Interuniversity Attraction Pole (IAP)

Ex-post Evaluation

P6/03

Statistical Analysis of Association and Dependence in Complex Data

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

Coordinator : I. Van Keilegom

April 30, 2010

Contents

1 Introduction

1.1 General information

The network is entitled

Statistical Analysis of Association and Dependence in Complex Data

and has project number P6/03.

There are 5 Belgian partners and 4 European partners in the network. The coordinator is from the French speaking part of Belgium, the four associate partners are from the Flemish part. Detailed information about the partners is given below.

Network coordinator:

1. Léopold Simar (2007-2008) and Ingrid Van Keilegom (UCL) (2009-)
   Institution: Université catholique de Louvain, Institut de statistique
   Budget: 900,000 €

Associate partners:

2. Iven Van Mechelen and Irène Gijbels (KUL-1)
   Institution: Katholieke Universiteit Leuven
   Budget: 650,000 €

3. Geert Verbeke (KUL-2)
   Institution: Katholieke Universiteit Leuven
   Budget: 400,000 €

4. Luc Duchateau (UG)
   Institution: Universiteit Gent
   Budget: 400,000 €

5. Noël Veraverbeke and Geert Molenberghs (UH)
   Institution: Universiteit Hasselt
   Budget: 400,000 €

European partners:

6. Anestis Antoniadis (UJF)
   Institution: Université Joseph Fourier, Grenoble, France
   Budget: 25,000 €

7. Paul Eilers (UU)
   Institution: Universiteit Utrecht, The Netherlands
   Budget: 25,000 €

8. Wenceslao González Manteiga (USC)
   Institution: Universidad de Santiago de Compostela, Spain
   Budget: 25,000 €

9. Mike Kenward (LSHTM)
   Institution: London School of Hygiene and Tropical Medicine, U.K.
   Budget: 25,000 €

1.2 History of the IAP network

During Phase V the following partners have participated to the project entitled 'Statistical techniques and modeling for complex substantive questions with complex data': the Belgian partners UCL, KUL-1, KUL-2, UH and ULB, and the European partners UJF and RWTH-Aachen (Aachen Technical University). With the exception of the ULB partner and the RWTH-Aachen partner, all of these partners take part in the present project. Moreover, one new Belgian group was added to the network (namely UG) and three new European partners were invited to join (UU, USC and LSHTM). More information on the accomplishments of the network during phase V (annual reports, overview report, ...) can be found on the web page of the network (http://www.stat.ucl.ac.be/IAP/PhaseV).

1.3 Summary of the objectives of the research project

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

The global objective of the network is to develop new models and methodological tools to do inference and to analyze these complex data structures. To achieve this goal, the network will be structured in five interlocking workpackages, devoted to different types of complex data structures, that will be studied within the network.

1. **Workpackage 1: Multivariate data with qualitative constraints**
   In statistical analysis, the quantity one wants to estimate or test a hypothesis about, often satisfies certain natural qualitative constraints, which one has to take into account, if one wants to make fully use of the nature of the data. Examples of constraints include boundaries (e.g. in frontier analysis), monotonicity, convexity, ellipticity or independence of unobserved components, unimodality and sparsity. Qualitative constraints also arise when using dimension reduction techniques, analyzing functional data or dealing with inverse problems. A wide spectrum of statistical techniques is required to analyze this type of complex data. Building further on the results obtained in the previous phase (Phase V) of the network, new challenging research questions will be studied in this area, like e.g. the estimation of stochastic boundaries, the use of dimension reduction techniques with incomplete data, etc.

2. **Workpackage 2: Temporally and spatially related data**
   Methods based on principal component analysis to forecast single variables on the basis of a large panel of time series are widely studied in the economic literature, and will be further

explored and compared. Also, the work on dynamic factor models, already initiated during Phase V of the network, will be further pursued. Another topic the network is very much interested in, is the study of non-stationary time series. The achievements in the domain of locally stationary time series, which have been extensively studied during Phase V, will be further developed with a new emphasis on goodness-of-fit tests, adaptive inference and timespace modeling. A unified approach will be taken in order to jointly address several types of non-stationarity (such as time-varying coefficients models, unit roots models,...).

3. **Workpackage 3: Incomplete data**
   Several types of incompleteness in the data arise in practice: missing data, censored data (right censoring, interval censoring, ...), truncated data, misclassified data, coarse data, ... The main focus will be on censored and missing data. In particular, the work on nonparametric estimation with censored data and on frailty models studied already in detail during Phase V, will be further pursued, and it will be studied how repeated data and survival data can be jointly modeled. The focus in the analysis of missing data will be on sensitivity analysis and on the combination of latent structures and mixed and mixture modeling ideas. Another research topic that the network is very much interested in, is the estimation of causal effects of observed exposures, measured with error, in randomised studies.

4. **Workpackage 4: Data with latent heterogeneity**
   Unobserved heterogeneity can be modeled in different ways. A natural and common way to model this heterogeneity is by means of mixed models, which have been extensively studied during Phase V. The gained expertise opens the door to study in particular mixed models with partially specified residual dependencies conditional on the values of the random effects, and generalized linear mixed models. For the latter model, flexible models for the random effects distribution will be investigated, like mixtures of normals to approximate a B-spline basis.

5. **Workpackage 5: High-dimensional and compound data**
   In many applications dealing with genomics, proteomics, metabolomics, etc., data to be dealt with typically include a very large number of variables. The information in such datasets often contains a lot of noise in the form of irrelevant information and masking variables. Additionally, the information can come from different types of sources. Major challenges for these datasets pertain to the detection of structure in high-dimensional problems, the filtering of noise and irrelevant pieces of information, multiple testing in the presence of a high number of variables, and drawing much stronger inferences by means of suitable combinations of different data pieces at hand. To deal with these challenges, suitable non and semiparametric techniques (including smoothing methods) will be developed, that can be used for noise reduction, appropriate dimension reduction and clustering techniques (including methods of mixture modeling) for high-dimensional two- and multiway data, and data fusion methods in which several pieces of multiblock or multiset data are jointly modeled.

Cross-links between the five workpackages will be established on at least three different levels, which are visually represented in Figure 1.

I. **Interlocking complexities in the data**
   In practical situations, often data are encountered that imply interlocking complexities as studied in several workpackages. Methods will be developed for dealing in an appropriate way with such compound complexities. This will require more than a mere concatenation of results as obtained from different workpackages, because when considering e.g. missing data in a multivariate context new kind of complexities will have to be dealt with.

