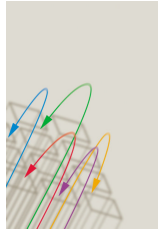


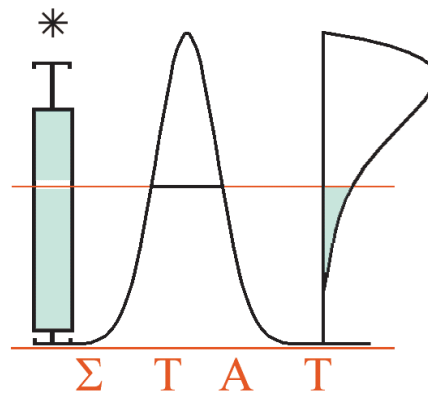
BELGIAN SCIENCE POLICY



**Progress Report 2007**  
**IAP-Network in Statistics**  
**Contract P6/03**

March 27, 2008

<http://www.stat.ucl.ac.be/IAP/PhaseVI>



# Contents

<b>1</b>	<b>Accomplished Research Projects</b>	<b>4</b>
1.1	Introduction and overview . . . . .	4
1.1.1	Introduction . . . . .	4
1.1.2	Overview . . . . .	4
1.2	Work package 1: Multivariate data with qualitative constraints . . . . .	6
1.2.1	Boundaries, frontiers, smooth versus non-smooth functions . . . . .	6
1.2.2	Inverse problems, deconvolution problems and images . . . . .	7
1.2.3	Semiparametric models, dimension reduction and efficiency studies . . . . .	7
1.2.4	Nonparametric inference and robust analysis . . . . .	8
1.2.5	Modeling and measuring of dependencies and copula functions . . . . .	10
1.2.6	Interactions with other work packages . . . . .	10
1.3	Work package 2: Temporally and spatially related data . . . . .	11
1.3.1	Discrete univariate time series under non standard assumptions . . . . .	11
1.3.2	Continuous processes . . . . .	12
1.3.3	Multivariate data . . . . .	12
1.3.4	Spatial data . . . . .	13
1.4	Work package 3: Incomplete data . . . . .	14
1.4.1	Complex modeling approaches for incomplete data . . . . .	14
1.4.2	Sensitivity analysis tools . . . . .	15
1.4.3	Censored survival data . . . . .	16
1.4.4	Interaction with other work packages . . . . .	19
1.5	Work package 4: Data with latent heterogeneity . . . . .	19
1.5.1	The assumption of normality of the random effects distribution . . . . .	19
1.5.2	Issues on computations . . . . .	20
1.5.3	Significance tests for the random effects distribution . . . . .	21
1.5.4	Various extensions of the mixed effects model . . . . .	21
1.5.5	Miscellaneous applications . . . . .	22
1.5.6	Interactions with other work packages . . . . .	23
1.6	Work package 5: Highdimensional and compound data . . . . .	23
1.6.1	Bioinformatics . . . . .	23
1.6.2	Data mining . . . . .	24
1.6.3	Psychometrics . . . . .	25
<b>2</b>	<b>Network Activities</b>	<b>27</b>
2.1	Web site and newsletter . . . . .	27
2.2	Scientific meetings . . . . .	27
2.2.1	Annual workshop . . . . .	27
2.2.2	Meetings . . . . .	28
2.3	Organization of the network: administrative meeting . . . . .	28
2.4	Collaborations, working groups and seminars . . . . .	29

2.4.1	Collaborations . . . . .	29
2.4.2	Working groups . . . . .	30
2.4.3	Seminars . . . . .	30
2.5	Short courses . . . . .	31
2.6	PhD and postdoctoral researchers . . . . .	32
<b>3</b>	<b>Technical Reports and Publications</b>	<b>33</b>
3.1	Université catholique de Louvain, UCL . . . . .	33
3.1.1	List of Technical Reports . . . . .	33
3.1.2	List of Publications . . . . .	36
3.2	Katholieke Universiteit Leuven, KUL-1 . . . . .	39
3.2.1	List of Technical Reports . . . . .	39
3.2.2	List of Publications . . . . .	41
3.3	Katholieke Universiteit Leuven, KUL-2 . . . . .	44
3.3.1	List of Technical Reports . . . . .	44
3.3.2	List of Publications . . . . .	45
3.4	Universiteit Gent, UG . . . . .	48
3.4.1	List of Technical Reports . . . . .	48
3.4.2	List of Publications . . . . .	48
3.5	Universiteit Hasselt, UH . . . . .	50
3.5.1	List of Technical Reports . . . . .	50
3.5.2	List of Publications . . . . .	51
3.6	Université Joseph Fourier, UJF–LMC–IMAG . . . . .	60
3.6.1	List of Technical Reports . . . . .	60
3.6.2	List of Publications . . . . .	60
3.7	Universiteit Utrecht, UU . . . . .	61
3.7.1	List of Publications . . . . .	61
3.8	Universidad de Santiago de Compostela, USC . . . . .	62
3.8.1	List of Technical Reports . . . . .	62
3.8.2	List of Publications . . . . .	63
3.9	London School of Hygiene and Tropical Medicine, LSHTM . . . . .	64
3.9.1	List of Technical Reports . . . . .	64
3.9.2	List of Publications . . . . .	64
3.10	List of joint publications . . . . .	65
3.10.1	List of Technical Reports . . . . .	65
3.10.2	List of Publications . . . . .	66

# 1 Accomplished Research Projects

## 1.1 Introduction and overview

### 1.1.1 Introduction

The research project has been built up around five work packages. Table 1 below gives the *main* contributors to each work package and indicates per package the partner that is coordinating the work.

Work package	Contributing partners
WP1: Multivariate data with qualitative constraints	UCL, KUL-1*, UH, UJF, UU, USC
WP2: Temporally and spatially related data	UCL*, KUL-1, UG, UJF, UU, USC
WP3: Incomplete data	UCL, KUL-1, KUL-2, UG, UH*, USC, LSHTM
WP4: Data with latent heterogeneity	UCL, KUL-1, KUL-2*, UH, UJF, UU, USC
WP5: Highdimensional and compound data	UCL, KUL-1, KUL-2, UG*, UH, UJF, UU, USC

Table 1: *Main contributors per work package, and coordinating partner per work package (indicated with a \*).*

In the subsections below we describe the progress that has been made in the various work packages. For each of the work packages we indicate interactions with research results in other packages. The references mentioned in the text (both Technical Reports and Publications) can be found at the end of this report.

### 1.1.2 Overview

The overall achievements of the research project can be summarized as follows:

- In WP1, new approaches for nonparametric stochastic frontiers are proposed, including robust frontier estimation and dynamics for panel data modeling. Appealing techniques for non-smooth curve estimation are developed, based on e.g. local linear fitting and on bridge regression splines. Practical problems in density deconvolution, and beyond, are studied. A lot of progress is made in semi- and nonparametric modeling. Some models also deal with censored and/or time dependent data, and others can handle very complex data such as mass spectroscopy data or heavy noisy data. New techniques for robust data analysis are developed, e.g., boxplots for skewed data, robust bootstrap methods, robust model selection. In studying dependencies, interesting results are noted, e.g. efficiency of rank-based estimators for bivariate extreme-value copulas, and the use of copulas for censored data and applications.
- The main achievements of the research network in correlated data (WP2) concerns four levels of increasing complexity: univariate discrete time series under non standard assumptions

(such as non stationarity, structural breaks and censoring), continuous processes (with a special focus on the Ratcliff diffusion model), the multivariate analysis of temporally related data and the statistical analysis of spatial (or 3-dimensional) data. At each level, important results are provided in terms of modeling and estimation, and are completed by specification tests. Each level of complexity is also strongly motivated by, and applied to examples of the real life (in various disciplines).

- The work of WP3 has placed emphasis on the dissemination and further development of models for incomplete and censored data, with focal points on clinical trials, epidemiological studies, surveys, and sociological and psychometric applications, often in conjunction with the propagation of proper methodology for longitudinal data, including latent structures, mixed-models, and meta-analytic methods, sometimes of high dimension, thus strongly interacting with WP1, WP4, and WP5. Also, ever increasing emphasis has been placed on sensitivity assessment tools for incomplete (longitudinal) data. All this together has resulted in a number of journal publications in 2007, as well as an edited volume and a monograph. Dissemination has also taken place through short courses in Belgium, Europe, and the economically developed and developing parts of the world, and through dedicated IAP seminars on survival analysis methods, organized in Belgium.
- In WP4, latent heterogeneity in the data has been explored in a variety of ways. The classical linear, generalized linear and non-linear mixed effects models have been generalized by allowing the random effects distribution to belong to a general class of continuous and/or discrete distributions. To this end, frequentist and Bayesian approaches have been employed and allowed to verify the impact of the correct specification of the random effects distribution on the estimation of various parts of the model. Latent heterogeneity has also been studied in survival models (including interval censoring and involving copula models), change point models, multivariate random effects, hierarchical and shared parameter models. New parameter estimation procedures have been developed as well as score tests for the random effects structure and techniques for prediction. Data from various disciplines (medicine, psychology, etc.) were taken to illustrate the performance of the various proposals.
- Research undertaken for highdimensional data and compound data (WP5), is split up in three disciplines: bioinformatics, data mining and psychometrics. In bioinformatics, both the analysis and experimental design of genetic data is studied. More specifically, new statistical analysis techniques are developed for micro-array data. In data mining, a substantial part of the research consists of the development of robust methods for highdimensional data. Furthermore, multivariate techniques are discussed and developed for NMR and gas phase chromatography data. In psychometrics, hierarchical classes models and multi-mode models are studied. Different statistical tools, such as simulated annealing, a fully Bayesian method, probabilistic models with a directed acyclic graph, and different dimension reduction techniques are used in this context.

## 1.2 Work package 1: Multivariate data with qualitative constraints

### 1.2.1 Boundaries, frontiers, smooth versus non-smooth functions

#### Frontier estimation: efficiency and productivity analysis

New and original approaches to estimation of nonparametric stochastic frontier models have been proposed in [68] and [75]. In [47] a new way of defining and estimating nonparametric frontier based on a nonstandard quantile function has been suggested, providing frontier estimators robust to extreme and outlying data points. In [49], conditional efficiency measures are extended to cover convex and non-convex production sets. These conditional measures were used in [40] and [39] to assess and compare the performance of universities in several countries of Europe. In [76], bootstrap allows to improve the inference in a two-stage semi-parametric model for production models, where nonparametric efficiency scores are regressed in a second stage on appropriate environmental factors. Bootstrap ideas are also used for testing for comparing weighted average of the efficiency of groups of firms in an industry. Bootstrap type inference on the total productivity factors has been done, and applied to analyze the economic evolution of Italian regions in [5]. Extensions of previous work on semi-parametric models for panel data allow today to introduce dynamic aspects in the models. Recent results in robust nonparametric frontier models have been collected in [48].

#### Estimation of non-smooth functions and estimation under qualitative constraints

Nonparametric estimation of regression curves with a possible non-smooth behavior at certain (unknown) points, via local linear fitting, is proposed in [140]. The method does the usual smoothing task in continuous regions but preserves at the same time the jumps. The basic ideas of the method have been further used in [92] for peak-preservation in regression curve estimation. Further discussions on smoothing and preservation of irregularities in curve estimation are given in [139]. The paper [363] considers nonparametric estimation of a one-dimensional piecewise-smooth function observed within white Gaussian noise on the interval. The paper proposes a two-step estimation procedure, where one first detects jump points by a wavelet-based procedure and then estimates the function on each smooth segment separately by bridge regression splines.

Nonparametric estimation of a monotone regression function is studied in [436]. The focus is on nonparametric estimation of a (more general) constrained regression function using penalized wavelet regression techniques. In particular, a penalized wavelet monotone regression estimator is investigated, and its optimal rates are established.

#### Boundary and set estimation, estimation of lengths and surfaces

Estimation of a set  $S$  from a random sample of points taken within  $S$  is considered in [414]. It is assumed that  $S$  is  $\alpha$ -convex, which means that a ball of radius  $\alpha$  can go around from outside the boundary of the set. In [404] a method for estimating the boundary length of a body  $G$  is proposed. It relies on a nonparametric estimator based on the information provided by a random sample, taken on a rectangle containing  $G$ , in which one is able to identify whether every point is inside or outside  $G$ . In [396] a method for estimating the boundary length of a body  $G$  is proposed, now assuming that a ball of radius  $\alpha$  can roll inside and outside the boundary of  $G$ . The hope is

to use this shape restriction to propose a new estimator for the boundary length.

### 1.2.2 Inverse problems, deconvolution problems and images

A case study in density deconvolution has been presented in [134]. Via this case study, more general, frequent problems in calculating integrals and optimizing objective functions have been addressed.

Progress has been made in research on the analysis of images. The paper [112] investigates compression algorithms applied to coefficients from nonlinear, adaptive, and overcomplete multi-scale data analyses, with specific attention to images. In [113] a specific procedure for adaptive refinement applied to discrete (or digital) data is investigated. A study of compression schemes applied to edge-oriented multiresolution analyses can be found in [156].

The paper [149] investigates an unbalanced continuous wavelet decomposition for fast optimal location of change points in cases with possibly low information intensity and high noise levels.

### 1.2.3 Semiparametric models, dimension reduction and efficiency studies

Lack-of-fit tests in a semiparametric model where the nonparametric part consists of detail coefficients in a wavelet decomposition are investigated in [87]. The doctoral research by Irène Gannaz (UJF) is concerned with the use of wavelet methods in semiparametric partially linear models. The work presents a wavelet thresholding based estimation procedure to estimate the components of the model by establishing a connection between an  $L_1$ -penalty based wavelet estimator of the nonparametric component and Huber's M-estimation of a standard linear model with outliers. The results are summarized in [369]. The paper [361] studies a composite likelihood approach to estimate the parameters of multivariate mixed Poisson models when the mixing distribution is a multivariate Gamma distribution. Consistency and asymptotic normality of the composite likelihood estimator are derived.

A new method for the binary classification problem is studied in [370]. It relies on empirical minimization of the hinge loss over an increasing sequence of finite-dimensional spaces. A suitable dimension is picked by minimizing the regularized loss, with the regularization term proportional to the dimension. An oracle-type inequality is established, ensuring adequate convergence properties.

The paper [362] proposes a semiparametric shock model for two dependent failure times where the risk indicator of one failure time plays the part of a time-varying covariate for the other one.

Important in regression are goodness-of-fit tests for (semi)parametric regression models. In [434], a novel testing approach is developed for testing whether the regression curve belongs to some parametric family of regression functions. The proposed test statistic measures the distance between the empirical distribution function of the parametric and of the nonparametric residuals. The extension to the case where the response variable is randomly right censored is given in [457]. In [81] goodness-of-fit tests for parametric regression functions are studied in the presence of time series errors, while [435] proposes an empirical likelihood test that is able to test the goodness-of-fit of a class of parametric and semiparametric regression models. The class includes as special cases fully parametric models, and semiparametric models, like the single-index and the partially linear model. A related problem is the comparison of regression curves coming from several populations,

and the development of goodness-of-fit tests for the form of the conditional variance function. These are studied in [456] and [51] respectively.

Simulation studies are an important tool in probability and statistics. In [360] the authors propose some new methods to simulate Gaussian fields that are effective, fast and valid not only for fractional Brownian fields, but for any Gaussian fields.

In some clinical, environmental or industrial trials, fixed-detection limits can result in a type-1 left-censored positive response variable. Since the number of censored observations is larger than expected, a zero-inflated Cox's regression model is assumed in [243]. The conditional probability of having a zero response is modeled as a logistic part while a Cox regression model is taken for the non-zero part of the response.

Fractional polynomials, basically being a set of (generalized) linear models, are a nice and flexible extension of classical polynomials. Typically, one selects the best-fitting model in this set of polynomials and proceeds as if no model selection were carried out. In [277] it is shown that model averaging using a set of fractional polynomials reduces bias and has better precision in estimating a safe level of exposure, as compared to an estimator from the selected best model.

#### 1.2.4 Nonparametric inference and robust analysis

##### **Nonparametric and semi-parametric inference and goodness-of-fit testing**

A two-sample test in functional data analysis, starting from discrete data, is studied in [60]. By choosing the same tuning parameter to produce each curve from its raw data, significant contributions to level inaccuracy and power loss can be avoided. In [79] backfitting and profile methods for general criterion functions, depending on a parameter of interest  $\beta$  and a nuisance function  $\theta$ , are studied. Employing different amounts of smoothing for each method to estimate  $\theta$ , produces two estimators of  $\beta$  with the same limiting distributions.

Mass spectroscopy data are rather complex. [367] studies functional mixed-effects models, with applications to such complex data in mind. The models inherit the flexibility of linear mixed-effects models in handling complex designs and correlation structures. Wavelet decompositions are used to model both fixed-effects and random-effects in the same functional space, which helps in interpreting the model as a functional data model. A contribution to the methodology available for extracting and analyzing signal content from protein mass spectrometry data is in [366]. After removing the noise by an invariant wavelet transform, the authors develop a background correction method based on penalized spline quantile regression and apply it to MALDI-TOF (matrix assisted laser deabsorbtion time-of-flight) spectra obtained from serum samples.

Studying heavy noisy data needs special attention since one cannot assume for example existence of moments (which is mostly done). In [426] the authors consider a nonparametric noisy data model where the unknown signal is assumed to belong to a wide range of function classes, including discontinuous functions and the additive noise distribution may have heavy tails but has zero median. They first use local medians to construct a Gaussian nonparametric regression model, and the resulting data being not equispaced, they apply a wavelet block penalizing procedure adapted to non equidistant designs to construct an estimator of the regression function.

Further developments on extensions of wavelet transforms are in [100]. This paper investi-



gates a specific class of second generation wavelet transforms that are suited for extension to data dependent refinement. In [137] a multiscale analysis of brush wear observations in copy systems is studied. A selective review article that synthesizes some recent work on “nonlinear” wavelet methods in nonparametric curve estimation and their role on a variety of applications is [364].

### **Multivariate data and robust analysis**

The paper [128] proposes several goodness-of-fit tests based on robust measures of skewness and tail weight. These are generalizations of the Jarque-Bera test based on the classical skewness and kurtosis, and are an alternative to the approach of Moors et al. using quantiles. An extension to the case of censored observations is in [240] using a reweighting scheme based on the Kaplan-Meier estimator. The same ideas can be applied to depth quantiles. This leads to regression quantiles for censored data which are robust to both outliers in the predictor and the response variable. The extreme value index is a well-known parameter to measure the tail heaviness of a distribution. In [163] a robust estimator of the tail index is proposed, by combining a refinement of the Pareto approximation for the conditional distribution of relative excesses over a large threshold with an integrated squared error approach on partial density component estimation. Most outlier detection rules for multivariate data are based on the assumption of elliptical symmetry of the underlying distribution. In [147] an outlier detection method which does not need the assumption of symmetry is proposed. This is a generalization of the Stahel-Donoho outlyingness, by using a robust measure of skewness as well. In a boxplot for continuous unimodal data, usually too many points are classified as outliers when the data are skewed. In [160] an adjustment of the boxplot that includes a robust measure of skewness in the determination of the whiskers is presented.

A review of recent developments on a bootstrap method for robust estimators which is computationally faster and more resistant to outliers than the classical bootstrap is in [238]. Applications are given for linear regression, location-scatter parameters, principal components, and discriminant analysis. [220] investigates the performance of robust estimates of multivariate location under non-standard data contamination models such as componentwise outliers. Influence functions of robust multivariate location estimates are derived under flexible contamination models and used to investigate the effect of propagation of outliers. It is also shown that high-breakdown affine equivariant estimators show poor breakdown behavior under componentwise contamination when the dimension is high. A multivariate extension of the median obtained by considering the median as the L1 location estimator is studied in [236]. This multivariate estimator yields simultaneously a location estimate and a shape/scatter estimate. The influence function and asymptotic variances are derived and compared with other estimators of location and shape. Robust and computationally efficient versions of forward selection and stepwise methods are introduced in [230]. The robustifications are obtained by replacing classical correlations by robust counterparts. The proposed robust methods are computationally very suitable and scalable to large high-dimensional datasets. [231] deals with robust model selection using robustified LARS algorithms to select the most important predictors when the number of candidate predictors is large. The technique is computationally very efficient and scalable. Bootstrap is used to obtain stable results. [446] provides a review of robust, high-breakdown methods that can deal with a substantial fraction of outliers in multivariate analysis.

A discussion that complements an overview of the history and development of robust statistics is in [237]. The importance of robust methods in mining large databases is the subject of an editorial contribution in [223].

