

PROGRAMME AND ABSTRACTS

7th CSDA 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 WG on Computational
and Methodological Statistics

<http://www.CMStatistics.org>

Computational and
Financial Econometrics

<http://www.CFEnetwork.org>



<http://www.qmul.ac.uk>



<http://www.bbk.ac.uk>



<http://www.lse.ac.uk>

E1111: Approximate Bayesian computation for bilinear processes*Presenter:* **Patricia de Zea Bermudez**, University of Lisbon, Portugal*Co-authors:* Kamil Turkman, Antonia Turkman

The class of bilinear models plays an important role in modeling nonlinearity for various reasons, such as the fact that it is a generalization of ARMA models. Under fairly general conditions, bilinear processes approximate finite order Volterra series expansions to any desired order of accuracy over finite time intervals. Volterra series expansion are a dense class within the class of nonlinear time series. Therefore, bilinear processes are also a dense class within nonlinear processes, approximating any nonlinear process to a desired level of accuracy. Due to their capacity of producing clusters of large values, bilinear models are often suggested. However, they are not frequently used in practice due to inference problems. ABC algorithms arise as ways to deal with problems associated with likelihood functions which are analytically difficult to handle or even intractable. The challenge is to find a set of statistics capable of representing the nonlinear dynamics of the system. Seven statistics are suggested, namely a portmanteau statistic that captures the linear time dynamics, a tail index estimator, which quantifies the tail heaviness and the extremal index that measures the degree of clustering of large values. These strategies are applied to several bilinear models for different i.i.d. innovation processes.

ES46 Room G16 CONTRIBUTIONS TO STATISTICAL ALGORITHMS AND SOFTWARE IN R**Chair: Cristian Gatu****E222: GEE for longitudinal ordinal data: Comparing R-repolr, R-ordgee, SAS-GENMOD, SPSS-GENLIN***Presenter:* **Nazanin Noorae**, University Medical Center Groningen, Netherlands*Co-authors:* Geert Molenberghs, Edwin van den Heuvel

To understand the relationship between covariates and outcome, generalized estimating equations (GEE) can be applied. Different approaches have been introduced to extend the GEE for ordinal outcomes and they have been implemented in statistical software. Comparisons of the implemented algorithm in the most frequently used packages (SAS-PROC GENMOD, SPSS-GENLIN, R-geepack and R-repolr) are provided using a simulation study. Multivariate logistic distributions with moderate to highly correlation are considered as the latent variable of ordinal outcomes. All the simulation is performed in the package copula in R. The performance of GEE is assessed on the bias and coverage probability of the parameter estimations and on the number of numerical convergence issues. The simulation study indicates substantial bias in the parameter estimates and numerical issues for data sets with small number of subjects. The unstructured correlation matrix requires larger numbers of subjects than independence and exchangeable correlation matrices to reduce the bias and numerical issues. The coverage probabilities of the parameters are satisfactory for the independence and exchangeable correlation matrix, but they are frequently liberal for the unstructured option. Based on the performance and the available options, preference is given to SPSS and R-repolr.

E886: Biomarker selection for omics data*Presenter:* **Ron Wehrens**, Fondazione Edmund Mach, Italy*Co-authors:* Pietro Franceschi

The R package BioMark provides several tools to define which variables are associated with class differences in data from fields like metabolomics and proteomics. The first group of tools uses Higher Criticism to define an optimal threshold between interesting and non-interesting variables. This can be applied to any statistic, be it a t value, a regression coefficient or something else, and is related to the expected distribution of p values under the null distribution. The second group of tools is based on stability selection, i.e. an assessment of how often specific variables are highlighted as interesting under perturbation of the data. This approach is especially attractive when the number of samples is larger than, say, ten per group. Also in this case, the strategy can be applied to any type of statistic. Using real and simulated data, the application and usefulness of these techniques will be shown.

E898: NH-Poisson: An R package for fitting and validating nonhomogeneous Poisson processes*Presenter:* **Ana C Cebrian**, University of Zaragoza, Spain*Co-authors:* Jesus Abaurrea, Jesus Asin

NH-Poisson is an R package which provides an assembly of tools for all the steps involved in data modeling using nonhomogeneous Poisson processes in one dimension. It includes functions for data preparation, maximum likelihood estimation, covariate selection and inference based on asymptotic distributions and simulation methods. It also provides specific methods for the estimation of Poisson processes resulting from a peak over threshold approach. The package also contains a wide tool kit for validation analysis, including several types of residuals and diagnostic techniques adapted from other statistical models. It supports functions for generating nonhomogeneous Poisson process trajectories. This package can be useful for anyone interested in modeling data using one-dimensional nonhomogeneous Poisson processes in any field.

E1099: Nonparametric inference for controlled branching processes: Expectation-maximization algorithm*Presenter:* **Carmen Minuesa**, University of Extremadura, Spain*Co-authors:* Miguel Gonzalez, Ines del Puerto

Controlled branching processes are stochastic growth population models in which the number of individuals with reproductive capacity in each generation is controlled by a random control function. The behaviour of these populations is strongly related to the main parameters of the offspring distribution. In practice, these values are unknown and their estimation is necessary. Usually it must be observed the whole family tree up to a given generation in order to estimate the offspring distribution. In this work, we deal with the problem of estimating the main parameters of the model assuming that the only observable data are the total number of individuals and progenitors in each generation. We set out the problem in a nonparametric framework and obtain the maximum likelihood estimator of the offspring distribution using the expectation-maximization algorithm. Finally, we show the accuracy of the algorithm by way of a simulated example.

ES121 Room B36 CONTRIBUTIONS IN TIME SERIES ANALYSIS I**Chair: Marc Hallin****E279: Model-free detection of structural breaks in time series***Presenter:* **Alexandra Piryatinska**, San Francisco State University, United States*Co-authors:* Boris Darkhovsky

Long time series are often generated by different mechanisms (probabilistic, deterministic, or mixed). In economic time series, the time when the underlying generating mechanism changes is called the moment of structural break (MSB). Traditionally, parameters of the models are used to detect the MSB. However, frequently no plausible model is known. In such cases, a novel, model-free approach for detection of the MSB is proposed. It is based on the concept of ϵ -complexity of continuous functions, and the non-parametric method of change-point detection. In particular, it was established that ϵ -complexity of a function satisfying the Hölder condition can be completely characterized by a pair of real numbers which are called ϵ -complexity coefficients. Under the assumption that structural breaks lead to a change of the expected values of the ϵ -complexity coefficients, a non-parametric method to estimate moments of such changes is proposed. Simulations results are given.