



Institute of Statistics

Annual Report

Year 2008

UCL
Université catholique de Louvain

Contents

1. Foreword	7
2. Staff	9
3. Research activities	15
3.1. Presentation	15
3.2. Research contracts	16
3.2.1. Research projects under contracts and cooperation projects	16
3.2.2. Applied research contracts	19
3.3. Reprints	20
3.4. Discussion papers	22
3.5. Books	24
3.6. Visitors	26
3.7. Statistics seminars	28
3.8. Organization of scientific meetings.....	30
3.9. Academic visits	33
3.10. Participation to conferences and statistics meetings	35
3.11. Academic teaching activities.....	39
4. Teaching Activities and educational activities	40
4.1. Bachelor level	40
4.2. Minor in statistics	40
4.3. New Master programs in statistics	40
4.4. Certificate	40
4.5. Other activities	40
4.6. Doctoral school	41
4.6.1. Presentation	41
4.6.2. Doctoral dissertations and PhD in progress	43
4.6.3. Doctoral seminars	45
4.6.4. Short courses	46
5. Support for the university and consulting activities	47
5.1. Presentation	47
5.2. SMCS activities	47
5.3. SMCS courses	47
5.4. Applied statistics workshops	49
6. Appendix	51
6.1. Abstracts of the discussion papers	51

1. Foreword

Since its founding in the early 1990s, the UCL Institute of Statistics has undoubtedly become the largest and most comprehensive statistics group within a single department in Belgium, and one of the largest in the EU. The Institute has as its purpose (i) to provide excellent education in probability and statistics, (ii) to extend the frontiers of knowledge in probability and statistics through pure and applied research, (iii) to collaborate with researchers in other disciplines to investigate important scientific issues, and (iv) to serve the UCL community in all areas related to probability and statistics.

In 2008 the teaching faculty comprised a full-time staff of 9 professors and over 20 teaching and research assistants. Several part-time and invited professors broadened the scope of expertise of the Institute in specific topics, like data mining or surveys. The Institute also accommodated post-doc researchers and short-term as well as long-term foreign visiting scholars. The administrative, library and technical staff consisted of 9 persons. It was the Institute's great loss that Christine Denayer, Executive Director, left to join the UCL central administration. We would like to take this opportunity to publicly thank Christine for her efforts to increase the administrative efficiency during the years she was here. Isabelle Petit has been appointed as the new Executive Director of our Institute in the Summer 2008.

The Institute of Statistics is responsible for all probability and statistics courses offered in the university. Students in many programmes, in various schools, take ancillary statistics courses taught by members of the Institute who have always placed strong emphasis on quality teaching. Apart from these courses taught in the different schools of the university, the Institute of Statistics continued strengthening its master and doctoral programmes. In the various directions offered, i.e., biostatistics, general statistics with specialisation in either methodological research or applications, an increasing number of students are registered and obtain degrees. Recently, the number of students in our core statistics classes has increased by 13% from 2006 to 2008. These excellent results are certainly due to our increasingly innovative and interactive teaching approach. Continuing professional development is also offered through various short courses and by a university certificate offered in collaboration with the Institute of continuing education, IUFC.

During 2008, the organization of the masters has been thoroughly re-examined. From the academic year 2009/10, the master programmes in statistics, biostatistics and actuarial sciences will be hosted by the Louvain School of Statistics, Biostatistics and Actuarial Sciences (LSBA), located within the Faculty of Science. It will join two program committees, one in statistics (comprising general and biostatistics), and one in actuarial sciences. This newly created school will benefit from synergy effects on the administrative level while keeping its strong links with the research centre of statistics. For a detailed account of the teaching activities of the Institute, see Section 4 of this report.

Members of the Institute have a wide variety of research interests, ranging from theoretical to applied topics and over a broad spectrum of methodological topics. The main research areas at the Institute develop along the three following main directions: (i) Mathematical Statistics (semi- and non-parametric statistics, Bayesian statistics, multivariate analysis, regression, mixture models, survival analysis, time series,...), (ii) Biostatistics, and (iii) Quantitative Analysis of Insurance and Finance Risks.

Members of the Institute actively collaborate with scientists from other disciplines on important research investigations. In particular, the application of probability and statistics to solving problems in insurance, pension and risk management is conducted at the Institute of Actuarial Sciences run jointly by the Louvain School of Management and the Institute of Statistics, combining a unique expertise in deep mathematical background as well as in business aspects. The activities in this field, including scientific publications, seminars, and research projects, are described in a separate report.

Research activity continued to flourish in 2008. During that, members of the Institute published 28 papers in refereed journals. Research to appear soon, or under evaluation, was reported in 34 discussion papers.

The Institute of Statistics is the coordinator of an IAP research network in statistics (Interuniversity Attraction Pole, Phase VI 2007-2011) on the theme "Statistical Analysis of Association and Dependence in Complex Data". The network involves nine research teams, of which five are Belgian and four are European, non-Belgian partners. The current network is a continuation of the network "Statistical techniques and modeling for complex substantive questions with complex data", financed from 2002 till 2006 by the Belgian Science Policy. In addition to the IAP, several research projects are funded by public and private bodies.

Members of the Institute serve on the editorial boards of several prestigious journals. These editorial roles for international journals held by Statistics staff is a clear mark of international recognition.

There are three regular seminars held at the Institute. At the Statistics Seminars, invited speakers present their research results. A diversity of subjects is presented at this seminar by foreign visitors of the Institute. From time to time, a joint statistics and econometrics seminar, organised in collaboration with CORE, takes place. The "Atelier de statistique appliquée/Applied Statistics Workshop" is organised by the Institute on a regular basis. It focuses on problem driven statistics, where a real world problem of substantial practical interest is treated. At the "Young Researchers Day - YRD" organized twice a year, PhD students in statistics present their recent research work. These doctoral seminars constitute an extra stimulant for PhD students and other young researchers.

During 2008, the structure of the university was fully re-designed through the "Plan de développement". According to this reform, the research in statistics will be conducted in the Research Centre of Statistics, Biostatistics and Actuarial Sciences within IMMAQ (Institut de recherche multidisciplinaire pour la modélisation et l'analyse quantitative - Institute of Multidisciplinary Research for Quantitative Modelling and Analysis). IMMAQ is a multisectorial institute associating researchers from three different research centers (CORE, IRES, and STAT-ACTU). These researchers develop and use in their various fields of expertise a coherent set of tools and methods for quantitative modelling and analysis.

The Graduate School in Statistics and Actuarial Science has been created in 2006, under the patronage of the "Fonds National de la Recherche Scientifique (FNRS)". It is also associated to the Graduate Colleges in Medicine, Agronomy, Economic Sciences, and Engineering, as Statistics and Actuarial Sciences have links to many other disciplines. The Graduate School gathers students from various French-speaking Belgian universities, and is chaired by the Institute of Statistics. It offers various activities, such as short courses, workshops and an annual colloquium, some of which held at UCL under the auspices of the Institute, others in partner universities.

The Institute of Statistics also offers a statistical consulting and computing service to the UCL community. Researchers of the other faculties receive advice concerning appropriate methodologies and suitable statistical packages for their specific problems. The scientific members and the computer scientists of the Institute are developing their knowledge about the evolution of statistical software and they give advice in this context. The Institute hosts the "Statistical application server" used daily by UCL researchers and students for their statistical computations. These services were merged in 2007 in a common task force, the SMCS service (for "Statistical methodology and computing support"). This initiative allowed to increase the quality of the assistance to the UCL community and the visibility of this activity. A detailed account of all the SMCS activities can be found in a separate specific report.

According to UCL "Plan de développement", the SMCS will become in 2009 a "technological platform" within IMMAQ.

We would like to thank the Université catholique de Louvain for supporting the activities of the Institute, as well as all the members of the Institute for developing a friendly, open and stimulating research and teaching environment. In the future, we hope to further increase the strengths of the Institute in education, research, and service to the community in the new structures of our university.

Louvain-la-Neuve, July, 2009

Michel Denuit
Chairman of the Institute of Statistics

Bernadette Govaerts
Responsible for the technological platform SMCS

Christian Hafner
Academic Secretary and Responsible for the Louvain School of Statistics, Biostatistics and Actuarial Science

Rainer von Sachs
Responsible for the Research Centre in Statistics, Biostatistics and Actuarial Science

2. Staff

Permanent Academic Members



Michel Denuit

Professor, Université catholique de Louvain (Institute of Statistics).
PhD Université Libre de Bruxelles, 1997.

Stochastic inequalities, mathematical risk theory, statistics applied to insurance.

Editor of the *Astin Bulletin* (since 2007), Associate Editor for *Insurance: Mathematics and Economics* (since 1999) and for *Methodology and Computing in Applied Probability* (since 2008).

Co-author of 9 books, including "*Modern Actuarial Risk Theory Using R*" (Springer, 2008), "*Actuarial Modelling of Claim Counts: Risk Classification, Credibility and Bonus-Malus Systems*" (Wiley, 2007), "*Actuarial Theory for Dependent Risk*" (Wiley, 2005), "*Construction de Tables de Mortalité Périodiques et Prospectives*" (Collection Audit-Actuariat-Assurance, Economica, 2005).

Chairman of the Institute of Statistics (2006-2009) and Academic Secretary of the Institute of Actuarial Sciences.

Homepage : <http://www.actu.ucl.ac.be/staff/denuit/mdenuit.html>



Dominique Deprins

Part-time professor, Université catholique de Louvain (Institute of Statistics), full-time professor, Facultés Universitaires Saint-Louis, Bruxelles.
PhD Université catholique de Louvain, 1989.

Epistemological work about Statistics and Probabilities : application in the field of mental health and, more generally, in the field of human sciences.

Academic secretary of Facultés Universitaires Saint-Louis (1999-2005), pedagogic adviser (since 2004), President of the Committee responsible for internal organization of education in mathematics, statistics, sciences and computer sciences (since 2000) at Facultés Universitaires Saint-Louis.

Member of the College of the O.P.P (Observatoire du Principe de Précaution).



Bernadette Govaerts

Professor, Université catholique de Louvain (Institute of Statistics).
PhD Université catholique de Louvain, 1987.

Industrial statistics, chemometrics, experimental design, statistical quality control, pre-clinical biostatistics, development, validation and monitoring of laboratory analytical methods, statistical consulting.

Bernadette Govaerts is responsible for the certificate of statistics, president of the executive committee of the Statistical Methodology and Computing Support service and president of the CORE-STAT-ACTU computing commission.

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/govaerts/govaerts.html>



Christian Hafner

Professor, Université catholique de Louvain (Institute of Statistics).

PhD Humboldt-Universität Berlin, Germany, 1996.

Non and semiparametric statistics, time series, volatility models, financial econometrics.

Associate editor of *Banking and Finance Review*; *Computational Statistics*; *International Review of Econometrics*; *Studies in Nonlinear Dynamics and Econometrics*

Books: "Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility", (Physica Verlag, 1998), "Introduction to the Statistics of Financial Markets" (joint with J. Franke and W. Haerdle; Springer Verlag, third edition 2009).

CORE fellowship (1996-1997). Research associate, Humboldt- Universität Berlin (1997-1999);

Head of the Econometrics group at Electrabel (1999-2002); Assistant professor, Erasmus Universiteit Coordinator for the Statistics Minor program, pedagogic adviser for the Master program in statistics, secretary of the jury of the statistics programs and responsible for Statistics seminars organization.

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/hafner>



Catherine Legrand

Professor, Université catholique de Louvain (Institute of Statistics).

PhD Universiteit Hasselt, 2006.

Biostatistics, design and analysis of clinical trials, survival data and frailty models, competing risks, repeated events.

Board member of the biostatistic section of the Belgian Statistical Society (since 2008). In charge of Prix Quetelet.

Pedagogic adviser for the Master in Biostatistics.



Christian Ritter

Part-time Professor, Université catholique de Louvain (Institute of Statistics).

PhD University of Wisconsin, Madison, USA, 1992.

Industrial statistics, statistical consulting, statistics in spreadsheets.

Creation of Ritter and Danielson Consulting sprl (2007) to provide R&D assistance.

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/ritter/Newwriter.html>



Johan Segers

Professor, Université catholique de Louvain (Institute of Statistics).
PhD Leuven, 2001.

Extreme value theory, dependence modelling via copulas, Edgeworth expansions. Copulas, curve estimation under shape constraints.

Associate editor of the *Journal of the Royal Statistical Society, Series B* (from 2005). Co-author of "*Statistics of Extremes : Theory and Applications*" (Wiley). Extramural fellow of CenTER, Tilburg University, The Netherlands (2006-2011). Secretary of the Section Mathematical Statistics of the Netherlands Society for Statistics and Operations Research (2005-2008). Member of the Institute of Mathematical Statistics and the Belgian

Statistical Society. Coordinator of the Erasmus program at the Institute of Statistics.

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/segers/index.html>



Léopold Simar

Professor, Université catholique de Louvain (Institute of Statistics).
Part-time professor (1974-2008), Facultés Universitaires Saint-Louis, Bruxelles.
PhD Université catholique de Louvain, 1974 (Docteur en Sciences Appliquées).

Frontier estimation, resampling methods, multivariate statistical techniques.

Associate editor, *Journal of Productivity Analysis* (since 2003), Member of the European Network of Excellence Prime (2004-2009), Associate partner of the Scuola Superiore Santa Anna, Pisa, Italy (project AQUAMETH), Dean, Faculté des Sciences Economiques, Sociales et Politiques, FUSL (1978 -1990), President of the Belgian Statistical Society (1999-2002),

Founder President of the Institute of Statistics, UCL (1992-2004).

Co-author with W. Härdle of "*Applied Multivariate Statistical*" (Springer-Verlag, 2003) and with C. Daraio of "*Advanced Robust and Nonparametric Methods in Efficiency Analysis*" (Springer, 2007). "Professore di Chiara Fama" 2006 and 2007, (Italian Ministry of Research, Scuola Superiore Santa Anna, Pisa & University of Pisa) and "Chaire d'Excellence Pierre de Fermat", 2008 and 2009, Région Midi-Pyrénées: Université des Sciences Sociales, Toulouse.

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/Simar/index.html>



Sébastien Van Bellegem

Professor, Université catholique de Louvain (Institute of Statistics).

PhD Université catholique de Louvain, 2003.

Habilitation in mathematics, Université de Toulouse, 2007.

Econometric theory and microeconometrics, semi and nonparametric statistics, locally stationary time series.

Associate editor of *Eurasian Review of Econometrics* (since 2006).

Academic secretary of the Institute of Statistics (2006-2009), President of the Jury for the

Master in Statistics.

Homepage : <http://www.vanbellegem.fr/research>



Ingrid Van Keilegom

Professor, Université catholique de Louvain (Institute of Statistics).

PhD Limburgs Universitair Centrum, 1998.

Fellow of the IMS (Institute of Mathematical Statistics), 2008, and Elected Member of the ISI (International Statistical Institute), 2004.

Non- and semiparametric regression, survival analysis, bootstrap methods, mathematical statistics, empirical likelihood methods, goodness-of-fit problems.

Associate editor of *The Annals of Statistics*, *Scandinavian Journal of Statistics*, *Annals of the Institute of Statistical Mathematics*, *Statistics and Probability Letters*, and *International Journal of Biostatistics*.

