# PROGRAMME AND ABSTRACTS

7th International Conference on

## Computational and Financial Econometrics (CFE 2013)

http://www.cfenetwork.org/CFE2013

and

6th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

## Computational and Methodological Statistics (ERCIM 2013)

http://www.cmstatistics.org/ERCIM2013

Senate House, University of London, UK
14-16 December 2013

ERCIM
European Research Consortium
for Informatics and Mathematics

ERCIM WG on Computational and Methodological Statistics
http://www.CMStatistics.org

Computational and Financial Econometrics
http://www.CFEnetwork.org

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

**ERCIM Working Group on**
**COMPUTATIONAL AND METHODOLOGICAL STATISTICS**

http://www.cmstatistics.org

## AIMS AND SCOPE

The working group (WG) CMStatistics focuses on all computational and methodological aspects of statistics. Of particular interest is research in important statistical applications areas where both computational and/or methodological aspects have a major impact. The aim is threefold: first, to consolidate the research in computational and methodological statistics that is scattered throughout Europe; second, to provide researches with a network from which they can obtain an unrivalled sources of information about the most recent developments in computational and methodological statistics as well as its applications; third, to edit quality publications of high impact and significance in the broad interface of computing, methodological statistics and its applications.

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

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

### Specialized teams

Currently the ERCIM WG has over 800 members and the following specialized teams

**BM:** Bayesian Methodology

**CODA:** Complex data structures and Object Data Analysis

**CPEP:** Component-based methods for Predictive and Exploratory Path modeling

**DMC:** Dependence Models and Copulas

**DOE:** Design Of Experiments

**EF:** Econometrics and Finance

**GCS:** General Computational Statistics WG CMStatistics

**GMS:** General Metholological Statistics WG CMStatistics

**HDS:** High-Dimensional statistics

**ISDA:** Imprecision in Statistical Data Analysis

**LVSEM:** Latent Variable and Structural Equation Models

**MCS:** Matrix Computations and Statistics

**MM:** Mixture Models

**MSW:** Multi-Set and multi-Way models

**NPS:** Non-Parametric Statistics

**OHEM:** Optimization Heuristics in Estimation and Modelling

**RACDS:** Robust Analysis of Complex Data Sets

**SAE:** Small Area Estimation

**SAET:** Statistical Analysis of Event Times

**SAS:** Statistical Algorithms and Software

**SEA:** Statistics of Extremes and Applications

**SFD:** Statistics for Functional Data

**SL:** Statistical Learning

**SSEF:** Statistical Signal Extraction and Filtering

**TSMC:** Times Series Modelling and Computation

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

**SPONSORS**

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

ELSEVIER
http://www.elsevier.com

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

cajAstur
http://www.cajastur.es

ERCIM WG on Computational and Methodological Statistics
http://www.CMStatistics.org

Computational and Financial Econometrics
http://www.CFEnetwork.org

**EXHIBITORS**

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

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

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

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

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

Timberlake Consultants (URL http://www.timberlake.co.uk)

**ENDORSED SOCIETIES & GROUPS**

ERCIM Working Group on *Computational and Methodological Statistics*

Computational and Financial Econometrics *CFEnetwork*

The Society for Computational Economics

International Statistical Institute

International Association for Statistical Computing

## Contents

**ES40 Room Court DEPENDENCE MODELLING: THEORY AND PRACTICE Chair: Ivan Kojadinovic**

E023: **On idempotent $D$-norms**
*Presenter:* **Michael Falk**, University of Wuerzburg, Germany
Replacing the spectral measure by a random vector $Z$ allows the representation of a multivariate max-stable distribution with standard negative margins via a norm, called $D$-norm, whose generator is $Z$. We investigate the set of all generators in detail. This approach towards multivariate extreme value distributions entails the definition of a multiplication type operation on the set of $D$-norms leading to idempotent $D$-norms. We characterize the set of idempotent $D$-norms. Iterating the multiplication provides a track of $D$-norms, whose limit exists and is again a $D$-norm. If this iteration is repeatedly done on the same $D$-norm, then the limit of the track is idempotent.

