Statistical Methods and Models for Complex Data

June 5-7, 2019

Venue: Department of Law, Economics, Management and Quantitative Methods (DEMM)
Piazza Arechi II – Palazzo De Simone
82100 Benevento, Italy

This workshop is supported by:

- *JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)*
- *Institute for Mathematical Science, Tokyo - Japan*
- *University of Sannio, Benevento – Italy*
- *Waseda Research Institute for Science and Engineering, Tokyo – Japan*

Statistical Methods and Models for Complex Data

June 5-7, 2019

Program

Venue: Department of Law, Economics, Management and Quantitative Methods (DEMM)
Piazza Arechi II – Palazzo De Simone
82100 Benevento, Italy

Scientific Committee:

Anna Clara Monti (chair), Fumiya Akashi, Marcella Corduas, Xiaoling Dou, Alessio Farcomeni, Luca Greco, Maria Iannario, Yan Liu, Simona Pacillo, Marco Riani, Masanobu Taniguchi.

This workshop is supported by:

- *JSPS KAKENHI Kiban (S) Grand-in-Aid No. 18H05290 (M. Taniguchi)*
- *Institute for Mathematical Science, Tokyo - Japan*
- *University of Sannio, Benevento – Italy*
- *Waseda Research Institute for Science and Engineering, Tokyo – Japan*

Wednesday JUNE 5, 2019 – *Morning*

11:15 – 11:45: Opening

Anna Clara Monti
Masanobu Taniguchi

Session I (11:45 – 12:45)

chaired by Tommaso Proietti

11:45 – 12:45 *Keynote speaker*:

Time Series Analysis under Non-Standard Settings
Masanobu Taniguchi (*Waseda University, Japan*)

12:45 – 14:00 *Lunch*

Wednesday JUNE 5, 2019 – *Afternoon*

Session II (14:00 – 15:30)

Chaired by Fumiya Akashi

14:00 – 14:30

Heterogeneous Component Multiplicative Error Models for Forecasting Trading Volumes
Giuseppe Storti (*University of Salerno, Italy*)
Joint work with A. Naimoli.

14:30 – 15:00

Robust Time Series Methods in Anti-Fraud
Domenico Perrotta (*Joint Research Center – European Committee – Ispra, Italy)*
Joint work with P.J. Rousseeuw, M. Riani, and M. Hubert

15:00 – 15:30

Robust Model Selection in Nonlinear Tme Series with Trend, Seasonality and Level Shift
Francesca Torti (*Joint Research Center – European Committee – Ispra, Italy)*
Joint work with G. Morelli

15:30 - 16:00 ***Coffee break***

Session III (16:00 – 17:00)

Chaired by Xiaoling Dou

16:00 – 16:30

The Use of Factorial Methods to Explore Complex Social Networks
Giancarlo Ragozini (*University of Naples Federico II, Italy)*

16:30 – 17:00

An Impartial Trimming Approach for Joint Dimension and Sample Reduction
Antonio Lucadamo (*University of Sannio, Italy*)
Joint work with P. Amenta e L. Greco

Thursday JUNE 6, 2019 – *Morning*

Session IV (9:00 – 11:00)

Chaired by Anna Clara Monti

9:00 – 10:00 ***Keynote speaker***

Robust and Consistent Variable Selection in High-dimensional Generalized Linear Models
Elvezio Ronchetti (*University of Geneva, Switzerland*)
Joint work with M. Avella-Medina

10:00 – 10:30

Robust Diagnostic Transformations of Responses that can be Positive or Negative
Marco Riani (*University of Parma, Italy*)
Joint work with A.C. Atkinson and Aldo Corbellini

10:30 – 11:00

Mode Trimming for Robust Estimation and Cluster Analysis
Alessio Farcomeni (*University of Rome La Sapienza, Italy*)
Joint work with F. Dotto, L. Garcìa-Escudero, A. Mayo-Iscar

11:00 – 11:30 ***Coffee break***

Session V (11:30 -13:00)