Ingrid Van Keilegom follows the developments on various research funding programmes for the Institute of Statistics. Coordinator of the IAP network in statistics, and responsible for the computing facilities at the Institute, and for the CORE/STAT library.

Holder of the European Research Council grant on "*M- and Z-estimation in semiparametric statistics: applications in various fields*".

Homepage : <http://www.stat.ucl.ac.be/ISpersonnel/vankeile/vankeile.html>



Rainer von Sachs

Professor, Université catholique de Louvain (Institute of Statistics).

PhD Heidelberg, Germany, 1991.

Mathematical statistics, nonparametric curve estimation, analysis of (nonstationary) time series, spectral density estimation, statistical signal processing, biomedical time series, financial time series, Wavelets and related localization methods.

President of the FNRS Graduate School in Statistics and Actuarial Sciences of the Communauté française, member of the IMS, Bernoulli Society, ISI (elected), Belgian Statistical Society and DMV (Fachgruppe Stochastik), President of the "Ecole doctorale thématique en Statistique et en Sciences actuarielles". Principal coordinator of the ARC project 07/12-002 "*Econometric modelling of multivariate financial time series*".

Homepage : <http://www.stat.ucl.ac.be/rvs>

Associate Academic Members

Luc Bauwens, Faculty of Economics, Social and Political Science
Patrick Bogaert, Faculty of Biological, Agronomic and Environmental Engineering
Pierre Devolder, Institute of Actuarial Sciences
Marie-Paule Kestemont, Faculty of Economics, Social and Political Science
Guy Lories, Faculty of Psychology
Annie Robert, Faculty of Medicine
Michel Verleysen, Faculty of Applied Mathematics

Invited Academic Members

Yves Berger
Céline Bugli
Libei Chen
Isabelle De Macq
Cédric Heuchenne
Philippe Lambert
Rosa Maria Crujeiras-Casais

Emeriti

Michel Mouchart
José Paris
Jean-Marie Rolin

Researchers and Doctoral Students

Louis Boulanger
Hilmar Böhm
Alexandra Daskovska
Rachida El Medhi
Nancy François
Jean-Marc Freyermuth
Gery Geenens
Gordon Gudendorf
Julien Hunt
Jonathan Jaeger
Paul Kounche
Thomas Meinguet
Giovanni Motta
Mathieu Pigeon
Olga Reznikova
Mohammed Rida Soumali
Réjane Rousseau
Maik Schwarz
Aleksandar Sujica
Majda Talamakrouni
Cédric Taverne
Bianca Teodorescu
Catherine Timmermans

Postdoctoral Researchers

Rosa Maria Crujeiras-Casais
Laurent Delsol
Stéphane Laurent

Associate Researchers

Taoufik Bouezmarni
Céline Bugli
Véronique Delouille
Maria Key Prato
Philippe Vanden Eeckaut

Visiting Researchers

Claudia Curi, University of Roma, Italy (20/09/08 - 11/10/08)
Christian Haedo, Argentina (30/10/08 - 01/02/09)
Hans Manner, University of Maastricht, The Netherlands (September 2008 - December 2008)
Nicolas Meyer, MCU-PH, France (01/10/08-30/09/09)

Doctors Honoris Causa

Luc Devroye, Mc Gill University, Montréal, Canada (2002)
Peter Hall, Australian National University, Canberra, Australia (1997)

Computer Scientists SGSI (Service général du système d'information)

Laurent Buset
Jean-Luc Marrion

Library

Alain Gillis

SMCS Staff (Support en Méthodologie et Calcul Statistique)

Alain Guillet
Nathalie Lefèvre (since September 2008)
Jean-Marie Zélis

Head of Unit (Administration)

Christine Denayer (until March 2008)
Isabelle Petit (since August 2008)

Administrative Staff

Anne-Marie Bellemans (until April 2008)
Marguerite-Marie Hanon
Sophie Malali
Monique Tanga

3. Research activities

3.1. Presentation

The research areas in which the members of the Institute of Statistics are working are diverse. The main areas of expertise are non- and semi-parametric regression techniques, time series analysis, survival analysis, medical and industrial statistics, extreme value analysis and statistics for the actuarial sciences.

In the context of non- and semi-parametric regression, the Institute is a leading expert in the area of frontier estimation, and particularly in the application of non- and semi-parametric approaches for this problem in the context of efficiency analysis. A lot of research is also carried out in the context of inverse problems in econometrics, and its applications in instrumental regression and deconvolution problems. The study of semi-parametric regression models (e.g. single index models, partial linear models) and of non-parametric location-scale models, is another area in which the Institute is taking a leading role. Denoising and functional image clustering by means of wavelets are also highly studied.

The analysis of time series is a second cornerstone of the research activities at the Institute. The focus lies on the modelling and analysis of non-stationary time series, multivariate (high-dimensional) time series, factor models, volatility models, spectral density estimation and goodness-of-fit methods. Also, applications in statistical signal processing, and biomedical, economic and financial time series are studied.

The analysis of data coming from medical or industrial studies is also a research topic to which much attention is paid at the Institute. Medical data are often subject to censoring (survival analysis). The non- and semi-parametric modelling of this type of data is studied in detail, both the asymptotics for these models, as the application to medical data. In the context of industrial statistics, the focus lies on experimental design and multicriteria optimisation with applications in drug discovery, and on the analysis and modelling of time intensity curves in sensometrics.

The development of methods for extreme value analysis is also of interest at the Institute : modelling of extremes in univariate and multivariate time series, and in particular in Markov chains.

The modelling of dependence by means of copulas, the development of stochastic inequalities (stochastic orders, stochastic extrema), and the application of statistical concepts to the research area of insurance are also highly studied.

3.2. Research contracts

3.2.1. Research projects under contracts and cooperation projects

This section discusses ongoing research projects and cooperation projects that are financed by outside agencies in the form of grants and contracts.

Statistical analysis of association and dependence in complex data - IAP Phase VI, (2007-2011)

Financing: Interuniversity Attraction Pole Programme, Belgian Science Policy, Brussels, Belgium

Coordinators: L. Simar and I. Van Keilegom

STAT researchers: H. Böhm, A. Daskovska, R. Crujeiras, J. Jaeger, S. Laurent and A. Sujica

Partner institutions: Katholieke Universiteit Leuven, Belgium
Universiteit Hasselt, Belgium
Universiteit Gent, Belgium
Université Joseph Fourier, Grenoble, France
Universiteit Utrecht, Netherlands
Universidad de Santiago de Compostela, Spain
London School of Hygiene and Tropical Medicine, UK

Project description:

One key aim of statistics is to analyse in an appropriate way the dependence and association present in a dataset. The data collected to analyse these dependence structures, are often of a complex nature and also the research questions are of an ever increasing complexity. This requires the construction of new models, or the adaptation of existing models which is a challenging task. The development of new methods and intensive interaction between experts will also be required to cope with these complex data. The global objective of the network is to develop new models and methodological tools to infer and to analyse these complex data structures.

Econometric modelling of multivariate financial time series (2007-2012)

Financing: Action de Recherche Concertée (ARC), Communauté Française de Belgique

Grant holders: L. Bauwens, Ch. Hafner, S. Van Belleghem and R. von Sachs

STAT researchers: A. Daskovska, J-M. Freyermuth and O. Reznikova

Project description:

This interdisciplinary research project deals with modelling, estimation and prediction of the dynamics and the temporal dependence in the mean and the variance-covariance structure of multivariate time series data arising in economical and financial applications. Particular emphasis is put on questions such as dimension reduction (factor approach, modelling of co-movements), non-stationary behaviour over time, modelling of structural breaks (regime-switching), volatilities with and without jump behaviour, etc. These questions are addressed by a number of econometricians and statisticians using and comparing a series of modern approaches in parametric, semi-parametric and non-parametric statistics. Applications to real data will help to assess the quality of the proposed models and estimation procedures.

"M- and Z-estimation in semiparametric statistics: applications in various fields" (2008-2012)

Financing: European Research Council under the European Community's Seventh Framework Programme ERC Grant, 2008-2012

Grant holder: I. Van Keilegom (individual grant)

STAT researcher: L. Delsol

Project description:

Semiparametric statistics is a branch of statistics that is, in comparison to the areas of fully parametric or nonparametric statistics, relatively unexplored and still in full development.

Semiparametric models offer a valid alternative for purely parametric ones, that are known to be sensitive to incorrect model specification, and completely nonparametric models, which often suffer from lack of precision and power. A drawback of semiparametric models so far is, however, that the development of mathematical properties under these models is often a lot harder than under the other two types of models. The present project tries to solve this difficulty partially, by presenting and applying a general method to prove the asymptotic properties of estimators for a wide spectrum of semiparametric models.

The objectives of this project are twofold. On one hand a general theory will be applied by Chen, Linton and Van Keilegom (2003) for a class of semiparametric Z-estimation problems, to a number of novel research ideas, coming from a broad range of areas in statistics. On the other hand it will be shown that some estimation problems are not covered by this theory. A more general class of semiparametric estimators (M-estimators called) will be considered and a general theory for this class of estimators will be developed. This theory will open new horizons for a wide variety of problems in semiparametric statistics.

Policies for research and innovation in the move towards the European research area, (2004-2009)
(Member of the European Network of Excellence PRIME)

Financing: European Community

Grant holder: L. Simar, Associate partner of the Scuola Superiore Santa Anna, Pisa, Italy

Project description:

Using advanced quantitative methods for evaluating the performance of Public Sector Research (AQUAMETH). Analysing in particular the system of the European universities and the availability of national data at the microlevel.

The "Helga and Wolfgang Gaul Stiftung", (2008-2009)

Financing: The Helga and Wolfgang Gaul Stiftung, Fakultät für Wirtschaftswissenschaften, Universität Karlsruhe (TH), Germany

Grant holder: L. Simar

Efficiency and Productivity of an Industry, (2009-2011)

Research grant and member of the research team: L. Simar

Partner institution: GREMAQ, Toulouse School of Economics, Agence Nationale de la Recherche (ANR), France

Functional Statistical Inference. New approaches in finance, environmental science, industry and economy (2005-2008)

Financing: Spanish Ministry of Education and Science

Partner: I. Van Keilegom

Principal investigator: Wenceslao González-Manteiga

Project description:

Many efforts have been devoted, in recent times, to the study of statistical inference on curves. In this project, the aim is to develop new theoretical methodologies in order to make inference on curves, with applications in different frameworks: finite populations, goodness-of-fit testing, neural networks, machine learning, functional data analysis, set estimation, spatial and spatio-temporal statistics and finance series. Many of the new methods to be developed, will be implemented for the first time. They may also represent modifications on previous versions and their validity will be tested by theoretical results and simulation studies. Applications will be made on different settings as environmental science (SO₂ predictions in a power plant, concentration of heavy metals in mosses in Galicia...), industry (control of

wastewater treatment by anaerobic processes) and finance (semi-parametric inference in portfolio design).

Adaptative regularization in nonparametric instrumental regression (2007-2009)

Financing: Fonds Spéciaux de Recherche (FSR)

Grant holder: S. Van Bellegem

Project description:

We focus on the nonparametric estimation of a regression function in endogenous models.

The regression function is identified through a set of instrumental variables. Stability problems occur in the estimation process, leading to an ill-posed inverse problem. The general objective is to develop adaptive methods of inversion (of regularization), based on the theory of adaptive, nonparametric, local inference.

Développement d'un outil automatisé de détection d'erreurs dans les données d'essais cliniques par le biais de méthodes statistiques et graphiques" (2008)

Promoters: Catherine Legrand (UCL), Nathalie Lefèvre (UCL), Marc Buysse (IDDI).

Project description :

This project aims at developing an objective method of quality control for data coming from clinical trials with the idea of detecting systematic errors. Developing such a method requires the use of appropriate statistical methods and developments of various graphical tools. Part of these statistical methods have already been implemented in a tool called SAFE for « Statistical Alerts For Errors », with the idea of being able to apply the various tests considered in a fully automatic way. Various types of errors and frauds can be observed in clinical trial data and therefore various statistical techniques will play a role and new statistical techniques will have to be developed to cover all the needs.

"Modern risk management models for insurance companies and pension funds" (2004-2009)

Financing: ARC Communauté française Wallonie-Bruxelles,

Grant holders: M. Denuit, P. Devolder, J-M. Rolin and Y. Smeers

Researchers: C. Courtois, D. Hainaut, A. Miller, J-P. Boucher, J. Trufin and B-L. Yerna

Project description:

1. Generalized life insurance policies
2. Risk theory under partial information
3. Unification of actuarial and financial pricing techniques
4. Risk management in incomplete markets

3.2.2. Applied research contracts

The Institute of Statistics is developing many collaborations within the Université catholique de Louvain and with several private firms in the field of applied statistics. In addition to seminars organized weekly (see point 3.7), the members of the Institute participate to research contracts in applied statistics and offer consulting services to other departments and institutions of the University (see point 3.2).

They also offer some courses on continued education (see point 4) at the University and in companies.

Implementation of a quality control plan for the monitoring of blood treatment processes for the Belgian Red Cross (2006-2007)

Financing: Belgian Red Cross

Grant holder: B. Govaerts

STAT researchers: A. Guillet

The Belgian Red Cross has to develop quality control procedures to monitor its blood treatment processes in order to be in conformity with the Belgian legislation. This project implies the adaptation of statistical quality control techniques like control charts, capability analysis and reception sampling to the particular problematic of blood treatment. The project aims at developing the necessary tools and implementing them in an adapted computing tool.

Analyzing the results of a designed experiment when the result is a curve

Financing: SANOFI-AVENTIS, France (2008-2009)

Grant holder: B. Govaerts

The goal of this collaboration project is to compare and implement different functional models to analyze the results of industrial experiments when the response is a curve their application to SANOFI problems of interest like chemical kinetic curves is also examined.

3.3. Reprints

RP0801 - Einmahl, J. H. J. and I. Van Keilegom
Specification tests in nonparametric regression.
Journal of Econometrics, 143, 88-102, 2008

RP0802 - Daouia, A., Florens, J.-P. and L. Simar
Functional convergence of quantile-type frontiers with application to parametric approximations.
Journal of Statistical Planning and Inference, 138, 708-725, 2008

RP0803 - Simar, L. and P. W. Wilson
Statistical inference in nonparametric frontier models: recent developments and perspectives.
The Measurements of Productive Efficiency, 2nd Edition, Harold Fried, C. A. Knox Lovell and Shelton Schmidt (eds), Oxford University Press, 2008

RP0804 - Linton, O., Sperlich, S. and I. Van Keilegom
Estimation of a semiparametric transformation model.
Annals of Statistics, 36, 2, 686-718, 2008

RP0805 - El Ghouch, A. and I. Van Keilegom
Non-parametric regression with dependent censored data.
Scandinavian Journal of Statistics, 35, 228-247, 2008

RP0806 - Ombao, H. and S. Van Bellegem
Evolutionary coherence of nonstationary signals.
IEEE Transactions on Signal Processing, 56, (6), 2259-2266, 2008

RP0807 - Van Bellegem, S. and R. von Sachs
Locally adaptive estimation of sparse evolutionary wavelet spectra.
The Annals of Statistics, 36, (4), 1879-1924, 2008.