### 1.2.5 Modeling and measuring of dependencies and copula functions

Bivariate extreme-value copulas are studied in [19]. The paper introduces rank-based versions of nonparametric estimators of bivariate extreme-value copulas, and establishes their asymptotic distributions. Surprisingly, even when the margins are known, a more efficient estimator arises if the information about the margins is ignored and ranks are used instead. For multivariate extremes, estimation of the dependence structure still presents a challenge and an interesting problem. In [12] a procedure for the bivariate case is presented that opens the road to a similar way of handling the problem in a multivariate setting. A method of moments estimator is proposed where a certain integral of a nonparametric, rank-based estimator of the stable tail dependence function is matched with the corresponding parametric version. Tail dependence copulas provide a natural perspective from which one can study the dependence in the tail of a multivariate distribution. For Archimedean copulas with continuously differentiable generators, regular variation of the generator near the origin is known to be closely connected to convergence of the lower tail dependence copulas to the Clayton copula. In [42] these characterizations are refined and extended to the case of generators which are not necessarily continuously differentiable.

In [427] the relationship between the error and the covariate in a nonparametric regression model is modeled by means of a copula. This modeling offers a useful alternative for the model assuming independence between the two variables, an assumption which is often violated in practice. [245] extends the conditional Koziol-Green model to dependent censoring. For this nonparametric model, weak convergence of the conditional distribution is shown and its efficiency over the general copula-graphic estimator is studied.

Dependencies can generally be described via association measures. For example, various association measures can be used to describe how “synchronous” two spike trains of neurons are. In [462] a new conditional synchrony measure is proposed. Focus is on the specification of a flexible marginal model for multivariate correlated binary data together with a pseudolikelihood estimation approach, to adequately and directly describe the measures of interest.

### 1.2.6 Interactions with other work packages

In a new edition of [61], new and advanced materials in applied multivariate analysis have been added, including some of the advanced techniques used in other **WPs** (e.g. heavy tailed distributions, copulas, support vector machine, classification and regression trees).

Several techniques discussed in this work package can deal with censored data and/or dependent data, and are directly linked with research under work packages **WP3** and **WP2** respectively.

Heterogeneity and approaches to deal with it is one of the important issues in **WP4**. The interactions with this WP are particularly through the study of generalized linear models and mixed models.

Research in WP1 paid attention to large data sets and robust methods as well as to variable selection and classification methods. As such obtained research results contribute to **WP5**.

### 1.3 Work package 2: Temporally and spatially related data

The main achievements of the research network in correlated data concern four levels of increasing complexity: univariate discrete time series under non standard assumptions, continuous processes, the multivariate analysis of temporally related data and the statistical analysis of spatial (or 3-dimensional) data.

#### 1.3.1 Discrete univariate time series under non standard assumptions

A first aspect of our research focused on the analysis of univariate discrete time series under non standard assumptions.

One source of complexity very often met in the applied sciences is the deviation from stationarity. In particular, structural breaks over time is a common feature in time series analysis and is one potential source of nonstationarity. In [445] the issue is addressed of how to detect possible structural breaks in both the conditional mean and variance functions of nonparametric regression models. By considering a regression model with errors that are martingale differences, both cases, independent data and time series data, are covered by the study. There are numerous nonparametric tests available in the literature, but none of them could be directly applicable to such cases where both the mean and variance may have discontinuities. In [445] nonparametric testing procedures are developed that simultaneously test for structural breaks in the conditional mean and the conditional variance. The asymptotic distribution of an adaptive test statistic is established, as well as its asymptotic consistency and efficiency. Simulations studies illustrate the performance of the adaptive testing procedure. The usefulness of the proposed adaptive test is illustrated on the analysis of financial time series.

Another problem encountered with real data is the problem of censoring. [13] consider a nonparametric regression model in which the data are dependent and the response is subject to random right censoring. They consider the problem of nonparametrically estimating the conditional quantile function in this model. The proposed method is based on a local linear fit using the check function approach. This paper is very much related to the research topics studied under **WP1** and **WP3**.

Goodness-of-fit of fit tests are also important to validate a model on data. In [11] the authors present a procedure to test for proportionality of the regression function and the scale function in a nonparametric regression model with dependent data. The test can be considered as a preliminary step to check the validity of certain time series models, like e.g. ARCH models. The test statistic is based on the difference between two nonparametric estimators of the distribution of the regression error.

### 1.3.2 Continuous processes

Due to the improvement in the data collection, temporal data are sometimes available into the form of a continuous process. A typical example is given by financial markets where microdata are available online. In order to reflect the continuous nature of the observed process, we need to develop more adapted models for correlation.

Important work has been done on the Ratcliff diffusion model that is used in psychology or social sciences. The model is a process account of the time course of human decision-making in two-choice tasks and therefore is also directly related to **WP4**. [114] adopts a Bayesian method for parameter estimation in the Ratcliff model. This method broadens the scope of psychological problems the models can address. In [159], a general method for performing diffusion model analyses on experimental data is presented. The authors show that by implementing design matrices, it becomes possible to fit models with parameters regressed onto predictors. A software tool intended to make the Ratcliff diffusion model for reaction time and accuracy data more accessible to experimental psychologists is proposed in [158]. The tool takes the form of a MATLAB toolbox, freely downloadable.

On this topic, specification tests have again been developed. In [85] a new test procedure is developed for the specification of the diffusion function in a class of semiparametric continuous-time financial econometric models. Asymptotic consistency results for the proposed test are established. The proposed theory is supported by good finite-sample properties.

### 1.3.3 Multivariate data

One important aspect of the research proposal was to study how to extend the univariate analysis to the analysis of multiple series. In this analysis, we also expect to give working methods for analyzing the correlation between the series. During 2007, the network produced a set of results in that direction.

Motivated by substantive questions in actuarial sciences, [440] addressed the problem of modeling life tables. Life table have two time axes: year and age. The Lee-Carter model is popular for modeling the dependence of mortality on these two variables and least squares estimation, using logs of mortalities, is the established method. [440] proposed an estimation based on Poisson regression and smoothness constraints.

In many financial applications, it is useful to concentrate on the multivariate modeling of the extreme of the time series. That issue is addressed in [34] and [2]. In [34], easily verifiable sufficient conditions are found for the extremes of a Markov chain with regularly varying stationary marginal distribution to behave asymptotically as a multiplicative random walk, called the tail chain. The distributions of the future and the past of this tail chain mutually determine each other through a kind of adjoint relation. An application of the theory yields the asymptotic distribution of the past and the future of the solution to a stochastic difference equation (squared ARCH processes being a special case) conditionally on the present value being large in absolute value. In [2], extreme values of a stationary, multivariate time series may exhibit dependence across coordinates and over time. The aim of this paper is to offer a new and potentially useful tool called tail process to describe and model such extremes. The key property is the following fact: existence of the tail

process is equivalent to multivariate regular variation of finite cuts of the original process. The theory is shown to be applicable with great ease to stationary solutions of stochastic autoregressive processes with random coefficient matrices, an interesting special case being a recently proposed factor GARCH model.

Panel, or longitudinal data, with temporal dynamics is a particular case of multivariate time series that is connected to **WP5**. That model is considered in [151]. The goal of that paper is to provide a method to overcome the computational burden associated with latent Markov models (when modeling both individual differences and temporal dynamics in panel data). The method is associated with a probabilistic model with a directed acyclic graph and applies transformations to the graph. When the panel contains a lot of individuals observed over a long period of time, different strategies of modeling and inference are needed. Using a double asymptotic framework (the number of individuals and the number of replications growing with sample size) [3] develop shrinkage for spectral analysis of multivariate time series of high dimensionality. They propose a new non-parametric estimator of the spectral matrix with two appealing properties: compared to the traditional smoothed periodogram the shrinkage estimator has a smaller  $L_2$ -risk and is numerically more stable due to a smaller condition number.

The multivariate analysis of continuous processes has also been studied in [107]. The authors present a diffusion model for the analysis of multivariate longitudinal data with continuous and possibly unbalanced measurement times. Data from a single person are modeled with an extension of an Ornstein-Uhlenbeck diffusion process.

### 1.3.4 Spatial data

A very high level of complexity is given by the analysis of spatial or 3-dimensional data, taking into account the important aspect of data described in the previous sections.

The extension of spectral analysis to spatial data is not straightforward and a lot of work is needed e.g. in order to calibrate a nonparametric estimator of the spatial spectral density. An important contribution to that topic has been achieved in [387] and [398]. In the first study, the authors analyze the asymptotic properties of smoothed nonparametric kernel spectral density estimators for the spatial spectral density of a continuous stationary spatial process. The implications for the edge effect bias of the choice of the kernel, bandwidth and spacing parameter in the design are also discussed, both for tapered and untapered estimators. The aim of the second paper is to study the properties of the multidimensional periodogram, under both cases of tapering and no tapering. It also presents some theoretical results regarding the expectation and the covariance of the multidimensional periodogram.

Real life examples are numerous. [380] show an application in genetics. In that study, a special type of spatial relationship is found on chromosomes. In tumors the normal pairing of chromosomes is disturbed, locally or on whole chromosomes, leading to 'chromosome copy numbers' unequal to 2. [380] propose and apply a high-resolution algorithm to estimate copy numbers from noisy data.

Still in medical science, a challenge in modern brain research is estimation of the diffusion vector from three-dimensional fields. [377] describe a smoothing method based on 3-dimensional tensor product  $P$ -splines.

Finally, having estimated spatial dependence structures, the question of comparing the spectral density and look at their potential differences is important for the practical use of the developed methods. In [402], a statistical technique is provided for checking whether the dependence structure of  $L$  spatial processes is the same. A test statistic is proposed for testing the hypothesis  $H_0 : f_1 = \dots = f_L$ , where each  $f_l$  denotes the spatial spectral density of each process, for  $l = 1, \dots, L$ . The test is based on a Cramér-von-Mises functional type test.

## 1.4 Work package 3: Incomplete data

The work on incomplete data can be broadly subdivided in: (1) complex modeling approaches for incomplete data; (2) sensitivity analysis tools; and (3) censored survival data.

### 1.4.1 Complex modeling approaches for incomplete data

Methods commonly used for the analysis of incomplete longitudinal clinical trial data rest on strong assumptions. Over the last decades, a number of full longitudinal data analysis methods have become available, such as the linear mixed model for Gaussian outcomes.

The book by [450] promotes the proper use of methodology for incomplete data in the context of clinical studies and, more broadly, in settings with life sciences data. The book synthesizes a large amount of work undertaken by the members of WP3, as well as methods from other international schools. It ranges from critiques on simple methods, via the promotion of broadly valid yet easy-to-use state of the art methods, to finally sensitivity analysis tools. [416] wrote a monograph sponsored partly by the UK National Health Service Coordinating Committee for Research Methodology. This monograph develops and expounds methods for the analysis of longitudinal clinical trial data with dropout. It covers methods for continuous and discrete data, including sensitivity analysis.

Several contributions have been made to encyclopedias and other reference works, such as by [451], who provide a general overview of the missing-data field. Also, discussion contributions have been produced, including by [453], at the occasion of a paper by Peter Diggle (Lancaster University) and colleagues, which discusses existing missing-data methodology and offer further classification schemes.

[307] wrote an editorial to the *Journal of the Royal Statistical Society, Series A*, dedicated to the state-of-the-art and future directions of research in incomplete data.

[425] show how hierarchical likelihood provides a convenient framework for fitting certain missing value models. First, a number of published generic missing value problems are examined which purport to show that hierarchical likelihood is inappropriate for missing data. The flaws in each of these published treatments of the problems are described, and correct hierarchical likelihood analyses are given in their place. Second, the approach is applied to that of modeling longitudinal continuous data with a nonrandom logistic dropout model. It is seen to be very much more efficient computationally than conventional likelihood optimization.

Specific attention goes out to multiple imputation. [418] show, through the reweighting of multiple imputation samples, analyses of incomplete data under nonrandom missing value mechanisms

can be carried out without formally fitting the nonrandom missingness model, a step which is typically computationally awkward. The approach uses particular properties of the logistic model for missingness, and it is shown that the analysis is robust to misspecification of certain aspects of this model. [419] provide a contemporary review of multiple imputation, focusing on issues whose importance has largely emerged in the last few years.

#### 1.4.2 Sensitivity analysis tools

Models for incomplete longitudinal and otherwise complex data under the more complex missingness not at random (MNAR) mechanism have gained considerable popularity. At the same time, cautionary remarks have been issued regarding their sensitivity to often unverifiable modeling assumptions. Consequently, there is evidence for a shift towards using ignorable methodology, supplemented with sensitivity analyzes to explore the impact of potential deviations of this assumption in the direction of missingness at random. A team from UH, jointly with Michael G. Kenward (LSHTM) published a paper [448] on fundamental issues regarding what is and what is not possible in terms of distinguishing between MAR and MNAR mechanisms. Precisely, they showed that for every MNAR model, there is an MAR ‘counterpart,’ exhibiting exactly the same fit to the observed data. A general overview of issues arising with incomplete hierarchical data are reviewed in [437], thereby bringing together elements of WP3 and **WP4**.

[142] propose an item response model for dealing with test speededness. The model consists of two random processes, a problem solving process and a random guessing process.

[175] propose a shared-parameter framework, typically used to jointly model longitudinal and survival data and to model incomplete data under a MNAR process, with a flexible random-effects distribution to study sensitivity in MNAR models. The same modeling framework is the subject of study in [324], who study the effect of miss-specifying the random effects distribution. They show that, as the number of repeated measurements per individual grows, the effect of random effects misspecification becomes minimal for certain parameter estimators.

[438] formulate a so-called latent-class mixture model that brings together aspects of all three commonly used model families for incomplete data, i.e., selection models, pattern-mixture models, and shared-parameter models. The paper also interacts. It offers an alternative basis for sensitivity analysis tools. [290] proposes an elegant and flexible modeling strategy for incomplete data under MNAR, when missingness is non-monotone. The latter qualification is important, since extremely little work had been done in this area.

The local influence diagnostics, proposed by Cook (1986), provide a flexible way to assess the impact of minor model perturbations on key model parameters’ estimates. [428] apply the local influence idea to the detection of test speededness in a model describing non-response in test data, and compare this local influence approach to the optimal person fit index proposed by Drasgow and Levine (1986), and the empirical Bayes estimate of the test speededness random effect. The measures provide information that can be used to supplement the classical test score in case the test was administered under a rather strict time constraint.

[424] outline a simple method for sensitivity analysis for randomized control trials with a single primary outcome. It is shown how a prior can be elicited from experts about expected differences

in response between completers and dropouts, and how this can be synthesized with the data using a pattern-mixture model. The authors give an approximate closed form for the posterior. The method is applied to a RT to improve the quality of peer review at the BMJ. Other applications and generalizations are sketched.

[289] study issues surrounding missingness, specific to the context of quality of life data from HIV clinical trials. [287] shows the effect of ignoring missing data in modeling the force of infection of the bovine herpes virus-1 in Belgian cattle and proposes the use of weighted generalized estimating equations with constrained fractional polynomials as a flexible modeling tool.

Also in the HIV area, Els Goetghebeur (UG) is co-author of a report ([228]) as member of the ‘Committee on Methodological Challenges in HIV prevention trials’ of the Board on Global Health, Institute of Medicine, of the US National Academies. Indeed, a string of recent large scale clinical trials in HIV prevention have failed to show positive results. This is in part due to an apparent lack of power created by a host of methodological complications. In this report the committee identifies key methodological problems in the field of microbicide and pre-exposure prophylaxis trials, and proposes ways to address them. For instance, in a chapter on adherence problems methods of causal inference are proposed and in a chapter on alternative designs they propose to conduct sequentially randomized trials to find optimal dynamic treatment regimes.

### 1.4.3 Censored survival data

Let us now turn to, possibly censored time-to-event data.

[251] proves a weak convergence result for the empirical copula process based on multivariate observations at fixed covariate values. The process involves smoothing in the covariate space. [291] studied a kernel distribution function estimator with a location dependent bandwidth. This modification reduces the bias term from order  $h^2$  to order  $h^4$ . [292] solved several two-boundary problems for a special Lévy process: the Poisson process with an exponential component. Variables studied are: first exit times, value of the overshoot, first entry times and value at entry time. [265] extended the Koziol-Green model to accommodate dependent censoring. Several asymptotic results are obtained for a new survival function estimator. [245] study weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring. [355] studied estimation problems in a system of two seemingly unrelated regression equations. Superiority is proved for Bayes and empirical Bayes estimators. [427] consider a non-parametric regression model in which it is assumed that the error is linked to the covariate  $X$  through a copula function. Via a pseudo-likelihood method, they estimate the parameter of the copula. As a result, they prove the asymptotic normality of this parameter. They also study the estimation of the conditional distribution of  $Y$  given  $X$ .

In some clinical, environmental, or industrial trials, fixed-detection limits can result in a type-1 left-censored positive response variable. However, [243] note that the number of censored observations is larger than expected and assume a zero-inflated Cox regression model. They model the conditional probability of having a zero response as a logistic model, while entertaining a Cox regression model for the non-zero part of the response.

[405] present a new model for cross-sectional lifetime data. The model is based on the length-



bias assumption, and it is adapted to situations in which several types of censoring may occur. The non-parametric maximum likelihood estimator of the survival function is derived.

For censored data nonparametric estimation of a density function, using ideas of presmoothing have been studied by [148]. Presmoothing of the Kaplan-Meier estimator has been shown to be more efficient. In the said manuscript, local linear estimation techniques are used in the pre-smoothing step, either directly or via a local logistic approach. This type of presmoothing is compared to a Nadaraya-Watson type of presmoothing which had been used before. An extensive simulation study revealed that the use of local logistic presmoothing leads to very good performance.

A more complex situation in survival analysis occurs when for censored data some failure indicators are missing (under different mechanism of missingness). In other words: for some of the observations it is not known whether the observation was censored or not. Ignoring the data for which the failure indicators are not known, and carrying out statistical analysis on the complete case data, leads to a loss in efficiency. In [141] a class of estimating functions is introduced for the regression parameter of the Cox proportional hazards model to allow unknown failure statuses on some study subjects. An adaptive estimator which achieves the minimum variance-covariance bound of the class is constructed. Similar methods are also developed for nonparametric estimation of the survival function of a homogeneous population and for the estimation of the cumulative baseline hazard function under the Cox model.

[381] consider the problem of prediction of expected time-to-death for individual cervix cancer patients. They use a multi-state proportional hazards model for survival. Predictions are obtained by simulation. The said model is also considered, from an estimation and computational angle, by [439].

In the context of (non)parametric regression with censored responses a number of contributions have been made by the UCL colleagues Heuchenne and Van Keilegom. First, [65] consider a polynomial regression model in which the response is subject to random right censoring and propose a new estimator of the regression coefficients in this model. The estimator extends the classical least squares estimator to censored data, and is based on preliminary non-parametric kernel estimation of the regression function. In [64], this estimator is extended to the case of any (nonlinear) parametric regression model. Second, [63] consider the estimation of the regression function under a non-parametric location-scale model. The estimator they propose is a Nadaraya-Watson type estimator, for which the censored observations are replaced by ‘synthetic’ data points estimated under this model. The estimator offers a useful alternative for the completely non-parametric estimator of the regression function, which usually behaves quite poorly when the censoring proportion is large.

[431] consider a semi-parametric and a non-parametric two-stage estimation approach to fit copula models for correlated infection times in the four udder quarters of dairy cows. They use a pseudo-likelihood ratio test to select an appropriate copula model from the power-variance copula family and propose a new bootstrap algorithm to obtain a  $p$ -value for this test.

One of the most popular ways of incorporating heterogeneity, association, and extra dispersion in time-to-event data is by way of so-called frailties, a particular form of random effect. [441] have written a book on this topic. [429] consider the similarities and differences between copulas

and frailty models as ways of accommodating the said features. The shared frailty model receives attention in [447], who consider the particular and non-standard problem of likelihood ratio and score testing. The issue arises because the null hypothesis lies on the boundary of the parameter space. The said problem has also been studied, in the context of linear mixed models, by [461].

[171] provides a nice overview of regression analysis of correlated interval-censored data, with particular emphasis on the accelerated failure time (AFT) modeling approach. [430] focus on the same modeling framework, and zoom in on the problem of assessing baseline and treatment-effect heterogeneity in the meta-analytic or multi-centric study settings. Furthermore, [188] studies the AFT setting from a Bayesian angle.

The simulation based approach for assessing the impact of measurement error termed, SIMEX, has been recently extended to regression models with a possibly misclassified categorical response and/or the covariates; it is termed MCSIMEX approach. To assess the importance of a regressor, not only its (corrected) estimate is needed, but also its standard error. Based on the original SIMEX approach a method which uses asymptotic expansions to estimate the asymptotic variance is developed by [192]. Further, [194], using the misclassification table for each binary score making up the count, develop an approach to estimating the misclassification probabilities with increased efficiency.

