



Workshop

“Model Selection, Nonparametrics and Dependence Modeling”

July 8 – 9, 2013, Ensai, Rennes, France

Organizing committee

Gerda Claeskens (KU Leuven)

Irène Gijbels (KU Leuven)

Paul Janssen (U. Hasselt)

Anneleen Verhasselt (U. Hasselt)

Valentin Patilea (CREST-Ensai, Rennes)



KU LEUVEN



Workshop “Model Selection, Nonparametrics and Dependence Modeling”

July 8 – 9, 2013, Ensai, Rennes, France

Program

Monday 8 July 2013

9:30 – 9:40	Opening		
<i>Chair: Paul Janssen</i>			
9:40 – 10:30	Christian Genest	A fresh look at extreme-value dependence modeling	p. 8
10:30 – 11:00	Marco Munda	Choosing a frailty distribution to model dependence in clustered survival data	p. 15
11:00 – 11:30	Coffee break		
11:30 – 12:20	Claudia Czado	Vine copulas and their applications	p. 7
12:20 – 12:50	Nathan Uyttendaele	Nonparametric reconstruction of the tree structure of a nested Archimedean copula	p. 16
12:50 – 13:50	Lunch		
13:50 – 14:20	Posters	Poster presenters are present	
<i>Chair: Anneleen Verhasselt</i>			
14:20 – 15:10	Aurore Delaigle	Nonparametric regression with homogeneous group testing data	p. 7
15:10 – 15:40	Dominik Liebl	Principal component analysis for dependent functional data on random domains	p. 14
15:40 – 16:10	Coffee break		
<i>Chair: Valentin Patilea</i>			
16:10 – 16:50	Min-Yeng Cheng	Nonparametric modeling of dynamical seasonality and trend with heteroscedastic and dependent errors	p. 6
16:50 – 17:20	Hamdi Raïssi	Testing instantaneous causality in presence of non constant unconditional covariance	p. 15
17:20 – 17:50	Lionel Truquet	A semi-parametric locally stationary ARCH Model	p. 16
19:45 –	Conference dinner		



Workshop

“Model Selection, Nonparametrics and Dependence Modeling”

July 8 – 9, 2013, Ensai, Rennes, France

Program

Tuesday 9 July 2013

<i>Chair: Irène Gijbels</i>			
9:00 – 9:50	Johanna Neslehova	Copula-based tests of independence for discrete or mixed data	p. 9
9:50 – 10:20	Axel Bücher	When uniform convergence fails: Non-smooth empirical copula processes with applications to estimation and testing	p. 10
10:20 – 10:50	Candida Geerdens	Copula based flexible modeling of associations between clustered event times	p. 10
10:50 – 11:15 Coffee break			
<i>Chair: Aurore Delaigle</i>			
11:15 – 11:45	Maarten Jansen	Information criteria for high-dimensional variable selection without shrinkage	p. 13
11:45 – 12:15	Inge Vrinssen	Robust variable selection using the nonnegative garrote method	p. 17
12:15 – 12:45	Klaus Herrmann	Studying the sum of two dependent random variables via copula modeling	p. 11
12:45 – 14:00 Lunch + Posters Poster presenters are present			
<i>Chair: Gerda Claeskens</i>			
14:00 – 14:50	Bruce E. Hansen	Model averaging, sieve regression, and minimax efficiency	p. 8
14:50 – 15:20	Paul Janssen	The Bernstein estimator of the copula function and of the copula density: large sample properties	p. 14
15:20 – 15:30 Short break			
<i>Chair: Maarten Jansen</i>			
15:30 – 16:00	David Hughes	Monotone local linear estimation of transducer functions	p. 12
16:00 – 16:30	Anneleen Verhasselt	P-splines quantile regression estimation in varying coefficient models	p. 16
16:30 – 16:40	Closing session		



Workshop

“Model Selection, Nonparametrics and Dependence Modeling”

July 8 – 9, 2013, Ensai, Rennes, France

Program

Posters

Roel Braekers	Modelling unbalanced clustered multivariate survival data via Archimedean copula functions.	p. 18
Sergio Calderón	Bayesian analysis of multivariate threshold autoregressive models with missing data	p. 19
Khalil El Waled	Nonparametric estimation problem for some time-periodic-drift stochastic differential equation	p. 19
Lore Dirick	The analysis of credit risk data: variable selection for a mixture cure model	p. 19
Takeshi Emura	Gene selection for survival data under dependent censoring, a copula-based approach	p. 20
Weiyu Li	Iterative estimation for conditional estimating equations	p. 21
Samuel Maistre	Testing for lack of fit in functional regression models	p. 21
Luísa Pereira	Almost sure central max-limit theorem for nonstationary random fields	p. 21
Eugen Pircalabelu	Model selection for graphical models using the focused information criterion	p. 22
Dominik Sznajder	Stochastic monotonicity tests	p. 23

Contents

Abstracts of invited presentations	6
Abstracts of contributed presentations	10
Abstracts of poster presentations	18

Abstracts of invited presentations

Nonparametric modeling of dynamical seasonality and trend with heteroscedastic and dependent errors

Ming-Yen Cheng

National Taiwan University, Taiwan

E-mail: cheng@math.ntu.edu.tw

Seasonality (or periodicity) and trend are features describing an observed sequence, and extracting these features is an important issue in many scientific fields. However, it is not an easy task for existing methods to analyze simultaneously the trend and dynamics of the seasonality such as time-varying frequency and amplitude, and the adaptivity of the analysis to such dynamics and robustness to heteroscedastic, dependent errors is not guaranteed. These tasks become even more challenging when there exist multiple seasonal components. We propose a nonparametric model to describe the dynamics of multi-component seasonality, and investigate the recently developed Synchronizing transform (SST) in extracting these features in the presence of a trend and heteroscedastic, dependent errors. The identifiability problem of the nonparametric seasonality model is studied, and the adaptivity and robustness properties of the SST are theoretically justified in both discrete- and continuous-time settings. Consequently we have a new technique for de-coupling the trend, seasonality and heteroscedastic, dependent error process in a general nonparametric setup. Results of a series of simulations are provided, and the incidence time series of varicella and herpes zoster in Taiwan and respiratory signals observed from a sleep study are analyzed. This is joint work with Yu-Chun Chen and Hau-Tieng Wu.