RP0808 - Gao, J., Gijbels, I. and S. Van Bellegem
Nonparametric simultaneous testing for structural breaks.
Journal of Econometrics, 143, 123-142, 2008

RP0809 - Böhm, H. and R. von Sachs
Structural shrinkage of nonparametric spectral estimators for multivariate time series.
Electronic Journal of Statistics, 2, 696-721, 2008

RP0810 - Fryzlewicz, P., Nason, G. P. and R. von Sachs
A wavelet-Fisz approach to spectrum estimation.
Journal of Time Series Analysis, 29, (5), 868-880, 2008

RP0811 - Florens, J.-P., Fougère, D. and M. Mouchart
Duration models and point processes.
The Econometrics of Panel Data, Mátỳàs, L. and P. Sevestre (eds), Springer-Verlag Berlin Heidelberg, 547-601, 2008

RP0812 - Govaerts, B., Dewé, W., Maumy, M. and B. Boulanger
Pre-study analytical method validation: comparison of four alternative approaches based on quality-level estimation and tolerance intervals.
Quality and Reliability Engineering International, 24, 667-680, 2008

RP0813 - Rousseau, R., Govaerts, B., Verleysen, M. and B. Boulanger
Comparison of some chemometric tools for metabonomics biomarker identification.
Chemometrics and Intelligent Laboratory Systems, 91, 54-66, 2008

RP0814 - Van Keilegom, I., González-Manteiga, W. and C. Sánchez-Sellero
Goodness-of-fit tests in parametric regression based on the estimation of the error distribution.
Test, 17, 401-415, 2008

RP0815 - Wang, L., Akritas, M. G. and I. Van Keilegom
An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models.
Journal of Nonparametric Statistics, 20, (5), 365-382, 2008

- RP0816** - Robert, C. Y. and J. Segers
Tails of random sums of a heavy-tailed number of light-tailed terms.
Insurance: Mathematics and Economics, 43, 85-92, 2008
- RP0817** - Einmahl, J. H. J., Krajina, A. and J. Segers
A method of moments estimator of tail dependence.
Bernoulli, 14, 1003-1026, 2008
- RP0818** - Fils-Villetard, A., Guillou, A. and J. Segers
Projection estimators of Pickands dependence functions.
The Canadian Journal of Statistics, 36, (3), 369-382, 2008
- RP0819** - Beuthe, M., Bouffieux, C., Krier, C. and M. Mouchart
A comparison of conjoint, multi-criteria, conditional logit and neural network analyses for rank-ordered preference data.
Recent Developments in Transport Modelling: Lessons for the Freight Sector, Ben-Akiva, M., Meersman, H. and E. Van de Voorde (eds), Emerald Group Publishing Limited, UK, 157-178, 2008
- RP0820** - Lebrun, P., Govaerts, B., Debrus, B., Ceccato, A., Caliaro, G., Hubert, P. and B. Boulanger
Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods.
Chemometrics and Intelligent Laboratory Systems, 91, 4-16, 2008
- RP0821** - Van Keilegom, I., Sánchez-Sellero, C. and W. González-Manteiga
Empirical likelihood based testing for regression.
Electronic Journal of Statistics, 2, 581-604, 2008
- RP0822** - Einmahl, J. and I. Van Keilegom
Tests for independence in nonparametric regression.
Statistica Sinica, 18, 601-616, 2008
- RP0823** - Heuchenne, C.
Strong uniform consistency results of the weighted average of conditional artificial data points.
Journal of Statistical Planning and Inference, 138, 1496-1515, 2008
- RP0824** - González-Manteiga, W., Heuchenne, C. and C. Sánchez-Sellero
Parametric conditional mean and variance testing with censored data.
Recent Advances in Applied Stochastic Models and Data Analysis. H. Skiadas, Christos, (ed.), N.p.: World Scientific Publishing Co. Pte Ltd, 2008
- RP0825** - Sougne, D., Heuchenne, C. and G. Hubner
The determinants of CDS prices: an industry-based investigation.
Credit Risk: Models, Derivatives and Management. Empirical Studies and Analysis. Wagner, Niklas, (ed.), Financial Mathematics Series, Volume 6., London, New York: Chapman & Hall / CRC, 85-96, 2008
- RP0826** - Li, S.-K., Simar, L., Wilson, P. W. and V. Zelenyuk
Introduction for the special issue on transitioning economics.
Journal of Productivity Analysis, 29, 77, 2008
- RP0827** - Park, B., Simar, L. and V. Zelenyuk
Local likelihood estimation of truncated regression and its partial derivatives: theory and application.
Journal of Econometrics, 146, 185-198, 2008
- RP0828** - Kneip, A., Simar, L. and P. W. Wilson
Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models.
Econometric Theory, 24, 1663-1697, 2008

3.4. Discussion papers

The abstracts of the discussion papers can be found in the Appendix.

- 0801** Geenens, G. and L. Simar
Nonparametric test for conditional independence in two-way contingency tables
- 0802** Geenens, G.
Explicit formula for asymptotic higher moments of the Nadaraya-Watson estimator
- 0803** Van Keilegom, I., de Uña-Alvarez, J. and L. Meira-Machado
Nonparametric location-scale models for successive survival times under dependent censoring
- 0804** Ojeda Cabrera, J. L. and I. Van Keilegom
Goodness-of-fit tests for parametric regression with selection biased data
- 0805** Böhm H. and R. von Sachs
Structural shrinkage of nonparametric spectral estimators for multivariate time series
- 0806** Daouia, A., Florens, J.-P. and L. Simar
Frontier estimation and extreme values theory
- 0807** Geenens, G. and L. Simar
Single-index modelling of conditional probabilities in two-way contingency tables
- 0808** Charpentier, A. and J. Segers
Tails of multivariate Archimedean copulas
- 0809** Fils-Villetard, A., Guillou, A. and J. Segers
Projection estimators of Pickands dependence functions
- 0810** Govaerts, B., Dewé, W., Maumy, M. and B. Boulanger
Pre-study analytical method validation: comparison of four alternative approaches based on quality level estimation and tolerance intervals
- 0811** Debrus, B., Lebrun, P., Ceccato, A., Caliaro, G., Govaerts, B., Olsen, B., Rozet, E., Boulanger, B. and P. Hubert
Use of ICA on HPCL-DAD data and high-order statistics to automatically achieve peak picking
- 0812** Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke
Empirical likelihood for non-smooth criterion functions
- 0813** San Martin, E., Jara, A., Rolin, J.-M. and M. Mouchart
On the analysis of Bayesian semiparametric IRT-type models
- 0814** Segers, J., Van Den Akker, R. and B.J.M. Werker
Improving upon the marginal empirical distribution functions when the copula is known
- 0815** Einmahl, H.J. and J. Segers
Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution
- 0816** Neumeyer, N. and I. Van Keilegom
Estimating the error distribution in nonparametric multiple regression with applications to model testing
- 0817** Jullion, A., Lambert, Ph. and F. Vandenhende
Estimation of receptor occupancy using varying coefficients models
- 0818** Ferraty, F., Van Keilegom, I. and P. Vieu
On the validity of the bootstrap in nonparametric functional regression

- 0819** Jullion, A. and Ph. Lambert
Extensions of Bayesian P-splines models for fitting PK curves
- 0820** Simar, L. and V. Zelenyuk
Stochastic FDH/DEA estimators for frontier analysis
- 0821** Daskovska, A., Simar, L. and S. Van Bellegem
Forecasting the Malmquist productivity index
- 0822** Bouezmarni, T., Rombouts, J. and A. Taamouti
Asymptotic properties of the Bernstein density copula for dependent data
- 0823** Teodorescu, B. and I. Van Keilegom
Goodness-of-fit test for generalized conditional linear models under left truncation and right censoring
- 0824** Beirlant, J., Joossens, E. and J. Segers
Second-order refined peaks-over-threshold modelling for heavy-tailed distributions
- 0825** Linton, O., Mammen, E., Nielsen, J.P. and I. Van Keilegom
Nonparametric regression with filtered data
- 0826** Russo, F., Wunsch, G. and M. Mouchart
Potential outcomes, counterfactuals, and structural modelling. Causal approaches in the social sciences
- 0827** Van Keilegom, I. and L. Wang
Semiparametric modeling and estimation of the dispersion function in regression
- 0828** Badin, L., Daraio, C. and L. Simar
Optimal bandwidth selection for conditional efficiency measures: a data-driven approach
- 0829** Eichler, M., Motta, G. and R. von Sachs
Fitting dynamic factor models to nonstationary time series
- 0830** Lopez, O., Patilea, V. and I. Van Keilegom
Single index regression models in the presence of censoring depending on the covariates
- 0831** Robert, C. Y., Segers, J. and C. A. T. Ferro
A sliding blocks estimator for the extremal index
- 0832** Crujeiras, R. M. and I. Van Keilegom
Least squares estimation of nonlinear spatial trends
- 0833** González-Manteiga, W., Pardo-Fernández, J. C. and I. Van Keilegom
ROC curves in nonparametric location-scale regression models
- 0834** González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero
Goodness-of-fit tests for censored regression based on artificial data points

3.5. Books

Härdle, W. and L. Simar (1993)

"Computer Intensive Methods in Statistics" – 175 p., Vol I, Statistics and Computing, Physica-Verlag, Berlin

Härdle, W., Klinke, S. and B.A. Turlach (1995)

"WploRE : An Interactive Statistical Computing Environment" – 387 p., Statistics and Computing, Springer-Verlag, New-York

Fan, J. and I. Gijbels (1996)

"Local Polynomial Modelling and its Applications" – Chapman and Hall: London

Hafner, C. (1998)

"Nonlinear Time Series Analysis with Applications to Foreign Exchange Rate Volatility" – 222 p., Physical Verlag Heidelberg

Florens, J.-P., Mouchart, M. and J.-M. Rolin (1999)

"Elements of Bayesian Statistics" – 544 p., Marcel Dekker: New-York

Kaas, R., Goovaerts, M.J., Dhaene, J. and M. Denuit (2001)

"Modern Actuarial Risk Theory" - Kluwer Academic Publishers, Dordrecht,

Hafner, C., Franke, J. and W. Härdle (2001)

"Einführung in die Statistik der Finanzmärkte" – 358 p., Springer-Verlag, New York

Wunsch, G., Mouchart, M. and J. Duchêne (2002)

"The Life Table : Modelling Survival and Death : Book series : European Studies of Population", vol. 11, Kluwer Academic Publishers, Dordrecht

Härdle, W. and L. Simar (2003)

"Applied Multivariate Statistical Analysis" - 486 p., Springer-Verlag, Berlin

Beirlant, J., Goegebeur, Y., Segers, J. and J. Teugels (2004)

"Statistics of Extremes: Theory and Applications" – John Wiley & Sons, New-York

Delwarde, A. and M. Denuit (2005)

"Construction de Tables de Mortalité Périodiques et Prospectives" - Audit-Actuariat-Assurances, Economica, Paris

Denuit, M., Dhaene, J., Goovaerts, M.J. and R. Kaas, (2005)

"Actuarial Theory for Dependent Risk: Measures, Orders and Models" - John Wiley & Sons, New York

Daraio, C. and L. Simar (2007)

"Advanced Robust and Nonparametric Methods in Efficiency Analysis: Methodology and Applications" Springer, New-York

Denuit, M., Maréchal, X., Pitrebois, S. and J.F. Walhin (2007)

"Actuarial Modelling of Claim Counts : Risk Classification, Credibility and Bonus-Malus Systems" Wiley, New York

Denuit M. and C. Robert (2007)

"Actuariat des Assurances de Personnes : Modélisation, Tarification et Provisionnement" Collection Audit-Actuariat-Assurances, Economica, Paris

Deprins, D. (2007)

"Le Pari de Pascal, une allégorie du Principe de Précaution", sur site de l'Observatoire du Principe de Précaution

Deprins, D., (2007)

"Quel homme se construit sous le paradigme de Précaution?", sur site de l'Observatoire du Principe de Précaution"

Härdle, W. and L. Simar (2007)

"Applied Multivariate Statistical Analysis" - Second Edition, 458 p., Springer-Verlag, Berlin

Florens, J-P., Mouchart, M. and J-M. Rolin **(2008)**
"Elements of Bayesian Statistics" - 544 p., Marcel Dekker, New York

Kaas, R., Goovaerts, M-J., Dhaene, J. and M. Denuit **(2008)**
"Modern Actuarial Risk Theory Using R.", Springer-Verlag, New York

3.6. Visitors

Short terms visitors

ASTON John, University of Warwick, United Kingdom
BACLIN Angélique, GSK, Belgium
BERTRAND Frédéric, Université Paul Sabatier, Strasbourg, France
BIGOT Jérémie, Université de Toulouse, France
BOUDT Kris, KUL, Leuven, Belgium
CAO Ricardo, University of Coruña, Spain
CHINE Karim, Cloud Era Ltd, United Kingdom
CHOROS Barbara, Humboldt-Universität, Berlin, Germany
CHRISTIANSEN Marcus C., Universität Rostock, Germany
CRUCIFIX Michel, Institut d'Astronomie et de Géophysique, UCL, Louvain-la-Neuve, Belgium
CRUJEIRAS Rosa Maria, Universidad de Santiago de Compostela, Spain
CURI Claudia, Università di Roma, Italy
DARAIIO Cinzia, Università di Pisa, Italy
DAVID Morgane, Solidpartners, Belgium
DAVISON Anthony, EPLF, Lausanne, Suisse
DELECROIX Michel, ENSAI Rennes, France
DELSOL Laurent, Université de Toulouse, France
DEO Rohit, Stern Business School, University of New York, USA
DETTE Holger, Universität Bochum, Germany
DRAYE Xavier, UCL, Louvain-la-Neuve, Belgium
DONY Julia, VUB, Brussels, Belgium
DUCROCQ Vincent, INRA, France
DUEMBGEN Lutz, University of Bern, Switzerland
DUQUESNE Marc, AXA, Belgium
EICHLER Michaël, University of Maastricht, The Netherlands
EINMAHL John, University of Tilburg, The Netherlands
GAMPE Jutta, Max-Planck Institute, Rostock, Germany
GILLAIN Christophe, AXA, Belgium
GONZALEZ-MANTEIGA, Wenceslao, Universidad de Santiago de Compostela, Spain
HALDRUP Niels, University of Aarhus, Denmark
HALLIN Marc, ULB, Brussels, Belgium
JANSSEN Paul, University of Hasselt, Belgium
JEONG Seok-Oh, Hankuk University, South Korea
JOHANNES Jan, University of Heidelberg, Germany
JONES Bradley, SAS, USA
KIRCH Claudia, University of Kaiserslautern, Germany
KNEIP Alois, University of Bonn, Germany
KREISS Jens-Peter, Technische Universität Braunschweig, Germany
KYUSANG Yu, University of Mannheim, Germany
LAFAYE de MICHEAUX Pierre, Université Pierre-Mendès-France, Grenoble , France
LANG Stefan, University of Innsbruck, Austria
LIPPI Marco, University of Roma "La Sapienza", Roma, Italy
MANNER Hans, University of Maastricht, The Netherlands
MASSONNET Goele, University of Hasselt, Belgium
MASUY-STROOBANT Godelieve, Institut de Démographie, UCL, Louvain-la-Neuve, Belgium

