caraiani, P., 31
Cardot, H., 35
Carfora, M., 124
Carissimo, A., 31
Carkovs, J., 8
Carlos, A., 6
Carnero, A., 33
Caron, F., 99
Caroni, C., 101
Carroll, R., 77
Carton, B., 106
Carvalho, C., 3
Casado, D., 32
Casalin, F., 81
Casarin, R., 3, 27
Castellanos, M., 32
Castelnuovo, E., 65
Castle, J., 133
Catani, P., 44
Catelan, D., 86
Cattaneo, M., 63
Cavalcante, L., 115
Cebrian, A., 100
Cerioli, A., 47
Chadha, J., 43
Chadwick, M., 8, 117
Chaganty, N., 48
Chan, J., 10, 41
Chang, S., 16
Chaouch, M., 35
Charemza, W., 25
Chatzikonstanti, V., 11
Chauveau, D., 63
Chen, C., 129
Chen, R., 16
Chen, W., 30, 131
Chernozhukov, V., 77
Chevalier, C., 61, 115
Chevallier, J., 130
Choi, J., 86
Choirat, C., 19
Chopin, N., 57
Chou, R., 26
Choudhary, V., 8
Choy, B., 105
Christiansen, C., 54
Christmann, A., 46
Christofides, C., 67
Chronopoulou, A., 82
Chuang, H., 132
Chung, K., 116
Ciemny, M., 10
Cipollini, A., 54
Cirillo, P., 34
Claudia, C., 6

Clement, E., 83
Clements, M., 133
Climent, F., 9, 40
Coelho, C., 80
Coffey, N., 60
Cogley, T., 65
Colubi, A., 13, 19, 20, 111
Conde, D., 36
Conde, S., 61
Cont, R., 84
Conti, A., 131
Conversano, C., 89
Cook, R., 75
Coolen, F., 18, 38
Coolen-Maturi, T., 38
Coppi, R., 78
Corbellini, A., 88
Coretto, P., 17
Cormann, U., 34
Cornea, A., 74
Cornelissen, J., 66, 86
Corradini, R., 70
Corsi, F., 43, 57, 93
Cortese, G., 77
Cosma, A., 118
Costa, T., 59
Costantini, M., 28
Costola, M., 84
Cotton, C., 75
Coulibaly, D., 106
Cozza, V., 78
Craig, S., 112
Creel, M., 94
Crespo Cuaresma, J., 67
Critchley, F., 100
Crossman, R., 18
Croux, C., 58, 66, 86
Cuerden, M., 75
Cuerpo, C., 66
Cumming, D., 118
Cutillo, L., 31, 124
Czado, C., 30, 47, 96

D'Ambrosio, A., 78
Dafni, U., 14
Daher, A., 112
Dahlhaus, R., 50
Dalla Valle, L., 3
Daouia, A., 126
Darolles, S., 68, 83, 84
Das, S., 127
Dash, G., 9, 91
Datta, S., 76
Davidov, O., 36
De Baets, B., 90, 111
De Battisti, F., 21
de Bruijn, B., 28
de Cooman, G., 38
de Haan, J., 24
de Innocentis, M., 25
De Iorio, M., 125
de Juan, A., 65
de Knijff, P., 24
de la Rosa de Saa, S., 78
De March, D., 13
De Rock, B., 58
de Rosa, R., 127
de Sousa, B., 33
de Una-Alvarez, J., 76, 88
De Waal, A., 71
De Waegenaere, A., 14
Dehling, H., 91, 101
Dehon, C., 58
Deistler, M., 131
del Barrio, E., 47
del Castillo, J., 11
Delle Monache, D., 65
Demos, A., 9, 29
Denoeux, T., 111, 112
Destercke, S., 19, 63
Detemple, J., 120
Dette, H., 48, 77
Dewachter, H., 55
Dhorne, T., 112
Di Jorio, L., 102
Dias, A., 121
Dias, S., 51
Diaz, I., 90
Diaz, S., 63
Didelez, V., 21
Dienstbier, J., 115
Diks, C., 1
Ding, S., 67
Dionisio, A., 80
Diop, A., 99
Distaso, W., 56
Doan, M., 69
Dobrev, D., 30
Domanin, M., 125
Donati, P., 133
Doncel, L., 85
Dong, A., 10
Dongmo Jiongo, V., 62
Donohue, M., 63
Dooley, C., 128
Dordonnat, V., 3
Dorta Guerra, R., 79
Doucet, A., 105
Drechsel, K., 95
Drost, F., 104
Duarte, E., 33
Dubecq, S., 7
Dubey, A., 93
Dubois, D., 78
Duchesne, P., 62
Dudek, A., 102
Dudek, J., 84
Dueker, M., 119
Duerre, A., 101
Dufays, A., 66
Dugundji, E., 7
Dumitrescu, S., 31
Dungey, M., 68
Dunson, D., 37, 75, 99
Dupuy, J., 99
Durairajan, T., 123
Durante, F., 48
Dureau, J., 14
Durham, G., 41
Durrant, B., 73
Dusseldorp, E., 89
Dyckerhoff, R., 101