Vine copulas and their applications

Claudia Czado

Technische Universität München, Germany

E-mail: cczado@ma.tum.de

Multivariate data sets often show complex dependency patterns. One approach is to model these data sets is copula based. While there are many bivariate copula families, multivariate copula families such as the elliptical or Archimedean copulas show not enough flexibility with regard to symmetry and tail dependence. The class of vine copulas tries to fill this gap. Vine copulas are built using bivariate building blocks. In particular the multivariate copula model consists out of three components. The first component is a sequence of trees identifying the associated copulas to bivariate conditional distributions, the second one their corresponding bivariate copula families and the third one the parameter values corresponding to the chosen bivariate families. I will introduce this pair copula construction of multivariate copula and discuss their parameter estimation and model selection of up to all three components. The usefulness of this model class will be illustrated through several real data analyses.

References:

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics* 44(2), 182–198.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In P. Jaworski, F. Durante, W. Härdle, and T. Rychlik (Eds.), *Copula Theory and Its Applications*. Berlin: Springer.

Czado, C., E. Brechmann, and L. Gruber (2013). Selection of vine copulas. In F. D. P. Jaworski and W. K. H (Eds.), *Copulae in Mathematical and Quantitative Finance*. Springer.

Further references on vines and their applications can be found at

<http://www-m4.ma.tum.de/forschung/vine-copula-models/#c662>

Nonparametric regression with homogeneous group testing data

Aurore Delaigle

University of Melbourne, Australia

E-mail: A.Delaigle@ms.unimelb.edu.au

Group testing data refer to binary data that are observed in an aggregated groupwise form. Typically, the variable of interest is the disease status of a patient, measured for example through a blood test. Instead of testing the blood of each patient in the study, the patients are pooled in groups, and only the pooled blood is tested for the disease. In those studies, it is often of interest to estimate the probability of having the disease given some observed covariates. We introduce new nonparametric estimators for this problem, under several possible grouping settings, and study the effect that grouping the data has on the quality of estimators.

A fresh look at extreme-value dependence modelling

Christian Genest

McGill University, Canada

E-mail: cgenest@math.mcgill.ca

Extreme-value dependence between continuous random variables is characterized by Pickands' dependence function. Several nonparametric techniques have recently been proposed to estimate this object, including a Bayesian approach. This talk will review their relative merits and will explore what classical functional data analytical methods have to say in this regard.

Model averaging, sieve regression, and minimax efficiency

Bruce Hansen

University of Wisconsin, USA

E-mail: behansen@wisc.edu

This talk will review recent developments in the theory of model averaging for regression estimation. We introduce an averaging nonparametric sieve regression estimator, which can achieve a lower integrated mean-squared error (IMSE) than least-squares estimators. We develop a uniform asymptotic approximation for the IMSE, introduce a cross-validation criterion for selection of the averaging weights, and show that that cross-validation weight selection is asymptotically IMSE equivalent to the infeasible best averaging sieve approximation. We also examine the asymptotic risk of nested least-squares averaging estimators when the averaging weights are selected to minimize a penalized least-squares criterion. We find conditions under which the asymptotic risk of the averaging estimator is globally smaller than the unrestricted least-squares estimator. For the Mallows averaging estimator under homoskedastic errors the condition takes the simple form that the regressors have been grouped in sets of four or larger. This condition is a direct extension of the classic theory of James-Stein shrinkage. This discovery suggests the practical rule that implementation of averaging estimators be restricted to models in which the regressors have been grouped in this manner. Our simulations show that this new recommendation results in substantial reduction in mean-squared error relative to averaging over all nested sub-models. Averaging between two nested models is equivalent to shrinkage. We explore the efficiency properties of shrinkage in high dimensional parametric models using a local asymptotic framework as in Hjort and Claeskens (2003). We show that if the shrinkage dimension exceeds two, the asymptotic risk of shrinkage estimators is strictly less than that of the MLE. This reduction holds globally in the parameter space. We also provide a new local asymptotic minimax efficiency bound. We show that shrinkage estimators asymptotically achieve this local asymptotic minimax bound when the shrinkage dimension is high. This theory is a combination and extension of standard asymptotic efficiency theory (Hájek, 1972) and local minimax efficiency theory for Gaussian models (Pinsker, 1980).

Copula-based tests of independence for discrete or mixed data

Johanna Neslehova

McGill University, Canada

E-mail: neslehova@math.mcgill.ca

Copula-based tests of independence for discrete or mixed data New statistics are proposed for testing the hypothesis that arbitrary random variables are mutually independent. These tests are consistent and well-behaved for any type of data, even for sparse contingency tables and tables whose dimension depends on the sample size. The statistics are Cramr-von Mises and Kolmogorov-Smirnov type functional of the empirical checkerboard copula. The asymptotic behavior of the corresponding empirical process will be characterized and illustrated; it will also be shown how replicates from the limiting process can be generated using a multiplier bootstrap procedure. As will be seen through simulations, the new tests are considerably more powerful than those based on the Pearson chi squared, likelihood ratio, and Zelterman statistics often used in this context.

Abstracts of contributed presentations

When uniform convergence fails: Non-smooth empirical copula processes with applications to estimation and testing

Axel Bücher, Johan Segers, Stanislav Volgushev
University of Bochum, Germany
E-mail: axel.buecher@rub.de

The empirical copula process plays a central role for statistical inference on copulas. Recently, Segers (Bernoulli, 2012) investigated weak convergence of this process with respect to the uniform norm under certain necessary smoothness conditions on the copula. In this talk, it is demonstrated that these conditions are not satisfied for several popular copula models. As a circumvention, we show that a weak convergence result can be obtained under substantially weaker smoothness conditions if the uniform norm is replaced by a suitably weaker, albeit sufficiently strong metric. Convergence with respect to this metric is equivalent to epi- and hypoconvergence of functions and, among other things, implies L^p convergence for any $p > 0$. We demonstrate the usefulness of the results by several statistical applications including minimum distance estimation and calculation of power curves for goodness-of-fit tests under local alternatives.

Segers, J. (2012). Asymptotics of empirical copula processes under non- restrictive smoothness assumptions. *Bernoulli*, 18(3):764–782.