MAUMY Myriam, Université Paul Sabatier, Strasbourg, France
MELARD Guy, ULB, Brussels, Belgium
MEYER Nicolas, MCU-PH, France
NESLEHOVA Johanna, ETH, Zürich, Switzerland
OMBAO Hernando, Brown University, Providence, USA
PARDO-FERNANDEZ, Juan Carlos, Universidade de VIGO, Spain
PARK Byeong, University of Seoul, South Korea
PLANCHON Viviane, Centre wallon de recherches agronomiques, Gembloux, Belgium
REISS Julian, Erasmus University of Rotterdam, The Netherlands
RESNICK Sidney, Cornell University, Ithaca, NY, USA
RIZOPOULOS Dimitris, Erasmus University, Rotterdam, The Netherlands
ROOTZEN Holger, University of Göteborg, Sweden
ROUEFF Christian, ENS Télécommunication, TSI, Paris, France
ROUSTANT Olivier, Ecole des Mines de Saint-Etienne, France
SCHWARZ Maik, University of Heidelberg, Germany
SPOKOINY Vladimir, Wierstrass Institute, Berlin, Germany
TADJUIDJE K. Joseph, University of Kaiserslautern, Germany
TERÄSVIRTA, Timo, Aarhus University, Denmark
THEIS Winfried, Unilever, the Netherlands
VERAVERBEKE Noël, University of Hasselt, Belgium
WEIHS Claus, University of Bonn, Germany
WILSON Paul, Clemson University, South Carolina, USA

3.7. Statistics seminars

February 08, 2008

Wenceslao Gonzalez Manteiga, Universidad de Santiago de Compostela, Spain - joint seminar with PAI
"Using SiZer map for testing problems"

John Einmahl, University of Tilburg, The Netherlands
"Asymptotics for the Hirsch index"

March 07, 2008 - Afternoon on "Extreme Value Theory"

Jutta Gampe, Max-Planck Institut für demografische Forschung, Rostock, Germany
"Estimating the best-practice mortality surface: a meta-analytic framework"

Holger Rootzen, Chalmers University of Technology, Göteborg, Sweden
"The multivariate generalized Pareto distribution: with an application to wind storm insurance"

March 14, 2008 - Afternoon on "Time Series Analysis"

Guy Mélard, ULB, Belgium
"A new recursive estimation for single input output models" - joint work with Abdelhamid Ouakasse

Christian Hafner, Institut de statistique, UCL, Belgium
"Efficient estimation of a multivariate multiplicative volatility model" - joint work with Oliver Linton

April 11, 2008 - Afternoon on "Causality"

Julian Reiss, Erasmus Universiteit of Rotterdam, The Netherlands
"Evaluating policy counterfactuals" - joint work with Nancy Cartwright

Guillaume Wunsch, UCL, Belgium
"Potential outcomes, counterfactuals, and structural modelling" - joint work with M. Mouchart and F. Russo

April 25, 2008

Stefan Lang, University of Innsbruck, Austria
"Hierarchical generalized structured additive regression"

May 16, 2008 - Afternoon on "Survival Analysis"

Catherine Legrand, Institut de statistique, UCL, Belgium
"A Bayesian approach to jointly estimate a random effect and a random interaction in a proportional hazards model. Application to the validation of a prognostic index in bladder cancer patients" - joint work with Paul Janssen, (Hasselt University), Luc Duchateau, (Ghent University), Richard Sylvester, (EORTC) and Vincent Ducrocq (INRA)

Noël Veraverbeke, University of Hasselt, Belgium
"Copulas in survival model"

September 24, 2008 - Afternoon on "Modelling Electricity Prices" - joint CORE/STAT seminars in the framework of the ARC project on time series

Niels Haldrup, University of Aarhus, Denmark
"A vector autoregressive model for electricity prices subject to long memory and regime switching"

Siem Jan Koopman, Vrije Universiteit Amsterdam, The Netherlands
"A multivariate model for hourly electricity prices: empirical evidence from two European markets"

October 10, 2008 - Afternoon on "Copulas" - joint CORE/STAT/ARC seminars

Claudia Kirch, University of Kaiserslautern, Germany
"Resampling methods in change-point analysis"

John Aston, University of Warwick, United Kingdom
"Statistical segmentation of time series through hidden Markov models"

October 29, 2008 - Afternoon on "Copulas"

Claudia Czado, Technische Universität München, Germany
"Statistical inference for pair-copula constructions of multiple dependence" - joint work with A. Min,

Alfonso Valdesogo, CORE, UCL, Belgium
"A canonical vine autoregressive model for large dimensions" - joint work with A. Heinen

October 31, 2008

Timo Teräsvirta, University of Aarhus, Denmark
"Conditional correlations models of autoregressive conditional heteroskedasticity with nonstationary GARCH equations"

Kris Boudt, KULeuven, Belgium
"Outlyingness weighted quadratic covariation"

November 14, 2008 - "Survival analysis" - joint STAT/IAP seminars

Dimitris Rizopoulos, University of Erasmus, Rotterdam
"Laplace approximations for joint models"

Goele Massonnet, Universiteit Hasselt, Belgium
"Contributions to frailty and copula modelling with applications to clinical trials and dairy cows data"

November 21, 2008 - joint STAT/ULB seminars

Marc Hallin, ULB, Belgium
"From distribution-freeness to semiparametric efficiency: sixty years of rank-based inference"

Ingrid Van Keilegom, Institut de statistique, UCL, Belgium
"Semiparametric modeling and estimation of the dispersion function in regression"

December 12, 2008

Rohit Deo, Stern Business School, New York, USA
"The Restricted Likelihood (REML) in Econometrics: estimation and likelihood based inference in autoregressive models, predictive regressions, dynamic panel data and co-integrated systems"

December 05, 2008

Kyusang Yu, Universität Mannheim, Germany
"Non- and semi-parametric additive regression for repeatedly measured data"

Julia Dony, VUB, Brussels, Belgium
"An empirical process approach to uniform in bandwidth consistency of kernel-type estimators"

3.8. Organization of scientific meetings

January 22 - 25, 2008: 10th edition of the European Symposium on Statistical Methods for the Food Industry

These days offer an ideal opportunity to statisticians, engineers, and users of statistical methods in food industries to meet and exchange experiences around stimulating themes like sensometrics, chemometrics, risk analysis and process control.

It was organized by the Institute of Statistics, with the collaboration of the Faculté d'Ingénierie Biologique, Agronomique et Environnementale of UCL, the Walloon Agricultural Research Centre and the Gembloux Agricultural University.

A special afternoon about introduction to sensory analysis and sensometrics was proposed to members of industries and research laboratories

Program:

January 22, 2008

Opening day: Introduction to sensory analysis and sensometrics for industry and research laboratories

January 23, 2008

Process Control Session:

Burger James, BurgerMetrics SIA, Latvia

"Multivariate analysis of hyperspectral images with application results related to spatial/spectral analysis of foods"

De Beer Th., Vercruyse P., Alleso M., Rantanen J., Vander Heyden Y., Remon J.-P., Vervaet C., Baeyens W., Ghent University, Belgium

"Implementation of a PAT system in a pharmaceutical freeze drying process"

Rozet Eric, Dewé W., Boulanger B., Ziemons E., Moonen F., Hubert Ph., Université de Liège, Belgium

"Simultaneous improvement of the predictive character of the validation data and estimation of measurement uncertainty"

Chemometrics I Session

François D., Krier C., Rossi F., Verleysen M., Université catholique de Louvain, Belgium

"Estimation de redondance pour le clustering de variables spectrales"

Naes T., Mage I., Jorgensen K., Mevik B.-H., Matforsk, Norway

"Simultaneous modelling of process variables and spectroscopic measurements taken at the raw material stage or during processing"

Reynes Ch., Sabatier R., Université de Montpellier I, France

"Recherche de pics caractéristiques d'un phénomène de maturation par algorithmes génétiques pour des spectres de masses"

Chemometrics II Session

Vert J.-P., Ecole des Mines de Paris, France

"Criblage virtuel et QSAR par méthodes à noyaux: des nouvelles méthodes d'apprentissage statistique pour prédire des propriétés de molécules à partir de leur structure 2D ou 3D"

Fernandez Pierna J.-A., Baeten V., Dardenne P., CRA Gembloux, Belgium

"Processing of food and feed spectroscopy data using support vector machines"

Kuentez V., Chavent M., Saracco J., Université de Bordeaux, France
"Une approche divisive de classification hiérarchique de variables quantitatives et qualitatives"

Lucadamo A., Gallo M., Università del Piemonte Orientale, A. Avogadro, Italy
"PARAFAC/CANDECOM analysis for compostional data"

Qannari E.-M., Hanafi M., Kohler A., Mazerolles G., ENITIAA/INRA, Nantes, France
"Analyse en composantes communes et poids spécifiques : nouvelles propriétés et extensions"

Sensometrics I Session

Sabatier R., Reynès C., Université de Montpellier I, France
"Deux extensions de l'ACP simple appliquées à l'analyse sensorielle"

Amenta P., University of Sannio, Italy
"Interpolative and predictive biplots applied to Co-intertia analysis"

January 24, 2008

Risk Analysis Session

Rousseau J., Université Paris Dauphine, France
"Elicitation multi-experts de lois a priori en statistique bayésienne"

Bollaerts K., Messens W., Aerts M., Grijspeerdt K., Boone I., Delhalle L., Dewulf J., Debussier E., Mintiens K., Hasselt University, Belgium
"Development of a modular quantitative microbial risk assessment to evaluate zoonotic risks in Belgium. Salmonellosis through consumption of pork as an example"

Briand B., Mercat-Rommens C., Ducharme G., IRSN, France
"Une méthodologie statistique pour expliquer les différents niveaux de contamination radioactive des végétaux"

Reese A., Lee R., CEFAS, UK
"Statistical issues relating to microbiological monitoring of selffish production areas"

Sensometrics II Session

Luciano G., Naes T., Tomic O., Brockhoff P., Nilsen A., Matforsk, Norway
"Monitoring of sensory panel performance by use of the PanelCheck"

Teillet E., Urbano C., Cordelle S., Schlich P., CNRS-UB-INRA, France
"Déterminer le goût de l'eau ne coule pas de source ..."

Zalila Z., Cuquemelle J., Chikh A., Penet C., Lorentz B., Deschamps D., Intellitech, France
"Contribution of xtractis methodology to the automatic extraction of robust fuzzy models. Application to the prediction of consumer liking and sensory evaluation, and to the optimization of product formulation"

Camminatiello I., D' Ambra L., Università di Napoli Federico II, Italy
"CATANOVA method for sensory data"

Lengard V., Mage I., Kermit M., CAMO Software AS, Norway
"L-PLS regression for consumer preference analysis of flavoured waters"

January 25th, 2008

Session Sensometrics III

Meullenet J.-F., Université d'Arkansas, USA

"A comparison of statistically based methods for the optimization of product formulation"

Dejean S., Combes S., González I., Larzul C., Cauquil L., Baccini A., Université Paul Sabatier, France

"Analyse canonique régularisée pour l'étude de relations entre mesures sensorielles et instrumentales de la qualité de la viande de lapin"

Josse J., Husson F., Pagès J., Agrocampus Rennes, France

"Test de significativité sur le coefficient RV. Application à l'évaluation de la stabilité des nappes dans le recueil par napping"

Meyers M.,

"On the definition of equivalence in Betabinomial models "

Blancher G., Mattei B., Oelhafen N., Adam C., Givaudan Schweiz AG, Switzerland

"Comparaison du tri libre et du tri hiérarchique. Application à l'étude des sauces"

Cariou V., Diaz E., Faye P., Qannari E.M., Vigneau E., Ould Bah A., ENITIAA/INRA, Nantes, France

"Segmentation de consommateurs sur la base de variables qualitatives"

Ledauphin St., Pommeret D., Qannari E.-M., Université Rennes 2, France

"Modélisation de la dégradation sensorielle des aliments par des chaînes de Markov cachées "

Brockhoff Per B., Sommer N.-A., Technical University of Denmark, Denmark

"Accounting for scaling differences in sensory profile data"

Pineau N., Hartmann C., Lenfant F., Martin N., Nestlé Research Center, Switzerland

"Temporal dominance of sensation : some statistical improvements of TDS curves"

3.9. Academic visits

The members of the Institute visited other institutions and most of them presented seminars.

January 2008

Léopold Simar, Institut d'Economie Industrielle (IDEI/GREMAQ), (6 months in 2008-2009) Toulouse School of Economics, France

February 2008

Sébastien Van Bellegem, "*Some inverse problems in econometrics*",
Econometrics seminar, Center for Operations Research and Econometrics, Louvain-la-Neuve, Belgium

March 2008

Christian Hafner, "*Efficient estimation of a multivariate multiplicative volatility model*",
ECARES, Université Libre de Bruxelles, Belgium

Léopold Simar, Department of Statistical Science, University of Bologna, Italy

Ingrid Van Keilegom, Department of Mathematics, University of Hamburg, Germany

April 2008,

Léopold Simar, Lectures on "*Efficiency analysis: the econometric approach*",
Department of Statistical Science, University of Bologna, Italy

Catherine Legrand, "*A new aspect of the statistical validation of the Allard prognostic index in bladder cancer*",
Presentation at the General Assembly of the Quetelet Society, joint with L. Duchateau, P. Janssen, V. Ducrocq, R. Sylvester, UCL, Belgium

Ingrid Van Keilegom, Department of Mathematics, University of Bochum, Germany

May 2008

Sébastien Van Bellegem, "*Some inverse problems in econometrics*",
Seminario de Estadística, Departamento de Estadística, Pontificia Universidad Católica de Chile, Santiago, Chile

Ingrid Van Keilegom, School of Mathematics, Georgia Institute of Technology, Atlanta, USA

June 2008

Cédric Heuchenne, "*Semiparametric inference based on the error distribution*",
Université de Liège, Belgium

Léopold Simar, Department of Economics, Clemson University, Clemson SC, USA

Ingrid Van Keilegom, Department of Statistics, University of Paul Sabatier, Toulouse, France

July 2008

Bernadette Govaerts, Etablissement Français du Sang, Besançon, France

Bernadette Govaerts, Sanofi-Pasteur, Lyon, France

Bernadette Govaerts, Sanofi-Aventis, Paris, France

Léopold Simar, Summer School on "*Advanced Quantitative Methods: Applied Multivariate Data Analysis, Bootstrapping and Productivity/Efficiency Measurements*",
Facoltà di Ingegneria, Università di Pisa, Italy

Ingrid Van Keilegom, INSA, Rennes, France

August 2008

Ingrid Van Keilegom, "*Empirical likelihood for non-smooth criterion functions*",
Department of Mathematics, University of Oslo, Norway

October 2008

Ingrid Van Keilegom, "*Empirical likelihood for non-smooth criterion functions*",
Department of Statistics, University of Paul Sabatier, Toulouse, France

Christian Ritter, Presentation of Data, Applied Statistics Workshop

Christian Ritter, Combining Software Tools to Practice Statistics

Léopold Simar, Facoltà di Ingegneria, Università di Pisa, Italy

November 2008

Ingrid Van Keilegom, "*Semiparametric modeling and estimation of the dispersion function in regression*",
Research Institute CentER, Tilburg University, The Netherlands