Eckley, I., 38
Eicher, T., 67
Eichler, M., 2, 86
El Asri, M., 58
Elsherif, M., 70
Emmanuelle, J., 68
Engsted, T., 42
Epprecht, C., 97
Erdemlioglu, D., 29
Erdogan, B., 43
Erlwein, C., 33
Ernst, E., 55
Escabias, M., 18
Escanciano, J., 34
Estebanez-Gallo, D., 36
Esteve, A., 59
Etxeberria, J., 12, 86
Eulert, M., 30
Everaert, G., 98
Exterkate, P., 86
Eyquem, A., 53

Fabrizi, E., 12
Fagan, G., 3
Fahmi, Y., 117
Fairchild, R., 4
Falk, M., 34
Farah, M., 116
Farcomeni, A., 114
Farrow, M., 20
Favaro, S., 76, 125
Febrero-Bande, M., 18, 32
Fei, F., 40
Feldkircher, M., 67, 97
Felsenstein, E., 131
Fengler, M., 97
Fenske, N., 48
Ferger, D., 91
Fermanian, J., 121
Fernandez, A., 19
Fernandez, M., 36
Fernandez-Macho, J., 118, 120
Fernandez-Perez, A., 132
Fernandez-Rodriguez, F., 118
Fernandez-Sanchez, M., 21
Fernandez-Val, I., 77
Ferrara, L., 133, 134
Ferrari, D., 13
Ferrari, P., 20
Ferraro, M., 13
Ferroni, F., 65
Feunou, B., 108
Figini, S., 31
Figueiredo, A., 60
Figueiredo, F., 14, 73
Filipova, K., 10
Fink, S., 91
Fiorentini, G., 81
Flournoy, N., 16
Fondeur, Y., 94
Fontaine, J., 7, 108
Forbes, C., 42
Forlin, M., 13
Fortiana, J., 59
Foschi, P., 97
Fox, E., 99
Fragiadakis, K., 23
Franceschini, C., 81
Francq, C., 25, 83, 104
Franses, P., 28
Frattarolo, L., 69
Freitas, A., 115
Fried, R., 50, 58, 62, 90, 101
Froger, J., 3
Fronczyk, K., 75
Fruehwirth, R., 33
Fruehwirth-Schnatter, S., 53, 57, 121
Fruth, J., 61
Fryzlewicz, P., 66
Fuentes-Garcia, R., 37, 76
Fuertes, A., 40, 121, 132
Fukumizu, K., 47
Funovits, B., 131
Furio, M., 9
Fusari, N., 93

Gabriel, C., 25
Gabrys, R., 33
Gaddah, A., 89
Gagliardini, P., 84, 134
Galeano, P., 32
Gallegati, M., 55
Galli, F., 118
Gamez, J., 19
Gamiz Perez, M., 116
Garcia Alonso, C., 21
Garcia, A., 60
Garcia-Barzana, M., 19
Garcia-Escudero, L., 47, 88, 111
Garcia-Ferrer, A., 65, 66
Garcia-Ligero, M., 22
Garcia-Martos, C., 59
Garcia-Perez, A., 47
Gareth, R., 74
Garratt, A., 81
Gastwirth, J., 113
Gatarek, L., 82
Gatto, R., 27
Gatumel, M., 123
Gautier, E., 72
Gefang, D., 121
Gel, Y., 15
Gelenbe, E., 103
Gelper, S., 58, 66
Genton, M., 127
Geppert, L., 21
Gerdrup, K., 27
Gerds, T., 77
Gerlach, R., 105, 129
Geweke, J., 41, 53
Ghahramani, Z., 99
Giannikis, D., 56
Giannone, D., 28, 53
Gil, M., 13, 78, 111
Gilmour, S., 72
Ginebra, J., 79
Ginsbourger, D., 61, 115
Giordano, F., 17, 134
Giraitis, L., 56
Glombek, K., 126
Gloter, A., 83