II. **Common modeling approaches**
   The study of dependence is a recurrent topic across all workpackages. In this regard different approaches will be taken to deal with dependence modeling, including the use of copula models, regression models that are based on e.g. various kinds of dimension reduction approaches, and random effects models, like e.g. generalized linear mixed models. Within the network, those different approaches will be compared, both on a theoretical level and on the level of analyses of several benchmarking datasets.

III. **Common methods and tools**
   The different workpackages will rely on a common set of tools, including techniques of kernel smoothing, semiparametric inference, Bayesian inference, optimization, randomization and bootstrap. Findings on these tools will be exchanged among workpackages, and generic results on these tools will be aimed at, allowing their use in a broad range of contexts.

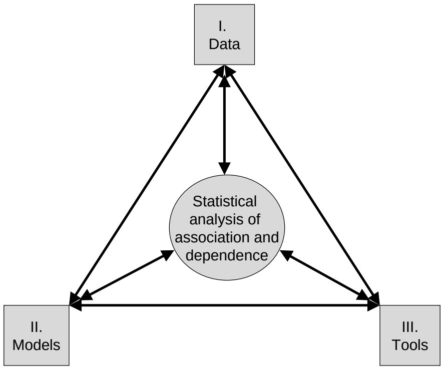

Figure 1: *Visual representation of the structure of the project.*

1.4 Summary of the objectives of the partnership

The importance of the IAP network for its partners is reflected in many different ways :

1. **Increased research activities and recognition of the partners of the network**
   Thanks to the network, there are more intensive and effective exchanges within a considerable part of the Belgian statistical community. The international position of the partners is strengthened by taking advantage of the complementarities in the international contacts,

the increased visibility of their work, etc. Moreover, the partnership also strengthens the possibilities to recruit suitable PhD students/researchers through joint advertising/operating on the international job market.

2. **Increased interaction between different fields in statistics**
   The researchers of the network are sometimes faced to similar problems but in different fields. An example (among many others) is given by the use of copula models, which is a recurrent topic in many fields in statistics. In this respect, each partner clearly benefits from the expertise of the whole network.

3. **More training opportunities for PhD students**
   The network offers more and richer training opportunities for young statistical researchers. For instance, many short courses are organized by the partners of the network. The announcements of these short courses are sent to all members of the network, and as a result the courses are attended by many of its PhD students.

2 Research results

In this section the scientific achievements of each of the five workpackages are outlined. For each workpackage we describe below briefly which results have been obtained and we refer to publications and/or discussion papers for more details about the obtained results. For each workpackage, we also summarize the main objectives and the main achievements in relation to the initial objectives, and we comment in case of deviations from the initial project workprogram.

2.1 Workpackage 1: Multivariate data with qualitative constraints

The contributing partners

As presented in the project description, contributions to this workpackage come mainly (but not exclusively) from the partners KUL-1, UCL, UH, UG and the European partners UJF, USC, and UU. A lot of collaborative research has been done between the partners as can be noticed from the lists of joint publications.

A summary description of the objectives

The main objectives of workpackage 1 in the project were:

(1) to develop powerful nonparametric techniques for estimating boundaries or frontiers; for estimating smooth or non-smooth (multivariate) functions/images; for estimation in inverse/deconvolution problems;

(2) to do inference in semiparametric regression models and to study semiparametric efficiencies;

(3) to study (robust) regression quantiles and multivariate extensions of these;

(4) to model dependencies and to test for independence; to study appropriate measures of dependencies, including analysis of robustness and efficiency.

The main achievements in relation to the initial objectives

Very substantial research work has been carried out, and without exaggeration it can be stated that the four main objectives have been met, and even more than that. For each of the four objectives we provide a summary of the main achievements.

(1) Powerful nonparametric techniques for estimating boundaries or frontiers:

> Bootstrap methods and the consistency of these methods are one of the major contributions in deterministic frontier models. Another major accomplishment is the study of partial frontier functions, their robustness properties and links between the different classes of partial frontiers on one hand and full frontiers on the other hand. In stochastic frontier estimation powerful local likelihood approaches have been developed.

Powerful nonparametric techniques for estimation in inverse/deconvolution problems:

> The density deconvolution problem has been studied intensively using kernel methods and wavelet methods, and several seminal insights into the inverse problems were gained and further exploited. Attention has been devoted to practical issues inherently to this and related problem.

Powerful nonparametric method for inference of smooth or non-smooth (multivariate) functions/images:

> Local linear fitting has been one of the major tools in developing good performing nonparametric methods for smooth or non-smooth functions or surfaces. Such methods have been developed for estimation of regression curves and surfaces, spectral densities and also irregular probability densities. Penalized regression techniques have been studied very intensively, with particular attention to the choice of the penalty function, mean and dispersion function estimation and the use of the technique in variable selection procedures. Also major contributions in empirical likelihood approaches are to be noted.

(2) Inference in semiparametric regression models and the study of semiparametric efficiencies.

> Major classes studied in this context are location-scale models and single-index models. Major contributions to the area include exploiting the estimation of the error structure in a regression model, semiparametric estimation of dispersion functions and several important contributions to nonparametric and semiparametric testing procedures.

(3) Study (robust) regression quantiles and multivariate extensions of these.

> Important new contributions in robust analysis of multivariate data are delivered: versions of boxplots that are more appropriate for skewed distributions; depth quantiles (also for censored data); robust methods for dimension reduction, such as robust Principal Component Analysis, and robust vector support machines methods; robust resampling techniques (such as robust bootstrapping), and robust clustering techniques.

(4) Modeling dependencies and testing for independence; study of appropriate measures of dependencies, including analysis of robustness and efficiency.

> Several groundbreaking results have been obtained related to this objective. Improved estimators for nonparametric copulas have been proposed, leading also to new test procedures for testing for certain type of dependencies. The concept of conditional copulas and associated conditional association measures were studied and these are very useful tools in many application areas. Many new aspects of extreme-value copulas have been studied and a particular focus was on efficiency of the proposed estimation methods.

Comments in case of deviations from the initial workprogram

There were no deviations from the initial workprogram. The whole workprogram described in workpackage 1 has been carried out.