Ingrid Van Keilegom, "*Semiparametric modeling and estimation of the dispersion function in regression*",
Institute of Statistics, Université catholique de Louvain, Belgium

Johan Segers, "*A gentle introduction to bivariate extreme-value copulas*",
Department of Mathematics, ETH Zürich, Switzerland

Christian Hafner, "*Efficient estimation of a dynamic semiparametric copula model*", Université de Liège,
Belgium

Léopold Simar, "*How to explain inefficiency in productivity analysis*", Statistics and Econometrics Seminar,
Toulouse School of Economics, France

December 2008

Léopold Simar, Fakultät für Wirtschaftswissenschaften, University of Karlsruhe, Germany

Léopold Simar, "*How to explain inefficiency in productivity analysis*",
Institute for Economic Policy Research (IWW), Karlsruhe University, Germany

Ingrid Van Keilegom, Department of Economics, University of Bonn, Germany

3.10. Participation to conferences and scientific meetings

The members of the Institute attended and/or participated to the following conferences:

January 2008

Bernadette Govaerts, invited as session chairman, International conference on design of industrial experiments, Antwerp, Belgium

Bernadette Govaerts, President of the organizing committee of AGROSTAT 2008: a conference of the French statistical society, Louvain-la-Neuve, Belgium

Léopold Simar, "*Performance of Universities: challenges and future developments*", AQUAMETH Workshop, University of Pisa, Italy

February 2008

Rainer von Sachs, "*Multiple change point detection by binary segmentation for time series*", invited, Mini-workshop on time series with sudden structural changes, Mathematisches Forschungsinstitut Oberwolfach, Germany

March 2008

Christian Hafner, "*Probability and Statistics*", invited, 8th German Open Conference at Aachen, Germany

April 2008

Bernadette Govaerts, 9th Chemometrics Symposium, participant, Gembloux, Belgium,

Sébastien Van Bellegem, "*A new likelihood ratio test of unit root*", Inference and Tests in Econometrics, A Tribute to Russell Davidson, GREQAM, invited, Marseille, France

May 2008

Thomas Meinguet, 2008 PhD Day of the Belgian Graduate School in Statistics, Brussels, Belgium

Johan Segers, Joint Meeting of the Society of Canada and the Société Française de Statistique, contributed, Ottawa, Canada

Ingrid Van Keilegom, "*Empirical likelihood for non-smooth criterion functions*", invited, joint Meeting of the Statistical Society of Canada and the Société Française de Statistique, Ottawa, Canada

Rainer von Sachs, "*Wavelets in statistics - a non-complete overview*", invited, Symposium: Wavelets in the presence and in the future; in honor of Ingrid Daubechies", University of Hasselt, Diepenbeek, Belgium

June 2008

Bernadette Govaerts, "*Implémentation d'un outil de suivi de qualité du processus de traitement du sang dans les sites de préparation des produits sanguins du service du sang de la Croix-Rouge de Belgique*", invited, Workshop on statistical tools for quality organized by MINITAB, Paris, France

Catherine Legrand, "*Single arm versus randomized studies*", Lung cancer in Croatia workshop, invited, Zagreb, Croatia

Catherine Legrand, "*Clinical trials Statistics for non-statisticians*" course, European Organisation for research and Treatment of Cancer, "*Sample size determination in phase III clinical trials*", invited, Belgium

Thomas Meinguet, 2008 Statistical modeling of extremes in data assimilation and filtering approaches, Strasbourg, France

Olga Reznikova, "*Efficient estimation of a semiparametric dynamic copula model*", 2nd International workshop on computational and financial econometrics (CFE'08), Institut d'Informatique, Université de Neuchâtel, Neuchâtel, Switzerland

Johan Segers, 2nd Brussels-Waseda seminar on time series and financial statistics, invited, Brussels, Belgium

Léopold Simar, "*Frontier estimation and extreme values theory*" and Panel discussion on "*The future of efficiency analysis in the real world*", member of the scientific committee and invited panelist, *North American productivity workshop V*, Stern School of Business, New York University, USA

Sébastien Van Bellegem, "*Hilbert scale regularization for functional estimation under moment conditions*", Cowles summer conference on operator methods and inverse problems in econometrics, invited, Yale University, New Haven, CT, USA

Ingrid Van Keilegom, member of the scientific committee and chairman of the First International Workshop on Functional and Operatorial Statistics, IWFOs 2008, University Paul Sabatier, Toulouse, France

Rainer von Sachs, International workshop on "*Recent Advances in Time Series Analysis*", co-organiser, Protaras, Cyprus

July 2008

Thomas Meinguet, 2008 International workshop on applied probability, Compiègne, France

Johan Segers, 2008 International workshop on applied probability, invited, Compiègne, France

Ingrid Van Keilegom, "*Empirical likelihood for non-smooth criterion functions*", 7th World Congress in Probability and Statistics, Singapore

August 2008

Christian Hafner, Econometric Society European Meeting, Milan, Italy

Sébastien Van Bellegem, "*Nonparametric Beta kernel estimator for long memory time series*", 62nd European Meeting of the Econometric Society, Milan, Italy

September 2008

Bernadette Govaerts, Non-clinical Statistics'08, member of the organizing committee, Leuven, Belgium

Stéphane Laurent, "*Statistique bayésienne objective*", Journée de probabilités de Grenoble, Institut Fourier, Grenoble, France

Olga Reznikova, "*Efficient estimation of a semiparametric dynamic copula model*", Workshop on Missing Information in Survival Analysis Beyond Right Censoring, Ghent University, Ghent, Belgium

Christian Ritter, Workshop on presentation of data, invited, ENBIS, Athens, Greece

Réjane Rousseau, "*Combination of independent component analysis and statistical modeling for the identification of metabonomic biomarkers in 1H-NMR spectroscopy*", Non-Clinical Statistics Conference, Leuven, Belgium

Johan Segers, IAP workshop "*Missing Information in Survival Data: Beyond Right Censoring*", contributed, Ghent, Belgium

Ingrid Van Keilegom, "*Semiparametric modeling and estimation of the dispersion function in regression*", invited, Celebration 20th Anniversary Master of Biostatistics, University of Hasselt, Belgium

Ingrid Van Keilegom, "*Missing Information in Survival Data Beyond Right Censoring*", member of the scientific committee and chairman, IAP Workshop, University of Ghent, Belgium

Rainer von Sachs, "*Missing Information in Survival Data Beyond Right Censoring*", IAP Workshop, University of Ghent, Belgium

Rainer von Sachs, International conference on "*Factor Structures for Panel and Multivariate Time Series Data*", co-authored invited talk, Maastricht, The Netherlands

October 2008

Bernadette Govaerts, "*Combination of independent component analysis and statistical modeling for the identification of metabonomic biomarkers in 1HH-NMR*", poster with Réjane Rousseau, Advances in Metabolic Profiling, Lisbon, Portugal

Bernadette Govaerts, "*Bubble : an automated 1H-NMR Signal processing method for metabonomics*", poster with Réjane Rousseau, Advances in Metabolic Profiling, Lisbon, Portugal

Stéphane Laurent, "*Frequentist validity of Bayesian inference for clinical trials*", 16th Annual Meeting of the Belgian Statistical Society, Namur, Belgium

Thomas Meinguet, 2008 Annual Meeting of the Belgian Statistical Society, Namur, Belgium

Olga Reznikova, "*Efficient estimation of a semiparametric dynamic copula model*", poster presentation, 16th Annual Meeting of the Belgian Statistical Society, Namur, Belgium

Ingrid Van Keilegom, member of the scientific committee of the 16th Annual Meeting of the Belgian Statistical Society, Namur, Belgium,

Rainer von Sachs, 16th Annual Meeting of the Belgian Statistical Society, Namur, Belgium

November 2008

Christian Hafner, "*Flexible modelling: smoothing and robustness*", KUL, Leuven, Belgium

Stéphane Laurent, "*Filtrations standard : retour sur les critères de Vershik et de confort*", Séminaire IRMA, Université de Strasbourg, Strasbourg, France

Catherine Legrand, "*Advanced topics in survival analysis*" Symposium, MSource, "*Parametric proportional hazards models with gamma frailty*" and "*Recurrent Events*", invited, Belgium

Johan Segers, "*Modelling multivariate dependence and extremes in finance*", invited, Oxford-Man Institute of Quantitative Finance, United Kingdom

Johan Segers, "*Flexible modelling: smoothing and robustness*", contributed, KUL, Leuven, Belgium

Sébastien Van Bellegem, "*Variables instrumentales en économétrie*", Journées METIS, invited, Université Paul Sabatier, Toulouse, France

Ingrid Van Keilegom, member of the scientific committee and chairman of the International Seminar on Nonparametric Inference, ISNI 2008, Vigo, Spain

Rainer von Sachs, "*Flexible modelling: smoothing and robustness*", co-authored poster presentation, International Workshop, KUL, Leuven, Belgium

3.11. Academic teaching activities

Participation to non-UCL teaching activities (e.g. guest professor, co-operation or Erasmus programs,...)

Bernadette Govaerts

Statistical and experimental design methods for industrial research, development and production
Jemeppe, Belgium
Client: Nanocyl

Statistical tools for quality control and improvement of blood treatment processes
CHU-ULG, Liège, Belgium

Cédric Heuchenne

Professor of Statistics at University of Liège, Ulg, Liège, Belgium

SUP Provisioning and Pricing Methodologies, jointly with Hübner, G., Lutgens, F., Muller, A.
Client, COFACE BELGIUM

Johan Segers

Extreme Value Analysis of Financial and Actuarial Data
KVBA-ARAB

Analyse van Extremen, jointly with John H. J. EINMAHL
Actuariel Instituut, Tilburg University, The Netherlands

Léopold Simar

Lectures on "Efficiency Analysis: the statistical approach", ERASMUS program
Department of Statistical Science, University of Bologna, Italy

Ingrid Van Keilegom

Survival data analysis course in the Master for Biostatistics
Guest professor at the University of Hasselt, Belgium

4. Teaching and educational activities

One major mission of the Institute of Statistics is educational and teaching services for the University. This activity concerns all levels of the statistical education and is described below. All professors and research assistants offer a strong investment in developing various pedagogical approaches that are adapted to the very large, heterogeneous public of the university. Moreover, because statistics is a very active field both in theory and applications, our courses are incessantly updated.

More specifically, our action in statistical education is organized at the following levels.

4.1. Bachelor level

The Institute of Statistics is responsible for advising all Faculties about the education in statistics and data analysis in their Bachelor programs. In most of these programs, the Institute makes proposals about the content and the level of the course, and our team is responsible for teaching and grading in these courses. In 2008, the Institute was involved in the Bachelor course of about 3.350 students, with a total of about 900 hours of teaching in statistics.

4.2. Minor in Statistics

Most students of UCL at the Bachelor level can choose an option in statistics called *Minor in statistics*. The goal of this program of 30 ECTS is to provide a basic knowledge in applied statistics that is useful in their specific field. This program is also a perfect preparation for the Master in statistics. Information about the minor in statistics can be found at: <http://www.uclouvain.be/prog-2008-minstat.html>

4.3. New Master programs in Statistics

In September 2007, the Institute launched its two new Master programs: one in General Statistics, and one in Biostatistics. The goal of the Master program is to attract quantitatively oriented students holding a Bachelor diploma to follow a two-year program in statistics. The Master in Statistics is declined in two directions: the "finalité approfondie" is research-oriented, whereas the "finalité spécialisée" is application-oriented.

Both the General Statistics and the Biostatistics programs present a combination of courses, seminars, and research projects within UCL or in collaboration with private organizations. They are tailored to attract students from any field with an interest in statistics. Within the General Statistics program, students can choose optional classes in various fields such as Biometrics, Data Management and Data Mining, Economics and Econometrics, Actuarial Sciences and Finance, Statistics in the Human Sciences, Science and Technology and Computer Science. In the Biostatistics program, one can choose from options in Bioinformatics, Epidemiology, and Biomedical Engineering.

The full program can be found at: www.uclouvain.be/prog-2008-stat2m.html and <http://www.uclouvain.be/prog-2008-bsta2m.html>.

4.4. Certificate

The Institute of Statistics also proposes a *Certificate in Statistics*, jointly organized with IUFC (*Institut Universitaire de Formation Continue*), which is mostly devoted to students having started a professional career and who want to improve or update their knowledge and technical skills in statistics. Six orientations are proposed in order to fit to the field and the level in statistics of the interested candidate: Elements of Statistics, Tools and Methods, Data Mining, Science and Technology, Biostatistics, and Advanced Methods. More details about the *Certificate* can be found at <http://www.uclouvain.be/formationcontinue-statistique.html>

4.5. Other activities

The Institute of Statistics has been active in various ways to further improve the level of teaching and the students' performance. In the following, a few examples of our activities can be found:

- The Institute participated in a project funded by the FDP (*Fonds de développement pédagogique*, FDP-VAE 2007/08) on the qualitative evaluation of the admission process via the *Valorisation des Acquis de l'Expérience* (VAE). The main objective of this project was to establish an expertise in the VAE admissions and was based on the following axes:
 1. qualitative evaluation of the admission procedures of the year 2007/08, both in terms of organisation and candidates
 2. Adjustment of the procedures and admission tools
 3. Accompanying pedagogical methods of candidates admitted by VAE
 4. Summary and distribution of the results to the community.

Three master programs were involved (MD, STAT, ESPO). The results were presented and discussed at the *Conseil de la Formation Continue* on February 20, 2009. The main promoter was F. de Viron (IUFC), with co-promoters E. Darras, B. Govaerts, O. Servais and C. Wouters.
- As in previous years, at the beginning of the academic year the Institute organized an introductory course in Computer Science and StatisticalS for students without much background in computational sciences and without having had access to computers and/or software. Mostly, these students are from developing countries. As far as software is concerned, basic features of Excel, SAS, R, and MatLab are discussed. We are grateful to Marc Feuerstein, one of our advanced Masters students, who accepted to teach this class and did so with great success.
- The Institute has signed a new agreement of collaboration with SAS-Institute Belgium for the years 2008-2012. By this convention, SAS-Institute supports:
 1. The training of UCL and Louvain Academy students in basic and advanced DATA Management with SAS software,
 2. The invitation of an external professor to teach annually a course in Data Mining within UCL programs
 3. The organisation of short courses on statistical methods based on SAS and SAS/JMP software opened to PhD students and external scientists (three or four per year).

4.6. Doctoral school

4.6.1. Presentation

In 2008, the Institute of Statistics continued to play the leading role in the administration and organization of the FNRS Graduate School in Statistics and Actuarial Sciences, an interuniversity school with partners coming from 9 departments of 5 French-speaking universities (ULB, UCL, ULg, FUNDP, FUSAG).

The "Bureau" of the School, composed of President (Rainer von Sachs, Institute of Statistics), Vice-President, Academic Secretary and Administrative Secretary (Sophie Malali, Institute of Statistics), meets regularly and is completed by two annual meetings of the Scientific Committee of the School, with representatives of both senior members and doctoral students of all partners of the School. Every academic member of the Institute of Statistics participates in the Graduate School, as well as every PhD student working at the Institute of Statistics on his or her thesis.