Gneiting, T., 41
Goegebeur, Y., 34
Goga, C., 35
Goicoa, T., 12, 86
Golightly, A., 74
Golosnoy, V., 42
Gombay, E., 37
Gomes, I., 73
Gomes, M., 14, 114
Gomez Mateu, M., 14
Gomez Melis, G., 14
Gomez, G., 32, 63
Gomez-Olmedo, M., 19
Gonzalez Davila, E., 79
Gonzalez, J., 32
Gonzalez, M., 23
Gonzalez-Davila, E., 70, 79
Gonzalez-Manteiga, W., 18
Gonzalez-Prieto, E., 66
Gonzalez-Rodriguez, G., 13, 20
Gonzalez-Yanes, A., 79
Gonzalo, J., 96
Goos, P., 72
Gordaliza, A., 47, 88, 111
Gorka, J., 69
Gosling, J., 51
Gourieroux, C., 7, 84
Grane, A., 60
Grassi, S., 104, 132
Grau-Carles, P., 85
Greco, F., 12
Greco, L., 80
Greselin, F., 99, 113
Gretton, A., 47
Gribisch, B., 42
Gribkova, S., 89
Griffin, J., 37
Groenen, P., 58
Grossmann, H., 72
Grothe, O., 93
Grubinger, T., 89
Gruen, B., 49, 67
Guan, B., 8
Gude, F., 16, 18
Guedj, B., 72
Guegan, D., 26, 105, 106
Guha, A., 75
Guidolin, M., 24
Guillen-Gosalbez, G., 33
Guillou, A., 34
Gulyas, L., 7
Gunter, U., 28
Gutierrez, C., 23

Hable, R., 46
Hadjiantoni, S., 116
Haerdle, W., 112
Haesbroeck, G., 126
Haghighi, M., 11
Halbleib, R., 43
Hall, S., 4
Hallin, M., 77
Halvorsen, J., 41
Han, A., 20
Han, S., 123
Hans, C., 37
Harari, O., 51
Harvey, A., 1, 59, 65, 107
Hass, L., 118
Hassani, B., 26
Haut, R., 63
Hautsch, N., 119
Haziza, D., 62
Heaney, R., 69
Hearne, L., 126
Hecq, A., 96
Heiden, M., 122
Henderson, D., 8
Henderson, R., 20
Hendry, D., 133
Hennig, C., 17
Henze, N., 87
Henzel, S., 122
Hermoso-Carazo, A., 22
Hernandez-Bastida, A., 21
Herve-Mignucci, M., 2
Hess, W., 10
Heuchenne, C., 77, 125
Hin, L., 97
Hinde, J., 60, 88, 128
Hizu, T., 108
Hlavka, Z., 87
Hoermann, S., 33
Hofert, M., 48
Hofmarcher, P., 67
Hollo, D., 69
Holzmann, H., 30
Hong, Y., 20
Hoogerheide, L., 24, 82
Hornik, K., 49
Hothorn, T., 48, 89
Huang, N., 26
Hudson, B., 105
Hudson, R., 4
Huellermeier, E., 78
Huesler, J., 34
Huh, J., 14
Hulliger, B., 62
Huskova, M., 87
Hussein, A., 38

Iacus, S., 91
Iania, L., 55
Iannino, M., 5
Ichino, M., 51
Ickstadt, K., 125
Ielpo, F., 7, 123, 130
Ilk, O., 75
Ilmonen, P., 46
Imoto, S., 103
Inacio, V., 18
Infante, E., 26
Ingrassia, S., 88, 99
Ippoliti, L., 50
Ishida, I., 107
Ishwaran, H., 15
Issler, J., 96

Jablonsky, P., 9
Jach, A., 50
Jacob Leal, S., 124
Jacob, P., 57
Jacobi, L., 121
Jakaite, L., 10
Jang, T., 42
Jannin, G., 84, 85
Janusevskis, J., 61
Jaruskova, D., 114
Javed, F., 66
Jeleskovic, V., 119
Jensen, M., 134
Jimenez, L., 33
Jimenez-Gamero, M., 20, 35, 113
Jimenez-Martin, J., 107
Jochmann, M., 81
Joe, H., 48
Johansen, S., 127
Jolani, S., 115
Jondeau, E., 84
Jones, E., 21
Jore, A., 27
Jouad, F., 26
Jouneau-Sion, F., 53
Jourdan, A., 60
Juneja, J., 131
Jung, R., 44
Jurczenko, E., 84
Jureckova, J., 77

