PROGRAMME AND ABSTRACTS

6th CSDA International Conference on

Computational and Financial Econometrics (CFE 2012)

http://www.cfe-csda.org/cfe12

and

5th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computing & Statistics (ERCIM 2012)

http://www.cfe-csda.org/ercim12

Conference Center "Ciudad de Oviedo", Spain

1-3 December 2012

ERCIM
European Research Consortium
for Informatics and Mathematics

UNIVERSIDAD DE OVIEDO

http://www.uniovi.es

Queen Mary
University of London

http://www.qmul.ac.uk

Depósito Legal: AS-03972-2012
ISBN: 978-84-937822-2-1

International Organizing Committee:
Ana Colubi, Christophe Croux, Erricos J. Kontoghiorghes and Herman K. Van Dijk.

CFE 2012 Co-chairs:
Andrew Harvey, Yasuhiro Omori, Willi Semmler and Mark Steel.

CFE 2012 Programme Committee:

A. Amendola, G. Amisano, F. Audrino, M. Billio, F. Canova, S. Chaudhuri, C.W.S. Chen, C. Francq, A. M. Fuertes, L. Khalaf, G. Koop, S.J. Koopman, C.-M. Kuan, R. Leon-Gonzalez, H. Lopes, M. McAleer, M. Paolella, H. Pesaran, J.-Y. Pitarakis, D.S.G. Pollock, P. Poncela, T. Proietti, M. Reale, J. Rombouts, S. Sanfelici, K. Sheppard, M.K.P. So, L. Soegner, G. Storti, E. Tzavalis, J.-P. Urbain, D. Veredas, M. Wagner, Q. Yao, P. Zadrozny, J.M. Zakoian and Z. Zhang.

CSDA Annals of CFE Managing Editors:
David A. Belsley, Erricos J. Kontoghiorghes and Herman K. Van Dijk.

ERCIM 2012 Co-chairs:
Gil Gonzalez-Rodriguez, Geoff McLachlan, Hannu Oja, Marco Riani and Stephen Walker.

ERCIM 2012 Programme Committee:

E. Ahmed, A.M. Alonso, A.C. Atkinson, A. Christmann, A. Delaigle, T. Denoeux, M. Falk, P. Filzmoser, P. Foschi, C. Gatu, L.A. Garcia-Escudero, R. Gerlach, M.I. Gomes, A. Gordaliza, A. Grane, X. He, M.D. Jimenez-Gamero, M. Hubert, T. Kneib, V. Leiva, A. Lijoi, J. Lopez-Fidalgo, A. Luati, G. Mackenzie, A. Mayo, S. Meintanis, I. Molina, D. Morales, D. Paindaveine, M.C. Pardo, S. Paterlini, I. Pruenster, E. Ronchetti, L. Ugarte, S. Van Aelst, I. Van Keilegom, P. Vieu, R. Zamar and A. Zeileis.

Local Organizer:

Department of Statistics, OR and DM, University of Oviedo, Spain

Contents

E389: **Robust segmentation through Hotelling T^2 test**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Andrea Cerioli, Anthony Atkinson

A form of robust segmentation using the forward search that allows the data to determine the number of clusters is described. Similarly to other existing robust methods of cluster analysis we also allow for outliers and so do not force all units to be clustered. Given that our analysis relies on forward plots of particular statistics, in order to provide sensitive inferences about the existence of clusters, it is necessary to augment such graphs with envelopes of the distributions of the statistics being plotted. We focus on a sort of Hotelling t-test where the two samples are made up respectively of the units which are inside and outside the subset at step m of the forward search. This allows for a robust split of each tentative group in two separate subgroups which could be useful also in hierarchical divisive clustering.

E527: **Constrained EM algorithms for Gaussian mixtures of factor analyzers**
Presenter: **Salvatore Ingrassia**, University of Catania, Italy
Co-authors: Francesca Greselin

Mixtures of factor analyzers are becoming more and more popular in the area of model based clustering of high-dimensional data. In data modeling, according to the likelihood approach, it is well known that the loglikelihood function may present spurious maxima and singularities and this is due to specific patterns of the estimated covariance structure. To reduce such drawbacks, we propose a constrained procedure which maintains the eigenvalues of the covariance structure into suitable ranges. Applications of this approach in robust clustering is then outlined.