E246: **When uniform weak convergence fails: Empirical processes for dependence functions via epi- and hypographs**
*Presenter:* **Johan Segers**, University of Louvain-la Neuve, Belgium
*Co-authors:* Axel Buecher, Stanislav Volgushev
For copulas whose partial derivatives are not continuous everywhere on the interior of the unit cube, the empirical copula process does not converge weakly with respect to the supremum distance. This makes it hard to verify asymptotic properties of inference procedures for such copulas. To resolve the issue, a new metric for locally bounded functions is introduced and the corresponding weak convergence theory is developed. Convergence with respect to the new metric is related to epi- and hypoconvergence and is weaker than uniform convergence. Still, for continuous limits, it is equivalent to locally uniform convergence, whereas under mild side conditions, it implies $L^p$ convergence. Even in cases where uniform convergence fails, weak convergence with respect to the new metric is established for empirical copulas and tail dependence processes. No additional assumptions are needed for tail dependence functions, and for copulas, the assumptions reduce to existence and continuity of the partial derivatives almost everywhere on the unit cube. The results are applied to obtain asymptotic properties of minimum distance estimators, goodness-of-fit tests and resampling procedures.

E095: **Generalized additive modelling for conditional copulas**
*Presenter:* **Valerie Chavez-Demoulin**, University of Lausanne, Switzerland
*Co-authors:* Thibault Vatter
A generalized additive modelling framework to t conditional copulas is proposed. This framework allows copula-based models to depend on covariates in a parametric, semi-parametric or non-parametric way. The method uses penalized log-likelihoods maximized through specific Newton-Raphson type algorithms. Simulations designed to study the numerical properties of the method indicate that it performs well. A real dataset is considered as an application.

E741: **Nonparametric tests for constancy of a copula**
*Presenter:* **Axel Buecher**, Ruhr-Universitaet Bochum, Germany
*Co-authors:* Ivan Kojadinovic
The modeling and estimation of stochastic dependencies by copulas has attracted an increasing attention over the last years in various fields of application. Most of the available statistical inference procedures are based on the implicit assumption that the copula of a multivariate time series remains constant over time. We present methods that allow us to test for this assumption, both for marginals that are identically distributed and for those whose distribution may change over time. The asymptotics of the test statistics rely on a general weak convergence result for the sequential empirical copula process based on time series data. In the case of observations that are strongly mixing, it is shown how a generalized multiplier bootstrap allows us to get approximate access to the quantiles of the limiting distribution.

E029: **Computational statistics in copula modeling: Why and how**
*Presenter:* **Marius Hofert**, ETH Zurich, Switzerland
*Co-authors:* Martin Maechler
In the highly non-linear world of copulas, computational challenges (high dimensions, numerical precision, parallel computing) arise quickly. More importantly, numerical issues may not be obvious from computed results and therefore such results may actually be misleading. This has even led to wrong conclusions drawn from erroneous simulation studies in the recent literature and poses a particular challenge for practitioners who want to apply copula modeling in business practice. After giving a short motivation on the subject, we briefly present a new R package which aims at simplifying statistical simulation studies and which carefully deals with important tasks such as parallel computing, seeding, catching of warnings and errors, and measuring run time.