[227] develop a general theory for the analysis of clustered data which enables consistent and asymptotically normal estimation of the effects of within-cluster exposures in the presence of cluster-level confounders. The results are used to study the validity of a popular proposal by Neuhaus and Kalbfleisch (1998) who separate the exposure effect into a within- and between-cluster component within a random intercept model. When regression models adjust for mediators on the causal path from exposure to outcome, the regression coefficient of exposure is biased for the direct effect of exposure when the association between post-exposure measurements and outcome is confounded by more than just the exposure. Structural nested direct-effect models (Robins, 1999) yield consistent estimators for the direct effect parameters, but are inefficient and can be extremely unstable when the intermediate variable is absolutely continuous. [458] develop direct-effect estimators that are not only more efficient, but also consistent under a less demanding model for a conditional expectation of the outcome.

[218] propose a new method for handling measurement error on observed exposures when the causal effect of exposure is estimated, using a structural mean model and instrumental variables approach. They show that the introduction of a second instrumental variable can allow to estimate the expected systematic error conditional on baseline covariates jointly with an asymptotically unbiased mean exposure effect estimate. To draw inference on the effect on mortality of acquiring an infection on a fixed day after acceptance into an intensive care unit (ICU), one wishes to use marginal structural allowing for measured time-dependent confounders. One must however also consider informative censoring of the survival time due to hospital discharge, instability of the inverse weighting procedure and last but not least the fact that exposure on a given day is not observed when ICU discharge comes prior to that day. [219] accommodate these problems by developing inference under a new class of marginal structural models which describe the hazard of death for patients if, possibly contrary to fact, they stayed in the ICU for at least a given number

of days and acquired infection or not on that day.

#### 1.4.4 Interaction with other work packages

All three papers by UCL-based researchers Heuchenne and Van Keilegom [65, 64, 63] are situated on the bridge between **WP1** and WP3. There generally is a strong interaction between WP3 and **WP4**. For example, Geert Verbeke (KUL2) and Geert Molenberghs (UH) are co-editors, jointly with Garrett Fitzmaurice (Harvard School of Public Health, Boston, MA) and Marie Davidian (North Carolina State University, Raleigh, NC), on the state-of-the art and recent advances in longitudinal data analysis (**WP4**), with an important focus on incomplete data and sensitivity analysis (WP3), too. The sensitivity analysis work is often to be situated within the context of mixed models or other modeling context for hierarchical data, thus placing itself at the crossroads of WP3 and **WP4**. Much of the work on longitudinal data (**WP4**) faces issues of incompleteness (WP3), as well. Likewise, a lot of the work on survival analysis incorporates latent structures (**WP4**), termed frailties in this context, and, at the same time, deals with censoring (WP3). Likewise, the book by [258] brings together aspects of WP3 and **WP4**. The work on the evaluation of surrogate markers from clinical studies is located at the interaction between WP3, **WP4**, and **WP5**.

## 1.5 Work package 4: Data with latent heterogeneity

### 1.5.1 The assumption of normality of the random effects distribution

The assumption of a normal random effects distribution has been extended by Ghidry et al. (2004) to a mixture of normal densities on a fine fixed grid with fixed standard deviations, whereby the mixing weights are estimated using a penalized approach. In [185] it is shown that the approach compares favorably to three competing approaches. Further, this technique has been applied to a variety of other models using a frequentist as well as a Bayesian estimation procedure. In the context of a meta-analysis, the approach was used to generalize the distribution of the baseline risk, see [184]. In [189], the random effects distribution of a generalized linear mixed effects model is assumed of this type. The same approach has been applied in a variety of survival applications using a Bayesian approach, see [188, 447, 190] for applications of this approach in accelerated failure time models whereby the random effects and/or the measurement error distribution is assumed to be of the above mixture type. [167] use the approach to derive a simple expression for association measures in bivariate survival models with interval-censored responses. In [175] the above mixture model for the random effects distribution is assumed for a shared parameter model combined with the Vertex-Exchange algorithm to estimate the parameters, while in [209] the distribution of the random effects distribution is completely unspecified. Mixtures are useful models for heterogeneous data, but parametric mixtures often show a relatively poor fit to the data. [375] combine smoothness and log-concave shapes to provide effective non-parametric models.

In [170], the normality assumption for the random effects distribution in a linear mixed model is replaced by a multivariate skew elliptical distribution and estimation is done using a Bayesian

approach. Non-parametric latent densities are assumed in [374] to describe the heterogeneity in discrete distributions, allowing for much a larger over-dispersion than the classical negative-binomial distribution.

A (Bayesian) Dirichlet process can be used to replace the parametric assumption of normality by a distribution which is stochastically centered on the normal distribution. This approach was applied in [169] to the same dental problem as in [190].

A two-part shared parameter model is suggested in [204] to account for the presence of excess zeros in a longitudinal binary outcome combined with a copula model for random effects. This allowed for examining the sensitivity of inference to various association structures.

The impact of misspecifying the random-effects distribution in generalized linear mixed models is the topic of interest in [300] and [303]. These findings underscore and further contribute to understanding the differences between linear and generalized linear mixed-effects models. In [324] the effect of misspecifying the random effects distribution in shared parameter models is studied. It is shown that as the number of repeated measurements per individual grows, the effect of random effects misspecification becomes minimal for certain parameter estimators.

In [135], the authors analyze the sum-score based formulation of the Rasch model. It is discussed from the point of view of pseudo-likelihood and of misspecified models. Exact relationships between the estimated standard errors for both models are derived.

### 1.5.2 Issues on computations

In [180], [443] and [211] a computational procedure is developed to obtain parameter estimates for a multivariate mixed model. A mixed model is used to describe the longitudinal profile of each marker. To account for the correlation between markers, the univariate mixed models are combined into a multivariate mixed model by specifying a joint distribution for the random effects using a pairwise modeling strategy.

A new type of Laplace approximation is proposed for the estimation of joint models for longitudinal and survival data when the random effects vector is high dimensional, see [174].

While Bayesian nonparametric methods are extremely powerful and have a wide range of applicability, software is lacking. In [186] a freely downloadable R package, DPpackage, is described considering Dirichlet Process, Polya Trees, mixtures of Triangular distributions, and Random Bernstein polynomials priors.

In the framework of estimating mixed models, in [152] a method is proposed that does not require a numerical evaluation of integrals. The proposal is a variational estimation method for mixed logistic models based on a lower bound approximation of the logistic model. In [459], a method is proposed to fit a crossed random-effects model with two levels and continuous outcomes for the analysis of continuous hierarchical data. A crossed random-effects model for binary data is also presented.

In [155] a maximum likelihood methodology is suggested to estimate the parameters of a one-dimensional stationary process of Ornstein-Uhlenbeck type that is constructed via a self-decomposable distribution. In [159] a general method for performing diffusion model analyses on experimental data is presented. The authors show that by implementing design matrices, it

becomes possible to fit models with parameters regressed onto predictors. A freely downloadable MATLAB toolbox is presented in [158] intended to make the Ratcliff diffusion model for reaction time and accuracy data more accessible to experimental psychologists. An iterative maximum a posteriori estimator in item response theory is proposed in [106] as an enhanced technique for estimating proficiency levels.

In meta-analysis with case-control groups it is of interest to study the variability of event probabilities in both groups. [373] proposed a novel way to estimate a non-parametric two-dimensional latent distribution.

[439] examined various estimation approaches for frailty models and compared these by using a simulation study. None of the approaches proves to be uniformly better than another.

### 1.5.3 Significance tests for the random effects distribution

[452] and [461] study the non-standard behavior of the score, Wald, and likelihood-ratio tests in hierarchical data when the null hypothesis is either situated on the boundary of the parameter space or constrained in a different way.

Local influence diagnostics (Cook, 1986) are applied in [428] in the detection of test speededness in a model describing non-response in test data. This approach is compared to an optimal person fit index approach (Drasgow and Levine, 1986) and an empirical Bayes estimate of the test speededness random effect.

### 1.5.4 Various extensions of the mixed effects model

A latent-variable model is suggested in [208] for longitudinal data with a bivariate ordinal outcome, whereby the fixed-effects parameters that are subject-specific retain their population averaged interpretation when properly scaled. A conditional mixed-effects model with crossed random effects is studied in [459]. [454] proposed a modeling framework for count data where normal random effects, included to capture the repeated nature of the data on the one hand, are combined with random effects, e.g., of a gamma type, included to capture overdispersion on the other hand. In [305], a novel method for fitting frailty models to survival data is considered. It consists of using a cumulative hazard based transformation to transform the data in such a way, that they can be analyzed using a linear mixed model.

A new class of models is suggested in [127] making use of copulas to deal with local item dependencies. This approach overcomes some of the problems associated with other local item dependency models.

A model is proposed, see [96], with a simultaneous latent structure for two of the three modes of a three-mode data set. This model helps in studying the correlational structure of two modes simultaneously. On the other hand, in [97] a model that simultaneously accounts for the three ways in three-way data is presented. Random effects are used to model between-actor variability, and structural relations between the linking variables are investigated.

Within the ANOVA context, the authors in [98] outline that hierarchical linear models provide a framework to analyze data from complex designs. Multivariate random effects may be helpful in analyzing interactions between fixed and (nested) random factors. Further, a diffusion model

is presented in [107] for the analysis of multivariate longitudinal with continuous and possibly unbalanced measurement times. Data from a single person are modeled with an extension of an Ornstein-Uhlenbeck diffusion process. A hierarchical model is suggested in [108] for regressing reaction time data onto covariates that avoids both aggregation and squared-error minimization. This model provides novel insights in lexical access. In [114] the authors consider a Ratcliff diffusion model, a well developed process account of the time course of human decision-making in two-choice tasks, and suggest to adopt a Bayesian method for parameter estimation. This method broadens the scope of psychological problems the models can address. An item response model is proposed in [142] for dealing with test speededness. The model consists of two random processes, a problem solving process and a random guessing process.

A nonparametric homoscedastic regression model in which the error is assumed to link to the covariate  $X$  is modeled in [427] through a copula function. Via a pseudo-likelihood method, an estimator of the copula parameter is proposed and its asymptotic normality proven. In [245] the copula approach is used to generalize the Koziol-Green model under dependent censoring.

In some clinical, environmental or industrial trials, fixed-detection limits can result in a type-1 left-censored positive response variable leading to a zero-inflated Cox regression model. The authors modeled in [243] the conditional probability of having a zero response as a logistic model while a Cox regression model was used for the non-zero part of the response.

### 1.5.5 Miscellaneous applications

For the analysis of discrete bounded-outcome score data, [195] suggested a latent random variable which logistically transformed scale has a normal distribution. The assumption of a latent heterogeneity simplified the calculations. An alternative approach is given in [206].

Based on the Signal-Tandmobiel study data, it was demonstrated that the interpretation of a model with respect to conditional independence can be drastically different on a latent scale from that of the manifest scale, see [210], [183] and [187]. Using the approach of Cecere et al. (2006) the variability of permanent tooth emergence is further examined in [193]. In diagnostic testing, [196] examined, via a case study, the impact of deviations from conditional independence.

Regeneration of blood vessels is a non-linear process, showing a quick start and gradual saturation. Using non-linear mixed models, this process has been studied in [383].

A large amount of work has been done in the context of the evaluation of surrogate markers in clinical studies ([253, 254, 255, 256, 257, 268, 270, 274, 296, 311, 267]). Since the said evaluation of surrogates involves at least two endpoints, i.e., the true endpoint and surrogate endpoints, and validation usually takes place within a multi-centric or otherwise meta-analytic context, the data invariably are of a hierarchical nature. This implies that hierarchical modeling ideas, such as mixed-effects models, are virtually always used. In some cases, either the surrogate endpoint, the true endpoint, or both endpoints, are longitudinal, rendering the hierarchy even more elaborate.

The application of mixed-effects models to clinical-trial data when interest lies in psychometric validation (reliability and generalizability) is studied in several papers ([295, 315, 348]). The advantage of such an approach is that existing data can be used, obviating the need for special-purpose studies. Also, one is not confined to the typical pre-post design, but rather entire sequences

of repeated measures can be used.

Specific applications of hierarchical models with random-effects and other latent structures include the use of random-effects models in estimating the force of infection from serological prevalence data ([322]) and the use of electroencephalogram data to discriminate between potential psychotropic compounds ([300]). Longitudinal data on neuron-firing experiment, conducted in monkeys, are the topic of interest in [462]. In [336] the coverage rate of vaccines is examined, while linear mixed models were applied to cDNA data in [284]. A zero-inflated negative binomial regression model is used to model the dmft-index expressing the caries experience in deciduous teeth, see [203].

### 1.5.6 Interactions with other work packages

Various contributions of WP4 pertain to modeling longitudinal data with a random effects structure. Longitudinal data suffer often from dropouts, patients missing a visit. Therefore many developments in this domain are closely related to the problem of missing data, the theme of **WP3**.

## 1.6 Work package 5: Highdimensional and compound data

Highdimensional data and compound data occur in different disciplines, and the methodological problems that need to be solved differ accordingly. Therefore, the research output in this work-package is split up in three domains: bioinformatics, data mining and psychometrics.

### 1.6.1 Bioinformatics

At UG, [241] propose robust and efficient tests and estimators for gene-environment/ gene-drug interactions in family-based association studies in which haplotypes, dichotomous/quantitative phenotypes and complex exposure/treatment variables are analyzed. Using causal inference methodology, they show that the tests and estimators are robust against unmeasured confounding due to population admixture and stratification, provided that Mendel's law of segregation holds and that the considered exposure/treatment variable is not affected by the candidate gene under study. [233] develop optimally cost-effective two-stage designs for marker selection exploiting the method of balanced testing. This novel testing approach relies on a measure of relative evidence involving p-values from both the perspective of the null and the alternative, to balance type I and type II errors rather than making selections driven primarily by the perspective of the null distribution. [225] use mixed model techniques to deal with sources of variability in the data-analysis of phenotyping experiments with transgenic rice. Somaclonal variability and insertion variability are separated from each other to assess the effect of the inserted gene and its variability in a correct way.

At UH, different methods for the analysis of micro-array data are developed. [299] consider several testing procedures for testing for a monotonic trend in dose response experiments, and applied them in the microarray data context. In particular, they combine the procedures with adjustments for across gene multiple testing, and address the issue of small sample size by using

resampling based inference. [350] consider a range of classification methods, and apply them to the problem of classifying samples based on microarray data. They compare the methods by using a simulation study. It appears that classical methods, like diagonal LDA, seem to perform very good as compared to other, more complicated techniques. [284] show how linear mixed effects models can be used to simultaneously normalize cDNA microarrays and to perform differential analysis based on the cDNA data.

At KUL-1, [162] present a novel algorithm for multidimensional unfolding that overcomes both general problems and problems that are specific for the analysis of gene expression data sets. This method offers a useful tool for preliminary explorations of microarray data.

At UU, [379] study the relationship between colon tumor grade and Single nucleotide polymorphisms (SNPs) data. SNPs are point mutations in DNA. They can be measured at a very large scale (up to half a million in one go). The goal of much high-throughput research is to find biomarkers. [382] use penalties in dose-response relationships to find the most influential variables in a large number of measurements.

At USC, [390] revise and comment on the work by Tyekucheva and Chiaromonte (2008), where the authors revise the different ways of solving the problem caused by the singularity of the estimated covariance matrix. They also propose a new estimator. Its performance is statistically compared and assessed with real data applications in the genomic context.

### 1.6.2 Data mining

The development of robust techniques in data mining is partially a joint research effort of researchers of UG and KUL-1. [446] review robust, high-breakdown methods that can deal with a substantial fraction of outliers in multivariate analysis. Location-covariance estimation, multiple and multivariate regression, discriminant analysis, principal components, and multivariate calibration are covered. [230] introduce robust and computationally efficient versions of forward selection (FS) and stepwise (SW) methods. The robustifications are obtained by replacing classical correlations by robust counterparts. The proposed robust methods are computationally very suitable and scalable to large high-dimensional datasets. [231] further describe robustified LARS algorithms to select the most important predictors when the number of candidate predictors is large. The technique is computationally very efficient and scalable. Bootstrap is used to obtain stable results. [223] publish an editorial on a special issue partly focusing on robustness for large data sets. This editorial explains the importance of robust methods in mining large databases. [238] review recent developments on a bootstrap method for robust estimators which is computationally faster and more resistant to outliers than the classical bootstrap. This fast and robust bootstrap method is, under reasonable regularity conditions, asymptotically consistent. Applications are given for linear regression, location-scatter parameters, principal components, and discriminant analysis. [132] present a general method to fit robust estimators in parametric models for huge data sets. The method offers distribution-free confidence intervals for the median of the predictions. The main focus is on general support vector machines (SVM) based on minimizing regularized risks. [136] describe the common problem of first and second order Rayleigh and Raman scatter when fitting Parallel Factor Analysis (PARAFAC) to three-way fluorescence excitation-emission data (EEM).



The scatter can be considered as an example of element-wise outliers. In this paper an automatic scatter identification method is developed based on robust statistical methods for high-dimensional data. [145] develop new techniques for model selection and model validation based on the Leave-one-out Cross-validation (LOO-CV) technique. Fast algorithms are presented for LOO-CV when using a high-breakdown method based on resampling, in the context of robust covariance estimation by means of the MCD estimator and robust principal component analysis. A robust PRESS curve is introduced as an exploratory tool to select the number of principal components.

At UH, [345] consider a technique for peak detection in mass spectra, based on a Poisson approximation of the distribution of different isotopic variants of a peptide. They show that, in its basic form, the technique allows to reliably detect peptides that do not contain any sulphur atoms, but fails to do so in the case when sulphur atoms are present. Furthermore, they develop modifications of the technique that allow to detect peptides also in the latter case.

At UCL, data mining in chemometrics is considered. [32] compare some chemometric tools for metabonomics biomarker identification. NMR-based metabonomics discovery approaches require statistical methods to extract, from large and complex spectral databases, biomarkers or biologically significant variables that best represent defined biological conditions. They explore the respective effectiveness of six multivariate methods: multiple hypotheses testing, supervised extensions of principal (PCA) and independent components analysis (ICA), discriminant partial least squares, linear logistic regression and classification trees. [26] optimize chromatographic analytical conditions. Gas phase chromatography generates high dimensional spectral data and is used daily in industry for the quantification of the concentrations of mixture components. They introduce an integrated methodology to maximize spectral peak separation and hence improve the quality of measurement results.

At UU, [378] develop techniques for mass spectra data which form yet another type of high-volume data. In medical circles great expectations are living about its use for prediction. In a carefully designed ‘shoot-out’, five classification methods for high-dimensional data were compared.

### 1.6.3 Psychometrics

High dimensional and compound data also occur frequently in the psychological literature. At KUL-1, different new techniques are developed to cope with these data complexities.

[130] introduce a new model to chart how individual differences in stimulus-response (S-R) profiles are caused by individual differences in the stimulus-mediating variables (S-M) link and/or by individual differences in the M-R link. An algorithm to fit CLASSI is described and evaluated. [131] evaluate the issue of local minima in hierarchical classes models by proposing a generic simulated annealing (SA) algorithm and three different types of random starts. Results support the use of a SA algorithm in combination with a random multistart procedure. [150] present a stochastic extension of the hierarchical classes model for two-way two-mode binary data. A fully Bayesian method for fitting the new models is presented and tools for model selection and model checking are proposed. [151] propose a method to overcome the computational burden associated with latent Markov models (when modeling both individual differences and temporal dynamics in panel data). The method is associated with a probabilistic model with a directed acyclic graph

and applies transformations to the graph. [153] investigate four possible model selection heuristics in the framework of multi-mode partitioning models for N-way N-mode data. Performance of these four heuristics is systematically compared in a simulation study. [154] propose different bootstrap strategies for estimating confidence intervals of the parameters in multilevel simultaneous component analysis framework. [378] present a multidimensional unfolding technique that is not prone to degenerate solutions and is based on multidimensional scaling of a complete data matrix. This technique leads to acceptable recovery of given preference structures and is relatively easy to implement. [164] extend the hierarchical classes family to two-way two-mode rating data (HICLAS-R). The authors highlight interesting relationships with classical real-valued two-way component analysis and with methods of optimal scaling. [165] propose a unifying model that implies a categorical and/or dimensional reduction of one or several modes of a multiway data set. The model encompasses a broad range of existing reduction models as special cases.

## 2 Network Activities

### 2.1 Web site and newsletter

All activities of the IAP-statistics network can be followed very closely from our web site. The address of the web site is

<http://www.stat.ucl.ac.be/IAP/PhaseVI>.

The web site contains e.g. the following information:

- Our logo
- Call for applications
- Description of the project
- List of scientific personnel working under the IAP project
- Downloadable member list
- Research activities (workshops, seminars, short courses,...)
- Downloadable technical reports, list of publications and list of books written by members of the network
- Annual reports and reports of scientific meetings
- Contact details

In addition an electronic newsletter is sent out every month to all IAP-members. In this newsletter, new activities (seminars, short courses, meetings, ...) are announced and a link to the appropriate web page is added for more details. The newsletter also contains a link to the updated list of publications and technical reports of the network.

