PROGRAMME AND ABSTRACTS

10th International Conference on

Computational and Financial Econometrics (CFE 2016)

http://www.cfenetwork.org/CFE2016

and

9th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2016)

http://www.cmstatistics.org/CMStatistics2016

Higher Technical School of Engineering, University of Seville, Spain

9 – 11 December 2016

ISBN 978-9963-2227-1-1

Technical Editors: Angela Blanco-Fernandez and Gil Gonzalez-Rodriguez.

International Organizing Committee:
Ana Colubi, Stella Hadjiantoni, M. Dolores Jimenez-Gamero, Erricos Kontoghiorghes, George Loizou and Herman Van Dijk.

CFE 2016 Co-chairs:
Peter Boswijk, Jianqing Fang, Alain Hecq, Michael Smith.

CFE 2016 Programme Committee:

Alessandra Amendola, Ansgar Belke, Monica Billio, Josep Lluis Carrion-i-Silvestre, Roberto Casarin, Veronika Czellar, Serge Darolles, Manfred Deistler, Jean-David Fermanian, Sylvia Fruehwirth-Schnatter, Marc Hallin, Matthew Harding, Jan Jacobs, Menelaos Karanasos, Daniel Kaufmann, Robert Kunst, Degui Li, Marco Lippi, Zudi Lu, Luis Filipe Martins, Gian Luigi Mazzi, Tucker McElroy, J Isaac Miller, Bent Nielsen, Yasuhiro Omori, Michael Owyang, Gareth Peters, Michael Pitt, Stephen Pollock, Tommaso Proietti, Artem Prokhorov, Anders Rahbek, Francesco Ravazzolo, Jeroen Rombouts, Eduardo Rossi, Christos Savva, Willi Semmler, Laura Spierdijk, Giuseppe Storti, Jonathan Stroud, Genaro Sucarrat, Jean-Pierre Urbain, Shaun Vahey, Helena Veiga, Carlos Velasco, Helga Wagner, Martin Wagner, Toshiaki Watanabe and Peter Winker.

CMStatistics 2016 Co-chairs:
Jianqing Fang, Jeng-Ming Chiou, Armelle Guillou and Mattias Villani.

CMStatistics 2016 Programme Committee:

Christophe Ambroise, Tatiyana Apanasovich, Moulinath Banerjee, Yves Berger, Mauro Bernardi, Eric Beutner, Roel Braekers, Eva Cantoni, Guang Cheng, Ming-Yen Cheng, Yu Cheng, Michael Daniels, Herold Dehling, Aurore Delaigle, Sebastian Doehler, Takeshi Emura, Michael Falk, Maria Brigida Ferraro, Roland Fried, Giuliano Galimberti, Laurent Gardes, Irene Gijbels, Ivette Gomes, Gil Gonzalez-Rodriguez, Przemyslaw Grzegorzewski, Christopher Hans, Marina Knight, Ivan Kojadinovic, Fumiyasu Komaki, Jens-Peter Kreiss, Chenlei Leng, Christophe Ley, Catherine Chunling Liu, Jesus Lopez-Fidalgo, Marloes Maathuis, Alireza Mahani, Alfio Marazzi, Giovanni Marchetti, Lola Martinez-Miranda, Geoffrey McLachlan, Paul McNicholas, Alicia Nieto-Reyes, Klaus Nordhausen, Davy Paindaveine, Markus Pauly, Juan Eloy Ruiz-Castro, Anne Ruiz-Gazen, Laura Sangalli, Robert Serfling, Leopold Simar, Stefan Sperlich, Jianguo Sun, Yanqing Sun, Valentin Todorov, Stefan Van Aelst, Ingrid Van Keilegom, Thomas Verdebout, Tim Verdonck, Daniel Vogel, Huixia Judy Wang, Jane-Ling Wang, Lan Wang, Nanny Wermuth, Liming Xiang, Hao Helen Zhang, Jacobo de Una-Alvarez and David van Dyk.