2.1.1 Boundaries, frontiers, and efficiency and productivity analysis

Estimation of a support of a (univariate or multivariate) density function, or estimation of a more general support is a topic of research in the network since the start. Also estimation of the boundary of a support is an intensive research topic in particular in the context of frontier estimation.

Estimation of general supports have been studied in various papers, among which [414] and [404].

In frontier estimation, the objective is to estimate from a sample of observed units (firms, banks, hospitals, . . .) the optimal production plans (optimal combinations of inputs/outputs) maximizing the outputs for a given combination of inputs or minimizing the inputs for a fixed level of the outputs. This optimal production frontier can then serve as a benchmark to evaluate and compare firms' performance by estimating their actual distance to the attainable optimal frontier. From a statistical point of view, the problem can be viewed as estimating the support (and its boundary) of a multivariate random variable in the input/output space. Two major approaches have been developed in the literature: *(i) deterministic frontier* models, where all the observed data are inside the unknown attainable production set, and *(ii) stochastic frontier* models where noise is allowed in the process and some observations can lay outside the production set. Parametric, semiparametric and nonparametric approaches have been investigated in both family of models. The main contributions in these fields can be summarized as follows.

(i) Deterministic frontier models

Envelopment estimators, DEA (Data Envelopment Analysis) and FDH (Free Disposal Hull), have been used extensively in the literature. Inference for these estimators is available and the bootstrap seems the only practical way to approximate their sampling distribution. [521] present a recent survey of latest developments and [1084] suggest an easy and efficient way to reduce the bias of the FDH estimators. The major theoretical contribution is [513] which provides the asymptotic theory of DEA estimators with a formal proof of the consistency of the bootstrap. The bootstrap can be based on a complex "double smoothing" procedure (smoothing of the estimated frontier and of a joint multivariate density in the input/output space) or by using subsampling. [1060] propose a simpler consistent bootstrap algorithm using the double smoothing approach and [1076] show how to select the optimal subsample size when subsampling is chosen. The latter is particularly useful when testing issues are considered (testing returns to scale or testing convexity in [1076] and testing restrictions in [1075]). A similar approach has been used in [78] to compare the efficiency of two groups of firms (like public or private,. . .). The particular case of CRS (constant returns to scale) impose some conical structure of the attainable set and the inference can be then improved. The theory for the full multivariate case is now available in [1150].

An important topic in efficiency analysis is to understand which external factor (like regulatory rules, environmental variables,. . .) may influence or explain the efficiency scores. These factors are not necessary under the control of the producer but may influence the production process. Most of the researchers have used two-stage approaches where in a second step, the DEA efficiency scores are regressed on these external factors. For the first time in this literature, [76] propose a Data Generating Process (DGP) where this approach can be used; in particular, the approach is shown to rely on a separability condition that previous literature had neglected. Here again, the bootstrap has to be used to improve the inference in the second step regression. Another approach is to use conditional efficiency scores (conditional on the value of these external factors). [49] propose a unifying approach allowing both convex and non-convex technologies. The asymptotic theory of these estimators is provided by [1108] and [1083] propose a data driven method to select the optimal bandwidth used for smoothing this external conditioning variable.

An other challenging issue when using envelopment estimators is the robustness to extremes and outliers. Related to this aspect, two concepts of partial frontiers have been introduced in the literature: an expected maximum output frontier of order m, and a conditional quantile-type frontier of order α. In [1164] the important question is investigated how the two families of partial production frontiers are linked. This paper also establishes some asymptotic results and contributes to fill gaps in the existing literature. The insights gained in the links between the two types of partial frontiers, brought up a completely new class of quantities and their estimation, studied in [1163]. The new concept of extremiles is in this paper applied in estimating frontier cost models. [47] develop in multivariate setup a robust version of the FDH estimator based on non-standard conditional order-α quantiles and [500] prove the functional convergence property of these estimators. A new challenging approach is to see if by selecting the order of these partial frontiers estimators, we can estimate the full frontier in a robust way by keeping also the asymptotic normality of the estimators. This has been done by investigating the link between frontier estimation and extreme value theory in [1128] for the order-α quantile estimator and in [1031] for the order-m estimators.

Such robust techniques have been described in the book [48]. Applications to European universities can be found in [40] and [39], and to transitioning economies in [515]. [1114] extend to a case of two frontiers the standard frontier model used in productivity analysis and eventually provides a new methodology for measuring a degree of market imperfection along with a measure of the bargaining power of the supply and demand sides.

When a panel of data is available, Malmquist indices are quite useful to investigate the evolution of the production process over time. [1129] show how the bootstrap can be used to produce prediction in this setup.

(ii) Stochastic frontier models

The challenge in this class of models is to identify the noise from the inefficiency when evaluating firms' performance. Even in a full parametric setup (parametric frontier function, strong parametric assumptions for the noise and the inefficiency distribution), the problem is still a challenge. [1122] indicate how bagging techniques (using bootstrap) are useful to improve the inference for individual efficiency scores.

In a full nonparametric setup, stochastic models are not identified, so a minimal set of assumptions is needed to allow more flexible models than the restrictive parametric ones. The first attempt in this direction is [68] where local maximum likelihood methods are proposed. In a different multivariate setup [75] propose a DGP allowing to introduce noise in DEA/FDH methodologies. This has been extended by [1151] to propose stochastic version of DEA/FDH estimators in a quite general setup. By using the results of [518], two stage approaches analyzed in [76] can also be extended to cover more flexible models for the second step regression.

Of course, when a panel of data is available one has more information to identify the noise. [73] analyze the properties of semiparametric efficient estimators in models with time dependence.

The problem of stochastic frontier models can in fact be seen as a deconvolution problem. A lot of research has been done by members of this (and the previous) network on the topics of density and regression deconvolution, and several breakthroughs have been realized in this area. Deconvolution problems or, more generally, inverse problems are hard problems, and frequent encountered difficulties (mainly related to practical implementation, issues related to optimization tasks and issues of numerical stability) have been reviewed and studied in [134] in a case study in density deconvolution.