Copula based flexible modeling of associations between clustered event times

Candida Geerdens, Paul Janssen
Hasselt University, Belgium
E-mail: candida.geerdens@uhasselt.be

Grouped survival data are characterized by the presence of correlation between event times within the same cluster. Copula models provide a way to account for this clustering. The focus within copula research is however often limited to the bivariate case, while the multivariate setting remains less explored. Following the idea of Joe and Hu (1996), the creation of a multivariate copula as a mixture of max-infinitely divisible bivariate copulae is investigated. The method allows to build multi-dimensional copulae with flexible and possibly symmetric dependence structures. It includes (in some way) other multivariate constructions like the regular Archimedean as well as the nested Archimedean copulae. The Joe and Hu-construction is applied

within the context of right-censored survival data and is fitted by a likelihood approach where the vast amount of copula derivatives present in the likelihood is approximated by finite differences. The numerical performance of this approximation is investigated in a small simulation study. An illustration on data from the field of veterinary medicine is included and supplemented with a model selection strategy to choose the most appropriate copula description for the data at hand.

Reference:

Joe, H. and Hu, T. (1996). Multivariate distributions from mixtures of max-infinitely divisible distributions. *Journal of Multivariate Analysis*, 57, 240 – 265.

Studying the sum of two dependent random variables via copula modeling

Klaus Herrmann, Irène Gijbels

KU Leuven, Belgium

E-mail: klaus.herrmann@wis.kuleuven.be

This talk focuses on the impact of stochastic dependence when applying a summation operation to components of a random vector. We therefore consider a random variable Z derived from a random vector $\mathbf{X} = (X_1, X_2)$ as a weighted sum over the components of \mathbf{X} , i.e. $Z = \omega_1 X_1 + \omega_2 X_2$, where ω_1 and ω_2 are given weights. While for special cases, such as independence or joint normality of \mathbf{X} , properties of Z are well known, we consider dependence structures introduced by arbitrary, absolutely continuous copulas. This allows us to study and model effects of dependence structures beyond the scope of standard models. One such application is portfolio allocation using copula-GARCH models, where our approach naturally extends the one given in [3]. If the components of \mathbf{X} are bounded from below, the distribution of Z can be computed using the (G)AEP algorithm of [1] and [2]. However, for applications in most fields, the assumption of boundedness is too restrictive. Our framework therefore relaxes this assumption and develops a full analysis for unbounded risks including an efficient numerical treatment comparable to the (G)AEP algorithm. We further show that this approach is connected to integration along a path in the unit square only determined by the margins of X_1 and X_2 . To bridge the gap to applications in finance, we also consider conditional expectations of the form $\mathbb{E}[Z|Z \leq z]$ as functions of z . When z is chosen as a quantile of Z , i.e. $z = F^{-1}(p)$ for p in $(0, 1)$, this resembles the expected shortfall commonly encountered in actuarial and financial literature. We illustrate the results presented during the talk with examples of bounded and unbounded margins and compare the competing approaches to standard numerical techniques such as Monte Carlo simulation and numerical integration.

References:

- [1] Arbenz, P., Embrechts, P. and Puccetti, G. (2011), The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. *Bernoulli*, 17(2), 562–591.
- [2] Arbenz, P., Embrechts, P. and Puccetti, G. (2012), The GAEP algorithm for the fast computation of the distribution of a function of dependent random variables. *Stochastics: An International Journal of Probability and Stochastic Processes*, 84(5-6), 569–597.
- [3] Huang, J., Lee, K., Liang, H. and Lin, W. (2009), Estimating value at risk of portfolio by conditional copula-GARCH method. *Insurance: Mathematics and Economics*, 45, 315–324.

Monotone local linear estimation of transducer functions

David Hughes

University of Liverpool, U.K.

E-mail: David.Hughes@liverpool.ac.uk

Local polynomial regression has received a great deal of attention in the past. It is a highly adaptable regression method when the true response model is not known. However, estimates obtained in this way are not guaranteed to be monotone. In some situations the response is known to depend monotonically upon some variables. Various methods have been suggested for constraining nonparametric local polynomial regression to be monotone. The earliest of these is known as the Pool Adjacent Violators algorithm (PAVA) and was first suggested by Brunk (1958). Kappenman (1987) suggested that a non-parametric estimate could be made monotone by simply increasing the bandwidth used until the estimate was monotone.

More recently Hall and Huang (2001) have suggested a tilting method based on changing the weights used in the estimating procedure. The weights are modified in such a way that they are as close as possible to the original weights. Finally Dette et al. (2006) have suggested a monotonicity constraint which they call the DNP method. Their method involves calculating a density estimate of the unconstrained regression estimate, and using this to calculate an estimate of the inverse of the regression function.

Fan, Heckman and Wand (1995) generalized local polynomial regression to quasi-likelihood based settings. Obviously such estimates are not guaranteed to be monotone, whilst in many practical situations monotonicity of response is required. In this talk we discuss how the above mentioned monotonicity constraint methods can be adapted to the quasi-likelihood setting. We are particularly interested in the estimation of monotone psychometric function and, more generally, biological transducer functions, for which the response is often known to follow a distribution which belongs to the exponential family. We consider some of the key theoretical properties of the monotone local linear estimators in the quasi-likelihood setting. We compare these four methods by means of a simulation study. We investigate a variety of response models, including binary, Poisson and Gaussian. In each study we calculate monotone estimates of the response curve using each method and compare their bias, variance, MSE and MISE. We also apply these methods to analysis of data from various hearing and vision studies. We show some of the deficiencies of using local polynomial estimates, as opposed to local likelihood estimates.

References:

- H. D. Brunk. On the estimation of parameters restricted by inequalities. *Ann. Math. Statist.*, 29:437-454, 1958.
- H. Dette, N. Neumeier and K. F. Pilz. A simple nonparametric estimator of a strictly monotone regression function. *Bernoulli*, 12(3): 469-490, 2006.
- J. Fan, N. E Heckman, and M. P. Wand, Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *Journal of the American Statistical Association*, 90(429):141-150, 1995.
- P. Hall and L.-S. Huang. Nonparametric regression subject to monotonicity constraints. *Ann Statist.*, 29(3):624-647, 2001.
- R. F Kappenman. Nonparametric estimation of dose response curves with application to ed50 estimation. *Journal of Statistical Computation and Simulation*, 28(1):13-28, 1987.