### 2.2 Scientific meetings

#### 2.2.1 Annual workshop

The annual workshop of 2007 entitled ‘Statistical inference for dependent data’ was organized by the University of Hasselt on April 26–27, 2007. There were 65 participants. Three two-hour lectures were given by the following main speakers: Roger Nelsen (Lewis and Clark College, USA), Patrick Gagliardini (Lugano, Switzerland) and Yuri Goegebeur (University of Southern Denmark). There were also 11 contributed talks.

The annual workshop of 2008 will be organized by the University of Ghent on September 18-20, 2008. It will focus on the cornerstone ‘Data’, which is one of the three cornerstones on which the project is based (the two others being ‘Models’ and ‘Methods’). The workshop is entitled ‘Missing information in survival data beyond right censoring’. The main speakers will be R. Chappell, J. Sun and J. Lawless. A social activity will be organized on September 20, in order to stimulate interactions (both scientifically and socially) in the network.

The annual workshop in 2009 and 2010 will be organized by respectively the KUL-2 and KUL-1 partner and will be focused on the cornerstone ‘Models’ respectively ‘Methods’. The final workshop in 2011 will be organized by the UCL partner.

### 2.2.2 Meetings

The following meetings were organized by the network in 2007:

- Kick-off meeting, February 9, 2007: an introductory meeting, aimed at getting to know better the different network members and their research areas. During this meeting, each partner presented briefly the various research topics studied by his group. There was also ample time for discussion and social contact. About 100 members attended the meeting. More information on the program of the kick-off meeting can be found at

[http://www.stat.ucl.ac.be/IAP/PhaseVI/archive\\_meeting.html](http://www.stat.ucl.ac.be/IAP/PhaseVI/archive_meeting.html)

- Meeting on ‘Mixture modeling’ organized by KUL-1, KUL-2, UH and the department of statistics of PUC (Chile), November 20-21, 2007. The meeting was attended by a big audience, and brought together researchers from different universities from the network working on this research topic.

## 2.3 Organization of the network: administrative meeting

The annual administrative meeting took place on 29 November 2007 at UCL. The meeting was attended by:

- For the Belgian Science Policy: C. Lejour
- For the follow-up committee: A. Davison (Ecole Polytechnique Fédérale de Lausanne (EPFL)), U. Gather (University of Dortmund) and T. Snijders (University of Oxford and University of Groningen)
- For the network: K. Bogaert (KUL-2), C. Cadarso-Suárez (USC), L. Duchateau (UG), I. Gijbels (KUL-1), L. Simar (promoter, UCL), F. Tuerlinckx (KUL-1), S. Van Bellegem (UCL), I. Van Keilegom (promoter, UCL), N. Veraverbeke (UH) and Ch. Denayer (administrative responsible, UCL).

During this meeting the following main issues were discussed: presentation of the follow-up committee; past and future scientific activities organized by the network; scientific collaborations; work valorization (web page, reports, ...); network organization, management and visibility.

The administrative meeting was followed in the afternoon by two seminars (see Subsection 2.4.3 below for more details).

## 2.4 Collaborations, working groups and seminars

### 2.4.1 Collaborations

The IAP network is working on a broad range of research topics in statistics. There is a large number of scientific collaborations within the network, as can be seen from the list of technical reports and publications (see Section 3, and in particular Subsection 3.10, where all joint technical reports and publications are collected). Below, we mention a few examples of ongoing collaborations between members of different teams of the network.

- Members of the KUL-1 (in particular Irène Gijbels) and of UJF (Anestis Antoniadis) have extensive collaborations on the development of semi-and nonparametric methods for complex data, such as heavy noisy data. One of the aspects of the work is exploring regularization techniques. I. Gijbels visited the UJF partner twice in 2007 (10-14 September and 7 December) to work on joint projects.
- Ingrid Van Keilegom (UCL), Noël Veraverbeke (UH) and Elisa Molanes (Univ. Carlos III, Madrid) are working on empirical likelihood methods for non-smooth criterion functions.
- Irène Gijbels (KUL-1), Noël Veraverbeke (UH) and Marek Omelka (IAP-postdoc working 6 months at KUL-1, followed by 6 months at UH) are working on local linear estimation methods of copula functions, with applications to goodness-of-fit tests and estimation of the Kendall process.
- Geert Molenberghs (UH) and Michael G. Kenward (LSHTM) completed a manuscript on the relationship between generalized estimating equations and fully specified models, which would be needed when a likelihood route were to be pursued.
- Geert Molenberghs (UH) and Geert Verbeke (KUL-2) are completing work on the unidentified portion of all coarsened-data models (e.g., missing data, censored data, grouped data) and models for augmented data (e.g., latent variables, latent classes, random-effects models, mixture models). They study the communality between both settings and propose the common term ‘enriched data’ to refer to both simultaneously.
- Ingrid Van Keilegom (UCL) has extensive collaborations with members of the USC partner (Wenceslao González-Manteiga and Carmen Cadarso-Suárez) on goodness-of-fit tests in (semi)-parametric regression, when the data are or are not subject to right censoring. They also work together on a project dealing with ROC-curves in regression, and a project on non-parametric estimation and inference for models for successive survival times. W. González-Manteiga visited the UCL (in February 2007), and I. Van Keilegom visited the USC partner twice in 2007 (10-16 June and 9-15 December) to work on these projects.
- Paul Janssen (UH), Luc Duchateau (UG) and PhD students of their respective research groups are working together on projects related to frailty models in survival analysis. They have written together a Springer book on that topic that will be published in 2008.

- Ingrid Van Keilegom (UCL) and Roel Braekers (UH) have finished a paper on flexible modeling, estimation and inference for nonparametric regression models, where the error term is assumed to be dependent on the covariates. The dependence is modeled by means of copulas.

#### 2.4.2 Working groups

Below are a few examples of active working groups in the network. They are an important tool to stimulate interactions between network partners, and to stay informed of the research achievements of other partners of the network.

- A working group on ‘Sensitivity analysis for longitudinal and incomplete data’ is active and populated with members of UH and KUL-2. Meetings take place roughly once a term. Each time one or two presentations are given and discussed at length. This keeps all group members up to date of ongoing research in the team, and it provides input on how specific research projects can be (re-)directed. The work is targeted at research problems, dissemination through education, consulting, and conferences. The scope encompasses modeling strategies for incomplete data and to sensitivity analysis for incomplete and coarse data. The group entertains contacts with the research community at large, the biopharmaceutical industry, and regulatory authorities, such as the Food and Drug Administration. In 2007, this working group met on February 27 (UH), May 7 (KUL-2) and October 1 (UH).
- A working group on ‘Competing risks in survival analysis’ has been set up, and consists of a group of about 15 researchers from UH, UG, UCL, as well as members from the EORTC (Brussels) and the University of Liège. The group consists of applied as well as more theoretical researchers. They gather more or less once every two months. So far, this working group met on November 20, 2007 (UCL), January 17, 2008 (EORTC) and March 19, 2008 (UH). The next meeting will be in May 2008 at UCL.
- A working group on ‘Evaluation of surrogate markers in clinical studies’ is active at Hasselt University. The point of gravity is located within UH, but a large number of researchers from other Belgian and foreign (France, UK, US) organizations and institutions participate on either a regular basis or in an *ad hoc* fashion. Often, visitors from abroad are received. The work is closely connected to the remit of WP4, with ramifications to WP3 and WP5.
- A working group on ‘Modeling dependencies and inference based on copulas’ has been set up, and consists for now of researchers from the KUL-1 and UH. Researchers from other partners are invited to join. The working group meets about once a month. So far, the group met on November 8, 2007 (KUL-1), December 17, 2007 (UHasselt), February 22, 2008 (KUL-1) and March 12, 2008 (KUL-1).

#### 2.4.3 Seminars

Each of the participating partners organizes on a regular basis statistics seminars at their universities. Announcements of these seminars are sent out to most Belgian statisticians, including those participating in the network.

Apart from the regular statistics seminars at the universities involved, several seminars have been organized by the network itself, around central themes of the network. They are on some occasions given by members of the network, in order to foster research interactions and exchange of ideas. These seminars are indicated by a star (\*).

- February 5, Juan Carlos Pardo-Fernández (University of Vigo), ‘Tests in nonparametric regression based on the error distribution’, at UCL
- \*February 5, Wenceslao González-Manteiga (USC), ‘Comparing spatial dependence structures’, at UCL
- February 23, Jérémie Bigot (Université Toulouse III), ‘Homeomorphic smoothing splines: a new tool for monotonizing a non-constrained estimator in nonparametric regression’, at UCL
- February 23, Geurt Jongbloed (Free University of Amsterdam), ‘Density estimation in the current status model’, at UCL
- \*April 20, Paul Eilers (UU), ‘Statistical classification for reliable high-volume genetic measurements’, at UCL
- \*April 20, Kristel Van Steen (UG), ‘Perspectives on genetic association screening, at UCL
- \*June 12, Ingrid Van Keilegom (UCL), ‘Estimation of a semiparametric transformation model’, at USC
- October 12, Mark van der Laan (University of California, Berkeley University), ‘Targeted maximum likelihood learning: Applications in discovery and randomized trials’, at KUL-1
- November 29, Anthony Davison (EPFL, Lausanne), ‘Likelihood inference for a problem in particle physics’, at UCL
- November 29, Ernesto San Martin (PUC Santiago, Chili), ‘On the statistical meaning of the parameters of skew-normal experiments’, at UCL

The latter two seminars were organized on the occasion of the administrative meeting of the network, that took place in the morning (see also Subsection 2.3 for more details). The speakers of these seminars are member of the follow-up committee (A. Davison) and former postdoc of the Phase V-IAP network (E. San Martin).

## 2.5 Short courses

Several short (intensive) courses have been organized within the framework of the IAP-statistics network. These courses were intended for all members of the network, and in particular (but not exclusively) for the PhD-students. The announcements were each time sent out to all members and posted on the web site. No (or reduced) registration fees were required for IAP-members.

A list of the short courses organized during the working year 2007 is given below.

- Short course on ‘Modeling dependence with copulas’ by Ch. Genest, 30-31 January, 1 February 2007, at UCL, jointly organized by the FNRS Graduate School in Statistics.
- Summer school on ‘Advanced nonparametric statistics’, May 21-24, 2007, including three short courses given by J. Opsomer, M. Wegkamp, and O. Linton, at UCL, jointly organized by the FNRS Graduate School in Statistics.
- A short course was given by Geert Molenberghs (UH) on ‘The statistical evaluation of surrogate endpoints in clinical trials,’ September 4, 2007, on the premises of the UCB company in Braine-l’Alleud, under auspices of the Biostatistics Section of the Belgian Statistical Society and the IAP Network.
- Geert Molenberghs (UH) and Geert Verbeke (KUL-2) offered a short course on ‘Mixed models and incomplete data,’ on November 22-23, 2007 at KUL, jointly organized by the ‘Interuniversity Graduate School of Psychometrics and Sociometrics’ and the IAP Network.

The summer school on ‘Advanced nonparametric statistics’ attracted a large number of PhD students, coming from several European countries, including a number of PhD students from the European partner of USC.

## 2.6 PhD and postdoctoral researchers

Below are a few examples of joint appointments of PhD and postdoctoral positions financed by the network, and of memberships of the jury of PhD theses at partner universities.

- Marek Omelka is an IAP postdoc coming from Prague University. He works for 6 months at KUL-1 (October 2007-April 2008) followed by 6 months at UH (April 2008 - October 2008). A working group has been set up around his work, and related projects.
- Doctoral thesis of Sébastien Da-Veiga, University of Toulouse, France, June 2007, promotor Anestis Antoniadis (UJF), ‘rapporteur’ Irène Gijbels (KUL-1).
- Doctoral thesis of Kris Bogaert, KUL-2, June 2007, promotor Emmanuel Lesaffre (KUL-2), member of the jury Irène Gijbels (KUL-1) and Paul Janssen (UH).
- Doctoral thesis Anouar El Ghouch, UCL, October 2007, promotor Ingrid Van Keilegom (UCL), member of the jury Irène Gijbels (KUL-1).
- Doctoral thesis of Irène Gannaz, UJF, December 2007, promotor Anestis Antoniadis (UJF), ‘rapporteur’ Irène Gijbels (KUL-1).



### 3 Technical Reports and Publications

Below we provide the scientific output related to the IAP-statistics network. We give both the Technical Reports and the Publications of network members in 2007:

- Technical Reports: all manuscripts that have been written in 2007, and have been submitted for publication to an international journal. These reports are also available on our web site:

[http://www.stat.ucl.ac.be/IAP/PhaseVI/publication\\_tr.html](http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_tr.html).

Each Technical Report has a number of the form TR07xxx (or TR08xxx for those of 2008), and we mention these reference numbers below. The web site also contains the pdf-file of many of the Technical Reports.

- List of Publications: all published papers in international journals in 2007 (with refereeing system), including also papers that are accepted for publication and are ‘in press’. We also include the books written by members of the network, that are published by international editors. See also the IAP-Statistics Reprints Series on our web site:

[http://www.stat.ucl.ac.be/IAP/PhaseVI/publication\\_reprint.html](http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_reprint.html)

for the published papers (reference numbers are of the form R07xxx), and

[http://www.stat.ucl.ac.be/IAP/PhaseVI/publication\\_books.html](http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_books.html)

for the books (reference numbers are of the form B07xxx).

#### 3.1 Université catholique de Louvain, UCL

##### 3.1.1 List of Technical Reports

- [1] Almeida, C. and M. Mouchart, Testing the normality of the latent variables in the polychoric correlation model, 2007. TR07010.
- [2] Basrak, B. and J. Segers, Regularly varying multivariate time series, 2007. TR07025.
- [3] Böhm, H. and R. von Sachs, Shrinkage estimation in the frequency domain of multivariate time series, 2007. TR07008.
- [4] Böhm, H. and R. von Sachs, Structural shrinkage of nonparametric spectral estimators for multivariate time series, 2008. TR08006.
- [5] Bonaccorsi, A., Curi, C., Daraio, C., Mancuso, P. and L. Simar, Analysing total factor productivity through a bootstrap-based nonparametric approach: the case of Italian regions, 2007. TR07035.
- [6] Bouezmarni, T., El Ghouch, A. and M. Mesfioui, Gamma kernels estimator for density and hazard rate of right censored data, 2007. TR07011.

- [7] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric regression, 2007. TR07032.
- [8] Charpentier, A. and J. Segers, Tails of multivariate Archimedean copulas, 2008. TR08013.
- [9] Daouia, A., Florens, J.-P. and L. Simar, Frontier estimation and extreme values theory, 2008. TR08008.
- [10] Debrus, B., Lebrun, P., Ceccato, A., Caliaro, G., Govaerts, B., Olsen, B., Rozet, E., Boulanger, E. and P. Hubert, Use of ICA on HPLC-DAD data and high-order statistics to automatically achieve peak picking, 2008. TR08017.
- [11] Dette, H., Pardo-Fernandez, J.-C. and I. Van Keilegom, Goodness-of-fit tests for multiplicative models with dependent data, 2007. TR07028.
- [12] Einmahl, J., Krajina, A. and J. Segers, A method of moments estimator of tail dependence, 2007. TR07031.
- [13] El Ghouch, A. and I. Van Keilegom, Local linear quantile regression with dependent censored data, 2007. TR07022.
- [14] Fils-Villetard, A., Guillou, A. and J. Segers, Projection estimators of Picklands dependence functions, 2008. TR08014.
- [15] Florens, J.-P., Johannes, J. and S. Van Bellegem, Identification and estimation by penalization in nonparametric instrumental regression, 2007. TR07029.
- [16] Geenens, G., Explicit formula for asymptotic higher moments of the Nadaraya-Watson estimator, 2008. TR08004.
- [17] Geenens, G. and L. Simar, Nonparametric test for conditional independence in two-way contingency tables, 2008. TR08003.
- [18] Geenens, G. and L. Simar, Single-index modelling of conditional probabilities in two-way contingency tables, 2008. TR08012.
- [19] Genest, C. and J. Segers, Rank-based inference for bivariate extreme-value copulas, 2007. TR07026.
- [20] 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, 2008. TR08016.
- [21] Johannes, J., Van Bellegem, S. and A. Vanhems, A unified approach to solve ill-posed inverse problems in econometrics, 2007. TR07030.
- [22] Jullion, A., Lambert, P. and F. Vandenhende, A non-parametric Bayesian method to smooth PET Time-Activity-Curves, 2007. TR07024.

- [23] Komárek, A., Lesaffre, E. and C. Legrand, Assessing baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution, 2007. TR07013.
- [24] Krier, C., Mouchart, M. and A. Oulhaj, Neural modelling of ranking data (with an application to stated preference data), 2007. TR07005.
- [25] Le Bailly de Tillegem, C. and B. Govaerts, A review of Quantitative Structure-Activity Relationship (QSAR) models, 2007. TR07027.
- [26] 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, 2007. TR07016.
- [27] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008. TR08023.
- [28] Mouchart, M., Russo, F. and G. Wunsch, Causality, structural modelling and exogeneity, 2007. TR07012.
- [29] Ojeda Cabrera, J.L. and I. Van Keilegom, Goodness-of-fit tests for parametric regression with selection biased data, 2008. TR08011.
- [30] Robert, C.Y. and J. Segers, Tails of random sums of a heavy-tailed number of light-tailed terms, 2007. TR07006.
- [31] Rolin, J.-M., Bayesian consistency of density estimators, 2007. TR07007.
- [32] Rousseau, R., Govaerts, B., Verleysen, M. and B. Boulanger, Comparison of some chemometric tools for metabonomics biomarker identification, 2007. TR07019.
- [33] San Martin, E. and J.-M. Rolin, Identifiability and estimability of parametric Rasch-type models, 2007. TR07004.
- [34] Segers J., Multivariate regular variation of heavy-tailed Markov chains, 2007. TR07002.
- [35] Van Keilegom, I., de Uña-Alvarez, J. and L. Meira-Machado, nonparametric location-scale models for successive survival times under dependent censoring, 2008. TR08002.
- [36] 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, 2007. TR07068.
- [37] Van Keilegom, I., Sánchez-Sellero, C. and W. González-Manteiga, Empirical likelihood based testing for regression, 2007. TR07023.

### 3.1.2 List of Publications

- [38] Almeida, C. and M. Mouchart, Bayesian encompassing specification test under not completely known partial observability. *Bayesian Analysis*, 2, 303-318, 2007. R07072.
- [39] Bonaccorsi, A., Daraio, C., Rätty, T. and L. Simar, Efficiency and university size: disciplinewise evidence from the European universities. *Hyvinvointipalvelujen Tuottavuus: Tulok-sia opintien varrelta*, VATT Publications, Helsinki, Finland, 2007. R07109.
- [40] Bonaccorsi, A., Daraio, C. and L. Simar, Efficiency and productivity in European univer-sities. Exploring trade-offs in the strategic profile, in Bonaccorsi, A. and C. Daraio, (eds.), *Universities as strategic units. Productivity and Efficiency Patterns in the European Univer-sity System*, Edward Elgar PRIME Collection, 2007. R07108.
- [41] Boulanger, B., Dewé, W., Gilbert, A., Govaerts, B. and M. Maumy-Bertrand, Risk manage-ment for analytical methods based on the total error concept: Conciliating the objectives of the pre-study and in-study validation phases. *Chemometrics and Intelligent Laboratory Systems*, 86, 198-207, 2007. R07073.
- [42] Charpentier, C. and J. Segers, Lower tail dependence for Archimedean copulas: characteri-zations and pitfalls. *Insurance: Mathematics and Economics*, 40, 525-532, 2007. R07074.
- [43] Charpentier, A. and J. Segers, Convergence of Archimedean copulas. *Statistics and Probab-ility Letters*, 78, 412-419, 2008. R08053.
- [44] Cortiñas Abrahantes, J., Burzykowski, T., Legrand, C., Janssen, P., Ducrocq, V. and L. Duchateau, Comparison of different estimation procedures for proportional hazards model with random effects, *Computational Statistics and Data Analysis*, 51, 3913-3930, 2007. R07208.
- [45] Cosma, A., Scaillet, O. and R. von Sachs, Multivariate wavelet-based shape preserving esti-mation for dependent observations. *Bernoulli*, 13, 301-329, 2007. R07052.
- [46] 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. R08006.
- [47] Daouia, A. and L. Simar, Nonparametric efficiency analysis: a multivariate conditional quan-tile approach. *Journal of Econometrics*, 140, 375-400, 2007. R07113.
- [48] Daraio, C. and L. Simar, *Advanced Robust and Nonparametric Methods in Efficiency Anal-ysis: Methodology and Applications*. Springer, New-York, 2007. B07001.
- [49] Daraio, C. and L. Simar, Conditional nonparametric frontier models for convex and non convex technologies: a unifying approach. *Journal of Productivity Analysis*, 28, 13-32, 2007. R07114.