2.1.2 Non- and semiparametric estimation of (irregular) curves and surfaces

Unknown functions and surfaces are not always smooth, and may show different types of irregularities in different regions. In a first paper (see [141]) it was shown how local linear fitting techniques could be used to estimate curves with jump discontinuities. The proposed method nicely compromises between smoothing on one hand and jump-preservation on the other hand. Theoretical properties of the method have been established, and its performance was illustrated on various examples. An extension of the method to estimating two-dimensional irregular regression surfaces was provided. This research is also linked to edge detection and image analysis. In [599] an overview is given on the use of local linear fitting for (univariate) smoothing while preserving irregularities. In [1195] those techniques are used to enhance a better estimation of curves near peaks. Curves with possibly different types of irregularities, namely jump and peak irregularities, were considered in [1170], and local linear techniques are used for their estimation. A review paper on locally weighted regression has been written upon invitation (see [1238]).

Another approach to deal with inference for curves with irregularities is to use penalized regression techniques with an appropriate choice of the penalty function. Some penalty functions typically are better suited for estimation of curves with irregularities than others. A specific penalty choice influences the optimization problem to be solved as well as the existence of a (unique) solution. In [1679] nonparametric regression estimation is considered in extended generalized linear models using penalized splines with non-quadratic penalties. This unified approach allows to study the optimization problem and asymptotic properties of the resulting estimator for a whole class of penalties. The papers [1240] and [1241] investigate penalized regression with a difference type of penalty and provide a data-driven procedure to choose the differencing order. Moreover, theoretical properties support this data-driven choice, and these results are among the few theoretical studies that have been carried out in this context. The powerful P-splines techniques is also used as a basic ingredient in a variable selection procedure for additive models that has been developed in [1640]. The combination of the P-splines technique together with the nonnegative garrote method extended to the context of a nonparametric functional additive model resulted into a variable selection with a very good performance. Estimation consistency and variable selection consistency of the procedure have been established.

[1239] consider the problem of under- or overdispersion estimation. Especially overdispersion of data is a well-known phenomenon when analyzing data. A unified approach to estimate nonparametrically mean and dispersion function is provided. A major contribution is that the flexible modeling allows for a combination of both situations: overdispersion or underdispersion, and moreover allows for even cross-overs among the situations as a function of the domain of a covariate. A starting point for the unified approach is a double exponential family of conditional distributions, which incorporates a dispersion parametric function into the usual exponential family. Flexible models are then obtained by extending these basic models. These extended models are quite appealing since their interpretability is similar to the popular generalized linear models. The developed procedure for mean and dispersion estimation has been used to analyze several data sets, including Italian abortion data. A specific example of a generalized linear model is the logistic regression model. Lasso logistic regression is used to analyze gene expression data in [1645]. Poisson regression models are another example of models that fit into a framework of generalized linear models. Estimation of the parameters in a multivariate mixed Poisson model is an important problem in image processing, especially for active imaging or astronomy. However, the classical maximum likelihood approach cannot be used since the corresponding masses cannot be expressed in a simple closed form. The paper [1569] studies a maximum pairwise likelihood approach when the mixing distribution is a multivariate Gamma distribution. This method is applied to change detection in low-flux images.

Bayesian smoothing ideas, and in particular Bayesian P-splines, have been used in a number of papers, with applications to pharmacokinetics. In preclinical and clinical experiments, pharmacokinetic (PK) studies are designed to analyze the evolution of drug concentration in plasma over time i.e. the PK profile. In this context, the estimation of pharmacokinetic parameters using adaptive Bayesian P-splines models is studied in [1109], while [482] estimate receptor occupancy using varying coefficients models. Extensions of Bayesian P-splines models for fitting PK curves is considered in [481].

In financial applications, one needs to deal with various irregular behaviors (e.g. of stock market prices, or stock indices). Among the key issues is the study of the volatility function. In financial mathematics and engineering, the erratic behavior is often modeled by using particular classes of stochastic processes, such as Lévy processes. A lot of work has been done here. See for example [1192], [1202], [1211]. The recent book [1259] gives a good survey of the use of Lévy jump processes in credit risk.

2.1.3 Nonparametric (location-scale) regression and semiparametric regression

[1649] model the relationship between the error and the covariate in a nonparametric regression model by means of a copula model. This offers an alternative for the model assuming independence between these two variables, since the latter assumption is often violated in practice. ROC curves are a useful tool to analyze the discrimination capability of a diagnostic variable in medical studies. In certain situations, the presence of a covariate related to the diagnostic variable can increase the discriminating power of the ROC curve. In [1687] the authors model the effect of the covariate over the diagnostic variable by means of nonparametric location-scale regression models, and they propose and study a new nonparametric estimator of the conditional ROC curve. In [1148] the estimation of the error distribution is considered in a nonparametric location-scale regression model with multivariate covariates. The estimator is based on multivariate local polynomial fits of the regression and variance function. Various applications to testing model assumptions in nonparametric multiple regression are also considered. Under the same model, [1116] consider change-point tests for the error distribution in nonparametric regression.

Single index models are quite useful to explore the effect of exogenous variables on the probabilities in two-way contingency tables. A general approach is developed in [1136] and its application for testing issues in the table is in [1135]. Gery Geenens presented his PhD in this field in June 2008. In [484], the authors first propose a new estimator of the joint distribution of a d-dimensional covariate X and a univariate response Y, that is subject to random right censoring. The estimator overcomes the common curse-of-dimensionality problem, by using a new dimension reduction technique. Second, they assume that the relation between X and Y is given by a single-index model, and propose a new estimator of the parameters in this model.

[79] study the backfitting and profile methods for general criterion functions that depend on a parameter of interest β and a nuisance function θ. They show that when different amounts of smoothing are employed for each method to estimate the function θ, the two estimation procedures produce estimators of β with the same limiting distributions.

Modeling heteroscedasticity in semiparametric regression can improve the efficiency of the estimator of the parametric component in the regression function, and is important for inference problems such as plug-in bandwidth selection and the construction of confidence intervals. In [1124] the authors propose a general method to estimate the dispersion function in a semiparametric way, and they obtain generic conditions under which the proposed estimator satisfies certain asymptotic properties.

[516] propose consistent estimators in a semiparametric transformation model, where the transformation is modeled parametrically, and all other components of the model are non- or semiparametric. The authors propose two methods for the estimation of the transformation parameter: maximizing a profile likelihood function or minimizing the mean squared distance from independence.