Information criteria for high-dimensional variable selection without shrinkage

Maarten Jansen

Université libre de Bruxelles, Belgium

E-mail: maarten.jansen@ulb.ac.be

The presented work (under revision with *Biometrika*) constructs information criteria based on penalized least squares and penalized likelihood for use in variable selection from observations in high-dimensional models. The penalties used in Mallows Cp and Akaike's information criteria provide an unbiased estimator for the loss or distance w.r.t. the unknown true model. This unbiasedness, however, only holds for the evaluation of the quality of a given model. When the criterion is used for optimization over a set of candidate models, the output is affected by the observational errors, resulting in additional bias. The mechanism behind this additional bias is that in high-dimensional full models it is easy for the observational errors to combine into an apparent but false significant parameter. Apart from the obvious, very significant parameters, these false positives present themselves as the most valuable candidates for selection into the model, whereas in reality they are the worst: their estimators, selected through optimization, carry more noise than an arbitrarily selected parameter. The discrepancy between the appearance of those variables as the best candidates and the reality of being the worst candidates, can be described and visualized as a mirror effect. The description of the mirror relies on an oracle variable selection that does not depend on the observational errors. The mirror effect by false positives can be reduced or even undone by shrinkage estimators, such as in soft-thresholding or lasso. Shrinkage, however, introduces new bias. Moreover, as shrinkage is tolerant for false positives, it leads to a considerable overestimation of the number of nonzeros. Therefore, we present a criterion for minimum loss variable selection without shrinkage. By exploration of the mirror effect, we demonstrate that careful minimization of the hard threshold loss, being aware of the mirror effect, results in much sparser models. Although hard thresholding is harder than soft thresholding (mathematically, computationally), our work illustrates that it yields much better minimum loss results. Our work also illustrates that optimization of the smoothing parameter (or, equivalently, optimization of the model size) is a problem of different nature than the variable selection for given model size. For the latter problem, shrinkage methods can be good approximations for the selection without shrinkage (which has a combinatorial complexity). The former problem, however, has no such quasi-equivalence between soft- and hard-thresholding.

The Bernstein estimator of the copula function and of the copula density: large sample properties

Paul Janssen, Jan Swanepoel, Noël Veraverbeke
Hasselt University, Belgium

E-mail: Paul.Janssen@uhasselt.be

Copulas and copula densities are building blocks to study dependence between components of a random vector. For many problems it is not evident how to select a parametric family of copulas to describe the data at hand and hence, nonparametric estimation of the copula and the copula density is an option. The Bernstein estimator of the copula and the copula density is one such nonparametric estimator that received attention in recent papers.

In the first part of the presentation we discuss the asymptotic properties of the Bernstein estimator of the copula function. We focus on almost sure consistency rates and on asymptotic normality. In a small simulation study we show the performance of the proposed estimator.

In the second part of the presentation we show the asymptotic normality of the Bernstein estimator of the copula density. Compared to existing results our theorem does not assume known marginal; this makes our theorem applicable for real data.

References:

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2012) Large sample behavior of the Bernstein copula estimator. *Journal of Statistical Planning and Inference*, 142, 1189–1197.

Janssen, P., Swanepoel, J. and Veraverbeke, N. (2013) A note on the asymptotic behavior of the Bernstein estimator of the copula density. Technical Report (submitted).

Principal component analysis for dependent functional data on random domains

Dominik Liebl, Alois Kneip
University of Cologne, Germany

E-mail: liebl@statistik.uni-koeln.de

Principal component analysis is a core method for analyzing functional data. Traditionally, it is assumed that each function of a given sample is directly observed—a situation rarely fulfilled in practice. Recent studies focus on the more realistic case of observing only (finitely many) noisy discretization points per function. Generally, there are two different approaches to do principal component analysis in this more challenging case. First, one can pre-smooth each function and estimate the covariance function from the pre-smoothed functions. Second, one can estimate the covariance function directly from the noisy discretization points. The latter possibility has the advantage to be applicable in very general data situations such as, e.g., sparse functional data with only a few observations per function.

We focus on the direct estimation of the covariance function from noisy discretization points. Motivated by our real data study of analyzing electricity spot prices, we consider the case of dependent functional data, where the noisy discretization points are only observed on random sub-parts of the whole domain. We extend and complement existing results regarding the estimation of the covariance function, the eigenfunctions, and eigenvalues by the derivation of the asymptotic bias and variance of the nonparametric bivariate local linear estimator. While the assumption of random sub-domains is rather particular, our theoretical results can be interpreted also for the usual case of a fix deterministic domain.

The statistical procedures outlined above are illustrated using German electricity market data. Specifically, we focus on time series of hourly electricity spot prices over the interesting time period from one year before to one year after Germany’s nuclear phase-out in March 2011.

Choosing a frailty distribution to model dependence in clustered survival data

Marco Munda, Catherine Legrand
Université catholique de Louvain, Belgium
E-mail: marco.munda@uclouvain.be

Frailty models are used to analyse clustered survival data. The frailty distribution provides a way to model the type of dependence between event times within a cluster. In this presentation, we propose a new diagnostic plot to assess the frailty distribution. We approach that problem from the fact that, although frailties are unobservable, the dependence structure that they impose on the data can be observed. We estimate the observed structure that we compare with model-based structures obtained from different frailty distributions. To measure association, we use quantile dependence coefficients. The method easily accommodates any cluster sizes and various censoring schemes.

Testing instantaneous causality in presence of non constant unconditional covariance

Hamdi Raïssi
INSA Rennes, France
E-mail: Hamdi.Raïssi@insa-rennes.fr

The problem of testing instantaneous causality between variables with time-varying unconditional covariance is investigated. It is shown that the classical tests based on the assumption of stationary processes must be avoided in our non standard framework. More precisely we underline that the standard test does not control the type I errors, while the tests with White (1980) and Heteroscedastic Autocorrelation Consistent (HAC) corrections can suffer from a severe loss of power when the covariance is not constant. Consequently a modified test based on a bootstrap procedure is proposed. The relevance of the modified test is underlined through a simulation study. The tests considered in this paper are also compared by investigating the instantaneous causality relations between US macroeconomic variables.

References:

White, H. (1980) A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48, 817–838.

A semi-parametric locally stationary ARCH Model

Lionel Truquet, Hamdi Raïssi, Valentin Patilea
Université de Rennes 1, France

E-mail: lionel.truquet@univ-rennes1.fr

We consider an ARCH(p) model with a time-varying intercept and constant coefficients for the lag variables. This model is a compromise between the time-varying ARCH models introduced by Dahlhaus and Subba Rao (2006) and the nonparametric local stationary model proposed by Vogt (2012). The time-varying intercept allows to capture volatility non stationarity, while the constant lags coefficients yield a constant autocorrelation for the squares of the process, like in a classical ARCH model. Inspired by the partially linear regression models, we define a class of weighted least squares estimators for the lag variables coefficients and we derive their parametric rate of convergence and the asymptotic normality. Moreover, we indicate what the optimal asymptotic variance in this class of estimators could be. The results are derived assuming only a square integrable marginal distribution. The problem of testing whether the lag variables coefficients are constant or not is also addressed. Simulation and real data applications illustrate the new modeling approach.