Chaired by Yan Liu

11:30 – 12:00

Robust Causality Test of Infinite Variance Processes
Fumiya Akashi (*Waseda University, Japan)*
Joint work with M. Taniguchi and A. C. Monti

12:00 – 12:30

Modelling Ordinal Time Series: An Integrated Approach
Domenico Piccolo (*University of Naples Federico II, Italy*)
Joint work with R. Simone and M. Corduas

12:30 – 13:00

Co-clustering Algorithms for Temporal Data represented by Histograms
Rosanna Verde (*University of Campania "Luigi Vanvitelli", Italy*)
Joint work with A. Balzanella

13:00 – 14:15 ***Lunch***

Thursday JUNE 6, 2019 - *Afternoon*

Session VI (14:15 – 15:45)

Chaired by Domenico Piccolo

14:15 – 14:45

Dependence Structures of the B-spline Copulas including the Bernstein ones
Xiaoling Dou (*Waseda University, Japan*)
Joint work with S. Kuriki, G. D. Lin and D. Richards

14:45 – 15: 15

P-splines based Time Series Clustering for Index-tracking Portfolio
Germana Scepi (*University of Naples Federico II, Italy)*
Joint work with A. D'Ambrosio, R. Mattera and M. Scaglione

15:15 – 15:45

Analysis of Censored Data under Skew GEV Distribution
Simona Pacillo (*University of Sannio, Italy*)
Joint work with P. Ribereau and E. Masiello

15:45 – 16:15 *Coffee Break*

Session VII (14:15 – 15:45)

Chaired by Marco Riani

16:15 – 16:45

Robust Link Functions for Ordinal Response Models
Anna Clara Monti (*University of Sannio*)
Joint work with V. Scalera and M. Iannario

16:45 – 17:15

Generalized Residuals for Outlier Detection in Ordinal Regression Models
Maria Iannario (*University of Naples Federico II*)
Joint work with A .C. Monti

Friday JUNE 7, 2019 – Morning

Session VIII (9:00 – 10:30)

Chaired by Elvezio Ronchetti

9:00 – 9:30

Robust Composite Estimators for Variance Component Models
Claudio Agostinelli (*University of Trento, Italy*)
Joint work with Victor J. Yohai

9:30 – 10:00

Robust Estimation of Multilevel Models: a Forward Search Approach
Luigi Grossi (*University of Verona, Italy*)
Joint work with A. Corbellini and F. Laurini

10:00 – 10:30

Labour Market Analysis through Robust Multilevel Models and Transformations
Aldo Corbellini (*University of Parma, Italy*)
Joint work with Marco Magnani and Gianluca Morelli

10:30– 11:00 ***Coffee break***

Session IX (11:00 – 12:00)

Chaired by Masanobu Taniguchi

11:00 – 11:30

Regularized Estimation of High Dimensional Auto- and Cross-Covariance Matrices
Tommaso Proietti *(University of Rome Tor Vergata, Italy)*
Joint work with A. Giovannelli

11:30 – 12:00

Prediction-based Parameter Estimation of Time Series
Yan Liu (*Kyoto University/RIKEN AIP, Japan*)

ABSTRACTS

Claudio Agostinelli
University of Trento, Italy

Robust Composite Estimators for Variance Component Models

Joint work with Victor J. Yohai

As occurs with many statistical models, Maximum Likelihood Estimates (MLE) are not robust for Variance Components models, i.e., MLE may be very much influence by a small fraction of outliers. A very interesting class of S-estimates for Linear Mixed Models was proposed by Victoria-Feser and Copt (2006). These estimates can be thought as a constrained version of the S-estimates for multidimensional location and scatter.

In the classical contamination model a fraction ε of the response vectors are replaced by outliers (Huber-Tukey or casewise contamination model). Victoria-Feser and Copt (2006) show that for this model the breakdown of these estimates is $\varepsilon^* = min(b, 1-b)$. Therefore if $b = 0.5$, we get $\varepsilon^* = 0.5$.