One of the objectives of the School is to offer PhD students a high-quality training to and by research, under the guidance of their supervisors and a follow-up committee.

In 2008, the Graduate School has organized an impressive number of short courses given by international visitors in various fields of methodological and applied statistics. These courses are at the heart of the training of our students to complete their education in statistics and its neighbouring fields, at an advanced level of graduate teaching ("3e cycle"). This offer is completed by regular research seminar series, young researchers days (among which the first "Journée doctorale" in May 2008 at the ULB, and the twice-yearly YRD at the Institute of Statistics), workshops and summer schools, some of which organized by our PhD students themselves. Students are encouraged to collaborate with international colleagues who regularly visit the Institute of Statistics, and to present their work in national and international meetings.

These activities give students the opportunity to work in a stimulating environment, open to discussion and exchanges, in order to develop inter-disciplinary work based on a common methodology in a variety of fields in statistics, biostatistics and actuarial sciences.

All information concerning the Graduate School activities can be found at the following address:
http://www.graduatecollegescience.be/edt_objectifs.php?lang=en&ecole=STATS.

Thanks to our administrative secretary maintaining this web page, relevant information is spread out quickly in order to reach all the members of the School.

4.6.2. Doctoral dissertations and PhD theses in progress

Doctoral Dissertations

Böhm, Hilmar

"Model Selection in principal component for multivariate time series"

Supervisor : Rainer von Sachs

Geenens, Gery

"Non- and semi-parametric models for conditional probabilities in two-way contingency tables"

Supervisor : Léopold Simar

Jullion, Astrid

"Adaptive Bayesian P-splines models for fitting time-activity curves and estimating associated clinical parameters in Positron Emission Tomography and Pharmacokinetic study"

Supervisor : Philippe Lambert

Le Bailly de Tillegem, Céline

"Statistical contribution to the virtual multicriteria optimisation of combinatorial molecules libraries and to the validation and application of QSAR models"

Supervisor : Bernadette Govaerts

Pitrebois, Sandra

"Bonus-malus scales and segmented tariffs"

Supervisors : Michel Denuit and Jean-François Walhin

Teodorescu, Bianca

"General conditional linear models with time-dependent coefficients under censoring and truncation"

Supervisor : Ingrid Van Keilegom

PhD theses in progress

Daskovska Alexandra

"Dynamical analysis of productivity index"

Supervisors: Léopold Simar, Sébastien Van Belleghem

El Mehdi Rachida

"L'analyse d'efficience des domaines de développement au Maroc"

Supervisor: Christian Hafner

François Nancy

"Statistical analysis of time intensity curves in sensory analysis"

Supervisor: Bernadette Govaerts, Philippe Lambert

Franco Bernard

"Development of statistical tools to test the equivalence between analytical measurement methods"

Supervisor: Bernadette Govaerts

Freyermuth Jean-Marc

"Tree-structured wavelet thresholding with applications in nonparametric curve estimation"

Supervisor: Rainer von Sachs

Gudendorf Gordon

"Extreme value analysis: modelling dependence between many variables"

Supervisor: Johan Segers

Hunt Julien

"Calcul stochastique en univers semi-markovien et applications financières"

Supervisor: Pierre Devolder

Jaeger Jonathan

"Functionnal estimation in system defined by differential equations using Bayesian smoothing methods"

Supervisors: Philippe Lambert, Catherine Legrand

Meinguet Thomas

"Extreme value theory for stochastic processes with continuous time and/or space coordinates"

Supervisor: Johan Segers

Motta Giovanni

"Factor modelling and nonparametric estimation for multivariate locally stationary time series"

Supervisor: Rainer von Sachs

Pigeon Mathieu

"Mixed regression models for insurance data with credibility updates and particular forms of censoring"

Supervisors: Michel Denuit

Reznikova Olga

"Adaptive modelling of the dependence in multivariate time series"

Supervisor: Christian Hafner

Rousseau Réjane

"Outils statistiques pour identification de biomarqueurs de toxicité métabonomiques"

Supervisors: Bernadette Govaerts, Michel Verleysen

Schwarz Maik

"Adaptive regularization in nonparametric instrumental regression"

Supervisors: Sébastien Van Belleghem (Université de Toulouse I), Rainer Dahlhaus (Université d'Heidelberg)

Soumali Mohammed Rida

"Asymptotic study of robustness properties of regression estimators in semiparametric regression models"

Supervisor: Ingrid Van Keilegom

Sujica Aleksandar

"Modeling and inference for dependent censoring mechanisms via location-scale regression models"

Supervisors: Ingrid Van Keilegom

Timmermans Catherine

"Statistical methods in the field of solar terrestrial relationships"

Supervisors: Rainer von Sachs, Véronique Delouille (ORB)

4.6.3. Doctoral seminars

February 1, 2008 - Young Researchers Day

Bianca Teodorescu, Statistics Institute, UCL, Belgium
"Goodness-of-fit tests for conditional models with time-dependent coefficients under censoring and truncation"

Rachida El Mehdi, Statistics Institute, UCL, Belgium
"Efficiency analysis application to financing of municipalities in Morocco"

Réjane Rousseau, Statistics Institute, UCL, Belgium
"Combination of independent component analysis and statistical modeling for the identification of metabonomic biomarkers in 1H-NMR spectroscopy"

Gery Geenens, Statistics Institute, UCL, Belgium
"Non- and semiparametric tests for conditional independence in two-way contingency tables"

Aurélie Miller, Institut Interdépartemental des Sciences Actuarielles, UCL, Belgium

Julien Hunt, Institut de statistique, UCL, Belgium
"Options and semi-Markov regime-switching"

May 26, 2008

Réjane Rousseau, Statistics Institute, UCL, Belgium
"Combination of independent component analysis and statistical modeling for the identification of metabonomic biomarkers in 1H-NMR spectroscopy",
Annual Doctoral Workshop of the Graduate School in statistics, ULB, Brussels, Belgium

September 26, 2008 - Young Researchers Day

Jean-Marc Freyermuth, Statistics Institute, UCL, Belgium
"Tree-structured wavelet estimation of spectra of replicated time series in a mixed effects model"

Catherine Timmermans, Statistics Institute, UCL, Belgium
"Using unbalanced Haar Wavelets for classification of time series"

Maik Schwarz, Statistics Institute, UCL, Belgium
"Estimating boundary points and noise level in deconvolution problems"

4.6.4. Short courses

"Statistical inference for curve estimation and classification" by Professor Lutz Dümbgen, University of Bern, Switzerland
January 21 and 22, 2008

"Goodness-of-fit testing in regression" by Professor Wenceslao González Manteiga, Universidad de Santiago de Compostela, Spain
February 12 and 15, 2008

"Local parametric estimation with applications in volatility estimation and risk management" by Professor Vladimir Spokoiny, Weierstrass Institute, Berlin, Germany
March 4-10, 2008

"L'interface R/C et la création de package R" by Professor Pierre Lafaye de Micheaux, Université de Grenoble, Grenoble, France
April 3-4, 2008

"Modeling with heavy tails and extremes" by Professor Sidney Resnick, Cornell University, Ithaca, NYC, USA
May 20 - 21, 2008

"Competing risks: a practical perspective" by Melania Pintilie, Ontario Cancer Institute / University Health Network, University of Toronto, Canada
Joint organisation between the FNRS Graduate School in Statistics and Actuarial Sciences, the FNRS Graduate School in Public Health, Health and Society, and The Biostatistics Section of the Belgian Statistical Society (SBS-BVS)
September 1-3, 2008

"Statistics in genetics" by Amy Anderson, Western Washington University, USA
Joint organisation between the FNRS Graduate School in Statistics and Actuarial Sciences, the FNRS Graduate School in Public Health, Health and Society, and the Society Adolphe Quetelet (Belgian region of the IBS).
September 16-18, 2008

"Nonlinear time series econometrics" by Professor Timo Teräsvirta, Aarhus University, Denmark
October 20-21, 2008

5. Support for the University and internal consulting activities

5.1. Presentation

One fundamental mission of the Institute of Statistics is to encourage the use of modern and “fit for purpose” statistical tools in the University’s various areas of applied research and stimulate interactions between researchers. In this context, the SMCS “Statistical methodology and computing service” has been created in 2007 (<http://www.stat.ucl.ac.be/SMCS/>). It offers an integrated, recognized and visible structure to three services offered by the Institute of Statistics in the past.

The consulting service gives advice and collaboration to researchers and members of the university community about the appropriate methodology and suitable statistical package for their specific problems.

The statistical software server allows the members of UCL to manage their data bases and perform their statistical computations and simulations on an efficient and well-maintained computing tool.

Training courses in statistical software are also proposed to inform researchers of existing statistical software and train them to their practical use.

In addition, in the context of its Master programs, the Institute of Statistics is also organising a Statistical Workshop. Twice every month, speakers are invited to present applications of statistical methods in various fields. Some presentations take also the form of tutorials related to subjects of interest for people working in applied statistics. These seminars welcome students in statistics but also other participants from university, business or industry.

5.2. SMCS activities

In 2008, the consulting service has offered numerous advices to members of the university community. Among them, more than 181 have been identified as “conseils” (advices of more than 2 hours of consulting work) and are listed on the consulting service web page: <http://www.stat.ucl.ac.be/SMCS>. Clients come from various faculties and departments of the university and application fields are very stimulating for the statistical consultants. The results of some substantial consulting activities are summarized in formal consulting reports. More details about these consulting activities may be found in the SMCS annual report.

5.3. SMCS courses

In 2008, the consulting service also organised 15 courses on the use of statistical software for the researchers and administrative members of UCL and Louvain Academy.

January, 2008

Cours d'initiation SAS Enterprise Guide by A. Guillet, C. Le Bailly de Tillegem

Cours d'initiation au langage R by A. Guillet, C. Le Bailly de Tillegem

Cours d'initiation au logiciel SPSS by B. Masuy

February and March, 2008

Formation SAS Advanced : SQL, Programming III, Macros by SAS Institute trainers

April 3-4, 2008

Short course sur l'interface R/C et la création de package R by P. Lafaye de Micheaux and J.-M. Freyermuth

May 8, 2008

Cours d'initiation au logiciel SAS/JMP by B. Jones, SAS Institute

June 2-6, 2008

Cours d'initiation SAS/BASE by B. Masuy, J.-M. Zélis

June 16-19, 2008

Pratique de la statistique avec SAS/Enterprise by A. Guillet

June 23-26, 2008

Cours d'initiation au langage R by A. Guillet

September 22 -25, 2008

Cours d'initiation au logiciel SPSS by B. Masuy

October 10 -14, 2008

Cours d'initiation aux logiciels Weft et NVivo pour une pratique de l'analyse by S. Vause, D. Fripiat and B. Masuy

November 13 -27, 2008

Initiation à la pratique de la statistique avec SPSS pour l'IEPR by N. Lefèvre, A. Guillet and R. Rousseau

November 25, 2008

Midis du SMCS : Démonstration de Sweave by St. Laurent

November 27, 2008

Initiation à la pratique de la statistique avec SPSS pour la FOPA by N. Lefèvre

December 15 -19, 2008

Cours d'initiation SAS/BASE by B. Masuy, J.-M. Zélis

5.4. Applied statistics workshops

February 15, 2008

Vincent Ducrocq, INRA, France

"Use of linear and nonlinear mixed models as tools to create genetic progress in dairy cattle populations"

Xavier Draye, UCL, Belgium

"Use of detailed genotypic information (molecular markers) in the dissection of genetic variance and the improvement of crops root architecture"

February 29, 2008

Winfried Theis, Unilever, The Netherlands

"Sparse sSampling for bioavailability studies in food research"

Michel Crucifix, Institut d'Astronomie et de Géophysique, UCL, Belgium

"Did antique civilisations save us from an ice age?"

April 16, 2008, - Joint Seminar CORE/STAT

Hans Manner, University of Maastricht, The Netherlands

"Testing for asset market linkages: a new approach based on time-varying copulas" joint work with Bertrand Candelon

Johanna Neslehova, ETH Zürich, Switzerland - (joint seminar with CORE)

"Archimedean copulas in high dimensions"

April 18, 2008

Godelieve Masuy-Stroobant, Institut de démographie, UCL, Belgium

"Qu'est-ce que GGP ? Generations & Gender Panel" joint work with UCL GGPS team (Sandrine de Waleffe, Thierry Eggerickx, Jean-Paul Sanderson and Christophe Vandeschrick)

Morgane David, Solidpartners, Belgium

"CCAF: the French reference survey about French food behaviours and consumptions" joint work with Bérengère Daxhelet

May 09, 2008

Bradley Jones, SAS, USA

"A new paradigm for the generation of designed experiments"

Olivier Roustant, Ecole des Mines de Saint-Etienne, France

"Applications of kriging modeling to the analysis of simulator codes. A 27D case study from nuclear engineering. Optimization problems. "

October 03, 2008

Christian Ritter, Institut de statistique, UCL, Belgium

"Presentation of quantitative information"

Christian Ritter, Institut de statistique, UCL, Belgium

"Combining software tools to practice Statistics"

October 24, 2008, - Après-midi "Statistique dans les Assurances"

Christophe Gillain, AXA Belgium

"Détermination des hypothèses biométriques et comportementales en vue de la projection de contrat d'assurance-vie"

Marc Duquesne, AXA Belgium

"Applications de la méthode Chain-Ladder en assurances IARD"

November 07, 2008

Karim Chine, Cambridge, UK

"Towards a federative, collaborative, user-centric and grid-enabled computational open platform"

Karim Chine, Cambridge, UK

"Distributed and web oriented data analysis applications in the cloud."

November 28, 2008

Viviane Planchon, CRA, Gembloux, Belgium

"Application des statistiques pour l'organisation et l'interprétation des essais inter-laboratoires (EIL)"

Angélique Baclin, GSK, Belgium

"« Rien ne sert de courir, il faut partir à point...»"

6. Appendix

6.1. Abstracts of the discussion papers

DP0801 - Nonparametric test for conditional independence in two-way contingency tables

Geenens, G. and L. Simar

Consider a two-way contingency table, built on two categorical variables R and S . Testing for independence between these two categorical variables is a well-known problem : the classical chi-square and likelihood ratio tests are used. Suppose now that for each individual of the sample we can also observe a set of p characteristics describing him. These extra covariates can provide more information on the dependence structure of the table, so that it could be interesting to test the independence between R and S conditionally to them. In this paper, we propose two nonparametric tests which generalize the chi-square test and the likelihood ratio test ideas to this case. The procedure is based on a kernel estimator of the conditional probabilities. The asymptotic law of the proposed test statistics is derived. The finite sample behaviour of the procedure is analysed through some Monte-Carlo experiments and the approach is illustrated with a real data example.

DP0802 - Explicit formula for asymptotic higher moments of the Nadaraya-Watson estimator

Geenens, G.

The Nadaraya-Watson estimator is certainly the most popular nonparametric regression estimator. The asymptotic bias and variance of this estimator, say $\hat{m}(x)$, are well known. Nevertheless, its higher moments are rarely mentioned in the literature. In this paper, explicit formulas for asymptotic higher moments, such as $\mathbb{E}((\hat{m}(x) - m(x))^r)$ or $\mathbb{E}((\hat{m}(x) - \mathbb{E}(\hat{m}(x)))^r)$, for any positive integer, are derived. Some examples illustrate the application of these new results, and a further application of them in the context of search of confidence intervals for nonparametric regression is suggested.