- [50] Delwarde, A., Denuit, M. and P.H.C. Eilers, Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized likelihood approach. *Statistical Modelling*, 7, 29-48, 2007. R07124.
- [51] Dette, H., Neumeyer, N. and I. Van Keilegom, A new test for the parametric form of the variance function in nonparametric regression. *Journal of the Royal Statistical Society - Series B*, 69, 903-917, 2007. R07054.
- [52] Einmahl, J.H.J., Krajina, A. and J. Segers, A method of moments estimator of tail dependence. *Bernoulli*, to appear, 2008. R08050.
- [53] Einmahl, J. and I. Van Keilegom, Specification tests in nonparametric regression. *Journal of Econometrics*, 143, 88-102, 2008. R08001.
- [54] Einmahl, J. and I. Van Keilegom, Tests for independence in nonparametric regression. *Statistica Sinica*, to appear, 2008. R08014.
- [55] El Ghouch, A. and I. Van Keilegom, Nonparametric regression with dependent censored data. *Scandinavian Journal of Statistics*, to appear, 2008. R08008.
- [56] El Ghouch, A., Van Keilegom, I. and I. McKeague, Empirical likelihood confidence intervals for dependent duration data. *Econometric Theory*, to appear, 2008. R08013.
- [57] Fils-Villetard, A., Guillou, A. and J. Segers, Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics*, to appear, 2008. R08051.
- [58] François, N., Govaerts, B. and C. Guyot-Declerck, Inferential non-centered principal curve analysis for the analysis of time intensity curves in sensory analysis. Methodology and application to beer astringency evaluation. *Journal of Chemometrics*, 21, 187-197, 2007. R07176.
- [59] Gao, J., Gijbels, I. and S. Van Bellegem, Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics, Special Issue on Specification Testing*, 143, 123-142, 2008. R08005.
- [60] Hall, P. and I. Van Keilegom, Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica*, 17, 1511-1532, 2007. R07057.
- [61] Härdle, W. and L. Simar. *Applied Multivariate Statistical Analysis*. Second Edition, 458p., Springer-Verlag, Berlin, 2007. B07002.
- [62] Haeusler, E. and J. Segers, Assessing confidence intervals for the tail index by Edgeworth expansions for the Hill estimator. *Bernoulli*, 13, 175-194, 2007. R07076.
- [63] Heuchenne, C. and I. Van Keilegom, Location estimation in nonparametric regression with censored data. *Journal of Multivariate Analysis*, 98, 1558-1582, 2007. R07053.
- [64] Heuchenne, C. and I. Van Keilegom, Nonlinear regression with censored data. *Technometrics*, 49, 34-44, 2007. R07056.

- [65] Heuchenne, C. and I. Van Keilegom, Polynomial regression with censored data based on preliminary nonparametric estimation. *Annals of the Institute of Statistical Mathematics*, 59, 273-297, 2007. R07147.
- [66] Hjort, N.L., McKeague, I.W. and I. Van Keilegom, Extending the scope of empirical likelihood. *Annals of Statistics*, to appear, 2008. R08007.
- [67] Komárek, A., Lesaffre, E. and C. Legrand, Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*, 26, 5457-5472, 2007. R07078.
- [68] Kumbhakar, S.C., Park, B.U., Simar, L. and E.G. Tsionas, Nonparametric stochastic frontiers: a local likelihood approach. *Journal of Econometrics*, 137, 1-27, 2007. R07112.
- [69] Linton, O., Sperlich, S. and I. Van Keilegom, Estimation of a semiparametric transformation model. *Annals of Statistics*, to appear, 2008. R08012.
- [70] Mouchart, M. and M. Vandresse, Bargaining powers and market segmentation in freight transport, *Journal of Applied Econometrics*, 22, 1295-1313, 2007. R07205.
- [71] Pardo-Fernández, J.C., Van Keilegom, I., and W. González-Manteiga, Testing for the equality of  $k$  regression curves. *Statistica Sinica*, 17, 1115-1137, 2007. R07055.
- [72] Pardo-Fernández, J.C., Van Keilegom, I. and W. González-Manteiga, Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics*, 35, 249-264, 2007. R07062.
- [73] Park, B., Sickles, R. and L. Simar, Semiparametric efficient estimation in dynamic panel data models. *Journal of Econometrics*, 136, 281-301, 2007. R07111.
- [74] Robert, C. Y. and J. Segers, Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance: Mathematics and Economics*, to appear, 2008. R08052.
- [75] Simar, L., How to improve the performances of DEA/FDH estimators in the presence of noise. *Journal of Productivity Analysis*, 28, 183-201, 2007. R07115.
- [76] Simar, L. and P. Wilson, Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, 136, 31-64, 2007. R07110.
- [77] Simar, L. and P.W. Wilson, Statistical inference in nonparametric frontier models: recent developments and perspectives. *The Measurement of Productive Efficiency*, 2nd Edition, Harold Fried, C.A. Knox Lovell and Shelton Schmidt editors, Oxford University Press, 2008. R08002.
- [78] Simar, L. and V. Zelenyuk, Statistical inference for aggregates of Farrell-type efficiencies. *Journal of Applied Econometrics*, 22, 1367-1394, 2007. R07116.
- [79] Van Keilegom, I. and R.J. Carroll, Backfitting versus profiling in general criterion functions. *Statistica Sinica*, 17, 797-816, 2007. R07071.

- [80] 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*, to appear, 2008. R08015.
- [81] Wang, L. and I. Van Keilegom, Nonparametric test for the form of parametric regression with time series errors. *Statistica Sinica*, 17, 369-386, 2007. R07075.

## 3.2 Katholieke Universiteit Leuven, KUL-1

### 3.2.1 List of Technical Reports

- [82] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets, 2007. TR07036.
- [83] Beirlant, J. and J.H.J. Einmahl, Asymptotics for the Hirsch index, 2007. TR07070.
- [84] Carr, P. and W. Schoutens, Hedging under the Heston with jump-to-default, 2007. TR07037.
- [85] Casas, I., Gao, J. and I. Gijbels, Specification testing in semiparametric diffusion models: theory and practice, 2007. TR07038.
- [86] Ceulemans, E. and H.A.L. Kiers, Discriminating between strong and weak structures in three-mode principal component analysis, 2007. TR07069.
- [87] Claeskens, G., Ding, H. and M. Jansen, Lack-of-fit tests in semiparametric mixed models, 2007. TR07039.
- [88] Debruyne, M. and M. Hubert, The influence function of the Stahel-Donoho covariance estimator of smallest outlyingness, 2007. TR07040.
- [89] Debruyne, M., Hubert, M. and J. Suykens, Model selection for kernel regression using the influence function, 2007. TR07041.
- [90] Debruyne, M., Hubert, M. and J. Van Horebeek, Detecting influential observations in kernel PCA, 2008. TR08001.
- [91] Depril, D., Van Mechelen, I. and B. Mirkin, Algorithms for additive clustering of rectangular data tables, 2007. TR07071.
- [92] Desmet, L. and I. Gijbels, Peak preserving regression using local linear fitting, 2007. TR07042.
- [93] Garcia, J., Goossens, S., Masol, V. and W. Schoutens, Lévy base correlation, 2007. TR07043.
- [94] Garcia, J., Goossens, S. and W. Schoutens, Let's jump together: pricing of credit derivatives: from index swaptions to CPPIs, 2007. TR07044.
- [95] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates, 2007. TR07072.

- [96] González, J., De Boeck, P. and F. Tuerlinckx, A double-structure structural equation model for three-mode data, 2007. TR07073.
- [97] González, J., Tuerlinckx, F. and P. De Boeck, Analyzing structural relations in multivariate dyadic binary data, 2007. TR07074.
- [98] González, J., Tuerlinckx, F., and P. De Boeck, Crossed random effects models for the analysis of complex designs, 2008. TR08019.
- [99] Hubert, M., Rousseeuw, P.J. and T. Verdonck, Robust PCA for skewed data, 2007. TR07045.
- [100] Jansen M., Refinement independent wavelets for use in adaptive multiresolution schemes, 2007. TR07046.
- [101] Jönsson, H. and W. Schoutens, Pricing of constant maturity credit default swaps under jump dynamics, 2007. TR07047.
- [102] Jönsson, H. and W. Schoutens, Single name credit default swaptions meet single sided jump models, 2007. TR07048.
- [103] Leenen, I. and E. Ceulemans, Three-way hierarchical classes: a comparison of INDCLAS and Tucker3-HICLAS models, 2007. TR07075.
- [104] Leoni, P. and W. Schoutens, Multivariate smiling, 2007. TR07049.
- [105] Madan, D.B. and W. Schoutens, Break on through to the single side, 2007. TR07050.
- [106] Magis, D. and G. Raïche, An iterative maximum a posteriori estimation of proficiency level to detect multiple local likelihood maxima, 2007. TR07076.
- [107] Oravecz, Z., Tuerlinckx, F. and J. Vandekerckhove, An Ornstein-Uhlenbeck model for multilevel multivariate repeated measurement data, 2007. TR07077.
- [108] Rouder, J.N., Tuerlinckx, F., Speckman, P., Lu, J. and P. Gomez, Navigating between Scylla and Charybdis: How to fit curves and lines to RT data, 2008. TR08020.
- [109] Schepers, J. and I. Van Mechelen, Uniqueness of real-valued hierarchical classes models, 2007. TR07078.
- [110] Schepers, J. and I. Van Mechelen, Capturing the nature of two-way interactions by means of a two-mode clustering method, 2008. TR08021.
- [111] Schepers, J., Van Mechelen, I. and E. Ceulemans, A real-valued model of hierarchical classes, 2007. TR07079.
- [112] Van Aerschot, W., Jansen, M. and A. Bulteel, Normal mesh based geometrical image compression, 2007. TR07051.
- [113] Van Aerschot, W., Jansen, M. and A. Bulteel, Normal mesh techniques for image contour approximation, 2007. TR07052.



- [114] Vandekerckhove, J., Tuerlinckx, F. and M.D. Lee, A Bayesian approach to diffusion process models of decision-making, 2008. TR08022.
- [115] Wilderjans, T., Ceulemans, E., and I. Van Mechelen, The CHIC model: a global model for coupled binary data, 2007. TR07080.
- [116] Wilderjans, T., Ceulemans, E., and I. Van Mechelen, Simultaneous analysis of coupled data blocks differing in size: a comparison of two weighting schemes, 2007. TR07081.

### 3.2.2 List of Publications

- [117] Albrecher, H., Ladoucette, S. and W. Schoutens, A generic one-factor Levy model for pricing synthetic CDOs. In *Advances in Mathematical Finance*, R.J. Elliot et al (eds), Birkhaeuser, 2007. R07048.
- [118] Albrecher, H., Mayer, Ph. and W. Schoutens, General lower bounds for arithmetic Asian option prices, *Applied Mathematical Finance*, to appear, 2007. R07047.
- [119] Albrecher, H., Mayer, Ph., Schoutens, W. and J. Tistaert, The little heston trap. *Wilmott Magazine, January Issue*, 83-92, 2007. R07046.
- [120] Antoniadis, A., Bigot, J. and I. Gijbels, Penalized wavelet monotone regression. *Statistics and Probability Letters*, 77, 1608-1621, 2007. R07045.
- [121] Antonio, K. and J. Beirlant, Actuarial statistics with generalized linear mixed models. *Insurance Mathematics and Economics*, 40, 58-76, 2007. R07044.
- [122] Antonio, K. and J. Beirlant, Issues in claims reserving and credibility: a semiparametric approach with mixed models. *The Journal of Risk and Insurance*, to appear, 2008. R08010.
- [123] Beirlant, J., Berlinet, A., and G. Biau, Higher order estimation at Lebesgue points, *Annals of the Institute of Statistical Mathematics*, to appear, 2008. R08009.
- [124] Beirlant, J., Figueiredo, F., Gomes, M.I. and B. Vandewalle, Improved reduced bias tail index and quantile estimators. *Journal of Statistical Planning and Inference*, to appear, 2008. R08011.
- [125] Beirlant, J., Glaenzel, A., Carbonez, A. and H. Leemans, Scoring research output using statistical quantile plotting. *Journal of Informetrics*, 1, 185-192, 2007. R07171.
- [126] Beirlant, J., Guillou, A., Villetard, A. and E. Delafosse, Estimation of the extreme value index and high quantiles under random censoring. *Extremes*, 10, 151-174, 2007. R07170.
- [127] Braeken, J., Tuerlinckx, F. and P. De Boeck, Copulas for residual dependencies. *Psychometrika*, 72, 393-411, 2007. R07137.
- [128] Brys, G., Hubert, M. and A. Struyf, Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, to appear, 2007. R07043.

- [129] Cariboni, J. and W. Schoutens, Pricing credit default swaps under Lévy models. *Journal of Computational Finance*, 10, 1-21, 2007. R07040.
- [130] Ceulemans, E. and I. Van Mechelen, CLASSI: A classification model for the study of sequential processes and individual differences therein. *Psychometrika*, to appear, 2007. R07138.
- [131] Ceulemans, E., Van Mechelen, I. and I. Leenen, The local minima problem in hierarchical classes analysis: an evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, 72, 377-391, 2007. R07139.
- [132] Christmann, A., Steinwart, I. and M. Hubert, Robust learning from bites for data mining. *Computational Statistics and Data Analysis*, 52, 347-361, 2007. R07039.
- [133] Debruyne, M., Hubert, M., Portnoy, S. and K. Vanden Branden, Censored depth quantiles. *Computational Statistics and Data Analysis*, 52, 1604-1614, 2008. R08016.
- [134] Delaigle, A. and I. Gijbels, Frequent problems in calculating integrals and optimizing objective functions: a case study in density deconvolution. *Statistics and Computing*, 17, 349-355, 2007. R07037.
- [135] Del Pino, G., San Martin, E., González, J. and P. De Boeck, On the relationships between sym score based estimation and joint maximum likelihood estimation. *Psychometrika*, to appear, 2007. R07140.
- [136] Engelen, S., Frosch Moller, S. and M. Hubert, Automatically identifying scatter in fluorescence data using robust techniques. *Chemometrics and Intelligent Laboratory Systems*, 86, 35-51, 2007. R07036.
- [137] Figarella, T. and M.H. Jansen, Brush wear detection by continuous wavelet transform. *Mechanical Systems and Signal Processing*, 21, 1212-1222, 2007. R07035.
- [138] Gao, J., Gijbels, I. and S. Van Bellegem, Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics, Special Issue on Specification Testing*, 143, 123-142, 2008. R08005.
- [139] Gijbels, I., Smoothing and preservation of irregularities using local linear fitting. *Applications of Mathematics*, to appear, 2008. R08017.
- [140] Gijbels, I., Lambert, A. and P. Qiu, Jump-preserving regression and smoothing using local linear fitting: a compromise. *The Annals of the Institute of Statistical Mathematics*, 59, 235-272, 2007. R07033.
- [141] Gijbels, I., Lin, D.Y. and Z. Ying, Non- and semi-parametric analysis of failure time data with missing failure indicators. *IMS Lecture Notes-Monograph Series*, 54, 203-223, 2007. R07032.
- [142] Goegebeur, Y., De Boeck, P., Wollack, J.A. and A.S. Cohen, A speeded item response model with gradual process change. *Psychometrika*, to appear, 2007. R07141.

- [143] Hidegkuti, I. and P. De Boeck, Between-group differences and taxometrics. *Psychological Reports*, 100, 211-230, 2007. R07142.
- [144] Hubert, M., Book review of "Data depth: robust multivariate analysis computational geometry and applications" by Liu, R., Serfling, R. and D.L. Souvaine, *Biometrics*, to appear, 2008. R08042.
- [145] Hubert, M. and S. Engelen, Fast cross-validation of high-breakdown resampling algorithms for PCA. *Computational Statistics and Data Analysis*, 5, 5013-5024, 2007. R07030.
- [146] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, to appear, 2007. R07029.
- [147] Hubert, M. and S. Van der Veeken, Outlier detection for skewed data. *Journal of Chemometrics*, to appear, 2007. R07028.
- [148] Jácome, A., Gijbels, I. and R. Cao, Comparison of the Nadaraya-Watson and local linear methods of presmoothed density estimation under censoring. *Computational Statistics*, to appear, 2008. R08018.
- [149] Jansen, M., Multiscale change point analysis in Poisson count data. *Chemometrics and Intelligent Laboratory Systems*, 85, 159-169, 2007. R07026.
- [150] Leenen, I., Van Mechelen, I., Gelman, A. and S. De Knop, Bayesian hierarchical classes analysis. *Psychometrika*, to appear, 2007. R07143.
- [151] Rijmen, F., Vansteelandt, K. and P. De Boeck, Latent class models for diary method data: parameter estimation by local computations. *Psychometrika*, to appear, 2007. R07144.
- [152] Rijmen, F. and J. Vomlel, Assessing the performance of variational methods for mixed logistic regression models. *Journal of Statistical Computation and Simulation*, to appear, 2007. R07145.
- [153] Schepers, J., Ceulemans, E. and I. Van Mechelen, Selecting among multi-code partitioning models of different complexities: a comparison of four model selection criteria. *Journal of Classification*, to appear, 2007. R07146.
- [154] Timmerman, M.E., Kiers, H.A.L., Smilde, A.K., Ceulemans, E. and J. Stouten, Bootstrap confidence intervals in multilevel simultaneous component analysis. *British Journal of Mathematical & Statistical Psychology*, to appear, 2007. R07148.
- [155] Valdivieso, L., Schoutens, W. and F. Tuerlinckx, Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type. *Statistical Inference for Stochastic Processes*, to appear, 2007. R07149.
- [156] Van Aerschot, W., Jansen, M. and A. Bulteel, Image compression using normal mesh techniques. *Curves and Surface Fitting*, Avignon 2006, Cohen, A., Merrien, J.L. and L. Schumaker, (eds), 266-275, 2007. R07025.

- [157] Vandekerckhove, J. and F. Tuerlinckx, MATLAB for behavioral scientists: a novice's guide to MATLAB. *Experimental Psychology*, to appear, 2007. R07152.
- [158] Vandekerckhove, J. and F. Tuerlinckx, Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, to appear, 2007. R07151.
- [159] Vandekerckhove, J. and F. Tuerlinckx, Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review*, 14, 1011-1026, 2007. R07150.
- [160] Vandervieren, E. and M. Hubert, An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, to appear, 2007. R07024.
- [161] Van Deun, K., Heiser, W.J. and L. Delbeke, Multidimensional unfolding by nonmetric multidimensional scaling of Spearman distances in the extended permutation polytope. *Multivariate Behavioral Research*, 42, 103-132, 2007. R07153.
- [162] Van Deun, K., Marchal, K., Heiser, W., Engelen, K. and I. Van Mechelen, Joint mapping of genes and conditions via multidimensional unfolding analysis. *BMC Bioinformatics*, 8, 181, 2007. R07154.
- [163] Vandewalle, B., Beirlant, J., Christmann, A. and M. Hubert, A robust estimator for the tail index of Pareto-type distributions. *Computational Statistics and Data Analysis*, 51, 6265-6268, 2007. R07023.
- [164] Van Mechelen, I., Lombardi, L. and E. Ceulemans, Hierarchical classes modeling of rating data. *Psychometrika*, to appear, 2007. R07155.
- [165] Van Mechelen, I. and J. Schepers, A unifying model involving a categorical and/or dimensional reduction for multimode data. *Computational Statistics and Data Analysis*, 52, 537-549, 2007. R07156.
- [166] Willems, P., Guillou, A. and J. Beirlant, Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function. *Journal of Hydrology*, 338, 221-236, 2007. R07022.

### **3.3 Katholieke Universiteit Leuven, KUL-2**

#### **3.3.1 List of Technical Reports**

- [167] Bogaerts, K. and E. Lesaffre, Modeling the association of bivariate interval censored data using the copula approach, 2007. TR07085.
- [168] Jara, A., Hanson, T., and E. Lesaffre, Robustifying generalized linear mixed models using mixtures of multivariate Polya trees, 2007. TR07015.
- [169] Jara, A., Lesaffre, E., De Lorio, M. and F. A. Quintana, Bayesian semiparametric inference for multivariate doubly-interval-censored data, 2008. TR08009.

- [170] Jara, A., Quintana, F.A. and E. San Martin, Linear effects mixed models with skew-elliptical distributions: a Bayesian approach, 2008. TR08010.
- [171] Komárek, A. and E. Lesaffre, The regression analysis of correlated interval-censored data: Review of accelerated failure time models with flexible distributional assumptions, 2007. TR07017.
- [172] Komárek, A. and E. Lesaffre, Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution, 2007. TR07018.
- [173] Komárek, A., Lesaffre, E. and C. Legrand, Assessing baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution, 2007. TR07013.
- [174] Rizopoulos, D., Verbeke, G. and E. Lesaffre, Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data, 2008. TR08005.
- [175] Tsonaka, R., Verbeke, G. and E. Lesaffre, Shared parameter models with a flexible random effects distribution to study sensitivity in non-ignorable missingness models, 2007. TR07001.