Alqallaf, Van Aelst, Zamar, and Yohai (2009) consider different contamination model for multivariate data: the independent contamination model (or cellwise contamination model). Alqallaf et al. (2009) showed that the breakdown point for the independent contamination model of S-estimate of multivariate location and scatter tends to *0* when $p \to \infty$. The same happens with other popular affine equivariance estimates. The S-estimates proposed by Victoria-Feser and Copt (2006) for Linear Mixed Models have a similar shortcoming: when $p \to \infty$, its breakdown point tends to *0*.

We discuss a new class of estimators for Linear Mixed Models. These estimators are based on S-estimates or τ-estimates of the Mahalanobis distances of two dimensional subvectors of the response variable and they can be thought as robust counterparts of the composite pairwise likelihood estimates proposed by Lindsay (1988). The introduced estimators have good robust behavior for both contamination models: the classical contamination model and the independent contamination model.

Fumiya Akashi
Waseda University, Japan

Robust Causality Test of Infinite Variance Processes

Joint work with Masanobu Taniguchi and Anna Clara Monti

This talk develops a robust causality test for time series models with infinite variance innovation processes. First, we introduce a measure of dependence for vector nonparametric linear processes and derive the limit distribution of the test statistic by Taniguchi et al. (1996) in infinite variance case. Second, we construct a weighted-version of the generalized empirical likelihood (GEL) test statistic, called the self-weighted GEL statistic in time domain. The limit distribution of the self-weighted GEL test statistic is shown to be a standard chi-squared one regardless of whether the model has finite variance or not. Some simulation experiments illustrate desired finite sample performance of the proposed method.

Aldo Corbellini
University of Parma, Italy

Labour Market Analysis through Robust Multilevel Models and Transformations

Joint work with Marco Magnani and Gianluca Morelli

The work presents a robust approach to labour share analysis. The estimate of labour share presents various complexities related to the nature of the datasets to be analysed. Typically, labour share is evaluated by using the discriminant analysis and linear or generalized linear models, that do not take into account the presence of missing values and possible outliers. Moreover, the variables to be considered are often characterized by a high dimensional structure. The proposed approach has the objective of improving the estimation of the model using robust regression techniques and data transformation.

Xiaoling Dou
Waseda University, Japan

Dependence Structures of the B-spline Copulas including the Bernstein ones

Joint work with Satoshi Kuriki, Gwo Dong Lin and Donald Richards

Using B-spline functions, we propose a class of copulas which includes the Bernstein ones (Baker's distributions). The range of correlation of the B-spline copulas is examined, and the Frechet-Hoeffding upper bound is proved to be attained when the number of B-spline functions goes to infinity. On the other hand, the B-spline is well-known as an order complete weak Tchebycheff system, from which the property of total positivity of any order follows for the maximum correlation case. This improves significantly the previous results about the Bernstein copulas. Besides, we also derive an elegant explicit formula for moments of the related B-spline functions on the right-half real line in terms of Stirling numbers of the second kind.

Alessio Farcomeni
University of Rome "Sapienza", Italy

Mode Trimming for Robust Estimation and Cluster Analysis

Joint work with Francesco Dotto, Luis-Angèl Garcìa-Escudero, Agustìn Mayo-Iscar

Several methods for robust estimation and cluster analysis are based on trimming the lowest contributions to the likelihood, plus constraints (e.g., bounding the ratio of the minimal and maximal eigenvalues of the estimated scatter matrix).

In this work we show that, surprisingly enough, removing the largest contributions to the likelihood leads to robust estimation as well. This strategy does not even need extra constraints on the parameter space. Furthermore, while being formally robust, under certain contamination schemes upper trimming procedures are more efficient than lower trimming procedures.

Ultimately, we propose flexible, efficient, and robust procedures based on splitting a fixed trimming proportion between the lowest and largest likelihood contributions.