2.1.4 Nonparametric and semiparametric testing procedures

Often bandwidths or other smoothing parameters in nonparametric testing procedures are chosen based on an estimation optimality criterion. The paper [630] investigates an optimal choice for the bandwidth parameter in a semiparametric testing context. The paper provides theoretical contributions as well as a practical selection rule. An important problem in regression is the development of goodness-of-fit tests for (semi)parametric regression models. In [998], 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 above paper, which considered the case of completely observed data, was extended by [457] to the situation where the response variable is subject to random right censoring. The latter paper is related to the work in WP3. In [81] goodness-of-fit tests for parametric regression functions are studied in the presence of time series errors (related to work in WP2), while [999] and [1091] propose two different empirical likelihood tests that are 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 variance function in regression. These problems are studied in [456] and [51] respectively. Both tests are based on the comparison of two estimators of the distribution of the errors, one under the null hypothesis and the other under the alternative hypothesis. [502] and [503] consider a nonparametric location-scale regression model, in which the error is independent of the covariate, and the regression and variance function are smooth but unknown functions. They construct tests for the validity of this model. The test proposed in [503] is based on the nonparametric estimation of the errors, whereas in [502] the test procedure is based on differences of neighboring responses.

The procedure proposed in [526] is motivated by recent developments in the asymptotic theory for analysis of variance when the number of factor levels is large. The similarity of the form of the test statistic to that of the classical F-statistic in analysis of variance allows easy and fast calculation.

2.1.5 Empirical likelihood

In the context of empirical likelihood (EL), a number of contributions have been made. An invited review paper on EL methods for regression, with discussion and rejoinder by the authors has been published recently (see [1092] and [1093]). The regression models considered in this review include parametric, semiparametric and nonparametric models. Both missing data and censored data are accommodated. In the context of inference for copula functions [1071] propose a jackknife EL method that offers important computational advantages with respect to the classical EL method. Copulas are also considered in [1663], who deal with a general theory for EL for non-smooth criterion functions. The method allows to deal with e.g. differences of quantiles, copulas or ROC curves, which were previously studied by means of a smoothed EL approach. Finally, in [1107] the scope of the EL methodology is extended in three directions: to allow for plug-in estimates of nuisance parameters in estimating equations, slower than $n^{1/2}$-rates of convergence, and settings in which there are a relatively large number of estimating equations compared to the sample size. A range of examples from survival analysis and nonparametric statistics are provided to illustrate the main results.

2.1.6 Multivariate data, robust analysis and nonparametric inference

Among the main research themes in recent years in the network, is the analysis of elliptically distributed multivariate data, possibly corrupted with outliers. A new deterministic algorithm for the Minimum Covariance Determinant estimator has been developed in [1174], and applied in a robust calibration framework in [1256]. For high-dimensional data, dimension reduction will often be applied by means of a principal component analysis (PCA) method. Work on robust PCA methods is reported on in WP5. A second research theme is the analysis of non-elliptical multivariate data. Classification methods for low- and high-dimensional skewed data are proposed in [1177] and [1246].

Building a sparse and stable linear model when a large number of candidate regressors is available, is far from trivial. The problem becomes even much more difficult if the data is not necessarily of high quality, but may contain all kinds of contamination. A robust model building strategy must yield reliable results for contaminated data, but must also be computationally efficient so that the result can be obtained in a reasonable amount of time. To this end, [230] and [231] developed computationally efficient techniques to robustly screen out the most relevant predictors from a large set of candidate regressors. In [230] robust versions of the well-known forward and stepwise variable selection algorithms are constructed, while in [231] a robust adaptation of the state-of-the-art but nonrobust least angle regression selection technique is developed.

Once the most important predictors have been identified, several candidate linear models can be investigated in more detail. To this end, [1414] have developed robust, fast to compute versions of cross-validation and bootstrap to robustly estimate the prediction performance of a linear model. [737] have proposed consistent model selection criteria that take into account both the model complexity and the prediction error of a linear model. The prediction error of the linear model is estimated by using a fast and robust bootstrap procedure, as reviewed in [738].

Most clustering techniques incorporate constraints on the shape of the clusters to make the clustering problem feasible. In [1381] the clusters are assumed to have linear structures. However, this constraint is relaxed by allowing that a fraction of the data does not belong to any of the clusters. Conditions for the existence and consistency of the solutions are given and a computationally feasible algorithm is proposed as well.

A review on the use of nonparametric methods and methods for robust analysis of data is given in [1200]. Further, many of the robust statistical analysis were collected in review papers. See [1244], [1417], [1203], [1255] and [1245].

In the second edition of the book [61], modern approaches of multivariate statistical tools (CART regression trees and SVM techniques) are illustrated. New sections on copulas and on heavy tails distribution have also been implemented. Most of the procedures and examples of the book, using R and Matlab programs, are available on the Web.

2.1.7 Wavelet decompositions, image analysis and multiscale methods

In many applications it is justified to assume a regression function to be monotone. A penalized wavelet estimation procedure for estimation of a monotone regression function was developed in [436]. A key issue here was to transfer the primary optimization problem into a dual optimization problem, which then can be solved more easily.

In [1646] 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 is 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.

Statistical analysis of images is another research area within the project. In [1644] a multiscale approach for statistical characterization of temporally and spatially heterogeneous functional images (arising from, e.g., medical data - MRI or NMR - or from satellite data) is presented.

In [45] the authors address the problem of estimating with wavelets a probability density and a cumulative probability function, respectively. It is highly non-trivial how to construct and develop estimation theory for multivariate wavelets under positivity and monotonicity constraints, respectively.

Another theme of research concentrates on the application of multiscale sparsity in various, non-standard situations. The use of the lifting scheme in statistical problems on irregular point sets has been investigated in [1208]. A further exploration incorporates kernel smoothing into a multiscale scheme using lifting. This is done in [1207]. Another direction is the use of adaptive lifting in geometrical data, such as images. See [1219]. Multiscale models are also used in model testing and variable selection in partially linear models, in [1234].

2.1.8 Functional data analysis

In [1235] the authors consider a functional nonparametric regression model, in which the response is univariate and the covariate is functional. They show the asymptotic validity of a naive and a wild bootstrap procedure under this model. An application to the construction of confidence intervals is also studied. [60] study two-sample tests in functional data analysis, starting from discrete data. It is shown that by choosing the same tuning parameter (e.g. bandwidth) to produce each curve from its raw data, significant contributions to level inaccuracy and power loss can be avoided.