Nonparametric reconstruction of the tree structure of a nested Archimedean copula

Nathan Uyttendaele, Johan Segers
Université catholique de Louvain, Belgium

E-mail: nathan.uyttendaele@uclouvain.be

One of the features inherent in nested Archimedean copulas, also called hierarchical Archimedean copulas, is their rooted tree structure. In this talk, a completely nonparametric method to estimate this structure is developed. Our approach consists in representing the rooted tree structure as a set of trivariate structures that can be individually estimated. Indeed, for any triple of variables there are only four possible rooted tree structures and, based on a sample, a choice can be made by performing comparisons between the three marginal empirical bivariate distributions of the triple. The set of estimated trivariate structures can then be used to build an estimate of the global rooted tree structure. This approach leads to an estimator that has reasonable properties, and its usefulness is outlined with an application example on log returns from six different companies.

P-splines quantile regression estimation in varying coefficient models

Anneleen Verhasselt
Hasselt University, Belgium

E-mail: anneleen.verhasselt@uhasselt.be

We consider the problem of quantile estimation in the context of varying coefficient models, i.e. our target is a certain quantile of the response instead of the mean in a flexible regression context. Varying coefficient models are a flexible extension of linear regression models, by the fact that the regression coefficients are functions that vary with for example time, instead of constants. We focus on quantile regression in varying coefficient models for longitudinal data using P-splines to estimate the regression coefficient functions. Our estimator has nice theoretical properties, such as consistency. We demonstrate the performance of our estimator on simulated and real data, and discuss data-driven choices of the smoothing parameters.

Robust variable selection using the nonnegative garrote method

Inge Vrinssen, Irène Gijbels

KU Leuven, Belgium

E-mail: inge.vrinssen@wis.kuleuven.be

We consider a multiple linear regression model with p covariates. Often there are only a few coefficients that are different from zero. Therefore Breiman proposed in [1] the nonnegative garrote method. This is a variable selection method that shrinks the least squares estimates and puts some of these coefficients equal to zero. The method has been further used in variable selection in additive regression models and varying coefficient models in [2] and [3]. But the nonnegative garrote method is not robust to outliers. In this talk we robustify the nonnegative garrote method. We can use existing robust methods such as the S-estimator [4], the τ -estimator, ..., to compute initial estimates of the coefficients. The main goal is thus to find robust methods to estimate the shrinkage factors and robust selection criteria to determine the value of the regularization parameter. Several robust estimation methods for the shrinkage factors and robust selection criteria for the regularization parameter are discussed. The performances of the different methods are investigated and compared via a simulation study. The methods are illustrated on a real data example.

References:

- [1] Breiman, L. (1995), Better subset regression using the nonnegative garrote. *Technometrics*, 51, 373-384.
- [2] Antoniadis, A., Gijbels, I. and Verhasselt, A. (2012a), Variable selection in varying coefficient models using P-splines. *Journal of Computational and Graphical Statistics*, 21, 638-661.
- [3] Antoniadis, A., Gijbels, I. and Verhasselt, A. (2012b), Variable selection in additive models using P-splines. *Technometrics*, 54, 425-438.
- [4] Rousseeuw, P. and Yohai, V. (1984), Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis*, edited by J. Franke, W. Hardle, and R.D. Martin, Lecture Notes in Statistics, 26, Springer Verlag, New York

Abstracts of poster presentations

Modelling unbalanced clustered multivariate survival data via Archimedean copula functions

Roel Braekers, Leen Prenen, Luc Duchateau

Universiteit Hasselt, Belgium

E-mail: Roel.Braekers@uhasselt.be

In an analysis of clustered multivariate survival data, two different types of models are commonly used that take the association between the different lifetimes into account: frailty models and copula models. Frailty models assume that conditional on a common unknown frailty term for each cluster, the hazard function of each individual within that cluster is independent. These unknown frailty terms with their imposed distribution are used to express the association between the different individuals in a cluster. Copula models on the other hand assume that the joint survival function of the individuals within a cluster is given by a copula function, evaluated in the marginal survival function of each individual. Hereby it is the copula function which describes the association between the lifetimes within a cluster.

A major disadvantage of the present copula models over the frailty models is that the size of the different clusters must be small and equal in order to set up manageable estimation procedures for the different parameters in this model. We describe in this talk a copula model for unbalanced clustered survival data. This is done by focusing on the class of Archimedean copula functions with completely monotone generators and exploiting the Laplace transform-expression of these generators to simplify the likelihood function. Hereby we note that the size of each cluster does not have to be equal anymore, and moderate to large cluster sizes are also allowed.

For this model, we develop one- and two-stage procedures to estimate the association parameter for the copula function. In the one-stage procedure, we consider a parametric model for the marginal survival functions while in the two-stage procedure we model the marginal survival functions by either a parametric or a non-parametric model.

As results, we show the consistency and asymptotic normality of the maximum likelihood estimators for the different parameters. We perform a simulation study to investigate the finite sample properties of this estimator and finally we illustrate this copula model on a real life data set in which we study the time until first insemination for cows which are clustered within different herds.

Bayesian analysis of multivariate threshold autoregressive models with missing data

Sergio Calderón, Fabio Nieto
Universidad Nacional de Colombia, Colombia
E-mail: sacalderonv@unal.edu.co

Bayesian analysis of multivariate thresholds autoregressive models (MTAR) with exogenous inputs and missing data is carry out. MCMC methods are used to get samples of the marginal posterior distributions, including threshold values and missing data. In order to identify autoregressive orders, we adapt the Bayesian variable selection method to the MTAR models. Number of regimes in the models is known. Simulation experiments and real data examples are presented.