Luigi Grossi
University of Verona, Italy

Robust Estimation of Multilevel Models: a Forward Search Approach

Joint work with Aldo Corbellini and Fabrizio Laurini

Robustness of standard regression models have been studied quite extensively. When repeated measures are available, the methodological framework is generalized to multilevel models, for which little is known in term of robustness, even in the simplest case of ANOVA. We present a sequential forward search algorithm for multilevel models that allows robust and efficient parameters estimation in presence of outliers, and it avoids masking and swamping.

The influence of outliers, if any is inside the dataset, will be monitored at each step of the sequential procedure, which is the key element of the forward search.

There are peculiar features when the forward search is applied to multilevel models. Such features pose new computational challenges, as some restrictions, that make the sub-models identifiable at every step, are required.

Preliminary results on simulated data have highlighted the benefit of adopting the forward search algorithm, which can reveal masked outliers, influential observations and show hidden structures. An application to real data is also illustrated, where trades of coffee to European countries are analyzed to identify outliers that might be linked to potential frauds.

Maria Iannario
Department of Political Sciences, University of Naples "Federico II", Italy

Generalized Residuals for Outlier Detection in Ordinal Regression Models

Joint work with Anna Clara Monti

In ordinal response models residual diagnostics is not commonly used for detecting outliers because generalized models for ordinal data, like cumulative models with proportional assumption, do not produce standard residuals. In this talk we analyse the generalized residuals introduced by Franses and Paap (2004) which provvide signaling tools for observations which are incoherent with the model and therefore potentially anomalous. We also discuss the more recent alternatives: the sign-based statistics (SBS) introduced by Li and Shepherd (2012) and the surrogate residuals introduced by Liu and Zhang (2017) which present nice proprieties even if they are mainly considered for checking the structure of the model, the choice of the link function and possible mixture populations or heteroscedasticity. Hence, we show the utility of the generalized residuals in robust estimation contexts.

Yan Liu
Kyoto University/RIKEN AIP, Japan

Prediction-based Parameter Estimation of Time Series

We discuss the parameter estimation problem of time series models from the perspective of the prediction problem. The prediction error and the interpolation error are regarded as a contrast function for the parameter estimation. The parameters are estimated by the minimum contrast estimation. The new contrast functions are not contained in the class of either location or scale disparities. The estimator is shown to be asymptotic consistent. The asymptotic distribution of the estimator depends on the assumptions on the stochastic process. In particular, the estimator is robust against the fourth order cumulants when the process is Gaussian. The Whittle estimator is asymptotically efficient in the sense that the family of parametric spectral densities is truly specified, while the new class contains robust members to the randomly missing observations from the stationary process.

Antonio Lucadamo
University of Sannio, Italy

An Impartial Trimming Approach for Joint Dimension and Sample Reduction

Joint work with Pietro Amenta e Luca Greco

A robust version of Reduced and Factorial k-means is proposed, that is based on the idea of trimming. Reduced and Factorial k-means are data reduction techniques well suited for simultaneous dimension reduction through PCA and clustering. The occurrence of data inadequacies can invalidate standard analyses. Actually, contamination in the data at hand can hide the underlying clustered structure of the data. An appealing approach to develop robust counterparts of Factorial and Reduced k-means is given by impartial trimming. The idea is to discard a fraction of observations that are selected as the most distant from the centroids.

Anna Clara Monti
University of Sannio, Italy

Robust Link Functions for Ordinal Response Models

Joint work with Valentino Scalera e Maria Iannario

Cumulative models are widespread in various fields to describe the dependence of an observed ordinal response on the subjects' covariates. Outlying covariates as well as incoherent responses may affect the reliability of the Maximum Likelihood estimators and that of the derived testing procedures. However the various link functions, which provide the relationship between the linear predictor and the cumulative probabilities, differ in terms of sensitivity to outliers, and a suitable link function can limit the impact of anomalous data in the estimation process. Consequently conditions are derived which allow to evaluate the properties of the link functions in terms of robustness, either when the covariate are outliers free or when extreme design points may occur among regressors. The purpose is to carry out robust inference through the usual likelihood function by an appropriate choice of the link function.