CMStatistics: ERCIM Working Group on
COMPUTATIONAL AND METHODOLOGICAL STATISTICS

http://www.cmstatistics.org

The working group (WG) CMStatistics 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 journal special issues. The Econometrics and Statistics (EcoSta) and Computational Statistics & Data Analysis (CSDA) are the official journals of the CMStatistics.

Specialized teams

Currently the ERCIM WG has over 1650 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 Methodological Statistics WG CMStatistics
GOF: Goodness-of-Fit and Change-Point Problems
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 website), or by email at info@cmstatistics.org.

CFEnetwork
COMPUTATIONAL AND FINANCIAL ECONOMETRICS

http://www.CFEnetwork.org

The Computational and Financial Econometrics (CFEnetwork) comprises a number of specialized teams in various research areas of theoretical and applied econometrics, financial econometrics and computation, and empirical finance. The teams contribute to the activities of the network by organizing sessions, tracks and tutorials during the annual CFEnetwork meetings, and by submitting research proposals. Furthermore the teams edit special issues currently published under the Annals of CFE. The Econometrics and Statistics (EcoSta) is the official journal of the CFEnetwork.

Specialized teams

Currently the CFEnetwork has over 1000 members and the following specialized teams

AE: Applied Econometrics
BE: Bayesian Econometrics
BM: Bootstrap Methods
CE: Computational Econometrics
ET: Econometric Theory
FA: Financial Applications
FE: Financial Econometrics
TSE: Time Series Econometrics

You are encouraged to become a member of the CFEnetwork. For further information please see the website or contact by email at info@cfenetwork.org.

EO1253: **Comparing robust fuzzy methods for clustering non-precise data**
Presenter: **Ana Belen Ramos-Guajardo**, University of Oviedo, Spain
Co-authors: Paolo Giordani

In many practical situations the data are not precise. The imprecision of the data can be managed by means of fuzzy sets. This type of data is characterized by a complex structure and, for this reason, there exist different kinds of contamination in this context. There are several proposals of robust methods for clustering fuzzy data. A type of 'robustification' is the use of medoids. Another approach consists in trimming the data. In detail, the outliers are trimmed and not used in the clustering procedure. A further proposal is to add a noise cluster, that is not a proper cluster, containing all the contaminated data. Finally, an alternative approach is the possibilistic one. In this case the membership degree is only based on the distance between the observation and the centroid. Hence, an outlier is characterized by having low membership degrees to all the clusters. We compare all the above mentioned methods by means of simulation and real-case studies in order to analyze their drawbacks and benefits.

EO0868: **Fuzzy two-mode clustering with polynomial fuzzifiers**
Presenter: **Maria Brigida Ferraro**, Sapienza University of Rome, Italy

A new two-mode clustering algorithm is proposed in a fuzzy framework. Two-mode clustering consists in simultaneously clustering modes (e.g. objects, variables) of an observed two-mode data matrix. An extension of the Fuzzy Double k-Means (FDkM) is addressed. The parameters of fuzziness in FDkM are introduced to obtain two fuzzy partitions. These parameters, analogously to the simple (one-mode) case of fuzzy k-means, are used to control the overlapping clusters by giving low membership degrees to objects/variables with unclear assignments. In this way, objects/variables are usually assigned to all clusters with non-zero membership degrees, preventing a hard classification for objects/variables that should be uniquely assigned to a single prototype. In order to overcome this kind of problem, polynomial fuzzifier functions are used. As particular cases, we obtain the FDkM and the hard Double k-Means. The adequacy of the proposal is checked by means of simulation and real-case studies.

EO115 Room 101 MIXTURE MODELS FOR COMPLEX DATA SETS — Chair: Geoffrey McLachlan

EO0935: **Maximum likelihood inference for mixtures of Gaussian regression models**
Presenter: **Giuliano Galimberti**, University of Bologna, Italy
Co-authors: Gabriele Soffritti

Finite mixtures of Gaussian linear regression models have been widely studied. They represent useful tools in applied statistics, whenever the interest is in studying the effect of a set of predictors on a response, in presence of unobserved sources of heterogeneity in the data. Despite their popularity, inferential procedures for the model parameters (and in particular for the component regression coefficients) have not been deeply investigated. The score vector and the Hessian matrix of the incomplete data log-likelihood for a finite mixture of multivariate Gaussian linear regression models are derived. Approximations for the standard errors of maximum likelihood (ML) estimators are obtained from these quantities. The properties of these standard error estimates are investigated through an extensive simulation study. Particular attention is devoted to the behavior of these estimates in presence of model misspecification.

