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
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

expressed analytically. It also provides for simple conditions under which the series is stationary and/or Markov, as well as being parsimonious. A parallel algorithm for computing the likelihood is given, along with a Bayesian approach for computing inference based on model averages over parsimonious representations of the copula. The model average estimates are shown to be more accurate in a simulation study. Two five-dimensional time series from the Australian electricity market are examined. In both examples, the fitted copula captures substantial asymmetric tail dependence, both over time and across elements in the series.

E1181: **Flexible sampling for the Gaussian copula extended rank likelihood model**
Presenter: **Ricardo Silva**, UCL, United Kingdom
Co-authors: Alfredo Kalaitzis

The framework of copula modeling has gained popularity due to its modular parametrization of joint distributions. Among other properties, copulas provide a recipe for combining flexible models for univariate marginal distributions with parametric families suitable for potentially high dimensional dependence structures. More radically, a previous extended rank likelihood approach bypasses learning marginal models completely when such information is ancillary to the learning task at hand as in, e.g., standard dimensionality reduction problems or copula parameter estimation. The main idea is to represent data by their observable rank statistics, ignoring any other information from the marginals. Inference is typically done in a Bayesian framework with Gaussian copulas, and it is complicated by the fact this implies sampling within a space where the number of constraints increases quadratically with the number of data points. The result is slow mixing when using off-the-shelf Gibbs sampling. We present an efficient algorithm based on recent advances on constrained Hamiltonian Markov chain Monte Carlo that is simple to implement and does not require paying for a quadratic cost in sample size.

ES111 Room B29 COMPUTATIONAL STATISTICS I Chair: Niels Richard Hansen

E871: **Estimation of the typical ranks of $2 \times 2 \times \cdots \times 2$ tensors through intrinsic dimension estimators**
Presenter: **Toshio Sakata**, Kyushu University, Japan
Co-authors: Toshio Sumi, Mitsuhiro Miyazaki

The tensor data, multi-way data, are recently applied to many field successfully. The rank of a tensor T is a measure of complexity of tensors. The maximal tensor rank has attracted many researchers' interest, however, statistically speaking, the typical rank is more important. The typical rank of a given type of tensors is defined to be an integer r such that the set of tensors with the rank r has a Lebesgue measure greater than 0. To determine a typical rank is difficult and attracted many researchers' concern, too. We investigate the possibility of determining the typical rank of $2 \times 2 \times \cdots \times 2$ tensors by estimating statistically the intrinsic dimension of the manifold formed of tensors of a rank less than or equal to r. In fact this is the situation that we have a manifold and points over the manifold and estimate its intrinsic dimension by the data points. This is the first such challenge. We will show by simulation that the maximum likelihood estimation of the intrinsic dimension is useful to estimate the typical rank. Several methods other than the maximum likelihood estimation will be also compared. Finally we note that our method is also applicable to the other types of tensors.

E1204: **Challenges in extremal index estimation through computer intensive procedures**
Presenter: **Manuela Neves**, University of Lisbon and CEAUL, Portugal
Co-authors: Dora Prata Gomes

In statistics of extremes there are a few parameters of particular interest characterizing the behavior of extreme or even rare events. Among these parameters the extremal index appears for dependent sequences. It measures the relationship between the dependence structure of the data and the behavior of the exceedances over high thresholds. It can be roughly defined as the reciprocal of the 'mean time of duration of extreme events' and needs to be adequately estimated not only by itself, but also because its influence in the estimation of other parameters of interest. Most semi-parametric estimators of this parameter show the same behavior: nice asymptotic properties but a high variance for small values of k, the number of upper order statistics used in the estimation and a high bias for large values of k. The Mean Square Error, a measure that encompasses bias and variance, usually shows a very sharp plot, needing an adequate choice of k. After reviewing some classical extremal index estimators, the emphasis is now given to explore new heuristic procedures, helped by computational techniques, for obtaining efficient estimates. A simulation study will illustrate the properties of the estimators and the performance of the adaptive algorithm proposed.

E887: **A comparison of computational approaches for maximum likelihood estimation of the Dirichlet parameters on high dimensional data**
Presenter: **Marco Giordan**, Fondazione Edmund Mach, Italy
Co-authors: Ron Wehrens

Likelihood estimates of the Dirichlet distribution parameters can be obtained only through numerical algorithms. Even though the Dirichlet distribution belongs to the Exponential Family, such algorithms can give estimates outside the correct range for the parameters. In addition, they can require a large amount of iterations to reach convergence. These problems can be exasperated if good starting points are not provided. We discuss several approaches looking at the trade-off between speed and stability. We illustrate the combinations of initialization and estimation methods using both real and simulated data of high dimension.

E858: **Two depth based strategies for robust estimation of a predictive distribution of a data stream**
Presenter: **Daniel Kosiorowski**, Cracow University of Economics, Poland

Data stream analysis can be informally defined as a computationally challenging sequence of stochastic process analyses conducted on-line basing on sliding window or windows from the process. They may lead to a sequence of investment decisions in case of an algorithmic trading, to dangerous behaviour detections in case of a city monitoring, to mining in astronomic data sets, etc. Data streams are generated by nonstationary and multiregime processes. Statistical procedures used for the data stream analysis should be nonparametric, be able to cope with "curse of dimensionality", be computationally feasible and robust but sensitive to changes of a majority of observations in the window (the regime of the underlying process detection). We present two depth based strategies for predictive distribution of the stream estimation. Our first strategy "dynamic robust rule of thumb" appeals to general considerations concerning Student depth. Our second strategy appeals to an idea of adjusted conditional distribution estimator previously proposed. We adjust this estimator using projection depth that protects us against outliers and inliers. We show properties of our proposals using multiregime econometric models with deterministic and stochastic transitions between regimes as well as empirical examples.