Kaban, A., 73
Kagkadis, A., 68
Kajiji, N., 9, 91
Kalli, M., 37
Kalogeropoulos, K., 14, 74
Kalotychou, E., 40
Kan, B., 22, 80
Kanj, S., 112
Kapetanios, G., 24, 106, 109
Karagrigoriou, A., 35
Karame, F., 94
Karlis, D., 99
Karlsen, H., 44
Karlsson, S., 67
Karyampas, D., 29
Kascha, C., 27
Kastner, G., 57
Kaszuba, B., 21
Kato, K., 132
Katzschner, T., 44
Kaymak, U., 114
Kazani, A., 111
Keblowski, P., 44
Kehrle, K., 5, 31
Kelly, D., 38
Khalaf, L., 106
Khalili, A., 15
Khovansky, S., 7
Kichian, M., 106
Killick, R., 38
Kim, D., 82
Kim, G., 126
Kim, H., 103
Kim, J., 102
Kim, Y., 126
Kingeski Galimberti, J., 120
Kirch, C., 87
Kleiber, C., 110
Kley, T., 77
Klimova, A., 46
Kloessner, S., 118
Kneib, T., 16, 33, 37, 48, 116
Kneip, A., 18
Knight, K., 114
Kobayashi, G., 108
Koch, D., 126
Koch, N., 54
Koenker, R., 48
Kokoszka, P., 33
Komarek, A., 49
Kontoghiorghes, E., 19, 116
Koo, B., 123
Koop, G., 41
Koopman, S., 68
Korobilis, D., 41
Kosmidis, I., 99
Kottas, A., 75, 116
Kourogenis, N., 94, 129
Kozumi, H., 108
Kraemer, W., 91
Kraus, D., 18
Krause, J., 5
Kremer, M., 69
Krieg, S., 62
Kriete-Dodds, S., 131
Krippner, L., 82
Kristensen, D., 81
Kristoufek, L., 133
Kronberger, G., 91
Kuan, C., 132
Kuhnt, S., 61
Kukuk, M., 31
Kunst, R., 28
Kuo, J., 67
Kvedaras, V., 95

La Rocca, M., 134
La Vecchia, D., 93
Lamari, W., 127
Langohr, K., 63
Laniado, H., 58
Larocque, D., 24
Lau, C., 25
Laubach, T., 7
Laurent, A., 102
Laurent, S., 29, 86, 130
Lavancier, F., 56
Lawrance, A., 123
Lawson, A., 86
Lazim, M., 98
Le Fol, G., 83, 84
Le Riche, R., 61
Lechevallier, Y., 51
Lecourt, C., 130
Ledford, A., 87
Lee, D., 87, 88
Lee, J., 103, 126
Lee, K., 4
Lee, S., 102
Lee, T., 123
Legrenzi, G., 108
Leipus, R., 56
Leisen, F., 3
Lenza, M., 28, 53
Leon-Ledesma, M., 65
Lepage, G., 83
Leton, E., 78

Levendorskii, S., 25
Levis, M., 4
Ley, E., 53
Li, D., 26
Li, G., 8
Li, W., 8
Liang, C., 131
Liboschik, T., 90
Liesenfeld, R., 42, 95
Lieshout, M., 101
Lijoi, A., 76, 99, 125
Lillo, R., 32, 36, 58
Lin, L., 129
Lin, S., 15
Lin, Y., 46
Linares-Perez, J., 22
Linton, O., 123
Lippi, M., 131
Lisi, F., 84
Liski, E., 100
Liu, G., 117
Liu, L., 10
Liu, S., 44
Liu, Y., 87
lo Cascio, I., 54
Lo Duca, M., 69
Loecker, F., 46
Lof, M., 109
Lombardia, M., 12
Lonnbark, C., 25
Loperfido, N., 81
Lopez, M., 78
Lopez, O., 89
Lopez-Fidalgo, J., 32, 79
Lopez-Pintado, S., 32
Lopez-Vizcaino, E., 12
Louca, C., 109
Louzis, D., 93
Lu, Z., 130
Luati, A., 2, 59
Lubiano, M., 78
Lucas, A., 68
Luciani, M., 6, 68, 131
Luger, R., 106
Lui, S., 95
Lunardon, N., 80
Lunde, A., 43
Lupi, C., 109
Lupu, R., 31
Lux, T., 42
Lyrio, M., 55