**ES52 Room Bedford ADVANCES IN COMPOSITIONAL DATA ANALYSIS AND RELATED METHODS I Chair: Karel Hron**

E859: **Zeros in compositional count data sets: The geometric Bayesian-multiplicative imputation**
*Presenter:* **Josep Antoni Martin-Fernandez**, University of Girona, Spain
*Co-authors:* Javier Palarea-Albaladejo, Karel Hron
A vector of counts is collected when the outcomes in each a number of identical and independent trials can fall in any of a fixed number of mutually exclusive categories. The analysis of this type of data is typically based on multinomial and Dirichlet models. These methods show important difficulties in those cases where the total sum of the vector is not of interest, suggesting that log-ratio methodology is a more general approach to the analysis of compositional count data. However, log-ratio methodology requires a preprocessing step where a proper treatment of the zeros is applied. In our approach we assume that zero values refer to unobserved positive values that may have been observed with a larger number of trials. According to this assumption, a treatment involving a Bayesian inference of the zero values and a multiplicative modification of the non-zero values is considered. This treatment offers the possibility to use valuable information from the personal knowledge of the analyst or from the data previously collected. In this way, an adequate strategy could be to consider a prior equal to the sample estimation of the expected value in the variable with a leave-one-out scheme. In other words, when we deal with a vector we are assuming that the other samples are our prior information. Because this approach is fully coherent with the properties of the geometric distribution of probability, we call it a geometric bayesian-multiplicative (GBM) replacement. An additional advantage of GBM replacement is that the posterior estimates do not depend on the number of categories, that is, conforms to the representation invariance principle. The performance of this technique is illustrated using real and simulated data sets.

E080: **N-way partial least squares for compositional data**
*Presenter:* **Michele Gallo**, University of Naples Orientale, Italy
Partial least squares (PLS) is a method for building regression models between independent and dependent variables. When a set of independent variables is measured on several occasions, the samples can subsequently be arranged in three-way arrays. In this case N-way partial least squares (N-PLS) can be used. N-PLS decomposes three-way array of independent variables and establishing a relation between the three-way array of independent variables and the array of dependent variables. Sometimes, the set of independent variables are parts of the same whole, thus each observation consists of vectors of positive values summing to a unit, or in general, to some fixed constant. When these data, known as compositional

data (CoDa), are analyzed by N-PLS, it is necessary to take into account the specific relationships between the parts that compositions are made of. The problems that potentially occur when one performs a N-way partial least squares analysis on compositional data are examined. A strategy based on the log-ratio transformations is suggested.

E389: **Left-censoring problems in compositional data sets**
*Presenter:* **Javier Palarea-Albaladejo**, Biomathematics and Statistics Scotland, United Kingdom
*Co-authors:* Josep Antoni Martin-Fernandez
Multivariate data representing part of a whole, usually called compositional data in statistics, are common in many fields of science-say chemical concentrations, food compositions, activity patterns, abundance of species, and so on. Their distinctive feature is that there is an inherent relationship between the parts, as they only convey relative information. This implies that ordinary statistical techniques are not generally adequate, and specialized statistical methods are required. Progress in compositional data analysis has been mostly driven by the log-ratio methodology. In this context, the focus is on imputation methods to deal with parts that include unobserved values falling below a certain threshold. Left-censoring problems often arise in experimental compositional data, either in the form of rounded zeros or values below certain limits of detection. We discuss the issue of introducing a number of methods that take into consideration the particular principles and nature of compositional data, from univariate non-parametric procedures to multivariate model-based approaches. These methods are supported with software that facilitates their practical implementation.

E863: **Compositional entropies in model based clustering**
*Presenter:* **Marc Comas-Cufi**, Universitat de Girona, Spain
*Co-authors:* Gloria Mateu-Figueras, Josep Antoni Martin-Fernandez
To cluster samples from a finite mixture density a model-based technique is recommended. Initially the cluster method selects the total number of mixture components. Assuming that the number of groups is less than or equal to the mixture components, the method hierarchically combines the components using an entropy criterion applied to posterior probabilities. Typically the criterion is based on the well-known Shannon entropy. In this work we show that any model-based cluster analysis applied to any type of data, not necessarily compositional, is enriched when the vector of posterior or individual's conditional probabilities (group memberships) are considered as elements of the simplex. In this way, entropy criterion based on the Aitchison distance and the compositional Kullback-Leibler divergence are introduced. Here the Aitchison distance between two compositions is defined as the Euclidean distance between the corresponding log-ratio coordinates. The compositional Kullback-Leibler divergence consists in a modification of the Jeffreys divergence. Both measures fulfil the two main compositional principles: scale invariance and subcompositional coherence. The performance of these compositional entropy criteria are evaluated and illustrated using real and simulated data sets.