DP0803 - Nonparametric location-scale models for successive survival times under dependent censoring

Van Keilegom, I., de Uña-Alvarez J. and L. Meira-Machado

Let (T_1, T_2) be gap times corresponding to two consecutive events, which are observed subject to random right-censoring, and suppose the vector (T_1, T_2) satisfies the nonparametric location-scale regression model $T_2 = m(T_1) + \sigma(T_1)\varepsilon$, where the functions m and σ are 'smooth', and ε is independent of T_1 . The aim of this paper is twofold. First, we propose a nonparametric estimator of the distribution of the error variable under this model. This problem differs from others considered in the recent related literature in that the censoring acts not only on the response but also on the covariate, having no obvious solution. On the basis of the idea of transfer of tail information (Van Keilegom and Akritas, 1999), we then use the proposed estimator of the error distribution to introduce nonparametric estimators for important targets such as: (a) the conditional distribution of T_2 given T_1 ; (b) the bivariate distribution of the gap times; and (c) the so-called transition probabilities. The asymptotic properties of these estimators are obtained. We also illustrate through simulations, that the new estimators based on the location-scale model may behave much better than existing ones.

DP0804 - Goodness-of-fit tests for parametric regression with selection biased data

Ojeda Cabrera, J. L. and I. Van Keilegom

Consider the nonparametric location-scale regression model $Y = m(X) + \sigma(X)\varepsilon$, where the error ε is independent of the covariate X , and m and σ are smooth but unknown functions. The pair $(X; Y)$ is allowed to be subject to selection bias. We construct tests for the hypothesis that $m(\cdot)$ belongs to some parametric family of regression functions. The proposed tests compare the nonparametric maximum likelihood estimator (NPMLE) based on the residuals obtained under the assumed parametric model, with the NPMLE based on the residuals obtained without using the parametric model assumption. The asymptotic distribution of the test statistics is obtained. A bootstrap

procedure is proposed to approximate the critical values of the tests. Finally, the finite sample performance of the proposed tests is studied in a simulation study, and the developed tests are applied on environmental data.

DP0805 - Structural shrinkage of nonparametric spectral estimators for multivariate time series

Böhm H. and R. von Sachs

In this paper we investigate the performance of periodogram based estimators of the spectral density matrix of possibly high-dimensional time series. We suggest and study shrinkage as a remedy against numerical instabilities due to deteriorating condition numbers of (kernel) smoothed periodogram matrices. Moreover, shrinking the empirical eigenvalues in the frequency domain towards one another also improves at the same time the Mean Squared Error (MSE) of these widely used nonparametric spectral estimators. Compared to some existing time domain approaches, restricted to i.i.d. data, in the frequency domain it is necessary to take the size of the smoothing span as "effective or local sample size" into account. While Böhm & von Sachs (2007) proposes a multiple of the identity matrix as optimal shrinkage target in the absence of knowledge about the multidimensional structure of the data, here we consider "structural" shrinkage. We assume that the spectral structure of the data is induced by underlying factors. However, in contrast to actual factor modelling suffering from the need to choose the number of factors, we suggest a model-free approach. Our final estimator is the asymptotically MSE-optimal linear combination of the smoothed periodogram and the parametric estimator based on an underfitting (and hence deliberately misspecified) factor model. We complete our theoretical considerations by some extensive simulation studies. In the situation of data generated from a higher-order factor model, we compare all four types of involved estimators (including the one of Böhm & von Sachs (2007)).

DP0806 - Frontier estimation and extreme values theory

Daouia, A., Florens, J.-P. and L. Simar

The production/econometric frontier is the locus of the optimal combinations of inputs and outputs. From a statistical point of view, it can be viewed as the upper surface of the support of a random vector under shape constraints. In this paper we investigate the problem of nonparametric monotone frontier estimation from an extreme-values theory perspective. This allows to revisit the asymptotic theory of the popular FDH estimator in a general setup, to derive new and asymptotically Gaussian estimators and to provide useful asymptotic confidence bands for the monotone boundary function. The study of the asymptotic properties of the resulting frontier estimators is carried out by relating them to an original dimensionless random sample and then applying standard extreme-values theory. The finite sample behaviour of the suggested estimators is explored through Monte-Carlo experiments. We also apply our approach to a real data set.

DP0807 - Single-index modelling of conditional probabilities in two-way contingency tables

Geenens, G. and L. Simar

When analyzing a contingency table, it is often worth relating the probabilities that a given individual falls into the different cells to a set of predictors. These conditional probabilities are usually estimated using appropriate regression techniques. In particular, in this paper, a semiparametric model is developed. Essentially, it is only assumed that the effect of the vector of covariates on the probabilities can entirely be captured by a single index, which is a linear combination of the initial covariates. The estimation is then twofold : the coefficients of the linear combination and the functions linking this index to the related conditional probabilities have to be estimated. Inspired by the estimation procedures already proposed in the literature for Single-Index regression models, four estimators of the index coefficients are proposed and compared, from a theoretical point-of-view, but also practically, with the aid of simulations. Estimation of the link functions is also addressed.

DP0808 - Tails of multivariate Archimedean copulas

Charpentier, A. and J. Segers

A complete and user-friendly directory of tails of Archimedean copulas is presented which can be used in the selection and construction of appropriate models with desired properties. The results are synthesized in the form of a decision tree: Given the values of some readily computable characteristics of the Archimedean generator, the upper and lower tails of the copula are classified into one of three classes each, one corresponding to asymptotic dependence and the other two to asymptotic independence. For a long list of single-parameter families, the relevant tail quantities are computed so that the corresponding classes in the decision tree can easily be determined. In addition, new models with tailor-made upper and lower tails can be constructed via a number of transformation methods. The frequently occurring category of asymptotic independence turns out to conceal a surprisingly rich variety of tail dependence structures.

DP0809 - Projection estimators of Pickands dependence functions

Fils-Villetard, A., Guillou, A. and J. Segers

Shape constraints on a functional parameter can often be formulated in terms of a closed and convex parameter set embedded in a real Hilbert space. This is the case, for instance, if the curve of interest is a Pickands dependence function of an extreme-value copula. The topic of this paper is the estimator that results when an arbitrary initial estimator possibly falling outside the parameter set is projected onto this parameter set. As direct computation of the projection is infeasible, the full parameter set needs to be replaced by a sieve of finite-dimensional subsets. Asymptotic properties of the initial estimator sequence in the Hilbert space topology transfer easily to those of the projected sequence and its finite-dimensional approximations.

DP0810 - Pre-study analytical method validation: comparison of four alternative approaches based on quality level estimation and tolerance intervals

Govaerts, B., Dewé, W., Maumy, M. and B. Boulanger

In industry and in laboratories, it is crucial to continuously control the validity of the analytical methods used to follow the products' quality characteristics. Validity must be assessed at two levels. The "pre-study" validation aims at demonstrating before use that the method will be able to achieve its objectives. The "in-study" validation is intended to verify, by inserting QC samples in routine runs, that the method remains valid over time. At these two levels, the analytical method will be claimed valid if it is possible to prove that a sufficient proportion of analytical results is expected to lie within given acceptance limits $[-\lambda, \lambda]$ around the nominal value.

This paper presents and compares four approaches to checking the validity of a measurement method at the pre-study level. They can be classified into two categories. In the first, a lower confidence bound for the estimated probability π of a result lying within the acceptance limits is computed and compared to a given acceptance level. Maximum likelihood and delta methods are used to estimate the quality level π and the corresponding estimator variance. Two approaches are then proposed to derive the confidence bound: the asymptotic maximum-likelihood approach and a method proposed by Mee [1]. The second category of approaches checks whether a tolerance interval for hypothetical future measurements lies within the predefined acceptance limits $[-\lambda, \lambda]$. β -expectation and β -content tolerance intervals are investigated and compared in this context.

These four approaches are illustrated on a bioanalytical HPLC-UV analytical process and compared through simulations.

DP0811 - Use of ICA on HPCL-DAD data and high-order statistics to automatically achieve peak picking

Debrus, B., Lebrun, P., Ceccato, A., Caliaro, G., Govaerts, B., Olsen, B., Rozet, E., Boulanger, B. and P. Hubert

One of the major difficulties within the context of the fully automated development of chromatographic methods consists in the automated detection of the peaks coming from complex matrices such as multicomponent pharmaceutical formulations or stability studies of these formulations. The same problem can also occur with plant materials or biological matrices. This step is thus critical and time-consuming, especially when Designs of Experiments (DOE) are used to generate chromatograms. The use of DOE leads to maximize the changes of the analytical conditions in order to cleverly explore an experimental domain. Unfortunately, this generally provides very different and "uncontrolled" chromatograms which can be hardly interpretable, complicating picking and peak tracking. In this context, numerical signal processing methods such as Independent Components Analysis (ICA) was investigated to solve this problem. The ICA principle assumes that the observed signal is the resultant of several phenomena (known as sources) and that all these sources are statistically independent. ICA is able to

estimate sources which most often seem judicious to represent the constitutive components of a chromatogram. In the present study, ICA was applied to HPLC-UV-DAD chromatograms and we showed that ICA allows differentiating noises and artifacts components from those of interest, by applying clustering methods based on high-order statistics computed on these components. Furthermore, on the basis of the described numerical strategy, it was also possible to rebuild a cleaned chromatogram easily legible. This represents a very significant advance towards our final objective, the fully automated development of liquid chromatography (LC) method.

DP0812 - Empirical likelihood for non-smooth criterion functions

Molanez Lopez, E.-M., Van Keilegom, I. and N. Veraverbeke

Suppose that X_1, \dots, X_n is a sequence of independent random vectors, identically distributed as a d -dimensional random vector X . Let $\mu \in \mathbb{R}^p$ be a parameter of interest and $v \in \mathbb{R}^q$ be some nuisance parameter. The unknown, true parameters (μ_0, v_0) are uniquely determined by the system of equations $E\{g(X, \mu_0, v_0)\} = 0$, where $g = (g_1, \dots, g_{p+q})$ is a vector of $p + q$ functions. In this paper we develop an empirical likelihood method to do inference for the parameter μ_0 . The results in this paper are valid under very mild conditions on the vector of criterion functions g . In particular, we do not require that g_1, \dots, g_{p+q} are smooth in μ or v . This offers the advantage that the criterion function may involve indicators, which are encountered when considering e.g. differences of quantiles, copulas, ROC curves, to mention just a few examples. We prove the asymptotic limit of the empirical log-likelihood ratio, and carry out a small simulation study to test the performance of the proposed empirical likelihood method for small samples.

DP0813 - On the analysis of Bayesian semiparametric IRT-type models

San Martin, E., Jara, A., Rolin, J.-M. and M. Mouchart

Motivated by the characteristics of two educational datasets, we study the Bayesian identification and consistency of semiparametric IRT-type models, where the uncertainty on the abilities' distribution is modeled using a prior distribution on the space of probability measures. We establish sufficient conditions for the identification and consistency in the Bernoulli and Poisson versions of the Rasch model. For unbounded count (resp. binary) responses the parameters are identified when a finite (resp. infinite) number of probes are available and they are consistently estimated when the number of subjects tends (resp. subjects and probes tend) to infinite. The validity of the findings as potential necessary conditions are evaluated using simulated data.

DP0814 - Improving upon the marginal empirical distribution functions when the copula is known

Segers, J., Van den Akker, R. and B.J.M. Werker

At the heart of the copula methodology in statistics is the idea of separating marginal distributions from the dependence structure. However, as shown in this paper, this separation is not to be taken for granted: in the model where the copula is known and the marginal distributions are completely unknown, the empirical distribution functions are semi-parametrically efficient if and only if the copula is the independence copula. Incorporating the knowledge of the copula into a nonparametric likelihood yields an estimation procedure which by simulations is shown to outperform the empirical distribution functions, the amount of improvement depending on the copula. Although the known-copula model is arguably artificial, it provides an instructive stepping stone to the more general model of a parametrically specified copula and arbitrary margins.

DP0815 - Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution

Einmahl, H.J. and J. Segers

Consider a random sample from a bivariate distribution function F in the max-domain of attraction of an extreme value distribution function G . This G is characterized by two extreme value indices and a spectral measure, the latter determining the tail dependence structure of F . A major issue in multivariate extreme value theory is the estimation of the spectral measure Φ_p with respect to the L_p norm. For every $p \in [1, \infty]$, a nonparametric maximum empirical likelihood estimator is proposed for Φ_p . The main novelty is that these estimators are guaranteed to satisfy the moment constraints by which spectral measures are characterized. Asymptotic normality of the estimators is proved under conditions that allow for tail independence. Moreover, the conditions are easily verifiable as we demonstrate through a number of theoretical examples. A simulation study shows substantially improved performance of the new estimators. Two case studies illustrate how to implement the methods in practice.

DP0816 - Estimating the error distribution in nonparametric multiple regression with applications to model testing

Neumeyer, N. and I. Van Keilegom

In this paper we consider the estimation of the error distribution in a heteroscedastic nonparametric regression model with multivariate covariates. As estimator we consider the empirical distribution function of residuals, which are obtained from multivariate local polynomial fits of the regression and variance functions, respectively. Weak convergence of the empirical residual process to a Gaussian process is proved. We also consider various applications for testing model assumptions in nonparametric multiple regression. The obtained model tests are able to detect local alternatives that converge to zero at $n^{-1/2}$ -rate, independent of the covariate dimension. We consider in detail a test for additivity of the regression function.

DP0817 - Estimation of receptor occupancy using varying coefficients models

Jullion, A., Lambert, Ph. and F. Vandenhende

In many applications of linear regression models, the regression coefficients are not regarded as fixed but as varying with another covariate named the effect modifier. A useful extension of the linear regression models are then varying coefficient models. To link the regression coefficient with the effect modifier, several methods may be considered. Here, we propose to use Bayesian P-splines to relate in a smoothed way the regression coefficient with the effect modifier. We show that this method enables a large level of flexibility: if necessary, adaptive penalties can be introduced in the model (Jullion and Lambert 2007) and linear constraints on the relation between the regression coefficient and the effect modifier may easily be added.

We provide an illustration of the proposed method in a PET study where we want to estimate the relation between the Receptor Occupancy and the drug concentration in the plasma. As we work in a Bayesian setting, credibility sets are easily obtained for receptor occupancy, which take into account the uncertainty appearing at all the different estimation steps.

DP0818 - On the validity of the bootstrap in nonparametric functional regression

Ferraty, F., Van Keilegom, I. and P. Vieu

We consider the functional nonparametric regression model $Y = r(\chi) + \varepsilon$, where the response Y is univariate, χ is a functional covariate (i.e. valued in some infinite-dimensional space), and the error ε satisfies $E(\varepsilon | \chi) = 0$. For this model, the point-wise asymptotic normality of a kernel estimator $\hat{r}(\cdot)$ of $r(\cdot)$ has been proved in the literature. In order to use this result for building point-wise confidence intervals for $r(\cdot)$, the asymptotic variance and bias of $\hat{r}(\cdot)$ need to be estimated. However, the functional covariate setting makes this task very hard. To circumvent the estimation of these quantities, we propose to use a bootstrap procedure to approximate the distribution of $\hat{r}(\cdot) - r(\cdot)$. Both a naive and a wild bootstrap procedure are studied, and their asymptotic validity is proved. The obtained consistency results are discussed from a practical point of view via a simulation study. Finally, the wild bootstrap procedure is applied to a food industry quality problem in order to compute point-wise confidence intervals.