Application of statistical analysis for functional data related to atmospheric observations is in [1562]. A retrieval algorithm that uses a statistical strategy based on dimension reduction is proposed. The algorithm has been applied to high resolution spectra measured by the Infrared Atmospheric Sounding Interferometer instrument to retrieve 5 atmospheric profiles of temperature, water vapor and ozone. The performance of the inversion strategy has been assessed by comparing the retrieved profiles to the ones obtained by co-locating in space and time profiles from the European Centre for Medium-Range Weather Forecasts analysis. Approaches for the statistical analysis of biomedical functional data are studied in [1561]. High levels of automation in proteomic profiling technologies that allow us to look at many proteins at once is lending to datasets that are increasing in size and dimension, challenging statisticians and biomedical quantitative professionals to develop appropriate methodologies to meet this demand. [1561] reviews work accomplished by the author and its collaborators in designing appropriate statistical tools for an efficient search for proteins and combinations of proteins that can be used as biomarkers to diagnose cancer, to provide more detailed prognostic information for individual patients, and even to identify which patients will respond to which treatments.

A challenging topic is the analysis of functional time series. This topic is obviously linked with work in WP2. Studies are devoted to multivariate functional time series clustering. In [1564] a method for effectively detecting patterns and clusters in high-dimensional time-dependent functional data is presented. It is based on wavelet-based similarity measures since wavelets are ideal for identifying highly discriminant local time and scale features. The contribution of each scale to the global energy is considered, in the orthogonal wavelet transform of each input function to generate a handy number of features that still makes the signals well distinguishable. The new similarity measure combined with an efficient feature selection technique in the wavelet domain is then used within more or less classical clustering algorithms to effectively differentiate among high dimensional populations. This work related to work for high-dimensional data in WP5. In [1567] a method is proposed to select the bandwidth for functional time series prediction. The idea is to calculate the empirical risk of prediction using past segments of the observed series and to select as value of the bandwidth for prediction the bandwidth which minimizes this risk.

2.1.9 Modeling and measuring of dependencies and copula functions

In nonparametric estimation of copulas, a considerable step forward has been made. An interesting investigation of an existing kernel estimator of a copula, using a local linear type of kernel, revealed that often assumptions are made (when establishing asymptotic results) that in facts rule out many of the commonly-used copula families. This interesting finding led to an improved kernel estimation of copulas, by introducing a shrinkage factor. These important results are published in [1666].

Often, the dependence structure between two (or more) random variables is itself influenced by another variable, a covariate. Such a dependence structure can be modeled via a so-called conditional copula function. Estimation of conditional copulas was discussed in [1643]. From the notion of conditional copula, the concepts of conditional association measures, such as conditional Kendall's tau or conditional Spearman's rho follow very naturally. These conditional association measures are very nice tools for describing conditional dependencies. Several illustrations of their use are provided in [1642].

Among the important question in modeling and measuring of dependencies, is the choice of an appropriate copula function or family. Some copulas exhibit interesting particular properties, such as positive quadrant dependence or negative quadrant dependence. When one has the additional knowledge that a copula satisfies for example a positive quadrant dependence property, then such qualitative knowledge should be taken into account in the estimation procedure. In [1641] testing procedures for testing the null hypothesis of positive quadrant dependence are developed. The test statistics are based on state-of-the-art estimation methods, developed by members of the KUL-1 and UH team. Moreover test statistics are built in various ways. The comparisons between the performances of the proposed tests provided a lot of interesting findings.

2.1.10 Extreme value theory and copulas

Extreme value theory is the basis for modeling extreme events. Contributions to extreme value theory include, among others, the papers [588], [589]. Special attention in extreme value theory goes to the estimation of the tail index. An improved reduced-bias tail index was introduced in [589]. How to estimate the extreme value index under random censoring has been studied in [126].

Extreme-value copulas constitute a family of models designed to describe dependence between extremes, see the survey paper in [1151]. These copulas are parametrized in terms of a functional parameter, called Pickands dependence function, which is a convex function on the unit simplex lying between certain bounds. How to enforce these constraints on a pilot estimator using a least-squares approach is the subject of [505]. Rank-based estimators of Pickands dependence functions are studied in [1099, 1047]. 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. Extensions to higher dimensions are treated in [1048].

Another quite popular class of copulas are the Archimedean ones. In [1090], building upon [42, 43], a complete and user-friendly directory of the tails of Archimedean copulas is presented which can be used in the selection and construction of appropriate models with desired properties. The results are synthesized in the form of a decision tree.

Rather than assuming a particular form of the dependence function, a more flexible approach is just to assume that the tails of the copula of the observables present some kind of regularity, or formally, that the copula is in the domain of attraction of an extreme-value copula. Statistical inference on tail dependence then can be reduced to estimation of the extreme-value attractor. This attractor can be estimated in many ways. Particularly attractive and efficient are the nonparametric maximum likelihood estimator in [1097] and the nonparametric Bayesian approach in [1049]. However, dimension reduction techniques can be shown to result in certain parametric models for which likelihood-based methods fail. For such cases, the method-of-moment estimator in [501] provides a convenient alternative.

Extrapolation outside the range of the data sample is possible only under some kind of regularity condition or shape constraint on the population distribution. Arguably the most convenient and flexible assumption is the one of regular variation and its extensions [1149]. The theory of second-order regular variation motivates an extension of the Pareto distribution as a universally applicable model for the tail of a distribution [1647]. The same theory allows an accurate assessment of the diversification benefits on tail-related risk measures such as value-at-risk when multiple risks are aggregated into a single one [1130].

2.1.11 Interactions with other workpackages

Workpackage 1 focuses mainly on developing powerful statistical techniques for inference for various complex data situations encountered in applications. Many of the topics dealt with in this workpackage are as such directly linked with topics in other work packages.

For example, several techniques discussed in this workpackage can deal with dependent data and/or censored data, and are directly linked with research under work WP2 and WP3 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, and extensions of these. Parts of the research done in workpackage 1 also includes comparisons of different approaches, such as for example a mixed-models random effects approach and another approach through flexible modeling of nonparametric functions.

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 and interact with topics in WP5.

2.2 Workpackage 2: Temporally and spatially related data

The contributing partners

Most Belgian partners (UCL, KUL-1, KUL-2) and most European partners (UJF, USC) have contributed to WP2.