E563: **Robust conditional tests for elliptical symmetry**
Presenter: **Isabel Rodrigues**, Technical University of Lisbon and CEMAT, Portugal
Co-authors: Ana Bianco, Graciela Boente

A robust procedure for conditional testing the hypothesis of elliptical symmetry is presented. That also provides a specialized test of multivariate normality. We derive the asymptotic behaviour of the robust conditional testing procedure under the null hypothesis of multivariate normality and under an arbitrary elliptically symmetric distribution. Their influence function is also studied. A simulation study allows us to compare the behaviour of the classical and robust test, under different contamination schemes.

ES01 Room 2 MATRIX COMPUTATIONS AND STATISTICS — Chair: Ahmed Sameh

E1060: **On some matrix based methods for analyzing networks and text**
Presenter: **E. Gallopoulos**, University of Patras, Greece
Co-authors: V. Kalantzis, E.-M. Kontopoulou, C. Bekas

Matrix based methods in areas such as sociometrics and text mining has been established for some time but in recent years the field has been growing with great strides in order to cope with the vast increase in the scale of social and other networks and the massive amounts of text available over the Web by attempting the best possible use of the capabilities of modern computer systems. We will present some of our work in network and text mining where matrix iterative methods and statistics meet. Specifically, we will show methods for estimating graph centrality suitable for large networks. We will also show some new developments on a software environment, the Text to Matrix Generator MATLAB Toolbox (TMG), for dimensionality reduction and its application in text mining.

E865: **Window estimation of the general linear model with singular dispersion matrix**
Presenter: **Stella Hadjiantoni**, Queen Mary, University of London, United Kingdom
Co-authors: Erricos John Kontoghiorghes

The re-estimation of the least squares solution after observations are deleted, known as downdating, is often required. The problem of deleting observations from the general linear model (GLM) with a possible singular dispersion matrix is considered. A new method to estimate the downdated GLM, updates the original model with the imaginary deleted observations. This results in a non-positive definite dispersion matrix which comprises complex covariance values. For the case of a non singular dispersion matrix, this updated GLM has been proved to give the same normal equations than the downdated model. The proposed updated GLM is formulated as a generalized linear least squares problem (GLLSP). The main computational tool is the generalized QR decomposition (GQRD) which is employed based on hyperbolic Householder transformations; however, no complex arithmetic is used in practice. The special structure of the matrices and previous computations are efficiently utilized.

E1046: **Computational strategies for non-negativity model selection**
Presenter: **Cristian Gatu**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Erricos John Kontoghiorghes

The problem of regression subset selection under the condition of non-negative coefficients is considered. The straightforward solution would be to estimate the corresponding non-negative least squares of all possible submodels and select the best one. A new computationally efficient procedure which computes only unconstrained least squares is proposed. It is based on an alternative approach to quadratic programming that derives the non-negative least squares by solving the normal equations for a number of unrestricted least squares subproblems. The algorithm generates a combinatorial tree structure that embeds all possible submodels. This innovative approach is computationally superior to the straightforward method. Specifically, it reduces the double exponential complexity to a a single traversal of a tree structure. The computational efficiency of the new selection strategy is further improved by adopting a branch-and-bound device that prunes non-optimal subtrees while searching for the best submodels. The branch-and-bound algorithm is illustrated with a real dataset. Experimental results on artificial random datasets confirm the computational efficacy of the new strategy and demonstrates its ability to solve large model selection problems that are subject to non-negativity constrains.

E1047: **A GSVD strategy for estimating the SEM and SURE models**
Presenter: **Mircea Ioan Cosbuc**, Alexandru Ioan Cuza University of Iasi, Romania
Co-authors: Cristian Gatu, Erricos John Kontoghiorghes

Computational strategies for estimating the Simultaneous Equations Model and the Seemingly Unrelated Regression Equations are proposed. The methods use the Generalized Singular Value Decomposition as the main numerical tool. They exploit the block diagonal and banded structure of the matrices involved in the factorization and do not presume that the covariance matrices are nonsingular. An efficient implementation of these methods is described and its performance is compared to classical methods that treat the models as Generalized Linear Models. Not only are the procedures faster for large systems, but they also preserve a banded structure of the data matrices, allowing the efficient storage of the sparse matrices involved. The algorithms make use of the LAPACK and BLAS functions through their C interfaces.