Ma, Y., 77
Mabrouk, A., 70
MacDonald, A., 87
MacEachern, S., 37, 99
Machado, L., 76
MacKenzie, G., 61, 62
Maechler, M., 48
Maharaj, E., 44
Maheu, J., 40, 134
Maillet, B., 84, 85
Maiti, R., 12
Makarova, S., 25
Malina, M., 80
Mamatzakis, E., 11
Mancino, M., 83
Mandal, S., 12
Manner, H., 6
Manzi, G., 20
Marazzi, A., 17
Marcellino, M., 28, 41, 106, 129, 133
Marin, J., 81
Marin-Sanguino, A., 33
Maringer, D., 131
Marques, F., 80
Marra, G., 116
Marsala, C., 102
Marshall, A., 112
Martin, G., 42
Martin, R., 50
Martin-Barragan, B., 36
Martinez, O., 11, 54
Martinez, R., 23
Martinez-Miranda, M., 116
Martinez-Moreno, J., 90
Marumo, K., 122
Maruri Aguilar, H., 51
Masegosa, A., 18, 19
Masten, I., 28
Mastromarco, C., 81
Matilla-Garcia, M., 94
Matran, C., 88, 111
Matthes, C., 65
Mayo Iscar, A., 88
Mayo-Iscar, A., 47, 111
Mayr, A., 48
Mayr, J., 122
Mazzi, G., 27, 29
McAleer, M., 107
McAlinn, K., 108
McCabe, B., 42
McCall, E., 112
McCrorie, R., 131
McElroy, T., 50
McNeil, A., 48
McNicholas, P., 15, 60
Medda, F., 132
Meerschaert, M., 25
Meier, P., 68
Meintanis, S., 23, 87
Meitz, M., 117
Melenberg, B., 14
Meligkotsidou, L., 55, 56
Mena, R., 76, 125
Menardi, G., 112
Mendes, V., 2, 9
Merika, A., 25
Merikas, A., 25
Mero, G., 83
Mesonnier, J., 7
Meyer, M., 36
Micale, V., 2
Michalopoulou, K., 111
Miffre, J., 132
Milas, C., 108
Militino, A., 12, 86
Millo, G., 110
Minotti, S., 88, 113
Mira, A., 57, 121
Miranda, E., 19, 63
Mirkov, N., 82
Mise, E., 2
Mitchell, H., 69
Mitchell, J., 81, 95
Mittnik, S., 55, 104, 107
Moauro, F., 27, 29
Mogliani, M., 133
Mohammad-Djafari, A., 80
Mohammadi, A., 100
Mohnot, R., 122
Molanes Lopez, E., 78
Molina Peralta, I., 62
Molina, M., 22
Momferatou, D., 28
Momparler, A., 40
Monfort, A., 6
Montes, I., 63
Monticini, A., 27
Moral, S., 19
Moral-Benito, E., 67
Morales, D., 12
Morana, C., 28
Moreira, A., 76
Moreira, C., 88
Moreno, J., 90
Moretti, G., 41
Morina, D., 75
Moro, R., 112
Morris, R., 116
Moser, M., 98
Mosler, K., 101
Mota, M., 22, 23
Motta, G., 6, 131
Moulines, E., 72
Moura, M., 120
Mpesmpeas, P., 128
Muehlenstaedt, T., 51
Mueller, U., 89
Mueller-Kademann, C., 43
Mukherjee, K., 126
Muniz, G., 49
Munoz, A., 32
Muradoglu, G., 4, 5
Murteira, J., 118
Mutl, J., 28
Muzzioli, S., 54

Nagata, S., 122
Nakajima, J., 108
Nakama, T., 13
Nakatsuma, T., 108
Naraidoo, R., 107
Neely, C., 29
Negri, I., 29
Neves, M., 73
Newell, J., 88, 128
Ng, J., 42
Nielsen, B., 42, 127
Nielsen, J., 19, 116
Nieto-Reyes, A., 36
Nikoloulopoulos, A., 48
Nipoti, B., 99
Nishihara, M., 124
Nishiyama, Y., 29
Nneji, O., 118
Noh, H., 116
Noirhomme, M., 51
Nolan, J., 25
Nordhausen, K., 100
Noriega, A., 117
Normandin, M., 24
Noureldin, D., 131
Novak, S., 69
Nunez-Anton, V., 16, 74

Ocal, N., 40
Odejar, M., 112
Oesch, C., 91
Oh, H., 126
Oja, H., 58, 100, 101
Ojeda Cabrera, J., 35
Okada, K., 114
Okubo, T., 80
Oliveira, M., 80
Oller, J., 16
Olmo, J., 3, 121
Olsen, E., 18
Omer, M., 24
Omori, Y., 108
Onorante, L., 10, 28
Orlando, A., 124
Ormeno, A., 65
Ortega, E., 129
Ortega, J., 11
Oruezabal, M., 32
Osiewalski, J., 70, 96
Osiewalski, K., 70, 96
Ossola, E., 134
Ouysse, R., 97
Owyang, M., 108
Oxley, L., 88, 132
Oya, K., 107
Oztek, M., 40