### 3.3.2 List of Publications

- [176] Beunckens, C., Molenberghs, G., Thijs, H. and G. Verbeke, Incomplete hierarchical data. *Statistical Methods in Medical Research*, 16, 1-36, 2007. R07016.
- [177] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, to appear, 2007. R07006.
- [178] De Wit, L., Putman, K., Lincoln, N., Baert, I., Angst, F., Berman, P., Beyens, H., Bogaerts, K., Brinkmann, N., Connell, L., Dejaeger, E., De Weerd, W., Jenni, W., Kaske, C., Komárek, A., Lesaffre, E., Leys, M., Louckx, F., Schuback, B., Schupp W., Smith, B., and H. Feys, Depression and anxiety in the first six months after stroke. A longitudinal multicenter study. *Disability and Rehabilitation*, to appear, 2008. R08036.
- [179] De Wit, L., Putman, K., Schuback, B., Komárek, A., Angst, F., Baert, I., Berman, P., Bogaerts, K., Brinkmann, N., Connell, L., Dejaeger, E., Feys, H., Jenni, W., Kaske, C., Lesaffre, E., Leys, M., Lincoln, N., Louckx, F., Schupp, W., Smith, B. and W. De Weerd, Motor and functional recovery after stroke. A comparison of four European rehabilitation centers, *Stroke*, 38, 2101-2107, 2007. R07202.
- [180] Fieuws, S., Verbeke, G., Maes, B. and Y. Van Renterghem, Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, to appear, 2007. R07090.
- [181] Fieuws, S., Verbeke G. and G. Molenberghs, Random-effects models for multivariate repeated measures. *Statistical Methods in Medical Research*, 16, 387-397, 2007. R07015.
- [182] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Advances in Longitudinal Data Analysis: A Handbook*, London: Chapman Hall/CRC, 2007. B07007.

- [183] Garcia-Zattera, A., Jara, A. and E. Lesaffre, Conditional independence of multivariate binary data with an application in caries research. *Computational Statistics and Data Analysis*, 51, 3223-3234, 2007. R07065.
- [184] Ghidey, W., Lesaffre, E. and T. Stijnen, Semi-parametric modeling of the distribution of the baseline risk in meta-analysis, *Statistics in Medicine*, 26, 5434-5444, 2007. R07204.
- [185] Ghidey, W., Lesaffre, E. and G. Verbeke, A comparison of methods for determining the random effects distribution of a linear mixed model, *Statistical Methods in Medical Research*, to appear, 2008. R08048.
- [186] Jara, A., Applied Bayesian non- and semi-parametric inference using DPpackage. *Rnews*, 7, 17-26, 2007. R07201.
- [187] Jara, A., Garcia-Zattera, M. and E. Lesaffre, Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics and Data Analysis*, 51, 5402-5415, 2007. R07063.
- [188] Komárek, A. and E. Lesaffre, Bayesian accelerated failure time model for correlated interval-censored data with a normal mixture as error distribution. *Statistics Sinica*, 17, 549-569, 2007. R07166.
- [189] Komárek, A. and E. Lesaffre, Generalized linear mixed model with a penalized Gaussian mixture as a random-effects distribution, *Computational Statistics and Data Analysis*, 52, 3441-3458, 2008. R08024.
- [190] Komárek, A. and E. Lesaffre, Bayesian accelerated failure time model with multivariate doubly-interval-censored data and flexible distributional assumptions, *Journal of the American Statistical Association*, to appear, 2008. R08025.
- [191] Komárek, A., Lesaffre, E. and C. Legrand, Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*, 26, 5457-5472, 2007. R07078.
- [192] Küchenhoff, L.H., Lederer, W. and E. Lesaffre, Asymptotic variance estimation for the missclassification SIMEX. *Computational Statistics and Data Analysis*, 51, 6197-6211, 2007. R07027.
- [193] Leroy, R., Cecere, S., Declerck, D. and E. Lesaffre, Variability in permanent tooth emergence sequences in Flemish children, *European Journal of Oral Sciences*, 116, 11-17, 2007. R07203.
- [194] Lesaffre, E., Küchenhoff, H., Mwalili, S. and D. Declerck, On the estimation of the misclassification table for finite count data with an application in caries research, *Statistical Modelling*, to appear, 2008. R08039.
- [195] Lesaffre, E., Rizopoulos, D. and S. Tsonaka, The logistic-transform for bounded outcome scores. *Biostatistics*, 8, 72-85, 2007. R07066.

- [196] Menten, J., Boelaert, M. and E. Lesaffre, Bayesian latent class models with conditionally dependent diagnostic tests: a case study, *Statistics in Medicine*, to appear, 2008. R08041.
- [197] Molenberghs, G., Beunckens, C., Thijs, H., Jansen, I., Verbeke, G., Kenward, M.G. and K. Van Steen, Analysis of incomplete data, In *SAS System for Clinical Trials II*, Dmitrienko, A., Chueng-Stein, C. and R. D'Agostino, (eds.), Cary: SAS Press, pp. 313-360, 2007. R07206.
- [198] Molenberghs, G. and E. Lesaffre, *Missing Data*. Wiley Encyclopedia of Clinical Trials, 2007. B07005.
- [199] Molenberghs, G. and G. Verbeke, Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, 61, 1-6, 2007. R07014.
- [200] Molenberghs, G. and G. Verbeke, Discussion of Diggle, P., Farewell, D. and Henderson, R.: "Analysis of longitudinal data with drop-out: objectives, assumptions, and a proposal". *Journal of the Royal Statistical Society, Series B*, 56, 542, 2007. R07107.
- [201] Molenberghs, G., Verbeke, G. and C. Demetrio, An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, to appear, 2007. R07096.
- [202] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensivity analysis, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.), London: Chapman Hall/CRC, 2007. R07210.
- [203] Mwalili, S., Lesaffre, E. and D. Declerck, The zero-inflated negative binomial regression model with correction for misclassification: an example in caries research, *Statistical Methods in Medical Research*, to appear, 2008. R08033.
- [204] Rizopoulos, D., Verbeke, G., Lesaffre, E. and Vanrenterghem, A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros, *Biometrics*, to appear, 2008. R08035.
- [205] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification, *Biometrika*, 95, 63-74, 2008. R08019.
- [206] Schacht, A., Bogaerts, K., Bluhmki, E. and E. Lesaffre, A new non-parametric approach for baseline covariate adjustment for two-group comparative studies, *Biometrics*, to appear, 2008. R08038.
- [207] Tibaldi, F., Verbeke, G., Molenberghs, G., Renard, D., Van den Noortgate, W. and P. De Boeck, Conditional mixed models with crossed-random effects. *British Journal of Mathematical and Statistical Psychology*, 60, 351-365, 2007. R07020.
- [208] Todem, D., Kim, K.M. and E. Lesaffre, Latent-variable models for longitudinal data with bivariate ordinal outcomes. *Statistics in Medicine*, 26, 1034-1054, 2007. R07064.

- [209] Tsonaka, R., Verbeke, G. and E. Lesaffre, Shared parameter models with a flexible random effects distribution to study sensitivity in non-ignorable missingness models, *Biometrics*, to appear, 2008. R08040.
- [210] Vanobbergen, J., Lesaffre, E., Garcia-Zattera, M.J., Jara, A., Martens, L. and D. Declerck, Caries patterns in primary dentition in 3-, 5-, and 7-year-old children: spatial correlation and preventive consequences. *Caries Research*, 41, 16-25, 2007. R07067.
- [211] Verbeke, G. and S. Fieuws, The effect of misspecified baseline characteristics on inference for longitudinal trends in linear models, *Biostatistics*, 8, 772-783, 2007. R07070.
- [212] Verbeke, G. and E. Lesaffre, *Repeated Measurements*. Wiley Encyclopedia of Clinical Trials, 2007. B07006.
- [213] Verbeke, G. and G. Molenberghs, What can go wrong with the score test? *The American Statistician*, 61, 289-290, 2007. R07001.
- [214] Verbeke, G. , Molenberghs G. and C. Beunckens, Formal and informal model selection with incomplete data, *Statistical Science*, to appear, 2007. R07182.

### **3.4 Universiteit Gent, UG**

#### **3.4.1 List of Technical Reports**

- [215] Goetgeluk, S., Vansteelandt, S. and E. Goetghebeur, Estimation of controlled direct effects, 2008. TR08018.
- [216] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015.
- [217] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples, 2007. TR07053.
- [218] Vansteelandt, S., Babanezhad, M. and E. Goetghebeur, Correcting instrumental variables estimators for systematic measurement error, 2007. TR07084.
- [219] Vansteelandt, S., Mertens, K., Suetens, C. and E. Goetghebeur, Marginal structural models for partial exposure regimes, 2008. TR08024.

#### **3.4.2 List of Publications**

- [220] Alqallaf, F., Van Aelst, S., Yohai, V.J. and R.H. Zamar, Propagation of outliers in multivariate data. *Annals of Statistics*, to appear, 2007. R07165.
- [221] Burez, J. and D. Van Den Poel, Separating financial from commercial customer churn: a modeling step towards resolving the conflict between the sales and credit department. *Expert Systems with Applications*, to appear, 2007. R07168.



- [222] Cortiñas Abrahantes, J., Burzykowski, T., Legrand, C., Janssen, P., Ducrocq, V. and L. Duchateau, Comparison of different estimation procedures for proportional hazards model with random effects, *Computational Statistics and Data Analysis*, 51, 3913-3930, 2007. R07208.
- [223] Croux, C., Gallopoulos, E., Van Aelst, S. and H. Zha, Editorial-Special Issue on Machine Learning and Robust Data Mining, *Computational Statistics and Data Analysis*, 52, 151-154, 2007. R07200.
- [224] Deschepper, E., Thas, O. and J.P. Ottoy, Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models, *Biometrics*, to appear, 2008. R08034.
- [225] De Wolf, J., Duchateau, L. and E. Schrevens, Dealing with sources of variability in the data-analysis of phenotyping experiments with transgenic rice. *Euphytica*, to appear, 2007. R07169.
- [226] Duchateau, L. and P. Janssen, *The Frailty Model*, 318 p., Springer, New York, 2008. B08001.
- [227] Goetgeluk, S. and S. Vansteelandt, Conditional generalized estimating equations for the analysis of clustered and longitudinal data, *Biometrics*, to appear, 2008. R08021.
- [228] Goetghebeur, E., *Methodological challenges in HIV prevention trials*, 290 p., The National Academies Press, Washington DC, 2008. B08002.
- [229] Hubert, M., Rousseeuw, P. J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, to appear, 2007. R07029.
- [230] Khan, J.A., Van Aelst, S. and R.H. Zamar, Building a robust linear model with forward selection and stepwise procedures. *Computational Statistics and Data Analysis*, 52, 239-248, 2007. R07119.
- [231] Khan, J.A., Van Aelst, S. and R.H. Zamar, Robust linear model selection based on least angle regression. *Journal of the American Statistical Association*, 102, 1289-1299, 2007. R07118.
- [232] Mehrzad, J., Dosogne, H., De Spiegeleer, B., Duchateau, L. and C. Burvenich, Bovine blood neutrophil acyloxyacyl hydrolase (AOAH) activity during endotoxin and coliform mastitis. *Veterinary Research*, 38, 655-668, 2007. R07199.
- [233] Moerkerke, B. and E. Goetghebeur, Optimal screening for promising genes in two-stage designs, *Biostatistics*, to appear, 2008. R08022.
- [234] Rasschaert, K., Verdonck, F., Goddeeris, B.M., Duchateau, L. and E. Cox, Screening of pigs resistant to F4 enterotoxigenic escherichia coli (ETEC) infection. *Veterinary Microbiology*, 123, 249-253, 2007. R07031.
- [235] Rayner, J.C.W., Thas, O. and B. De Boeck, A generalised Emmerson recurrence relation, *Australian and New Zealand Journal of Statistics*, to appear, 2008. R08037.

- [236] Roelant, E. and S. Van Aelst, An L1-type estimator of multivariate location and shape. *Statistical Methods and Applications*, 15, 381-393, 2007. R07121.
- [237] Rousseeuw, P.J. and S. Van Aelst, A survey of robust statistics - Comment. *Statistical Methods and Applications*, 15, 290-292, 2007. R07120.
- [238] Salibian-Barrera, M., Van Aelst, S. and G. Willems, Fast and robust bootstrap. *Statistical Methods and Applications*, to appear, 2007. R07167.
- [239] Tian, L., Cai, T.X., Goetghebeur, E. and L.J. Wei, Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika*, 94, 297-311, 2007. R07117.
- [240] Van Hoek, I., Vandermeulen, E., Duchateau, L., Lefebvre, H.P., Croubles, S., Peremans, K., Polis, I. and S. Daminet, Comparison and reproducibility of plasma clearance of exogenous creatinine, exo-iohexol, endo-iohexol, and Cr-51-EDTA in young adult and aged healthy cats. *Journal of Veterinary Internal Medicine*, 21, 950-958, 2007. R07038.
- [241] Vansteelandt, S., DeMeo, D., Su, J., Smoller, J., Murphy, A.J., McQueen, M., Celedon, J., Weiss, S.T., Silverman, E.K. and C. Lange, Testing and estimating gene-environment interactions in family-based association studies, *Biometrics*, to appear, 2008. R08020.

## 3.5 Universiteit Hasselt, UH

### 3.5.1 List of Technical Reports

- [242] Anderluh, J. and T. Kadankova, Double-sided knock-in calls in an exponential compound Poisson framework, 2008. TR08007.
- [243] Braekers, R. and P.D. Markel, Zero-inflated Cox regression with a detection limit, 2007. TR07034.
- [244] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric regression, 2007. TR07032.
- [245] Gaddah, A. and R. Braekers, Weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring, 2007. TR07033.
- [246] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates, 2007. TR07072.
- [247] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015.
- [248] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples, 2007. TR07053.
- [249] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008. TR08023.

- [250] Molenberghs, G., Beunckens, C., Sotto, C. and M.G. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit, 2007. TR07083.
- [251] Veraverbeke N., Weak convergence of the empirical copula process with covariates, 2007. TR07003.

### 3.5.2 List of Publications

- [252] Aelvoet, W., Windey, F., Molenberghs, G., Temmerman, M., Van Reeppts, P. and J.-M. Foidart, Screening for inter-hospital differences in cesarean section rates in low-risk deliveries using administrative data. Can it contribute to quality of care initiatives? *Biomedical Central Health Services Research*, to appear, 2007. R07181.
- [253] Alonso, A. and G. Molenberghs, Surrogate marker evaluation from an information theory perspective. *Biometrics*, 63, 180-186, 2007. R07018.
- [254] Alonso, A. and G. Molenberghs, Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective. *Statistical Methods in Medical Research*, 63, 180-186, 2007. R07021.
- [255] Alonso, A., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H., Shkedy, Z., Tibaldi, F., Cortiñas Abrahantes, J. and M. Buyse, Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints using R. *Clinical Trials*, to appear, 2007. R07180.
- [256] Alonso, A., Molenberghs, G., Burzykowski, T., Renard, D., Geys, H., Shkedy, Z., Tibaldi, F., Cortiñas Abrahantes, J. and M. Buyse, Response to Qu and Case's letter about Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics*, 63, 1-2, 2007. R07179.
- [257] Assam, P., Tilahun, A., Alonso, A. and G. Molenberghs, Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints. *Clinical Trials*, to appear, 2007. R07101.
- [258] Auget, J.-L., Balakrishnan, N., Mesbah, M. and G. Molenberghs. *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis*, New-York, Birkhäuser, 2007. B07003.
- [259] Beunckens, C., Molenberghs, G., Thijs, H. and G. Verbeke, Incomplete hierarchical data. *Statistical Methods in Medical Research*, 16, 1-36, 2007. R07016.
- [260] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, to appear, 2007. R07006.
- [261] Beunckens, C., Sotto, C. and G. Molenberghs, A simulation study comparing weighted estimating equations with multiple imputation based estimating equations for longitudinal binary data, *Computational Statistics and Data Analysis*, to appear, 2007. R07005.

- [262] Bilau, M., Matthys, C., Baeyens, W., Bruckers, L., De Backer, G., Den Hond, E., Keune, H., Koppen, G., Nelen, V., Schoeters, G., Van Larebeke, N., Willems, J.L. and S. De Henauw, Dietary exposure to dioxin-like compounds in three age groups: results from the Flemish environment and health study, *Chemosphere*, 70, 584-592, 2008. R08047.
- [263] Bollaerts, K., Aerts, M., Faes, C., Grijspreedt, K. Dewulf, J. and K. Mintiens, Human salmonellosis: estimation of dose-illness from outbreak data, *Risk Analysis*, to appear, 2008. R08046.
- [264] Bollaerts, K., Aerts, M., Ribbens, S., Van Der Stede, Y., Boone, I. and K. Mintiens, Identification of salmonella high risk pig-herds in Belgium using semi-parametric quantile regression, *Journal of the Royal Statistical Society, Series A*, 171, 1-16, 2008. R08045.
- [265] Braekers R. and N. Veraverbeke, A conditional Koziol-Green model under dependent censoring, *Statistics and Probability Letters*, to appear, 2007. R07211.
- [266] Burzykowski, T., Buyse, M., Piccart-Gebhart, M.J., Sledge, G., Carmichael, J., Luck, H.-J., Mackey, J.R., Nabholz, J.-M., Paridaens, R., Biganzoli, L., Jassem, J., Bontenbal, M., Bonnetterre, J., Chan, S., Atalay Basaran, G. and P. Therasse, Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate endpoints in metastatic breast cancer, *Journal of Clinical Oncology*, to appear, 2008. R08028.
- [267] Burzykowski, T., Buyse, M., Yothers, G., Sakamoto, J. and D. Sargent, Exploring and validating surrogate endpoints in colorectal cancer, *Lifetime Data Analysis*, 14, 54-64, 2008. R08026.
- [268] Buyse, M., Burzykowski, T., Carroll, K., Michiels, S., Sargent, D., Miller, L.L., Elfring, G.L., Pignon, J.-P. and P. Piedbois, Progression-free survival is a surrogate for survival in advanced colorectal cancer. *Journal of Clinical Oncology*, 25, 5218-5224, 2007. R07161.
- [269] Claeskens, G., Nguti, R. and P. Janssen, Likelihood ratio and score tests for a shared frailty model: a non-standard problem, *TEST*, to appear, 2008. R08023.
- [270] Collette, T., Buyse, M. and T. Burzykowski, Are prostate-specific antigen changes valid surrogates for survival in hormone refractory prostate cancer ? A meta-analysis is needed ! *Journal of Clinical Oncology*, 25, 5673-5674, 2007. R07163.
- [271] Cortiñas Abrahantes, J., Aerts, M., van Everbroeck, B., Saegerlan, C., Berkvens, D., Geys, H., Mintiens, K., Roels, S. and P. Cras, Classification of sporadic Creutzfeldt-Jacob disease based on clinical and neuropathological characteristics. *European Journal of Epidemiology*, 22, 457-465, 2007. R07183.
- [272] Cortiñas Abrahantes, J., Burzykowski, T., Legrand, C., Janssen, P., Ducrocq, V. and L. Duchateau, Comparison of different estimation procedures for proportional hazards model with random effects, *Computational Statistics and Data Analysis*, 51, 3913-3930, 2007. R07208.