A summary description of the objectives

WP2's objectives are summarized as follows:

> "Methods based on principal component analysis to forecast single variables on the basis of a large panel of time series are widely studied in the economic literature, and will be further explored and compared. Also, the work on dynamic factor models, already initiated during Phase V of the network, will be further pursued. Another topic the network is very much interested in, is the study of semiparametrically efficient distribution-free methods for the analysis of volatility, based on quantile autoregression methods, and for the analysis of unit root problems, that play a major role in the modeling of economic and financial processes. Finally, the achievements in the domain of local stationary time series, which have been extensively studied during Phase V, will be further developed, with a new emphasis on time-space modeling, on-line estimation and automatic model building."

The main achievements in relation to the initial objectives

Most objectives have been met. First, dynamic factor models have been studied and further developed. Second, spatially related data have been modeled and new estimation procedures are investigated. Third, temporally related data under violations of stationarity and homogeneity assumptions have been studied, and publicly available software has been developed. The development of feasible models allowing non stationarity (link to WP1), censored data (link to WP3) or characterizing complex psychological data (link to WP4) have also been considered. Fourth, new methods have been proposed and investigated using wavelets for signal processing and structural break analysis. Finally, the extensions to multiple series or processes have necessitated the development of new models and estimation procedures.

Comments in case of deviations from the initial project workprogram

Due to the (unfortunate) non-selection of one of the original partners (ULB) to the network, one of the originally planned research topics, namely the development of semiparametric distribution-free approaches to unit root problems, has not been further developed by the network. Nevertheless, several other new research lines have been initiated by individual partners and jointly by several partners of the network. Of course, these new research lines need to be developed further, even if many results are available and have been published already in the international literature.

2.2.1 Overview

Main contributions in the context of univariate correlated data have been made in case of non stationarity due to structural breaks or time-varying spectra. In case of stationarity new developments include statistical inference for time series with irregular spectra, for their extremes, under censoring, for non linear processes and for continuous time processes. An important aspect was the development and investigation of feasible models for complex financial or psychological temporally related data. Extensions to multivariate temporally related data have been studied. New methods are developed in continuous time multilevel models including Bayesian inference. Correlation between multiple series have been modeled and statistically analyzed. New developments concern dynamic conditional correlation model, multivariate extensions of volatility models and modeling of extremes of multiple series. New estimation procedures have been developed in case of violations of stationarity and homogeneity assumptions. Finally, detection and modeling spatial correlation has been investigated using, for example, spectral techniques. New estimation procedures are also proposed in nonlinear regression with spatially correlated errors.

Temporally and spatially related data are encountered in several sciences and their statistical analysis has a long history. However, the development and investigation of models for temporal or spatio-temporal correlated data under non standard assumptions, in discrete or continuous times and in one or higher dimensions, still encounters many important but unsolved problems. The main achievements of the research network for temporally and spatially related data can be subdivided into three categories: complex univariate temporally related data under non standard assumptions, multivariate temporally related data and spatially related data.

2.2.2 Complex univariate temporally related data 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 [990] 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. The proposed adaptive test is illustrated on the analysis of financial time series. In [523] a pointwise adaptive estimator of the time-varying spectrum has been proposed for an explicit wavelet-based model of local stationarity allowing for the spectrum to change very suddenly in time. The behavior of the estimator is studied in homogeneous and inhomogeneous regions of the wavelet spectrum. The authors of [1073] treat the case of locally stationary long-memory processes. In a semiparametric approach, they estimate the time-varying long-memory parameter of an otherwise fully non parametric specification of a time-varying spectral density by some log-regression wavelet-based estimator.

Another type of complexity often present in real data is due to the irregularity of the spectral density function of a stationary time series. In [507], a new approach to wavelet threshold estimation of spectral densities of stationary time series is studied. An alternative method is proposed to address the problem of heteroscedasticity of the minimax mean-square risk for a class of spectral densities, including those of very low regularity. Nonparametric estimation of a spectral density with improved estimation at the peaks is dealt with in [1171]. The proposed method relies on local linear fitting techniques. These techniques are also the basic ingredient for estimating nonparametrically the volatility function in [1160]. The discussed estimation method allows for a discontinuous behavior of both the mean and the variance function, and relates to the topic studied in [990]. Local polynomial fitting (in a non time series setup) is among the methods studied in detail and applied in various contexts in WP1.

[1142] consider nonparametric estimation for dependent data, where the observations do not necessarily come from a linear process. The authors study density estimation and also discuss associated problems in nonparametric regression using the 2-mixing dependence measure. In particular the results under the 2-mixing are compared with those derived under the assumption that the process is linear. Time series can also be studied from the perspective of their largest values, the spikes, and the temporal aggregation of these spikes into clusters. In [1119], a new estimator is proposed for the extremal index, a measure of the degree of clustering of extremes.

Sequences of random variables which exhibit certain dependence structures have been studied theoretically in several papers. In [1561] the maximum entropy extension of a partially specified autocovariance sequence of a periodically correlated process is considered. The sequence may be specified on a non-contiguous set. A method is given which solves the problem completely; it gives the positive definite solution when it exists and reports that it does not exist otherwise. The method is numerically reliable even when the solution is 'almost' semidefinite, and also works when only positive semidefinite extension(s) exist. In [1573] the authors prove a central limit theorem for linear triangular arrays under weak dependence conditions. The result is then applied to the study of dependent random variables sampled by a Z-valued transient random walk. An application to parametric estimation by random sampling is also provided. In [1574] a random walk in random scenery is considered. Under a weak dependence assumption on the scenery a functional limit theorem is proven. A specific class of unbounded functions of intermittent maps are studied in [1570] and limit theorems and moment inequalities are established in this setup. A recursive algorithm for the computation of the first and second order derivatives of the entropy of a periodic autoregressive process with respect to the autocovariances is provided in [1556] Moreover, an implementation of the algorithm is available as an R package. A new correction method for temperature series at a daily time scale, relying on a direct non-linear cubic spline regression method has been introduced in [1560]. The authors show that the proposed method is able to correct mean of the series, but also high order quantiles and moments of the perturbed series. Finally, an open source software, based on R language is provided. The package also includes implementation of some other existing methods.