Packalen, M., 8
Padberg, M., 95
Pahikkala, T., 90
Paiardini, P., 29
Paindaveine, D., 46
Pakel, C., 133
Palm, F., 130
Pan, Q., 113
Panaretos, V., 18
Panayidou, K., 116
Panchenko, V., 1
Pandey, S., 8
Panopoulou, E., 4, 44, 55
Pantelidis, T., 44
Paolella, M., 5, 31
Papageorgiou, C., 8, 67
Papageorgiou, I., 79
Papaspiliopoulos, O., 57, 74, 120
Pappa, E., 54
Pardo Fernandez, J., 34
Pardo, M., 62
Pardo-Fernandez, J., 125
Parent, I., 3
Park, B., 126
Park, T., 103
Parmeter, C., 8
Patrick, D., 68
Paul, F., 74
Paustian, M., 53
Pavlidis, E., 42, 54, 117
Pavlidis, N., 54

Paya, I., 42, 117
Payne, K., 112
Pazos, E., 74
Peddada, S., 36
Peel, D., 42, 117
Peersman, G., 124
Pelenis, J., 94
Pelizzon, L., 69, 85
Pelka, M., 102
Peluso, S., 43, 57
Pena, D., 66
Penaranda, F., 24
Peng, D., 62
Peng, L., 74
Penikas, H., 119
Pereira, M., 97
Perez Amaral, T., 107
Perez Garrido, B., 62
Perez-Alvarez, S., 32
Perez-Espartero, A., 33
Perez-Quiros, G., 59
Perna, C., 134
Perra, S., 32
Perrotta, D., 88
Pesaran, H., 109
Pestana, D., 114
Peters, G., 10, 105, 129
Petralia, F., 75
Petrone, S., 76
Pfeiffer, K., 89
Phamhi, D., 31
Philip, D., 68
Philippe, A., 56
Picek, J., 115
Picheny, V., 115
Piras, G., 110
Pires, A., 100
Pitt, M., 120
Pittis, N., 94
Plastira, S., 4
Polak, P., 5
Polanco-Martinez, J., 120
Poli, I., 13
Pollock, S., 2
Polymenis, A., 70
Poncela, P., 2, 59, 66
Poncelet, P., 102
Porqueddu, M., 129
Pougaza, D., 80
Pourahmadi, M., 49
Pourroy, M., 106
Pozo, C., 33
Praskova, Z., 50
Prastyo, D., 112
Prata Gomes, D., 73
Preda, C., 18
Primicer, G., 53
Proietti, T., 2, 27
Prono, T., 69
Provost, S., 128
Pruenster, I., 76, 125
Puig, P., 15, 75
Pujol, O., 59
Punzo, A., 88, 99, 113
Putintseva, M., 5

Queen, C., 128
Quiros, A., 32
Quost, B., 111

Rackauskas, A., 57
Radice, R., 116
Rahbek, A., 104
Rahoui, G., 26
Rakkestad, K., 98
Raknerud, A., 98
Rakotomarolahy, P., 105
Ramalho, E., 10, 11, 118
Ramalho, J., 10, 11, 118
Ramasso, E., 127
Ramos Guajardo, A., 13
Ramos, A., 22
Ramprasath, L., 123
Ramsay, J., 35
Ramsey, J., 55
Ranaldo, A., 104
Ranjan, R., 41
Rao, J., 15
Ravazzolo, F., 3, 27
Rayner, J., 87
Rea, W., 132
Reale, M., 88, 132
Refenes, A., 93
Reiss, R., 34
Renault, M., 3
Renne, J., 6, 7
Restaino, M., 96
Reverter, F., 16
Riani, M., 47, 111
Ricci, L., 6
Richard, P., 96
Richet, Y., 115
Rifqi, M., 102
Rigby, R., 49
Rindisbacher, M., 120
Rittler, D., 2
Ritz, C., 18
Rivas-Lopez, M., 79
Robinson, P., 1
Robinzonov, N., 104
Roca-Pardinas, J., 15
Rockinger, M., 84
Rodero Cosano, M., 21
Rodrigues, V., 33
Rodriguez Alvarez, M., 15
Rodriguez, J., 59
Rodriguez-Diaz, J., 17
Rodriguez-Girondo, M., 16, 76
Rodriguez-Muniz, L., 90
Roldan, A., 90
Roldan, C., 90
Roman-Roman, P., 20
Rombaut, M., 127
Rombouts, J., 66, 109
Romo, J., 32, 36, 58
Rompolis, L., 109
Ronchetti, D., 68
Ronchetti, E., 127
Rosa-Gonzalez, F., 70
Rossi, E., 104, 122
Roustant, O., 50
Ruckdeschel, P., 33
Rudas, T., 46
Rueda, C., 36
Ruggeri Cannata, R., 29
Ruggiero, M., 76
Ruiz de Villa, M., 16
Ruiz, E., 2, 33
Ruiz-Fuentes, N., 101
Ruiz-Gazen, A., 100
Rumi, R., 19
Ruspini, E., 13
Ruwet, C., 47