Simona Pacillo
University of Sannio, Italy

Analysis of Censored Data under Skew GEV Distribution

Joint work with Pierre Ribereau and Esterina Masiello

This work addresses the problem of estimating, in the presence of censored data, the parameters of the SGEV distribution, an extension of the generalized extreme value (GEV) distribution obtained by introducing a parameter which regulates the skewness (P. Ribereau et. al, 2016). The estimation is carried out by the maximum likelihood method and the probability-weighted moment method. The proposed idea is evaluated by simulation study.

Domenico Perrotta
European Commission, Joint Research Centre, Ispra (VA), Italy

Robust Time Series Methods in Anti-Fraud

Joint work with P.J. Rousseeuw , M. Riani, and M. Hubert

The protection of the financial interests of the European Union (EU) is enshrined in the founding treaties and is a priority for the European institutions. The Joint Research Centre (JRC) of the European Commission works in this context with the European Anti-fraud Office and its partners in the EU member states. Fraudulent activities that cost European taxpayers money and that the JRC tries to detect using statistical methods, include the evasion of import duties, deflection of trade, mis-declaration of product origin, undervaluation, mis-description of goods at import, irregularities in payments of export refunds, trade-based money laundering.

The patterns to be detected include upward spikes in trade flows, i.e. sudden, unexpected, unprecedented increases in trade flows at a point in time, more structural changes such as level shifts, and price outliers, i.e. trade flows with price significantly larger or smaller from the prices for comparable trade flows. For this, robust statistics is applied on a daily basis with lot of success and huge financial impact. This contribution illustrates a new framework introduced by Rousseeuw et al. [2018] for detecting outliers (isolated or consecutive) and level shifts in short time series that may have trend and seasonal patterns, possibly relevant in our anti-fraud context.

The time series of interest refer to monthly trade volumes of products imported in the EU in a period that may cover several years. There are several thousands of relevant combinations of a product at fraud risk, a country of origin and a country of destination to analyze each month. This requires an automatic and computationally efficient approach that is able to report accurate information on outliers and the positions and amplitudes of level shifts.

The framework is based on a parametric approach to estimate level shifts that differs from the nonparametric smoothing methods in Fried and Gather [2007] or robust methods for REGARIMA models (Bianco et al. [2001]). The approach combines ideas from the FastLTS algorithm for robust regression with alternating least squares. The tool is complemented by the "double wedge plot", a new graphical display which indicates outliers and potential level shifts that can be looked at whenever the automatic monitoring system detects a significant level shift.

We illustrate the framework on real trade time series but the properties are carefully studied using the well-known airline data (Box and Jenkins [1976]), contaminated versions of it, and data simulated according to our model. Software for this new framework is available in the MATLAB FSDA toolbox (http://fsda.jrc.ec.europa.eu or http://rosa.unipr.it/fsda.html).

Domenico Piccolo
University of Naples Federico II, Italy

Modelling Ordinal Time Series: an Integrated Approach

Joint work with Rosaria Simone and Marcella Corduas

Most of surveys concerning ordinal evaluations about private products, public services, voting intentions, etc. are repeated during time in order to check for trends and for the effectiveness of policy actions.

Then specific approaches should be pursued to take into account both the discrete nature of data and their intrinsic serial correlation.

This paper introduces dynamic models for the intrinsic components of rating data, with focus on a specific form of serial dependence in their structure. The proposal is discussed on the basis of time series derived from the ISTAT survey concerning price expectations in Italy.