Nonparametric estimation problem for some time-periodic-drift stochastic differential equation

Khalil El Waled, Dominique Dehay
IRMAR Rennes, France
E-mail: khalil.elwaled@gmail.com

This work deals with the problem of estimation of the drift coefficient function for a diffusion process following the model

$$d\xi_t = f(t)\xi_t dt + dB_t$$

where $f : \mathbb{R} \rightarrow \mathbb{R}$ is a periodic continuous function, when the process is observed through continuous time $[0, T]$ as $T \rightarrow \infty$, and the period $P > 0$ is known. For this purpose, we construct a kernel estimator of $f(\cdot)$ which is consistent. Its rate of convergence which depends on the signum of $\int_0^P f(t)dt$. We establish that the convergence of this estimator is uniform with respect to $f(\cdot)$ varying in different classes of periodic functions, and we investigate the asymptotic minimax risk for this problem.

References:

- [1] Dehay, D., (2013) Parameter maximum likelihood estimation problem for time-periodic-drift Langevin type stochastic differential equations, (submitted).
- [2] Dehay, D., El Waled, K. (2013) Nonparametric estimation problem for a time-periodic signal in a periodic noise, *Statistics and Probability Letters* 83, 608-615.
- [3] Dehay, D., El Waled, K. (2013) Nonparametric estimation problem for a time-periodic-drift Langevin type stochastic differential equation (work in progress).

The analysis of credit risk data: variable selection for a mixture cure model

Lore Dirick, Gerda Claeskens
KU Leuven, Belgium
E-mail: Lore.Dirick@kuleuven.be

In recent research, several survival analysis techniques (originally mainly used in medical science) are being used to analyze credit loan information. The advantage of this method compared to previously used methods of credit scoring (e.g. cluster analysis, logistic regression) is that, when using survival analysis, one is able to predict *when* creditors will default and not only *whether* they will default. In the credit risk context, the survival function $S(t) = P(T > t)$

can be interpreted as the probability that some customer will still be repaying his loan at timepoint t .

Despite the fact that there are certain analogies between standard survival and survival in a credit loan context, there are also differences that might make the standard survival analysis approach inappropriate for the analysis of credit data. The main problem is that, typically, a very high proportion of credit data is right-censored, not only because the customer default is not observed during the observation period, but simply because default does not take place in the entire loan lifetime. Because of this, Tong et al. (2012) use a mixture cure approach to analyze the credit risk of a specific customer. The idea behind such models is that the population of loan applicants comprises two subpopulations, one that contains applicants that are susceptible to default (hence will default eventually), and another one that contains applicants that are not susceptible, or immune, and will never default. The susceptibility of a certain loan applicant is modeled by the so-called *incidence* model part, using logistic regression. Survival times of the susceptible subpopulation part are consequently modeled by the *latency* model part using proportional hazards regression. Recently, an R-package was introduced by Cai et al. (2012) to estimate such semi-parametric mixture models.

Using a mixture cure model involves working with a separate parameter vector for each of the two model parts. These two covariate vectors may or may not contain the same elements, which suggests that performing model selection on these kind of models could be very useful. However, hardly any attempts to perform model selection have been made in previous research. The use of widespread model selection criteria such as Akaike's information criterion (AIC) and the Bayesian information criterion (BIC) for those mixture cure models is not straightforward, because the exact likelihood can not be computed because there is a problem of missing information. We develop and present an AIC for mixture cure models, using the idea of Cavanaugh and Shumway (1998) of calculating complete data log likelihoods performing the (supplemented) EM-algorithm.

References:

- Tong, E. N. C., Mues, C., and Thomas, L. C., 2012: Mixture cure models in credit scoring: if and when borrowers default. *Eur. J. Oper. Res.*, **218**, 132–139.
- Cai, C., Zou, Y., Peng, Y. and Zhang, J. (2012). smcure: An R-package for estimating semi-parametric mixture cure models. *Comput. Meth. Prog. Bio.*, **108**, 1255–1260.
- Cavanaugh, J. E. and Shumway, R. H. (1998). An Akaike information criterion for model selection in the presence of incomplete data. *J. Stat. Plan. Infer.*, **67**, 45–65.

Gene selection for survival data under dependent censoring, a copula-based approach

Takeshi Emura, Yi-Hau Chen

Graduate Institute of Statistics, National Central University, Taiwan

E-mail: emura@stat.ncu.edu.tw

Dependent censoring arises in biomedical studies where the survival outcome of interest is censored by competing causes. In survival data with microarray gene expressions, a gene selection or gene ranking based on the Cox regression analyses has been used extensively, which however are valid under the independent censoring assumption. In this talk, the first objective is to study the effect of dependent censoring on the gene selection procedure. Here, we model the joint distribution of survival outcome and the competing cause viagcopulasand study the potential bias due to dependent censoring. The second objective is to utilize the copula model to develop an alternative gene selection procedure. The proposed procedure adjusts for the

dependent censoring and outperforms the existing method when the dependent censoring indeed exists. Simulations and data analyses demonstrate the usefulness of our proposal.

Iterative estimation for conditional estimating equations

Weiyu Li, Valentin Patilea
CREST- Ensai, Rennes, France
E-mail: Weiyu.LI@ensai.fr

Many statistical and econometric models could be written under the form of conditional estimating equations, also called of conditional moment equations. In the classical approach for estimating parameters identified by such restrictions, one replaces the conditional moments by a sufficiently rich finite set of unconditional moments and applies the generalized method of moments (GMM). However, the GMM approach does not guarantee consistency since the parameters are not necessarily identified by a finite set of marginal moments. Motivated by this aspect, several recent articles proposed alternative approaches that preserve consistency. Herein we consider an estimation approach for conditional estimating equations that is called smooth minimum distance (SMD) and is based on the optimization of a nonlinear contrast. We introduce an iterative version of SMD based on a quadratic approximation of the contrast. At any step of the iteration, the estimate has an explicit form and therefore the new method could be easily implemented. We present an extensive empirical study of the new method. In particular we compare it with classical methods (least squares, maximum likelihood, GMM).

Testing for lack of fit in functional regression models

Samuel Maistre, Valentin Patilea
CREST- Ensai, Rennes, France
E-mail: samuel.maistre@ensai.fr

We consider regression models with a response variable taking values in a Hilbert space, of finite or infinite dimension, and hybrid covariates. That means there are two sets of regressors, one of finite dimension and a second one functional with values in a Hilbert space. The problem we address is the test on the effect of the functional covariates. This problem occurs in many situations: testing the effect of the functional covariate in a semi-functional partial linear regression with scalar responses, significance test for functional regressors in nonparametric regression with hybrid covariates and scalar or functional responses, testing the effect of a functional covariate on a scalar or functional outcome. We propose a new test based on univariate kernel smoothing. The test statistic is asymptotically standard normal under the null hypothesis provided the smoothing parameter tends to zero at a suitable rate. The one-sided test is consistent against any fixed alternative and detects local alternatives a la Pitman approaching the null hypothesis at suitable rate. In particular we show that neither the dimension of the outcome nor the dimension of the functional covariates influences the theoretical power of the test against such local alternatives.