Sadurskis, K., 8
Saikkonen, P., 117
Saint Pierre, P., 89
Sainz, J., 85
Salakoski, T., 90
Salehi-Rad, M., 100
Salini, S., 21
Salmeron, A., 19
Salvador, B., 36
Salvador, E., 3
Samworth, R., 73
Sanchez, M., 59
Sanchez-Pla, A., 16
Sanchez-Rubio, J., 32
Sanfelici, S., 83
Sangalli, L., 35, 125
Sanjel, D., 128
Sanso-Navarro, M., 3
Santos-Martin, M., 17
Santucci de Magistris, P., 104, 122, 132
Saporta, G., 18
Sarda, P., 18
Sarlin, P., 130
Sass, J., 94
Savin, I., 7
Savva, C., 55, 109
Sbordone, A., 65
Scaillet, O., 134
Scarrott, C., 87, 88
Schafer, M., 125
Schauer, M., 74
Schaumburg, E., 30
Schenk-Hoppe, K., 107
Schetinin, V., 10
Scheufele, R., 95
Scheuring, S., 114
Schmid, F., 93
Schnieders, J., 93
Schoch, T., 62
Schoelkopf, B., 47
Scholtens, B., 24
Schroeder, T., 132
Schuessler, R., 30
Schwaab, B., 68
Schweizer, D., 118
Secchi, P., 35, 125
Segers, J., 93
Sejdinovic, D., 47
Sekhposyan, T., 108
Selland Kleppe, T., 95
Semmler, W., 55
Sensini, L., 96
Sentana, E., 24
Seo, B., 123
Seri, R., 19
Sermaidis, G., 74
Serrat, C., 79
Sevi, B., 130
Sezer, A., 80
Sguera, C., 32
Shah, R., 73
Shang, H., 120
Shephard, N., 43, 131
Sheppard, K., 43, 131
Shevchenko, P., 105
Shi, Y., 67
Shibata, T., 124
Shin, Y., 81
Shirota, S., 108
Siciliano, R., 78
Siliverstovs, B., 95
Silva Filho, O., 119
Simon, C., 127
Sinova, B., 111
Sirichand, K., 4
Slanzi, D., 13
Slouma, M., 117
Smeekes, S., 98
Smeets, M., 62
Smith, J., 41
Smola, A., 47
Soares, M., 93
Sobotka, F., 48, 116
Soegner, L., 28, 94
Sokolinskiy, O., 1
Solaro, N., 20
Solgi, R., 121
Solibakke, P., 120
Som, A., 37
Song, L., 47
Song, Y., 40
Sorribas, A., 33
Sousa, I., 79
Sousa, J., 114
Spangl, B., 33
Spazzini, F., 122
Speekenbrink, M., 49
Sriperumbudur, B., 47
Starica, C., 119
Stasinopoulos, D., 49
Steel, M., 53
Stehlik, M., 16
Steinberg, D., 51
Steinebach, J., 50
Stella, F., 91
Stentoft, L., 109
Stephan, P., 42
Steude, S., 31
Stoeber, J., 96
Stogiannis, D., 101
Storti, G., 43
Straetmans, S., 86
Stramer, O., 74
Strauss, O., 63
Sucarrat, G., 81
Surgailis, D., 56
Symeonaki, M., 111