EO1206: **A fuzzy version of robust mixtures of Gaussian factor analyzers**
Presenter: **Francesca Greselin**, University of Milano Bicocca, Italy
Co-authors: Agustin Mayo-Iscar, Luis Angel Garcia-Escudero

Clustering aims at dividing a data set into groups or clusters that consist of similar data. Fuzzy clustering accepts the fact that the clusters or classes in the data are usually not completely well separated and thus assigns a membership degree between 0 and 1 for each cluster to every datum. We introduce a robust method for fuzzy clustering based on mixtures of Gaussian Factor analyzers. We illustrate our theoretical considerations by simulations and applications to real data. A comparison with probabilistic clustering is also provided.

EO1042: **Fast model-based clustering of functional data via Gaussian mixture models**
Presenter: **Hien Nguyen**, University of Queensland, Australia

Functional data analysis has become ubiquitous in recent years. It provides an effective framework for statistical modelling of infinite-dimensional functional objects, which occur frequently in practice. A popular paradigm for model-based clustering of functional data is via linear-basis filtering and mixture of mixed effects models. We demonstrate that in some situation, such a paradigm reduces to the simple Gaussian mixture model, for which there are numerous fast and free computational tools available. A demonstration of our clustering approach is performed on data from the calcium imaging of a zebrafish brain.

EO435 Room 305 A ROBUST STATISTICAL MODELLING — Chair: Alfio Marazzi

EO1027: **Robust functional principal components for sparse data**
Presenter: **Graciela Boente**, Universidad de Buenos Aires and CONICET, Argentina
Co-authors: Matias Salibian-Barrera, Jane-Ling Wang

Functional principal components analysis allows us to obtain parsimonious predictions for each trajectory in the sample. The problem of robustly estimating functional principal components when there are only a few observations per curve available will be discussed. Specifically, assume that we observe $X_i(t_{ij})$, $i=1,\dots,n$, $j=1,\dots,n_i$, where the n_i's can be small for all curves. Many available methods to estimate functional principal components rely on a smoothing step of the observed trajectories, and thus require many observations per curve. A notable exception is the conditional expectation approach (PACE) which estimates the covariance function by smoothing the sparsely available cross-products being able to "combine information" from many curves. A first attempt at protecting this approach from potential outliers by using a robust smoother on the cross-products does not work because the distribution of the cross-products is generally asymmetric. We exploit the linear structure of the conditional distribution of $X_i(t)|X_i(s)$ as a function of $X_i(s)$ for elliptical processes to obtain robust estimators of the scatter function of the underlying random process. We report some numerical experiments comparing the performance of the resulting estimates and existing alternatives.

EO0849: **Robust semi-parametric estimators: Missing data and causal inference**
Presenter: **Eva Cantoni**, University of Geneva, Switzerland
Co-authors: Xavier De Luna

Situations are considered where we aim at estimating location and scale parameters from a distribution law of interest, from which a random sample has been drawn. We introduce semi-parametric estimators, which are able to deal simultaneously with two common challenges within this general context: (i) not all observations from the random sample intended are available (incomplete data due to dropout, selection, potential outcomes framework), and (ii) some of the available observations in the sample may be contaminated (generated by a nuisance distribution, outliers). Under an assumption of ignorable missingness, popular semi-parametric estimators of the parameters of interest are augmented inverse probability weighted (AIPW, doubly robust) estimators. They use two auxiliary models, one for the missingness mechanism, and another for an outcome of interest, both given observed covariates. AIPW estimators are then robust to misspecification of one of these two models (but not both simultaneously - a so-called double robustness property). We introduce versions of AIPW, which provide, moreover, robustness to contamination