E940: **Classical and robust principal component analysis for density functions using Bayes spaces**
*Presenter:* **Klara Hruzova**, Palacky University, Czech Republic
*Co-authors:* Karel Hron, Matthias Templ, Peter Filzmoser, Gianna Serafina Monti
Probability densities are used to describe relative contributions of Borel sets of real line to the overall probability of support of a random variable. This is one of the reasons for considering density functions as functional compositional data with a constant-sum-constraint equal to one. For this type of data it was necessary to construct a new Hilbert space, called Bayes space, that accounts for specific properties of densities as functional data carrying relative information. Recently, there have been several approaches to analyze statistically density functions in an ordinary way, e.g. using functional principal component analysis. The aim of this contribution is to show a different approach for the mentioned case of functional principal component analysis where densities are transformed using a functional version of centred logratio transformation (clr) for compositional data. Finally, the methodology will be applied to real-world data set from official statistics.

## ES60 Room Torrington DESIGN OF EXPERIMENT IN INDUSTRY Chair: Kalliopi Mylona

E082: **Designed experiments for semi-parametric models and functional data with a case-study in tribology**
*Presenter:* **David Woods**, University of Southampton, United Kingdom
*Co-authors:* Christopher Marley, Susan Lewis
Experiments with functional data are becoming ubiquitous in science and engineering with the increasing use of online monitoring and measurement. Each run of the experiment results in the observation of data points that are realised from a smooth curve. Although large quantities of data may be collected from each run, it may still only be possible to perform small experiments with a limited number of runs. We describe a statistical methodology for an example from Tribology, concerning the wear-testing of automotive lubricants. Here, we investigated how lubricant properties and process variables affected the shape of a functional response measuring wear. Novel techniques were developed for the initial design of a screening study where the levels of some of the factors could not be set directly. A two-stage semi-parametric modelling approach was applied, using a varying coefficient model and principal components. New methods for the design of follow-up experiments for such models were also developed and applied. In addition to the new methodology, we present conclusions from the case study about which factors had substantial effects, and how they influenced the shape of the wear curves.

E508: **Optimal design of choice experiments with partial profiles**
*Presenter:* **Peter Goos**, Universiteit Antwerpen, Belgium
*Co-authors:* Daniel Palhazi Cuervo, Roselinde Kessels, Kenneth Sorensen
To limit the choice task complexity in choice experiments involving many attributes, some authors recommend the use of partial profiles. In the special case where respondents are indifferent between different alternatives, D-optimal designs are known for many scenarios. For more realistic cases in which respondents are not indifferent, no D-optimal designs are known, and experimenters have to resort to heuristic design construction algorithms. A new algorithm is presented to construct locally optimal partial-profile designs as well as Bayesian optimal partial-profile designs in these more realistic cases. The excellent performance of the new algorithm is demonstrated by showing that its results match the known D-optimal designs in case respondents are indifferent.

E659: **Minimum setup criterion to select designs with hard-to-change factors**
*Presenter:* **Andre Pinho**, Federal University of Rio Grande do Norte, Brazil
*Co-authors:* Linda Ho, Carla Vivacqua
Considering fierce competition, companies always seek strategies to improve their time to market. Although new simulation technologies are available, physical prototype testing still remains an important step in the product development cycle. It is common that prototypes are composed of several parts, with some more difficult to assemble than others. Moreover, typically, there is only one piece available of each part and a large number of different setups. Under these conditions, designs with randomization restrictions become attractive approaches. Considering this scenario, a new and additional criterion to construct split-plot type designs is presented. Designs with a small number of setups of the more difficult parts, which are especially useful for screening purposes, are discussed. The proposal of the minimum setup (MS) criterion is driven by real applications and represents an alternative to directly accommodate practical or operational requirements, disregarded by other criteria. The process