DP0819 - Extensions of Bayesian P-splines models for fitting PK curves

Jullion, A. and Ph. Lambert

In clinical experiments, the evolution of a product concentration in tissue over time is often under study. Different products and tissues may be considered. For instance, one could analyse the evolution of drug concentration in plasma over time. One could also observe the evolution of radioactivity uptakes in different regions of the brain during a PET scan (Positron Emission Tomography). Such evolutions are named, generically, pharmacokinetic curves (PK curves).

Some clinical measures of interest are derived from PK curves. For instance, when analysing the evolution of drug concentration in the plasma, PK parameters such as the area under the curve (*AUC*), the maximal concentration (C_{max}) and the time at which it occurs (t_{max}) are usually calculated. In a PET study, one could measure receptor occupancy (RO) in some regions of the brain. These clinical measures could be badly estimated if the observed PK curves are noisy. The objective of this paper is to provide some extensions of the Bayesian P-splines model for fitting this type of curves. Two extensions are provided. The first one introduces adaptive penalties in the Bayesian P-splines model. Jullion and Lambert (2007) already proposed an adaptive Bayesian P-splines model. Here, two alternative specifications relying on multivariate normal (Nelsen 1999; Sklar 1959) and Archimedean copulas (Genest and MacKay 1986) are proposed. Simulations show that the best performer to fit PK-like curves is the model of Jullion and Lambert (2007).

In the second extension, two solutions are proposed to deal with heterogeneous variance. In clinical studies, the variance is often not constant but varies either with the mean response or with time. The first solution relies on the Yeo-Johnson transformation while the second expresses the variance as a parametric function of the mean response. Simulations have been performed on PK-like curves, to compare these two extensions with the Bayesian P-splines model applied on the original and on the log-transformed data, in three different scenarios. In the first one, the conditional variance increases with the conditional mean. In the second one, the variance is homogeneous while in the third one, it increases with the independent variable, x .

It turns out that, in the first scenario, the best performer is the model where the variance is a parametric function of the conditional mean followed by the Bayesian P-splines model applied on the logarithm of the data. The simulation performed from scenario 2 shows that, when the variance of the data is constant, the Bayesian P-splines model applied on the original data is the best performer. Finally, in the third scenario, when the variance increases with x , the best choice is the Yeo-Johnson transformation model.

These two extensions can be combined to obtain a satisfactory fit of PK curves.

DP0820 - Stochastic FDH/DEA estimators for frontier analysis

Simar, L. and V. Zelenyuk

In this paper we extend the work of Simar (2007) introducing noise in nonparametric frontier models. We develop an approach that synthesizes the best features of the two main methods in the estimation of production efficiency. Specifically, our approach first allows for statistical noise, similar to Stochastic Frontier Analysis (even in a more flexible way), and second, it allows modelling multiple-inputs-multiple-outputs technologies without imposing parametric assumptions on production relationship, similar to what is done in non-parametric methods (DEA, FDH, etc. . .). The methodology is based on the theory of local maximum likelihood estimation and extends recent works of Park, Kumbhakar, Simar and Tsionas (2007) and Park, Simar and Zelenyuk (2006).

Our method is suitable for modelling and estimation of the marginal effects onto inefficiency level jointly with estimation of marginal effects of input. The approach is robust to heteroskedastic cases and to various (unknown) distributions of statistical noise and inefficiency, despite assuming simple anchorage models. The method also improves DEA/FDH estimators, by allowing them to be quite robust to statistical noise and especially to outliers, which were the main problems of the original DEA/FDH. The procedure shows great performance for various simulated cases and is also illustrated for some real data sets.

DP0821 - Forecasting the Malmquist productivity index

Daskovska, A., Simar, L. and S. Van Belleghem

The Malmquist Productivity Index (MPI) suggests a convenient way of measuring the productivity change of a given unit between two consequent time periods. Until now, only static approach for analyzing the MPI was available in the literature. However, this approach hides a potentially valuable information given by the evolution of productivity over time. In this paper, we introduce a dynamic procedure for forecasting the MPI. We compare several approaches and give credit to an approach based on the assumption of circularity. Because the MPI is not circular, we present a new decomposition of the MPI, in which the time-varying indices are circular. Based on that decomposition, a new working dynamic forecasting procedure is proposed and illustrated. To construct prediction intervals of the MPI, we extend the bootstrap method in order to take into account potential serial correlation in the data. We illustrate all the

new techniques described above by forecasting the productivity index of 17 OCDE countries, constructed from their GDP, labor and capital stock.

DP0822 - Asymptotic properties of the Bernstein density copula for dependent data

Bouezmarni, T., Rombouts, J. and A. Taamouti

Copulas are extensively used for dependence modeling. In many cases the data does not reveal how the dependence can be modeled using a particular parametric copula. Nonparametric copulas do not share this problem since they are entirely data based. This paper proposes nonparametric estimation of the density copula for α -mixing data using Bernstein polynomials. We study the asymptotic properties of the Bernstein density copula, i.e., we provide the exact asymptotic bias and variance, we establish the uniform strong consistency and the asymptotic normality.

DP0823 - Goodness-of-fit test for generalized conditional linear models under left truncation and right censoring

Teodorescu, B. and I. Van Keilegom

Consider a semiparametric time-varying coefficients regression model of the following form: $\phi(S(z|X)) = \beta(z)^t X$, where ϕ is a known link function, $S(\cdot|X)$ is the survival function of a response Y given a covariate X , $X = (1, X, X^2, \dots, X^p)$ and $\beta(z) = (\beta_0(z), \dots, \beta_p(z))^t$ is the unknown vector of regression coefficients. This model reduces for special choices of ϕ to e.g. the additive hazards model or the Cox proportional hazards model with time dependent coefficients. The response is subject to left truncation and right censoring. An omnibus goodness-of-fit test is developed to test whether the model fits the data. A bootstrap version, to approximate the critical values of the test, is proposed and proved to work from a practical point of view as well. The test is also applied to real data.

DP0824 - Second-order refined peaks-over-threshold modelling for heavy-tailed distributions

Beirlant, J., Joossens, E. and J. Segers

Modelling excesses over a high threshold using the Pareto or generalized Pareto distribution (PD/GPD) is the most popular approach in extreme value statistics. This method typically requires high thresholds in order for the (G)PD to fit well and in such a case applies only to a small upper fraction of the data. The extension of the (G)PD proposed in this paper is able to describe the excess distribution for lower thresholds in case of heavy tailed distributions. This yields a statistical model that can be fitted to a larger portion of the data. Moreover, estimates of tail parameters display stability for a larger range of thresholds. Our findings are supported by asymptotic results, simulations and a case study.

DP0825 - Nonparametric regression with filtered data

Linton, O., Mammen, E., Nielsen, J.P. and I. Van Keilegom

We present a general principle for estimating a regression function nonparametrically allowing for a wide variety of data filtering, e.g., repeated left truncation and right censoring. Both the mean and the median regression case are considered. The method works by first estimating the conditional hazard function or conditional survivor function and then integrating. We also investigate improved methods that take account of model structure such as independent errors, and show that such methods can improve performance when the model structure is true. We establish the pointwise asymptotic normality of our estimators.

DP0826 - Potential outcomes, counterfactuals, and structural modelling. Causal approaches in the social sciences

Russo, F., Wunsch, G. and M. Mouchart

This paper examines the potential outcome model developed by Rubin and its counterfactual underpinnings as developed by Lewis. Though a major contribution of Rubin's potential outcome model has been to stress the importance of the design stage, we recall the main methodological and epistemological flaws of this approach. We argue that the study of causes and effects does not necessarily require counterfactuals, once a structural modelling framework, as the one developed here, is adopted. Our approach emphasises and spells out the role of background knowledge, marginal-conditional decomposition, and of stability for providing a causal explanation of a given phenomenon.

DP0827 - Semiparametric modeling and estimation of the dispersion function in regression

Van Keilegom, I. and L. WANG

Modeling heteroscedasticity in semiparametric regression can improve the efficiency of the estimator of the parametric component in the regression function, and is important for inference problems such as plug-in bandwidth selection and the construction of confidence intervals. However, the literature on exploring heteroscedasticity in a semiparametric setting is rather limited. Existing work is mostly restricted to the partially linear mean regression model with a fully nonparametric variance structure. The nonparametric modeling of heteroscedasticity is hampered by the *curse of dimensionality* in practice. Moreover, the approaches used in existing work need to assume smooth objective functions, therefore exclude the emerging important class of semiparametric quantile regression models.

To overcome these drawbacks, we propose a general semiparametric location-dispersion regression framework, which enriches the currently available semiparametric regression models. With our general framework, we do not need to impose a special semiparametric form for the location or dispersion function. Rather, we provide easy to check sufficient conditions such that the asymptotic normality theory we establish is valid for many commonly used semiparametric structures, for instance, the partially linear structure and single-index structure. Our theory permits non-smooth location or dispersion functions, thus allows for semiparametric quantile heteroscedastic regression. We demonstrate the proposed method via simulations and the analysis of a real data set.

DP0828 - Optimal bandwidth selection for conditional efficiency measures: a data-driven approach

Badin, L., Daraio, C. and L. Simar

In productivity analysis an important issue is to detect how external (environmental) factors, exogenous to the production process and not under the control of the producer, might influence the production process and the resulting efficiency of the firms.

Most of the traditional approaches proposed in the literature have serious drawbacks. An alternative approach is to describe the production process as being conditioned by a given value of the environmental variables (Cazals, Florens and Simar, 2002, Daraio and Simar, 2005). This defines conditional efficiency measures where the production set in the input \times output space may depend on the value of the external variables.

The statistical properties of nonparametric estimators of these conditional measures are now established (Jeong, Park and Simar, 2008). These involve the estimation of a non standard conditional distribution function which requires the specification of a smoothing parameter (a bandwidth). So far, only the asymptotic optimal order of this bandwidth has been established. This is of little interest for the practitioner. In this paper we fill this gap and we propose a data-driven technique for selecting this parameter in practice. The approach, based on a Least Squares Cross Validation procedure (LSCV), provides an optimal bandwidth that minimizes an appropriate (weighted) integrated Squared Error (ISE). The method is carefully described and exemplified with some simulated data with univariate and multivariate environmental factors. An application on real data (performances of Mutual Funds) illustrates how this new optimal method of bandwidth selection works in practice.

DP0829 - Fitting dynamic factor models to nonstationary time series

Eichler, M., Motta, G. and R. von Sachs

Abstract. Factor modelling of a large time series panel has widely proven useful to reduce its cross-sectional dimensionality. This is done by explaining common co-movements in the panel through the existence of a small number of common components, up to some idiosyncratic behaviour of each individual series. To capture serial correlation in the common components, a dynamic structure is used as in traditional (uni- or multivariate) time series analysis of second order structure, i.e. allowing for infinite-length filtering of the factors via dynamic loadings. In this paper, motivated from economic data observed over long time periods which show smooth transitions over time in their covariance structure, we allow the dynamic structure of the factor model to be non-stationary over time, by proposing a deterministic time variation of its loadings. In this respect we generalise existing recent work on static factor models with time-varying loadings as well as the classical, i.e. stationary, dynamic approximate factor model. Motivated from the stationary case, we estimate the common components of our dynamic factor model by the eigenvectors of a consistent estimator of the now time-varying spectral density matrix of the underlying data-generating process. This can be seen as time-varying principal components approach in the frequency domain. We derive consistency of this estimator in a "double-asymptotic" framework of both cross-section and time dimension tending to infinity. A simulation study illustrates the performance of our estimators.

DP0830 - Single index regression models in the presence of censoring depending on the covariates

Lopez, O., Patilea, V. and I. Van Keilegom

Consider a random vector $(X', Y)'$, where X is d -dimensional and Y is one-dimensional. We assume that Y is subject to random right censoring. The aim of this paper is twofold. First we propose a new estimator of the joint distribution of $(X', Y)'$. This estimator overcomes the common curse-of-dimensionality problem, by using a new dimension reduction technique. Second we assume that the relation between X and Y is given by a single index model, and propose a new estimator of the parameters in this model. The asymptotic properties of all proposed estimators are obtained.

DP0831 - A sliding blocks estimator for the extremal index

Robert, C. Y., Segers, J. and C. A. T. Ferro

In extreme value statistics for stationary sequences, blocks estimators are usually constructed by using disjoint blocks because exceedances over high thresholds of different blocks can be assumed asymptotically independent. In this paper we focus on the estimation of the extremal index which measures the degree of clustering of extremes. We consider disjoint and sliding blocks estimators and compare their asymptotic properties. In particular we show that the sliding blocks estimator is more efficient than the disjoint version and has a smaller asymptotic bias. Moreover we propose a method to reduce its bias when considering sufficiently large block sizes.

DP0832 - Least squares estimation of nonlinear spatial trends

Crujeiras, R. M. and I. Van Keilegom

The goal of this work is to study the asymptotic and finite sample properties of an estimator of a nonlinear regression function when errors are spatially correlated, and when the spatial dependence structure is unknown. The proposed method is based on a weighted nonlinear least squares approach, taking into account the spatial covariance. Weak consistency of the regression parameters estimator is derived, along with its asymptotic Gaussian limit. The behaviour of the proposed estimator is illustrated with a simulation study, considering different correlation structures in \mathbb{R}^2 and a more general case including a spatial covariate. The method is also applied to two real data cases.

DP0833 - ROC curves in nonparametric location-scale regression models

González-Manteiga, W., Pardo-Fernández, J. C. and I. Van Keilegom

The receiver operating characteristic curve (ROC curve) is a tool of extensive use to analyse the discrimination capability of a diagnostic variable in medical studies. In certain situations, the presence of a covariate related to the diagnostic variable can increase the discriminating power of the ROC curve. In this article we model the effect of the covariate over the diagnostic variable by means of nonparametric location-scale regression models. We propose a new nonparametric estimator of the conditional ROC curve and study its asymptotic properties. We also present some simulations and an illustration to a data set concerning diagnosis of diabetes.

DP 0834 - Goodness-of-fit tests for censored regression based on artificial data points

González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero

Suppose the random vector (X, Y) satisfies the regression model $Y = m(X) + \sigma(X)\varepsilon$, where $m(\cdot) = E(Y \cdot)$, $T^2(\cdot) = \text{Var}(Y \cdot)$ and ε is independent of X . The response Y is subject to random right censoring and the covariate X is completely observed. New goodness-of-fit testing procedures for m and $\sigma^2(\cdot)$ are proposed.

They are based on an integrated regression function technique which uses artificial data points constructed with the method of Heuchenne and Van Keilegom (2007b).

Weak convergence of the resulting processes is obtained and their finite sample behaviour is compared via simulations with the method of Stute, González-Manteiga and Sánchez-Sellero (2000).