- [273] Decin, L., Hony, S., de Koter, A., Molenberghs, G., Dehaes, S. and F. Markwick-Kemper, The variable mass loss of the AGB star WX Psc as traced by the CO  $J = 1 - 0$  through 7 - 6 lines and the dust emission. *Astronomy and Astrophysics*, 475, 233-242, 2007. R07104.
- [274] de Gramont, A., Buyse, M., Cortiñas Abrahantes, J., Burzykowski, T., Quinaux, E., Cervantes, A., Figer, A., Lledo, G., Flesh, M., Mineur, L., Carola, E., Etienne, P.L., Ribera, F., Chirivella, I., Perez-Staub, N., Louvet, C., Andre, T., Tabah-Fisch, I. and C. Tournigand, Reintroduction of oxaliplatin is associated with improved survival in advanced colorectal cancer. *Journal of Clinical Oncology*, 25, 3224-3229, 2007. R07159.
- [275] De Jonge, E., Yigit, R., Molenberghs, G., Straetmans, D. and W. Ombelet, Predictors of oligoamenorrhea at 1-year follow-up in premenopausal women using a levonorgestrel-releasing intrauterine system. *Contraception*, 76, 91-95, 2007. R07091.
- [276] Duchateau, L. and P. Janssen, *The Frailty Model*, 318 p., Springer, New York, 2008. B08001.
- [277] Faes, C., Aerts, M., Geys, H. and G. Molenberghs, Model averaging using fractional polynomials to estimate a safe level of exposure. *Risk Analysis*, 27, 111-123, 2007. R07003.
- [278] Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suarez, C. Acuña, C. and M. Cano, A flexible method to measure synchrony in neuronal firing, *Journal of the American Statistical Association, Case Studies*, to appear, 2008. R08043.
- [279] Fieuws, S., Verbeke G. and G. Molenberghs, Random-effects models for multivariate repeated measures. *Statistical Methods in Medical Research*, 16, 387-397, 2007. R07015.
- [280] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Advances in Longitudinal Data Analysis: A Handbook*, London: Chapman Hall/CRC, 2007. B07007.
- [281] Fitzmaurice, G. and G. Molenberghs, Historical perspective, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, (eds.), London: Chapman Hall/CRC, 2007. R07207.
- [282] Geys, H., Faes, C. and P. Catalano, Joint models for continuous and discrete longitudinal data, In *Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.), Chapman Hall/CRC, 2007. R07217.
- [283] Gyselaers, W., Molenberghs, G. and W. Ombelet, Interlabor venous flow is different between left and right kidney in uncomplicated third trimester pregnancy. *Gynecologic and Obstetric Investigation*, to appear, 2007. R07007.
- [284] Haldermans, P., Shkedy, Z., Van Sanden, S., Burzykowski, T. and M. Aerts, Using linear mixed models for normalization of cDNA microarrays. *Statistical Applications in Genetics and Molecular Biology*, 6, 19, 2007. R07158.
- [285] Hens, N., Aerts, M., Shkedy, Z., Kimani, P.K., Kojouharova. M., Van Damme, P. and P. Beutels, Modeling age-time dependent incidence rates of hepatitis B and estimating the prevalence and force of infection using generalized additive models. *Epidemiology and Infection*, to appear, 2007. R07061.

- [286] Hens, N., Aerts, M., Shkedy, Z., Theeten, H., Van Damme, P. and Ph. Beutels, Modelling multi-sera data: the estimation of new joint and conditional epidemiological parameters, *Statistics in Medicine*, to appear, 2007. R07213.
- [287] Hens, N., Faes, C., Aerts, M., Shkedy, Z., Mintiens, K., Laevens, H. and F. Boelaert, Handling missing data when modeling the force of infection: a population averaged approach. *Journal of Agricultural, Biological, Environmental Statistics*, 12, 498-513, 2007. R07059.
- [288] Hens, N., Gay, N., Shkedy, Z., Aerts, M. and J. Edmunds, A comparison of heterogeneity in the acquisition of varicella zoster virus and parvovirus B19 for five different European countries. *European Journal of Public Health*, 17, 58-59, 2007. R07186.
- [289] Huan-I. Chan, Wu, A.W., Finnern, H.W., Thijs, H., Gath, J.C. and D.L. Fairclough, Handling missing quality of life data in HIV clinical trials - what is practical ? *Quality of Life Research*, to appear, 2007. R07193.
- [290] Jansen, I. and G. Molenberghs, A flexible marginal modeling strategy for non-monotone missing data. *Journal of the Royal Statistical Society, Series A*, 171, 1-27, 2007. R07094.
- [291] Janssen, P., Swanepoel, J. and N. Veraverbeke, Modifying the kernel distribution estimator towards reduced bias. *Statistics*, 41, 93-103, 2007. R07077.
- [292] Kadankova, T. and N. Veraverbeke, On several two-boundary problems for a particular class of Lévy processes. *Journal of Theoretical Probability*, 20, 1073-1085, 2007. R07164.
- [293] Kellen, E., Zeegers, M. P., Dirx, M., Houterman, S., Droste, J., Lawrence, G., Truyers, C., Bruckers, L., Molenberghs, G. and F. Buntinx, Occurrence of both bladder and prostate cancer in five cancer registries in Belgium, The Netherlands and the United Kingdom. *European Journal of Cancer*, to appear, 2007. R07004.
- [294] Keune, H., Loots, I., Bruckers, L., Bilau, M., Koppen, G., Van Larebeke, N., Schoeters, G. and V. Nelen, Monitoring environment, health and perception: an experimental survey on health and environment in Flanders, Belgium. *Risk Perception and Social Trust - IJENVI Special Issue*, to appear, 2007. R07189.
- [295] Laenen, A., Alonso, A. and G. Molenberghs, A measure for the reliability of a rating scale based on longitudinal clinical trial data. *Psychometrika*, to appear, 2007. R07011.
- [296] Lassere, M., Johnson, K., Boers, M., Tugwell, P., Brooks, P., Simon, L., Strand, V., Conaghan, P., Ostergaard, M., Maksymowych, W.P., Landewe, R., Bresnihan, B., Tak, P.P., Wakefield, R., Mease, P., Bingham, C.O.III, Hughes, M., Altman, D., Buyse, M., Galbraith, S. and G. Wells, Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. *Journal of Rheumatology*, 34, 607-615, 2007. R07196.
- [297] Lassere, M., Johnson, K., Hughes, M., Altman, D., Buyse, M., Galbraith, S. and G. Wells, Simulation studies of surrogate endpoint validation using single trial and multitrial statistical approaches. *Journal of Rheumatology*, 34, 616-619, 2007. R07197.

- [298] Lin, D., Shkedy, Z., Burzkowski, T., Aerts, M., Göhlmann, H.W.H., De Bondt, A., Perera, T., Geerts, T. and L. Bijmens, Classification of trends in dose-response microarray experiments using information theory selection methods. *Journal of the American Statistical Society*, to appear, 2007. R07187.
- [299] Lin, D., Shkedy, Z., Yekutieli, D., Burzykowski, T., Göhlmann, H.W.K., De Bondt, A., Perera, T., Geerts, T. and L. Bijmens, Testing for trend in dose-response microarray experiments: comparison of several testing procedures, multiplicity and resampling-based inference. *Statistical Applications in Genetics and Molecular Biology*, 6, 6, 2007. R07162.
- [300] Litière, S., Alonso, A. and G. Molenberghs, Type I and Type II error random-effects misspecification in generalized linear mixed models. *Biometrics*, 63, 1038-1044, 2007. R07012.
- [301] Litière, S., Alonso, A., Molenberghs, G. and H. Geys, The impact of misspecified random-effects distribution on the estimation and performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, to appear, 2007. R07103.
- [302] Lousbergh, D., Claes, E., Op De Beeck, L., Rummens, J.L., Vanden Brande, J., Faes, C., Bruckers, L., Molenberghs, G., Dhollander, Kellen, E., Hensen, K., Lathouwers, D., Meekers, E., and F. Buntinx, Ten years of cancer in the Belgian province of Limburg: 1996-2005, Hasselt, LIKAS, 2007. B07008.
- [303] Maringwa, J. T., Faes, C., Geys, H., Aerts, M., Teuns, G. and L. Bijmens, On the use of historical control studies in pre-clinical safety studies. *Journal of Biopharmaceutical Statistics*, 17, 493-509, 2007. R07051.
- [304] Maringwa, J., Geys, H., Shkedy, Z., Faes, C., Molenberghs, G., Aerts, M., Van Ammel, K., Teisman, A. and L. Bijmens, Application of semiparametric mixed-models and simultaneous confidence bands in a cardiovascular safety experiment with longitudinal data. *Journal of Biopharmaceutical Statistics*, to appear, 2008. R08049.
- [305] Massonnet, G., Janssen, P. and T. Burzykowski, Fitting conditional survival models to meta-analytic data by using a transformation towards mixed-effects models, *Biometrics*, to appear, 2008. R08029.
- [306] Meroc, E., De Clercq, K., Mintiens, K., Faes, C., Cortiñas, J., Aerts, M., Ducheyne, E., Hendrickx, G. and P. Mellor, Role of human interventions, In *EFSA Report on Epidemiology Analysis of the 2006 Bluetongue Virus Serotype 8 Epidemic in North-Western Europe*, 2007. R07216.
- [307] Molenberghs, G., Editorial: What to do with missing data ? *Journal of the Royal Statistical Society, Series A*, 170, 861-863, 2007. R07002.
- [308] Molenberghs, G., Norman Breslow's leadership for the international biometric society. *Life-time Data Analysis*, to appear, 2007. R07097.

- [309] Molenberghs, G., Beunckens, C., Sotto, C. and M. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, 70, 1-18, 2007. R07098.
- [310] Molenberghs, G., Beunckens, C., Thijs, H., Jansen, I., Verbeke, G., Kenward, M.G. and K. Van Steen, Analysis of incomplete data, In *SAS System for Clinical Trials II*, Dmitrienko, A., Chueng-Stein, C. and R. D'Agostino, (eds.), Cary: SAS Press, pp. 313-360, 2007. R07206.
- [311] Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A. and M. Buyse, The meta-analytic framework for the evaluation of surrogate endpoints in clinical trials, *Journal of Statistical Planning and Inference*, 138, 432-449, 2008. R08044.
- [312] Molenberghs, G. and G. Fitzmaurice, Incomplete data: Introduction and overview, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.), London: Chapman Hall/CRC, 2007. R07209.
- [313] Molenberghs G. and M.G. Kenward, *Missing Data in Clinical Studies*. John Wiley, New York, 2007. B07004.
- [314] Molenberghs, G. and E. Lesaffre, *Missing Data*. Wiley Encyclopedia of Clinical Trials, 2007. B07005.
- [315] Molenberghs, G., Vangeneugden, T. and A. Laenen, Estimating reliability and generalizability from hierarchical biomedical data. *Journal of Biopharmaceutical Statistics*, 17, 1-33, 2007. R07019.
- [316] Molenberghs, G. and G. Verbeke, Discussion of Diggle, P., Farewell, D. and Henderson, R.: "Analysis of longitudinal data with drop-out: objectives, assumptions, and a proposal". *Journal of the Royal Statistical Society, Series B*, 56, 542, 2007. R07107.
- [317] Molenberghs, G. and G. Verbeke, Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, 61, 1-6, 2007. R07014.
- [318] Molenberghs, G., Verbeke, G. and C. Demetrio, An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, to appear, 2007. R07096.
- [319] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensivity analysis, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.), London: Chapman Hall/CRC, 2007. R07210.
- [320] Moons, E., Wets, G. and M. Aerts, Nonlinear models for determining mode choice. *Lecture Notes in Computer Science*, Springer, Berlin/Heidelberg, 183-194, 2007. R07087.
- [321] Mossong, J., Hens, N., Friederichs, V., Davidkin, I., Broman, M., Litwinska, B., Siennicka, J., Trzcinska, A., Van Damme, P., Beutels, P., Vyse, A., Shkedy, Z., Aerts, M., Gabutti, G. and S. Salmaso, Parvovirus B19 infection in 5 European countries: sero-epidemiology, force of infection and maternal risk of infection. *Epidemiology and Infection*, to appear, 2007. R07185.



- [322] Namata, H., Meroc, R., Aerts, M., Faes, C., Cortiñas Abrahantes, J., Imbrechts, H., and K. Mintiens, Salmonella in Belgian laying hens: an identification of risk factors. *Preventive Veterinary Medicine*, to appear, 2007. R07194.
- [323] Namata, H., Shkedy, Z., Faes, C., Molenberghs, G., Theeten, H., Van Damme, P. and P. Beutels, Estimation of the force of infection from current status data using generalized linear mixed models. *Journal of Applied Statistics*, 34, 923-939, 2007. R07105.
- [324] Parzen, M., Lipsitz, R., Fitzmaurice, G., Ibrahim, J.G., Troxel, A., and G. Molenberghs, Pseudo-likelihood methods for the analysis of longitudinal binary data subject to nonignorable non-monotone missingness, *Journal of Data Science*, 5, 1-21, 2007. R07092.
- [325] Piccart-Gebhart, M.J., Burzykowski, T., Buyse, M. Sledge, G., Carmichael, J., Luck, H.-J., Mackey, J.R., Nabholz, J.-M., Paridaens, R., Biganzoli, L., Jassem, J., Bontenbal, M., Bonnetterre, J., Chan, S., Atalay Basaran, G. and P. Therasse, Effects of taxanes alone or in combination with anthracyclines as first-line therapy of patients with metastatic breast cancer, *Journal of Clinical Oncology*, to appear, 2008. R08027.
- [326] Pryseley, A., Tilahun, A. , Alonso, A. and G. Molenberghs, Information-theory based surrogate marker evaluation from several randomized clinical trials with continuous true and binary surrogate endpoints, using R, *Clinical Trials*, to appear, 2007. R07212.
- [327] Rafter, J., Bennett, M., Caderni, G., Clune, Y., Hughes, R., Karlsson, P.C., Klinder, A., O’Riordan, M., O’Sullivan, G.C., Pool-Zobel, B., Rechkemmer, G., Roller, M., Rowland, I., Salvadori, M., Thijs, H., Van Loo, J., Watzl, B. and J.K. Collins, Dietary synbiotic reduces cancer risk factors in polypectomised and colon cancer patients. *The American Journal of Clinical Nutrition*, 85, 488-496, 2007. R07192.
- [328] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification, *Biometrika*, 95, 63-74, 2008. R08019.
- [329] Robaey, G., De Bie, J., Wichers, M.C., Bruckers, L., Nevens, F., Michielsen, P., Van Ranst, M. and F. Buntinx, Early prediction of major depression in chronic hepatitis C patients during peg-interferon alpha-2b treatment by assessment of vegetative-depressive symptoms after four weeks. *World Journal of Gastroenterology*, 13, 5736-5740, 2007. R07190.
- [330] Sariachvili, M., Droste, J., Dom, S., Wieringa, M., Vellinga, A., Hagendorens, M., Bridts, C., Stevens, W., Sprundel, M.V., Desager, K. and J. Weyler, Is breast feeding a risk factor for eczema during the first year of life ? *Pediatric Allergy Immunology*, 18, 410-417, 2007. R07088.
- [331] Shkedy, Z., Decin, L., Molenberghs, G. and C. Aerts, Estimating stellar parameters from spectra using a hierarchical Bayesian approach. *Monthly Notices to the Royal Astronomical Society*, 377, 120-132, 2007. R07009.
- [332] Smeets, K., Ruytinx, J., Van Belleghem, F., Semane, B., Lin, D., Vangronsveld, J. and A. Cuypers, Critical evaluation and statistical validation of a hydroponic culture system for *Arabidopsis thaliana*. *Plant Physiology and Biochemistry*, to appear, 2007. R07188.

- [333] Specenier, P.M., Van Den Weyngaert, D., Van Laer, C., Weyler, J., Van Den Brande, J., Huizing, M.T., Dyck, J., Schrijvers, D. and J.B. Vermorken, Phase II feasibility study of concurrent radiotherapy and gemcitabine in chemo-naïve patients with squamous cell carcinoma of the head and neck: long-term follow-up data. *Annals of Oncology*, 18, 1856-1860, 2007. R07068.
- [334] Speybroeck, N., Marcotty, T., Aerts, M., Dolan, T., Williams, B., Lauer, J., Molenberghs, G., Burzykowski, T., Mulumba, M. and D. Berkvens, Titrating *Theileria parva*: single stocks against combination of stocks, *Experimental Parasitology*, to appear, 2008. R08030.
- [335] Sterna, J. and T. Burzykowski, The assessment of the usefulness of hemilaminectomy without fenestration in the treatment of thoracolumbar disc disease in chondrodystrophic dogs. *Polish Journal of Veterinary Sciences*, 10, 165-172, 2007. R07195.
- [336] Theeten, H., Hens, N., Vandermeulen, C., Depoorter, A.M., Roelants, M., Aerts, M., Hoppenbrouwers, K. and P. Van Damme, Infant vaccination coverage in 2005 and predictive factor for incomplete or invalid vaccination in Flanders, Belgium: an EPI survey. *Vaccine*, 25, 4940-4948, 2007. R07060.
- [337] Thijssen, S., Cuypers, A., Maringwa, J. T., Smeets, K., Horemans, N., Lambrichts, I. and E. Van Kerkhove, Low cadmium exposure triggers a biphasic oxidative stress response in mice kidneys. *Toxicology*, 236, 29-41, 2007. R07050.
- [338] Thijssen, S., Lambrichts, I., Maringwa J. T. and E. Van Kerkhove, Changes in expression of fibrotic markers and histopathological alterations in kidneys of mice chronically exposed to low and high Cd doses. *Toxicology*, 238, 200-210, 2007. R07049.
- [339] Tibaldi, F., Renard, D. and G. Molenberghs, Accounting for variability in individual hierarchical clinical trial data. *Pharmaceutical Statistics*, to appear, 2007. R07099.
- [340] Tibaldi, F., Verbeke, G., Molenberghs, G., Renard, D., Van den Noortgate, W. and P. De Boeck, Conditional mixed models with crossed-random effects. *British Journal of Mathematical and Statistical Psychology*, 60, 351-365, 2007. R07020.
- [341] Tilahun, A., Assam, P., Alonso, A., and G. Molenberghs, Flexible surrogate marker evaluation from several randomized clinical trials with continuous endpoints, using R and SAS. *Computational Statistics and Data Analysis*, 51, 4152-4163, 2007. R07008.
- [342] Tilahun, A., Assam, P., Alonso, A. and G. Molenberghs, Information-theory based surrogate marker evaluation from several randomized clinical trials with binary endpoints, using SAS. *Journal of Biopharmaceutical Statistics*, to appear, 2007. R07093.
- [343] Toussaint, J.F., Assam, P., Caij, B., Dekeyser, F., Knapen, K., Imberechts, H., Goris, N., Molenberghs, G., Mintiens, K. and K. De Clercq, Uncertainty of measurement for competitive and indirect ELISAs. *Scientific and Technical Review of the OIE*, to appear, 2007. R07095.

- [344] Tremmery, S., Buitelaar, J. K., Steyaert, J., Molenberghs, G., Feron, F. J. M., Kalff, A. C., Hurks, P. P. M., Hendriksen, J. G. M., Van Zeben, T. M. C. B., Vles, J. S. H. and J. Jolles, The use of health care services and psychotropic medication in a community sample of 9 year old school ADHD children. *European Child and Adolescent Psychiatry*, to appear, 2007. R07010.
- [345] Valkenburg, D., Assam, P., Thomas, G., Krols, L., Kas, K. and T. Burzykowski, Using a Poisson approximation to predict the isotopic distribution of sulphur-containing peptides centric proteomic approach. *Rapid Communications in Mass Spectrometry*, 21, 3387-3391, 2007. R07160.
- [346] Vandeloo, M., Bruckers, L. and J. Janssens, Effects of lifestyle in the onset of puberty as determinant for breast cancer. *European Journal of Cancer Prevention*, 16, 17-25, 2007. R07191.
- [347] Van Den Bruel, A., Aertgeerts, B., Bruyninckx, R., Aerts, M. and F. Buntinx, Signs and symptoms for the diagnosis of serious infections in children: a prospective study in primary care. *British Journal of General Practics*, 57, 538-546, 2007. R07184.
- [348] Vangeneugden, T., Molenberghs, G., Laenen, A., Alonso, A. and H. Geys, Generalizability in non-Gaussian longitudinal clinical trial data based on generalized linear mixed models. *Journal of Biopharmaceutical Statistics*, to appear, 2007. R07100.
- [349] Van Goethem, M., Schelfout, K., Kersschot, E., Colpaert, C., Verslegers, I., Biltjes, I., Tjalma, W.A., De Schepper, A., Weyler, J. and P.M. Parizel, MR Mammography is useful in the preoperative locoregional staging of breast carcinomas with extensive intraductal component. *European Journal of Radiology*, 62, 273-282, 2007. R07069.
- [350] Van Sanden, S., Lin, D., and T. Burzykowski, Performance of classification methods in a microarray setting: a simulation study. *Biocybernetics and Biomedical Engineering*, 27, 15-28, 2007. R07157.
- [351] Van Sanden, S., Lin, D. and T. Burzykowski, Performance of gene selection and classification methods in microarray setting: a simulation study, *Communications in Statistics - Simulation and Computation*, 37, 409-424, 2008. R08031.
- [352] Verbeke, G. and G. Molenberghs, What can go wrong with the score test? *The American Statistician*, 61, 289-290, 2007. R07001.
- [353] Verbeke, G. , Molenberghs G. and C. Beunckens, Formal and informal model selection with incomplete data, *Statistical Science*, to appear, 2007. R07182.
- [354] Wang, L. and N. Veraverbeke, Bayes prediction based on right censored data. *Statistical Papers*, to appear, 2007. R07177.
- [355] Wang, L. and N. Veraverbeke, MSE superiority of Bayes and empirical Bayes estimators in two generalized seemingly unrelated regressions, *Statistics and Probability Letters*, 78, 109-117, 2008. R08032.