Almost sure central max-limit theorem for nonstationary random fields

Luísa Pereira
University of Beira Interior, Portugal
E-mail: lpereira@ubi.pt

Recently, a new field of limit theorem has been discovered, which is called almost sure central limit theorem (ASCLT). Many results have been obtained for independent and dependent

random variables. The first results on ASCLT, reported in the papers of Brosamler (1988), Schatte (1988) and Lacey and Philipp (1990), dealt with partial sums of independent and identically distributed (i.i.d.) random variables. For an i.i.d. sequence $\{X_n\}_{n \in \mathbb{N}}$ with zero mean, unit variance and partial sum $S_k = \sum_{i=1}^k X_i$, $k \geq 1$, the simplest version of the ASCLT states that $\frac{1}{\log n} \sum_{k=1}^n \mathbb{1}_{\{S_k \leq \sqrt{k}x\}} \rightarrow \Phi(x)$ a.s. for any fixed $x \in \mathbb{R}$, where a.s. means almost surely, $\mathbb{1}_A$ denotes the indicator function of the event A and $\Phi(x)$ is the standard normal distribution function. Later on, Fahrner and Stadmüller (1998), Cheng, Peng and Li (1998) and Berkes and Cski (2001) extended the ASCLT for the maximum of i.i.d. random variables. Let $\{X_n\}_{n \in \mathbb{N}}$ be an i.i.d. sequence, and let $M_k = \max_{1 \leq i \leq k} X_i$ denote the partial maximum, $k \geq 1$. If there exist normalizing constants $a_k > 0$, $b_k \in \mathbb{R}$ and a nondegenerate distribution function $G(x)$ such that $P(M_n \leq a_n x + b_n) \rightarrow G(x)$, then we have $\frac{1}{\log n} \sum_{k=1}^n \frac{1}{k} \mathbb{1}_{\{M_k \leq a_k x + b_k\}} \rightarrow G(x)$ a.s., for any continuity point x of G . It is well known that $G(x)$ must be of the same type as the extreme value distribution $G(x) = \exp\{-(1 + \gamma x)^{-1/\gamma}\}$, where is the so-called extreme value index. For almost sure convergence related to the normalized maximum of stationary or nonstationary normal sequences see Csáki and Gonchigdanzan (2002), Zouxiang and Nadarajah (2011) and Chen and Lin (2006). More recently ASCLT was extended to the maximum of stationary normal random fields by Choi (2010). In this paper we consider the maximum of a particular class of nonstationary random fields and obtain its almost sure limit theorem. Random field theory has recently received increasing attention since its applications are extremely numerous and diverse which includes image analysis, atmospheric sciences and geostatistics, among others

Model selection for graphical models using the focused information criterion

Eugen Pircalabelu, Gerda Claeskens, Lourens Waldorp
 KU Leuven, Belgium

E-mail: eugen.pircalabelu@kuleuven.be

A new method for model selection for Gaussian directed acyclic graphs (DAG) and Markov networks, with extensions towards ancestral graphs, is constructed to have good mean squared error properties, for both low and high dimensional settings ($n < p$). The method is based on the focused information criterion (FIC) and unlike the traditional AIC or BIC, the FIC allows for selecting individual models, tailored to a specific purpose (the focus), as opposed to attempting an identification of a single model that should be used for all purposes.

Working under a local misspecification framework which assumes that the true model is in a ‘neighborhood’ of the least complex model one is willing to assume, we define a *focus parameter*, i.e. $\mu(\theta_0, \gamma_0 + \delta/\sqrt{n})$ as a function of the parameters of the model density, and potentially of a user-specified vector of covariate values, for any particular node in the graph $G(E, V)$. Subsequent steps involve specifying a collection of models and a (un-)penalized objective function which is optimized for parameters corresponding to the prespecified models. The FIC estimates $\text{MSE}(\hat{\mu}_S)$ for each of a collection of models S , and selects the model with the smallest value. The focus of the research (i.e. the purpose of the model), directs the selection and different focuses may lead to different selected models. In this way, one can obtain better selected models in terms of MSE, than obtained from a global model search. For this application of FIC for structure learning in graphs, the focus is the expected value of a variable, reflecting interest in discovering a topology of the graph that produces a low MSE for this focus.

An analysis on real and simulated data suggests that the FIC scoring approach is able to identify graphical models with better MSE and mean squared prediction values than obtained from existing methods, such hill-climbing using AIC and BIC, PC, SIN and the graphical Lasso or Chow-Liu algorithms.

By the application of FIC on fMRI data, it is shown that using FIC one has at disposal a

powerful method to study evolutions over time of networks constructed to study the functional connectivity between brain regions. Moreover, the method identifies more clearly important brain regions which seem to be highly connected with others, acting as ‘hubs’ or ‘informational gateways’.

Stochastic monotonicity tests

Dominik Sznajder, Irène Gijbels

KU Leuven, Belgium

E-mail: Dominik.Sznajder@wis.kuleuven.be

Stochastic monotonicity is a dependence concept which originates from the work of Tukey (1958) and Lehmann (1959), where it was called complete regression dependence. This name refers to the underlying idea that all conditional quantiles are monotonic functions (in the same direction). In particular it implies, what is nowadays called, regression monotonicity, i.e., monotonicity of $E(Y|X = x)$ as a function of x . The stochastic monotonic relation between random variables is of particular interest in financial, insurance and econometric studies. For recent developments in testing for stochastic monotonicity and a broad overview of applications see [5] and [1].

Stochastic monotonicity is a very strong dependence concept. It implies Tail Monotonicity, which further implies Quadrant Dependence, studied in [2] and [4] respectively.

The general methodology for a testing procedure is based on smooth constrained copula bootstrapping. It is motivated by the fact that often dependence between random variables in a random vector can be expressed as a feature of the underlying copula function. We propose to build a test statistic as functional violation measures based on the empirical copula estimator. Furthermore, the statistical inference is based on the bootstrapped distribution of the test statistic. This requires a resampling scheme under the null hypothesis and we propose a smooth constrained nonparametric copula estimation procedure as a remedy. It is based on local polynomial smoothing of the initial constrained estimator and on transforming its partial derivatives by a rearrangement technique.