Taamouti, A., 108
Tachmazidou, I., 13
Tafin Djoko, D., 119
Tarrant, W., 26

Taskinen, S., 101
Tawn, J., 87
Tedongap, R., 108
Teh, Y., 99
Tenorio de Carvalho, F., 51
Tenreiro, C., 34
Teran, P., 90
Thas, O., 87
Theodossiou, P., 55
Thieler, A., 58, 62
Thorsrud, L., 27
Todorov, V., 17
Tokdar, S., 37
Tomassi, C., 17
Torres-Ruiz, F., 20
Torti, F., 47
Trabelsi, C., 102
Trandafir, P., 12
Trapani, L., 69
Triantafyllopoulos, K., 123
Trinca, L., 72
Trivisano, C., 12
Troffaes, M., 19, 38
Troiano, L., 90
Tsai, P., 72
Tsiotas, G., 105
Tsybakov, A., 72
Tuerk, D., 86
Turk, D., 2
Tworek, S., 80
Tyler, D., 100
Tzavalis, E., 56, 109
Tzavidis, N., 62

Uddin, M., 107
Ugarte, L., 12, 86
Urbain, J., 98
Urbina, J., 11

Vacha, L., 6, 29
Vaello-Sebastia, A., 97
Vahey, S., 81
Vaida, F., 63
Valero, J., 75
Van Aelst, S., 111
Van Bellegem, S., 126
Van Camp, A., 38
van den Burg, G., 58
Van den Hout, A., 49
van der Meulen, F., 74
van Dijk, D., 1
van Dijk, H., 3, 24, 82
van Dijk, K., 24
Van Eyden, R., 71
Van Keilegom, I., 34, 77, 116
Van Mechelen, I., 89
van Zanten, H., 74
Vasileva, K., 4
Vatne, B., 98
Veerhuis, P., 129
Vegas, E., 16
Veiga, A., 97
Veiga, H., 122
Velo, G., 93, 104
Venditti, F., 129
Venetis, I., 11
Ventosa-Santaularia, D., 117
Ventura, L., 80
Veredas, D., 6, 68
Vermorken, M., 132
Vert, J., 73
Verwaeren, J., 111
Vicente, J., 80
Viens, F., 82
Villa, S., 91
Violante, F., 109
Visser, I., 49
Voev, V., 43
Vogel, D., 101
Vogelsang, T., 28
Volgushev, S., 48
Vollmer, M., 125
von Sachs, R., 66
Vonta, I., 35
Vosvrda, M., 29
Vrontos, I., 55, 56
Vrontos, J., 56
Vrontos, S., 55, 56

Wade, S., 76
Waegeman, W., 90, 111
Wagalath, L., 84
Wagner, H., 121
Wagner, M., 28
Wagner, S., 118
Walden, A., 49
Waldmann, E., 37
Walker, S., 1, 76
Wallis, K., 41
Walter, G., 38
Wan Yaacob, W., 98
Wang, J., 36, 100, 105
Wang, S., 20
Wang, W., 16
Wang, Y., 130
Ward, C., 118
Waters, A., 43
Wei, Y., 77
Weiss, G., 95
Welfe, A., 44
Werker, B., 104
Wessels, R., 3
West, M., 108
Wichitaksorn, N., 105
Wieczorek, J., 125
Wied, D., 91
Wiencierz, A., 63
Wild, C., 46
Wilkinson, D., 74
Williamson, S., 99
Wintenberger, O., 83
Wiper, M., 59
Wisniewski, T., 42
Wit, E., 100
Woerner, J., 83
Wohlrabe, K., 104
Woitek, U., 81
Wolff, R., 122
Wong, W., 16
Wu, C., 30
Wuthrich, M., 10
Wynn, H., 17, 61

Xanthopoulos, S., 93
Xia, X., 130
Xiao, H., 130
Xie, M., 73
Xu, J., 61
Xu, R., 63
Xu, W., 113
Xuriguera, R., 113

Yamamoto, Y., 82
Yang, M., 72
Yang, Y., 14
Yap, B., 98
Yazici, B., 22, 80
Yazici, C., 115
Yeh, J., 132
Yelou, C., 106
Yerlikaya-Ozkurt, F., 115
Yohai, V., 17
Young, A., 1
Youssef, N., 17, 61
Yu, K., 129
Yun, M., 132

Zaffaroni, P., 9
Zahradnik, P., 120
Zakamulin, V., 130
Zakoian, J., 25, 83, 104
Zdenek, M., 41
Zedan, C., 133
Zeggini, E., 13
Zeileis, A., 89
Zellner, A., 82
Zemlys, V., 95
Zerhouni, N., 127
Zerilli, P., 4
Zeugner, S., 67, 97
Zhang, D., 46
Zhang, M., 46
Zhang, X., 120
Zhao, X., 26, 88
Zhao, Y., 130
Zhelonkin, M., 127
Zhigljavsky, A., 27
Zhylyevskyy, O., 7
Ziegelmann, F., 119
Zowghi, M., 11