Tommaso Proietti
University of Rome Tor Vergata, Italy

Regularized Estimation of High Dimensional Auto- and Cross-Covariance Matrices

Joint work with Alessandro Giovannelli

The estimation of the (auto-and) cross-covariance matrices of respectively a stationary random process plays a central role in prediction theory and time series analysis. In the univariate framework, we proposed an estimator based on regularizing the sample partial autocorrelation function, via a modified Durbin-Levinson algorithm that receives as an input the banded and tapered sample partial autocorrelations and returns a consistent and positive definite estimator of the autocovariance matrix. We discuss multivariate generalizations, based on a regularized Whittle algorithm, shrinking the lag structure towards a finite order vector autoregressive system (by penalizing the partial canononical correlations), on the one hand, and shrinking the cross-sectional covariance towards a diagonal target, on the other. As the shrinkage intensity increases, the multivariate system converges to a set of unrelated univariate processes. We illustrate the merits of the proposal with respect to the problem of out of sample prediction and the estimation of the spectral density of sea surface temperature time series.

Giancarlo Ragozini
University of Naples Federico II, Italy

The Use of Factorial Methods to Explore Complex Social Networks

Nowadays, networks describe the underlying model of numerous phenomena observed in many different fields, ranging from biology and physics, to social and human science. Networks often present complex structures. They can be both multi-modal and multi-relational. Moreover, each relationship can be observed across time occasions.
In this paper, first we will introduce the main concepts related to complex social networks, and next, we will discuss how the complexity of networks can be treated by means of suited factorial methods. In particular, we will present how to adapt Multiple Correspondence Analysis to analyze and graphically represent two-mode networks; how to use of multiple factor analysis to deal with time-varying two-mode networks, and how to exploit the DISTATIS method for the analysis of Multiplex one-mode networks

Finally, we will illustrate the performance of these techniques through several applications to real data related to collaboration networks in different contexts: technological districts, theatrical coproduction, academic collaboration.

Marco Riani
University of Parma, Italy

Robust Diagnostic Transformations of Responses that can be Positive or Negative

Joint work with Anthony C. Atkinson and Aldo Corbellini

The parametric family of power transformations to approximate normality analyzed by Box and Cox can be applied only to positive data. Yeo and Johnson generalized this transformation to allow for the inclusion of zero and negative response values, which arise, in our examples, in data on the profitability of investment funds and firms.

The talk describes the use of constructed variables to provide an approximate score statistic for the transformation which, like a test in standard regression, is based on aggregate properties of the data. Robust analysis with the forward search provides a series of subsets of the data of increasing size, ordered by closeness to the fitted model for each subset size. The "fan plot" of the statistics for these subsets against subset size clearly indicates the effect of individual observations, especially outliers, on the estimated transformation parameter.

In some cases it is not clear that positive and negative observations require the same transformation. The score test is extended to determine whether positive or negative observations require distinct transformations. This procedure leads to an informative extended fan plot, including a test that the two transformations are the same. Extra insight into data structures in the examples comes from brushing linked plots. If time allows, there will be some discussion of the distributions of the test statistics.

Elvezio Ronchetti
University of Geneva, Switzerland

Robust and Consistent Variable Selection in High-dimensional Generalized Linear Models

Joint work with Marco Avella-Medina

Generalized linear models are popular for modelling a large variety of data. We consider variable selection through penalized methods by focusing on resistance issues in the presence of outlying data and other deviations from assumptions. In particular, we discuss the connections between robustness, sparsity, and oracle properties and the extension of basic robustness concepts to the high-dimensional setting. Specifically, we highlight the weaknesses of widely-used penalized M-estimators, propose a robust penalized quasi-likelihood estimator, and show that it enjoys oracle properties in high dimensions and is stable in a neighborhood of the model. We illustrate its finite-sample performance on simulated and real data.

Germana Scepi
University of Naples Federico II, Italy

P-splines based Time Series Clustering for Index-tracking Portfolio

Joint work with Antonio D'Ambrosio, Raffaele Mattera and Marco Scaglione

The growing interest in studying phenomena changing over time have brought to the development of many clustering techniques which deal with time course data, making available many new techniques and methods suitable for time series application. Recently, clustering techniques based on P-spline smooths have been used in order to deal with financial time series without data pre-processing.