The proposed methodology is generic and can be applied to other dependence concepts, which can be expressed as shape constraints on the copula function. The presented results are described in [6] and [3] and refer to previous work on other dependence structures in [2] and [4].

References:

- [1] Delgado, M.A., Escanciano, J.C. (2012). Distribution-free tests of stochastic monotonicity. *Journal of Econometrics*, 170, 68–75.
- [2] Gijbels, I., Sznajder, D. (2013). Testing tail monotonicity by constrained copula estimation. *Insurance: Mathematics and Economics*, 52(2), 338–351.
- [3] Gijbels, I., Sznajder, D. (2011). Constrained copula estimation in stochastic monotonicity testing. Manuscript.
- [4] Gijbels, I., Sznajder, D. (2013). Positive quadrant dependence testing and constrained copula estimation. *The Canadian Journal of Statistics*, 41(1), 36–64.
- [5] Lee, S., Linton, O., Whang, Y.-J. (2009). Testing for stochastic monotonicity. *Econometrica*, 77(2), 585–602.
- [6] Sznajder, D. (2011). Copula-based testing for dependence structures. PhD Thesis.

List of Participants

Family Name	First Name	Affiliation	Email
Barbu	Vlad Stefan	Université de Rouen	barbu@univ-rouen.fr
Braekers	Roel	Hasselt University	roel.braekers@uhasselt.be
Bücher	Axel	Katholieke Universiteit Leuven	axel.buecher@rub.de
Cadre	Benoît	ENS Cachan Bretagne, IRMAR	cadre@bretagne.ens-cachan.fr
Calderón	Sergio	Universidad Nacional de Colombia	sacalderonv@unal.ac.co
Charkhi	Ali	Katholieke Universiteit Leuven	ali.charkhi@kuleuven.be
Cheng	Ming-Yen	National Taiwan University	cheng@math.ntu.edu.tw
Claeskens	Gerda	Katholieke Universiteit Leuven	gerda.claeskens@kuleuven.be
Comets	Emmanuelle	INSERM	emmanuelle.comets@inserm.fr
Czado	Claudia	Technische Universität München	cczado@ma.tum.de
Dehay	Dominique	Université Rennes 2, IRMAR	dominique.dehay@univ-rennes2.fr
Delaigle	Aurore	The University of Melbourne	A.Delaigle@ms.unimelb.edu.au
Delyon	Bernard	Université Rennes 1, IRMAR	bernard.delyon@univ-rennes1.fr
Dirick	Lore	Katholieke Universiteit Leuven	Lore.Dirick@kuleuven.be
El Waled	Khalil	Université Rennes 2	khalil.elwaled@gmail.com
Emura	Takeshi	National Central University, Taiwan	emura@stat.ncu.edu.tw
Favre Martinoz	Cyril	CREST-Ensaï & IRMAR	cyril.favremartinoz@ensai.fr
Fromont Renoir	Magalie	CREST-Ensaï & IRMAR	mfromont@ensai.fr
Geenens	Gery	University of New South Wales, Sydney	ggeenens@unsw.edu.au
Geerdens	Candida	Hasselt University	candida.geerdens@uhasselt.be
Genest	Christian	McGill University	cgenest@math.mcgill.ca
Gijbels	Irène	Katholieke Universiteit Leuven	irene.gijbels@wis.kuleuven.be
Hansen	Bruce	University of Wisconsin	behansen@wisc.edu
Herrmann	Klaus	Katholieke Universiteit Leuven	Klaus.Herrmann@wis.kuleuven.be
Hristache	Marian	CREST-Ensaï	marian_hristache@ensai.fr
Hughes	David	University of Liverpool	dmhughes@liv.ac.uk
Jansen	Maarten	Université Libre de Bruxelles	maarten.jansen@ulb.ac.be
Janssen	Paul	Hasselt University	paul.janssen@uhasselt.be
Li	Weiyu	CREST-Ensaï & IRMAR	weiyu.li@ensai.fr
Liebl	Dominik	Universities of Cologne	dliebl@uni-bonn.de
Lopez	Olivier	Université Pierre et Marie Curie, LSTA	olivier.lopez0@upmc.fr
Maistre	Samuel	CREST-Ensaï & IRMAR	samuel.maistre@ensai.fr
Massiot	Gaspar	CREST-Ensaï	gaspar.massiot@ensai.fr
Munda	Marco	Katholieke Universiteit Leuven	marco.munda@uclouvain.be
Neslehova	Johanna	McGill University	johanna@math.mcgill.ca
Patilea	Valentin	CREST-Ensaï & IRMAR	patilea@ensai.fr
Pereira	Luísa	University of Beira Interior	lpereira@ubi.pt
Perthame	Emeline	Agrocampus-Ouest, IRMAR	emeline.perthame@agrocampus-ouest.fr
Pircalabelu	Eugen	Katholieke Universiteit Leuven	Eugen.Pircalabelu@kuleuven.be
Raïssi	Hamdi	IRMAR-INSA de Rennes	hamdi.raïssi@insa-rennes.fr
Rao	Sunil	University of Miami	jrao@biostat.med.miami.edu
Rouvière	Laurent	CREST-Ensaï & IRMAR	laurent.rouviere@ensai.fr
Saumard	Matthieu	IRMAR-INSA de Rennes	matthieu.saumard@gmail.com
Sznajder	Dominik	Katholieke Universiteit Leuven	dominik.sznajder@wis.kuleuven.be
Truquet	Lionel	CREST-Ensaï & IRMAR	lionel.truquet@ensai.fr
Uyttendaele	Nathan	Katholieke Universiteit Leuven	na.uytten@gmail.com
Verhasselt	Anneleen	Hasselt University	anneleen.verhasselt@uhasselt.be
Vimond	Myriam	CREST-Ensaï & IRMAR	mvimond@ensai.fr
Vrinssen	Inge	Katholieke Universiteit Leuven	inge.vrinssen@wis.kuleuven.be



The Organizers gratefully acknowledge the support and sponsorship of the Workshop by:
Ensaï, CREST, KU Leuven
and
Research Foundation Flanders (FWO), IRMAR, INSA de Rennes, Rennes Métropole

The Workshop is organized under the patronage of The French Statistical Society

The Workshop is organized in celebration of the International Year of Statistics

