PROGRAMME AND ABSTRACTS

9th International Conference on

Computational and Financial Econometrics (CFE 2015)

http://www.cfenetwork.org/CFE2015

and

8th International Conference of the
ERCIM (European Research Consortium for Informatics and Mathematics) Working Group on

Computational and Methodological Statistics (CMStatistics 2015)

http://www.cmstatistics.org/CMStatistics2015

Senate House & Birkbeck University of London, UK
12 – 14 December 2015

ISBN 978-9963-2227-0-4

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

Contents

EO178 Room Montague ROBUST STATISTICAL MODELLING Chair: Alfio Marazzi

EO0740: **Higher-order adjustments of the signed scoring rule root statistic**
Presenter: **Laura Ventura**, University of Padova, Italy
Co-authors: Valentina Mameli, Monica Musio

Proper scoring rules can be used as an alternative to the full likelihood, when the aim is to increase the robustness. Proper scoring rule inference is usually based on the first-order approximations to the distribution of the scoring rule estimator or of the scoring rule ratio test statistic. However, several examples illustrate the inaccuracy of first-order methods, even in models with a scalar parameter, when the sample size is small or moderate. Analytical higher-order asymptotic expansions for proper scoring rules, generalizing results for likelihood quantities but allowing for the failure of the information identity, have been previously discussed. However, the calculation of the quantities involved in the analytical adjustments of the signed and signed profile scoring rule root statistic is cumbersome, even for simple models. The aim is to discuss the alternative approach to higher-order adjustments, based on a parametric bootstrap.

EO0807: **Robust state space models**
Presenter: **Eva Cantoni**, University of Geneva, Switzerland
Co-authors: William Aeberhard, Chris Field, Joanna Mills Flemming, Ximing Xu

In fishery science, state space models are often used because they explicitly include both process error in the dynamics and observation error in the data. The estimation of the parameters of the model is usually performed by maximum (marginal) likelihood. But the use of robust statistical methodologies is essential because the data being modelled in fisheries science and management may be subject to large measurement error. In addition, the state and observation processes are only approximations with the consequence that there may well be outlying observations or deviating substructure. We propose a way to robustify the estimation of the parameters of a state space model, which modifies the objective function to introduce robustness. The evaluation of the new marginal objective function being computationally challenging, we use software implementing automatic differentiation to reach our goal. We show simulation results supporting our proposal and applications to areal dataset.

EO1100: **Robust estimation for mixtures of Gaussian factor analyzers**
Presenter: **Alfonso Gordaliza**, Universidad de Valladolid, Spain
Co-authors: Luis Angel Garcia-Escudero, Francesca Greselin, Salvatore Ingrassia, Agustin Mayo-Iscar

Mixtures of Gaussian factors are powerful tools for modeling an unobserved heterogeneous population, offering at the same time dimension reduction and model-based clustering. Unfortunately, the high prevalence of spurious solutions and the disturbing effects of outlying observations, along maximum likelihood estimation, open serious issues. We consider restrictions for the component covariances, to avoid spurious solutions, and trimming, to provide robustness against violations of normality assumptions of the underlying latent factors. A detailed AECM algorithm for this new approach is presented. Simulation results and an application to the AIS dataset show the aim and effectiveness of the proposed methodology.

EO1323: **Robust Bayesian regression**
Presenter: **Marco Riani**, University of Parma, Italy
Co-authors: Anthony Atkinson, Aldo Corbellini

The forward search provides a flexible and informative form of robust regression. We describe two ways of introducing prior information into the regression model used in the search, either through fictitious observations or through prior distributions of the parameters. The relationship between the two methods is established. The extension to the forward search is not entirely straightforward, requiring weighted regression. Forward plots are used to exhibit the effect of correct and incorrect prior information on inferences. Analysis of the 546 observations of the Windsor housing data shows the presence of several outliers and illustrates the effect of incorrect prior specification.

EO268 Room SH349 ANALYSIS OF DATA FROM COMPLEX SURVEYS Chair: Paola Vicard

EO0822: **Empirical likelihood approaches under complex sampling**
Presenter: **Yves Berger**, University of Southampton, United Kingdom

Data are often collected with unequal probabilities from stratified population. Empirical Likelihood is widely used in mainstream statistics. We propose a new empirical likelihood approach for sample data selected with unequal probabilities. In this situation, the standard empirical likelihood approach cannot be applied. Under a set of regularity conditions, the empirical log-likelihood function has an asymptotic chi-squared distribution. The proposed approach does not rely on variance estimates, re-sampling or joint-inclusion probabilities, even when the parameter of interest is not linear and does not have a normal distribution. An alternative approach is the pseudoempirical log-likelihood function which is not entirely appealing from a theoretical point of view, because it relies on a parameter (the design effect) which need to be estimated. A previous approach does not rely on design effect, and can be more accurate than the adjusted pseudoempirical approach. Standard confidence intervals based on variance estimates may give poor coverages, when normality does not hold. This can be the case with skewed data and outlying values. The proposed empirical likelihood confidence interval has good coverages and balanced tail errors even when the sampling distribution of the point estimator is not normal.

EO0610: **Resampling from finite populations: An empirical process approach**
Presenter: **Pier Luigi Conti**, Sapienza University of Rome, Italy
Co-authors: Daniela Marella, Fulvia Fulvia Mecatti

In sampling finite populations, several resampling schemes have been proposed. The common starting point is that, despite its excellent asymptotic properties, Efron's original bootstrap only works for i.i.d. data. This condition is not met in sampling finite populations, because of the dependence among units due to the sampling design. Hence, adaptations are needed to account for the non i.i.d. nature of data. Different versions of the standard bootstrap algorithm have been proposed in the literature. A new class of resampling procedures for finite populations is defined. Such a class appears to provide a unified framework that allows for encompassing other resampling algorithms already proposed. Its main theoretical justification is based on asymptotic, large sample arguments: the probability distribution of the original statistic and its approximation based on resampling converge to the same limit. Technically speaking, it is shown that a "finite population version" of the empirical process and its "resampled form" weakly converge to the same limiting Gaussian process. In a sense, this justification is similar to those given for classical bootstrap.

EO0619: **PC algorithm for complex survey data via resampling**
Presenter: **Daniela Marella**, University Roma Tre, Italy
Co-authors: Paola Vicard

The PC algorithm is one of the main methods for learning the structure of a Bayesian network from sample data. The algorithm uses conditional independence tests for model selection in graphical modeling and it is based on assumption of independent and identically distributed observations (i.i.d). The i.i.d. assumption is almost never valid for sample surveys data since most of the commonly used survey designs employ stratification and/or cluster sampling and/or unequal selection probabilities. The impact of complex design on i.i.d. based procedures can be very severe leading