Starting by this proposal, we introduce a strategy dealing with a traditional problem in quantitative finance, as the construction of an index-tracking portfolio. The index-tracking is a passive form of investment and it faces the problem of reproducing the performance of a stock market index by considering a portfolio of assets comprised in the index itself. For this reason, tracking index strategies have become popular because they offer attractive risk-return profiles at low costs.

According to this strategy, an investor faces a lower costs than a full replication strategy because he doesn't purchase all the stocks of a particular market index but only a part of it.

Indeed, the goal of the proposed investment strategy is to build a portfolio of assets which is co-integrated with the index and reproduces as close as possible the return structure of the selected index, the so called "tracker fund".

In the following, after a brief description of both the method and the faced problem, we provide an explanation of the portfolio building strategy. Some experimental results, based on experiments conducted on several data sets with different dissimilarity measures, are also presented.

Giuseppe Storti
University of Salerno, Italy

Heterogeneous Component Multiplicative Error Models for Forecasting Trading Volumes

Joint work with Antonio Naimoli

We propose a novel approach to modelling and forecasting high-frequency trading volumes, revisiting the Component Multiplicative Error Model of Brownless et al. (2011) by a more flexible specification of the long-run component which is based on a Heterogeneous MIDAS polynomial structure. This uses an additive cascade of MIDAS polynomial filters moving at different frequencies in order to reproduce the changing long-run level and the persistent autocorrelation structure of high-frequency trading volumes. The merits of the proposed approach are illustrated by means of an application to six stocks traded on the XETRA market in the German Stock Exchange.

Masanobu Taniguchi
Waseda University, Japan

Time Series Analysis under Non-Standard Settings

This talk delivers a series of recent developments for time series analysis under non-standard settings. Concretely the following subjects are addressed:

(i) Non-regular estimation for discontinuous spectra.

(ii) Jackknifed Whittle estimation.

(iii) Higher order asymptotic theory for semi-parametric spectral estimation.

(iv) Asymptotics of realized volatility with ARCH(∞) microstructure noise.

(v) Model selection for contiguously specified spectral family.

The results show unusual aspects of the asymptotic theory, and open a new horizon for time series.

Francesca Torti
European Commission, Joint Research Centre, Ispra (VA), Italy

Robust Model Selection in Nonlinear Time Series with Trend, Seasonality and Level Shift

Joint work with Gianluca Morelli

Outliers and structural changes are commonly encountered in economic time series. The presence of these extraordinary events can easily mislead the conventional time series analysis procedures resulting in erroneous conclusions, especially if the starting point in which the change in level takes place is unknown. Widespread recognition of the importance of level shifts or structural breaks in time series has motivated a great deal of econometric research. In this paper we provide a robust uni_ed framework which enables to treat, outliers, unknown level shifts, changes in the seasonal pattern and model selection in a integrated environment.

Rosanna Verde
University of Campania Luigi Vanvitelli, Italy

Co-clustering Algorithms for Temporal Data Represented by Histograms

Joint work with Antonio Balzanella

This paper focuses on co-clustering algorithms for histogram data and their application to temporal data. We consider a dataset made of multiple high dimension time series recorded by a sensor network and propose to summarize non-overlapping batches of each time series through histograms. This allows to reduce the data dimensionality keeping the information about the data generation process and characteristic values such as the mean, the standard deviation, or the quantiles. We introduce two co-clustering algorithms for analyzing such kind of histogram based dataset which extend the double k-means algorithm. The first proposed algorithm, named ”distributional double Kmeans (DDK)”, is based on the L2 Wasserstein distance and represents each block of units and variables, through a histogram centroid. The second algorithm, named adaptive distributional double Kmeans (ADDK), is an extension of DDK with automated variable weighting. By means of an application to real sensor data, we show that the proposed algorithms are able to get a partitioning of series which allows to discover groups of sensors recording similar data over time; to get an overview of the evolution in the monitored phenomenon, by looking at the partition of the variables; to evaluate the contribution of time periods to the definition of the optimal partitioning, by looking at the variable weights.