- [356] Wouters, K., Ahnaou, A., Cortiñas, J., Molenberghs, G., Geys, H., Bijmens, L. and H. I. M. Drinkenbrug, Psychotropic drug classification based on sleep-wake behavior of rats. *Applied Statistics*, 56, 223-234, 2007. R07013.
- [357] Zelefsky, J.R., Srinivasan, M., Kundu, A., Lietman, T., Whitcher, J.P., Wang, K., Buyse, M. and E.T. Cunningham, Hookworm infestation as a risk factor for Mooren’s ulcer in South India. *Ophthalmology*, 144, 450-453, 2007. R07198.

## 3.6 Université Joseph Fourier, UJF–LMC–IMAG

### 3.6.1 List of Technical Reports

- [358] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets, 2007. TR07036.
- [359] Bernard-Michel, C., Gardes, L. and S. Girard, Gaussian regularized sliced inverse regression, 2007. TR07020.
- [360] Brouste, A., Istas, J. and S. Lambert-Lacroix, On Gaussian random fields simulations, 2007. TR07009.
- [361] Chatelain, F., Lambert-Lacroix, S. and J.-Y. Tournier, Composite likelihood estimation for multivariate mixed Poisson models generated by gamma intensities, 2007. TR07014.
- [362] Letué, F., A semiparametric shock model for a pair of event-related dependent censored failure times, 2007. TR07021.

### 3.6.2 List of Publications

- [363] Abramovich, F., Antoniadis, A. and M. Pensky, Estimation of piecewise-smooth functions by amalgamated bridge regression splines, *Shankya*, 69, 1-27, 2007. R07219.
- [364] Antoniadis, A., Wavelet methods in statistics: some recent developments and their applications. *Statistics Surveys*, 1, 16-55, 2007. R07042.
- [365] Antoniadis, A., Bigot, J. and I. Gijbels, Penalized wavelet monotone regression. *Statistics and Probability Letters*, 77, 1608-1621, 2007. R07045.
- [366] Antoniadis, A., Bigot, J., Lambert-Lacroix S. and F. Letué, Nonparametric pre-processing methods and inference tools for analyzing time-of-flight mass spectrometry data, *Current Analytical Chemistry*, 3, 127-147, 2007. R07220.
- [367] Antoniadis, A., and T. Sapatinas, Estimation and inference in functional mixed-effects models, *Computational Statistics and Data Analysis*, 51, 4793-4813, 2007. R07218.
- [368] Brouste, A., Istas, J. and S. Lambert-Lacroix, On Gaussian random fields simulations, *Journal of Statistical Software*, to appear, 2007. R07221.

- [369] Gannaz, I., Robust estimation and wavelet thresholding in partial linear models, *Statistics and Computing*, 17, 293-310, 2007. R07222.
- [370] Zwald, L. and G. Blanchard, Finite dimensional projection for classification and statistical learning. *IEEE Transactions in Information Theory*, to appear, 2007. R07041.

### 3.7 Universiteit Utrecht, UU

#### 3.7.1 List of Publications

- [371] Delwarde, A., Denuit, M. and P.H.C. Eilers, Smoothing the Lee-Carter and Poisson Log-bilinear models for mortality forecasting: a penalized likelihood approach. *Statistical Modelling*, 7, 29-48, 2007. R07124.
- [372] De Zwart, I.M., Haans, J.J., Verbeek, P., Eilers, P. H., de Roos, A. and A.A. Masclee, Gastric accommodation and motility are influenced by the barostat device: assessment with Magnetic Resonance Imaging. *American Journal of Physiology in Gastrointestinal and Liver Physiology*, 292, 208-214, 2007. R07135.
- [373] Eilers P.H.C., Data exploration in meta-analysis with smooth latent distributions. *Statistics in Medicine*, 26, 3358-3368, 2007. R07123.
- [374] Eilers, P.H.C., Ill-posed problems with counts, the composite link model, and penalized likelihood. *Statistical Modelling*, 7, 263-278, 2007. R07136.
- [375] Eilers, P.H.C. and M.W. Borgdorff, Non-parametric log-concave mixtures. *Computational Statistics and Data Analysis*, 51, 5444-5451, 2007. R07126.
- [376] Haans, J.J., Paridaans, N.P., Eilers, P.H., Doornbos, J., De Roos, A. and A.A. Masclee, Effect of sildenafil on gastric function assessed with magnetic resonance imaging. *Gastroenterology*, 132, 375-375, 2007. R07133.
- [377] Heim, S., Fahrmeir, L., Eilers, P.H.C. and B.D. Marx, Space-varying coefficient models for brain imaging. *Computational Statistics and Data Analysis*, 51, 6212-6228, 2007. R07125.
- [378] Hendriks, M.M., Smit, S., Akkermans, W.L., Reijmers, T.H., Eilers, P.H., Hoefsloot, H.C., Rubingh, C.M., de Koster, C.G., Aerts, J.M. and A.K. Smilde, How to distinguish healthy from diseased ? Classification strategy for mass spectrometry-based clinical proteomics. *Proteomics*, 7, 3672-3680, 2007. R07128.
- [379] Lips, E., de Graaf, E., Tollenaar, R., van Eijk, R., Oosting, J., Szuhai, K., Karsten, T., Nanya, Y., Ogawa, S., Van De Velde, C., Eilers, P., Van Wezel, T. and H. Morreau, Single nucleotide polymorphism array analysis of chromosomal instability patterns discriminates rectal adenomas from carcinomas. *Journal of Pathology*, 212, 269-277, 2007. R07131.
- [380] Oosting, J., Lips, E.H., Van Eijk, R., Eilers, P.H., Szuhai, K., Wijmenga, C., Morreau, H. and T. Van Wezel, High-resolution copy number analysis of paraffin-embedded archival tissue using SNP BeadArrays. *Genome Research*, 17, 368-376, 2007. R07132.

- [381] Pieterse, Q.D., Kenter, G.G., Eilers, P.H. and J.B. Trimbos, An individual prediction of the future (disease-free) survival of patients with a history of early-stage cervical cancer, multistate model. *International Journal of Gynecological Cancer*, to appear, 2007. R07130.
- [382] Vandenhende, F., Eilers, P.H.C., Ledent, E., Renard, D. and F. Tibaldi, Joint detection of important biomarkers and optimal dose-response model using penalties. *Statistics in Medicine*, 26, 4876-4888, 2007. R07127.
- [383] Van Weel, V., Toes, R., Seghers, L., Deckers, M.M., de Vries, M.R., Eilers, P.H., Sipkens, J., Schepers, A., Eefting, D., Van Hinsbergh, V.W., Van Bockel, J.H. and P.H. Quax, Natural killer cells and CD4+ T-cells modulate collateral artery development. *Arteriosclerosis, Thrombosis and Vascular Biology*, 27, 2310-2318, 2007. R07129.
- [384] Zijlmans, H.J.M.M.A., Fleuren, G.J., Baelde, H.J., Eilers, P.H., Kenter, G.G. and A. Gorter, Role of tumor-derived proinflammatory cytokines GM-CSF, TNF-alpha, and IL-12 in the migration and differentiation of antigen-presenting cells in cervical carcinoma. *Cancer*, 109, 556-565, 2007. R07134.

## 3.8 Universidad de Santiago de Compostela, USC

### 3.8.1 List of Technical Reports

- [385] Boente, G., Pérez-González, A. and W. González-Manteiga, Robust nonparametric estimation with missing data, 2007. TR07054.
- [386] Cao, R. and W. González-Manteiga, Goodness-of-fit tests for conditional models under censoring and truncation, 2007. TR07055.
- [387] Crujeiras, R. M., and Fernández-Casal, On the estimation of the spectral density for continuous spatial processes, 2007. TR07056.
- [388] Febrero, M., Galeano, P. and W. González-Manteiga, Outlier detection in functional data by depth measures with application to identify abnormal NOx levels, 2007. TR07057.
- [389] Fernández-De Castro, B. and W. González-Manteiga, Boosting for real and functional samples. An application to an environmental problem, 2007. TR07058.
- [390] González-Manteiga, W. and R. M. Crujeiras, Comments on: Augmenting the bootstrap to analyze high dimensional genomic data, 2007. TR07060.
- [391] González-Manteiga, W., Lombardía, M. J., Molina, I., Morales, D. and L. Santamaría, Bootstrap estimation of the mean squared error of an EBLUP in mixed linear models for small areas, 2007. TR07059.
- [392] Lombardía, M. J. and S. Sperlich, Semiparametric inference in generalized mixed effects models, 2007. TR07061.

- [393] Martínez-Miranda, M. D., Raya-Miranda, R., González-Manteiga, W. and A. González-Carmona, A bootstrap local bandwidth selector for additive models, 2007. TR07062.
- [394] Matías, J. M., Febrero-Bande, M., González-Manteiga, W. and J. C. Reboredo, Boosting GARCH and neural networks for the prediction of heteroskedastic time series, 2007. TR07063.
- [395] Menezes, R., García-Soidán, P. and M. Febrero, A kernel variogram estimator for clustered data, 2007. TR07064.
- [396] Pateiro-López, B. and A. Rodríguez-Casal, Length and surface area estimation under smoothness restrictions, 2007. TR07065.
- [397] Pérez-González, A., Vilar-Fernández, J. M. and W. González-Manteiga, Asymptotic properties of local polynomial regression with missing data and correlated errors, 2007. TR07066.
- [398] Porcu, E., Crujeiras, R. M., Mateu, J. and W. González-Manteiga, On the second-order properties of the multidimensional periodogram for regularly spaced data, 2007. TR07067.
- [399] 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, 2007. TR07068.
- [400] Van Keilegom, I., Sánchez-Sellero, C. and W. González-Manteiga, Empirical likelihood based testing for regression, 2007. TR07023.

### 3.8.2 List of Publications

- [401] Cao, R. and W. González-Manteiga, Goodness-of-fit tests for conditional models under censoring and truncation. *Journal of Econometrics*, 143, 166-190, 2008. R08003.
- [402] Crujeiras, R. M., Fernández-Casal, R., and W. González-Manteiga, Comparing spatial dependence structures using spectral density estimators. *Environmetrics*, 18, 793-808, 2007. R07079.
- [403] Cuevas, A., Febrero, M. and R. Fraiman, Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, 22, 481-496, 2007. R07080.
- [404] Cuevas, A., Fraiman, R. and A. Rodríguez-Casal, A nonparametric approach to the estimation of lengths and surface areas. *Annals of Statistics*, 35, 1031-1051, 2007. R07081.
- [405] de Uña-Álvarez, J. and A. Rodríguez-Casal, Nonparametric estimation from length-biased data under competing risks. *Computational Statistics and Data Analysis*, 51, 2653-2669, 2007. R07082.
- [406] Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suarez, C. Acuña, C. and M. Cano, A flexible method to measure synchrony in neuronal firing, *Journal of the American Statistical Association, Case Studies*, to appear, 2008. R08043.

- [407] Febrero, M., Galeano, P. and W. González-Manteiga, A Functional Analysis of NOx levels: location and scale estimation and outlier detection. *Computational Statistics*, 22, 411-427, 2007. R07083.
- [408] González-Manteiga, W, Lombardía, M. J., Molina, I., Morales, D. and L. Santamaría, Estimation of the mean squared error of predictors of small area parameters under a logistic mixed model. *Computational Statistics and Data Analysis*, 51, 2720-2733, 2007. R07084.
- [409] González-Manteiga, W, and P. Vieu, Editorial - Introduction to the special issue on statistics for functional data. *Computational Statistics and Data Analysis*, 51, 4788-4792, 2007. R07085.
- [410] Menezes, R., García-Soidán, P. and M. Febrero, A kernel variogram estimation for clustered data. *Scandinavian Journal of Statistics*, 35, 18-37, 2008. R08004.
- [411] Molina, I., Saei, A. and M. J. Lombardía, Estimation of unemployment rates in small areas by a multinomial logit mixed model. *Journal of the Royal Statistical Society, Series A*, 170, 975-1000, 2007. R07086.
- [412] Pardo-Fernández, J. C., Van Keilegom, I., and W. González-Manteiga, Testing for the equality of  $k$  regression curves. *Statistica Sinica*, 17, 1115-1137, 2007. R07055.
- [413] Pardo-Fernández, J. C., Van Keilegom, I. and W. González-Manteiga, Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics*, 35, 249-264, 2007. R07062.
- [414] Rodríguez-Casal, A. Set estimate under convexity type assumptions. *Annales de l'I.H.P. - Probabilités et Statistiques*, 43, 763-774, 2007. R07089.
- [415] 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*, to appear, 2008. R08015.

### **3.9 London School of Hygiene and Tropical Medicine, LSHTM**

#### **3.9.1 List of Technical Reports**

- [416] Carpenter, J.R. and M.G. Kenward, Missing data in clinical trials - a practical guide, 2007. TR07082.
- [417] Molenberghs, G., Beunckens, C., Sotto, C. and M.G. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit, 2007. TR07083.

#### **3.9.2 List of Publications**

- [418] Carpenter, J.R., Kenward, M.G. and I.R. White, Sensitivity analysis after multiple imputation under missing at random - a weighting approach. *Statistical Methods in Medical Research*, 16, 259-275, 2007. R07173.



- [419] Kenward, M.G. and J.R. Carpenter, Multiple imputation: current perspectives. *Statistical Methods in Medical Research*, 16, 199-218, 2007. R07174.
- [420] Molenberghs, G., Beunckens, C., Sotito, C. and M. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, 70, 1-18, 2007. R07098.
- [421] Molenberghs, G., Beunckens, C., Thijs, H., Jansen, I., Verbeke, G., Kenward, M.G. and K. Van Steen, Analysis of incomplete data, In *SAS System for Clinical Trials II*, Dmitrienko, A., Chueng-Stein, C. and R. D'Agostino, (eds.), Cary: SAS Press, pp. 313-360, 2007. R07206.
- [422] Molenberghs, G. and M.G. Kenward. *Missing Data in Clinical Studies*. John Wiley, New York, 2007. B07004.
- [423] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensivity analysis, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.), London: Chapman Hall/CRC, 2007. R07210.
- [424] White, I.R., Carpenter, J.R., Evans, S. and S. Schroter, Eliciting and using experts opinions about non-response bias in randomized controlled trials. *Clinical Trials*, 4, 125-139, 2007. R07172.
- [425] Yun, S., Lee, Y. and M.G. Kenward, Using hierarchical likelihood for missing data problems. *Biometrika*, 94, 905-919, 2007. R07175.

### 3.10 List of joint publications

#### 3.10.1 List of Technical Reports

- [426] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets, 2007 (KUL-1, UJF). TR07036.
- [427] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric regression, 2007 (UCL, UH). TR07032.
- [428] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates, 2007 (KUL-1, UH). TR07072.
- [429] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008 (UG, UH). TR08015.
- [430] Komárek, A., Lesaffre, E. and C. Legrand, Assessing baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution, 2007 (UCL, KUL-2). TR07013.
- [431] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples, 2007 (UG, UH). TR07053.

- [432] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008 (UCL, UH). TR08023.
- [433] Molenberghs, G., Beunckens, C., Sotto, C. and M.G. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit, 2007 (UH, LSHTM). TR07083.
- [434] 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, 2007 (UCL, USC). TR07068.
- [435] Van Keilegom, I., Sánchez-Sellero, C. and W. González-Manteiga, Empirical likelihood based testing for regression, 2007 (UCL, USC). TR07023.

### 3.10.2 List of Publications

- [436] Antoniadis, A., Bigot, J. and I. Gijbels, Penalized wavelet monotone regression. *Statistics and Probability Letters*, 77, 1608-1621, 2007 (KUL-1, UJF). R07045.
- [437] Beunckens, C., Molenberghs, G., Thijs, H. and G. Verbeke, Incomplete hierarchical data. *Statistical Methods in Medical Research*, 16, 1-36, 2007 (KUL-2, UH). R07016.
- [438] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, to appear, 2007 (KUL-2, UH). R07006.
- [439] Cortiñas Abrahantes, J., Burzykowski, T., Legrand, C., Janssen, P., Ducrocq, V. and L. Duchateau, Comparison of different estimation procedures for proportional hazards model with random effects, *Computational Statistics and Data Analysis*, 51, 3913-3930, 2007 (UCL, UG, UH). R07208.
- [440] Delwarde, A., Denuit, M. and P. H. C. Eilers, Smoothing the Lee-Carter and Poisson log-bilinear models for mortality forecasting: a penalized likelihood approach. *Statistical Modelling*, 7, 29-48, 2007 (UCL, UU). R07124.
- [441] Duchateau, L. and P. Janssen, *The Frailty Model*, 318 p., Springer, New York, 2008 (UG, UH). B08001.
- [442] Faes, C., Geys, H., Molenberghs, G., Aerts, M., Cadarso-Suarez, C. Acuña, C. and M. Cano, A flexible method to measure synchrony in neuronal firing, *Journal of the American Statistical Association, Case Studies*, to appear, 2008 (UH, USC). R08043.
- [443] Fieuws, S., Verbeke G. and G. Molenberghs, Random-effects models for multivariate repeated measures. *Statistical Methods in Medical Research*, 16, 387-397, 2007 (KUL-2, UH). R07015.
- [444] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Advances in Longitudinal Data Analysis: A Handbook*, London: Chapman Hall/CRC, 2007 (KUL-2, UH). B07007.

- [445] Gao, J., Gijbels, I. and S. Van Bellegem, Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics, Special Issue on Specification Testing*, 143, 123-142, 2008 (UCL, KUL-1). R08005.
- [446] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, to appear, 2007 (KUL-1, UG). R07029.
- [447] Komárek, A., Lesaffre, E. and C. Legrand, Assessing Baseline and treatment effect heterogeneity for survival times between centers using a random effects accelerated failure time model with flexible error distribution. *Statistics in Medicine*, 26, 5457-5472, 2007 (UCL, KUL-2). R07078.
- [448] Molenberghs, G., Beunckens, C., Sotto, C. and M. Kenward, Every missing not at random model has got a missing at random counterpart with equal fit. *Journal of the Royal Statistical Society, Series B*, 70, 1-18, 2007 (UH, LSHTM). R07098.
- [449] Molenberghs, G., Beunckens, C., Thijs, H., Jansen, I., Verbeke, G., Kenward, M.G. and K. Van Steen, Analysis of incomplete data, In *SAS System for Clinical Trials II*, Dmitrienko, A., Chueng-Stein, C. and R. D'Agostino, (eds.), Cary: SAS Press, pp. 313-360, 2007 (KUL-2, UH, LSHTM). R07206.
- [450] Molenberghs, G. and M.G. Kenward. *Missing Data in Clinical Studies*. John Wiley, New York, 2007 (UH, LSHTM). B07004.
- [451] Molenberghs, G. and E. Lesaffre, *Missing Data*. Wiley Encyclopedia of Clinical Trials, 2007 (KUL-2, UH). B07005.
- [452] Molenberghs, G. and G. Verbeke, Likelihood ratio, score, and Wald tests in a constrained parameter space. *The American Statistician*, 61, 1-6, 2007 (KUL-2, UH). R07014.
- [453] Molenberghs, G. and G. Verbeke, Discussion of Diggle, P., Farewell, D. and Henderson, R.: "Analysis of longitudinal data with drop-out: objectives, assumptions, and a proposal". *Journal of the Royal Statistical Society, Series B*, 56, 542, 2007 (KUL-2, UH). R07107.
- [454] Molenberghs, G., Verbeke, G. and C. Demetrio, An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, to appear, 2007 (KUL-2, UH). R07096.
- [455] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensivity analysis, In *Advances in Longitudinal Data Analysis: A Handbook*, Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke (eds.) , London: Chapman Hall/CRC, 2007 (KUL-2, UH, LSHTM). R07210.
- [456] Pardo-Fernández, J.C., Van Keilegom, I., and W. González-Manteiga, Testing for the equality of  $k$  regression curves. *Statistica Sinica*, 17, 1115-1137, 2007 (UCL, USC). R07055.
- [457] Pardo-Fernández, J. C., Van Keilegom, I. and W. González-Manteiga, Goodness-of-fit tests for parametric models in censored regression. *Canadian Journal of Statistics* 35, 249-264, 2007 (UCL, USC). R07062.

- [458] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification, *Biometrika*, 95, 63-74, 2008 (KUL-2, UH). R08019.
- [459] Tibaldi, F., Verbeke, G., Molenberghs, G., Renard, D., Van den Noortgate, W. and P. De Boeck, Conditional mixed models with crossed-random effects. *British Journal of Mathematical and Statistical Psychology*, 60, 351-365, 2007 (KUL-2, UH). R07020.
- [460] 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*, to appear, 2008 (UCL, USC). R08015.
- [461] Verbeke, G. and G. Molenberghs, What can go wrong with the score test? *The American Statistician*, 61, 289-290, 2007 (KUL-2, UH). R07001.
- [462] Verbeke, G. , Molenberghs G. and C. Beunckens, Formal and informal model selection with incomplete data, *Statistical Science*, to appear, 2007 (KUL-2, UH). R07182.