Censoring of time series data is another problem sometimes encountered with real data. [1098] and [504] consider a nonparametric regression model in which the data are dependent and the response is subject to random right censoring. The former paper considers the problem of nonparametrically estimating the conditional quantile function in this model, while in the latter paper the problem of estimating the conditional expectation of a known transformation of the response is considered. Both papers are very much related to the research topics studied under WP1 and WP3. On the other hand, the construction of confidence intervals and goodness-of-fit tests are also important to validate models on data. In [1131] three types of confidence intervals are developed for a general class of functionals of a survival distribution based on censored dependent data. The confidence intervals are constructed via asymptotic normality, the empirical likelihood (EL) method, and the blockwise EL method. A procedure to test for proportionality of the regression function and the scale function in a nonparametric regression model with dependent data is presented in [1096]. The test can be considered as a preliminary step to check the validity of certain time series models, like e.g. ARCH models.

A lot of work is done in the network in developing models for financial data. A brief report on this work is given in WP1, in a global modeling context using stochastic processes as a main tool. However, the GARCH and stochastic volatility (SV) models are two competing, well-known and often used models to explain the volatility of financial series. [1104] consider a closed form estimator for a stochastic volatility model, derive its asymptotic properties and propose a set of simple, strongly consistent decision rules to compare the ability of the GARCH and the SV model. The results are illustrated by an application to stocks in the Dow Jones industrial average index. [1590] develop and evaluate new algorithms based on GARCH models, neural networks and boosting techniques, designed to model and predict heteroscedastic time series. The behavior of the proposed algorithms is evaluated over simulated data and over the Standard and Poors 500 Index returns series, resulting in frequent and significant improvements in relation to the ARMA-GARCH models.

Another domain where time series in discrete time arise naturally is that of ability testing: A set of items (e.g., questions) is given to a sample of examinees (e.g., pupils or students) with the instruction to solve them. Almost always there is a time limit, and this induces time pressure. In [1686] a probabilistic model is developed for testing speediness, and it is investigated whether it is possible to detect persons that are particularly sensitive to time pressure using local influence diagnostics. In addition, [561] link the idea of test speediness to differential item functioning (meaning that items may have different properties in different groups of people). In another line of work, an algorithm is developed that overcomes the computational burden associated with latent Markov models. In [613], a method is proposed that associates the models in question with a directed acyclic graph and that applies transformations to this graph to arrive at an efficient inferential algorithm (see also WP5).

In preclinical and clinical experiments, pharmacokinetic (PK) studies are designed to analyze the evolution of drug concentration in plasma over time i.e. the PK profile. Some PK parameters are estimated in order to summarize the complete kinetic profile of the drug. [1109] propose a Bayesian nonparametric model based on P-splines, in case the data are sparse and noisy, where typically two difficulties arise with classical estimation methods. The first one is related to the choice of a suitable compartmental model given the small number of data available in preclinical experiment for instance. Second, nonlinear methods can fail to converge. An extension to a hierarchical setting where concentrations from different subjects are available is also presented.

Due to the improvement in the data collection, it is sometimes more natural to consider the temporal data as a realization of a continuous time process. Typical examples are given by financial markets, where microdata are available online, by environmental data, where pollutant are measured continuously over the day, or by medical data, where functional images like MRI or NMR are recorded. A lot of work is done in the network in developing models for functional data, which is one research topic studied under WP1. However, in order to reflect the continuous nature of the observed data more adapted diffusion models have also been developed.

Regarding single level continuous time models, contributions are made with regard to decision making. Models in this area are based on diffusion processes, which are real-valued continuous time

Markov processes. More specifically, in Ratcliff diffusion model a decision process is subdivided into several components, the core of which is a diffusion process. In a first step, [159] have casted the Ratcliff diffusion model in a statistical regression framework, constructed and tested algorithms to fit this model. Along with this theoretical work, [617] developed a freely downloadable Matlab toolbox to allow experimental psychologists to perform Ratcliff diffusion analyses in an easy way. While the previous contributions all relied on a frequentist framework, [618] have also developed Bayesian inferential tools for the Ratcliff diffusion model; this considerably broadens the scope of problems the models can address. An application of this Bayesian inference method (dealing with practice effects on the diffusion model parameters) is described in [1197]. In addition to the Ratcliff diffusion model, inference for a more general type of diffusion models are studied in [1218] that represent a one-dimensional stationary process of the Ornstein-Uhlenbeck type.

2.2.3 Multivariate temporally related data

One important aspect of the research proposal was an extension of the univariate analysis to the case of multiple time series or continuous time processes.

Often a researcher does not collect data from a single subject, but from multiple subjects. This may be interesting from a substantive point of view (e.g., when studying individual differences) or advantageous from a statistical point of view (e.g., when one needs to pool information from different sources to arrive at more stable results). There are clear links between this work and WP4. In a first line of research, a multilevel or hierarchical Ornstein-Uhlenbeck process model has been developed to analyze irregularly spaced multivariate time series from different subjects simultaneously. The basic model has been described in [1214] together with a Bayesian inference algorithm. Subsequently, the model has been expanded to allow for measurement error and for time-varying covariates, see [575]. In both studies, the models have been applied to data from diary studies measuring affective quality. Because of the complexity of the Bayesian inference algorithm, [1179] have investigated whether one could not make use of simpler algorithms that had been developed before to estimate linear mixed models. It turns out that there are some exact and approximate relations between the models, but that the Ornstein-Uhlenbeck process allows for some interesting sources of random variability that the linear mixed model cannot incorporate. Also the diffusion-based models for decision making have been extended hierarchically, so as to allow for analyses of experiments with many subjects and fewer decision trials. Given the complexity of some stimuli (e.g., words, pictures, etc.), it is not possible to make subjects undergo numerous identical trials; this makes a hierarchical model a necessity. The hierarchical models have been developed in a Bayesian context because of the computational tools that can be applied within this framework. In a first study, a hierarchical Ratcliff diffusion model is presented and explained from a psychometric point of view, [1181]. This has led to a modeling framework that is highly flexible and easy to work with for applied researchers. Next, the model has been applied in [1254] to a speeded semantic categorization task in which both subjects and items were random (i.e., a crossed random effects model).

Another main contribution is the development of working methods for analyzing the correlation between multiple series and its application to financial data. Correlation mixtures of elliptical copulas arise when the cross-sectional correlation parameter of a bivariate time series is driven itself by a latent random process. In [1066], both penultimate and asymptotic tail dependence of such copulas are found to be much larger than for ordinary elliptical copulas with the same unconditional correlation. Furthermore, for Gaussian and Student t-copulas, tail dependence at sub-asymptotic levels is generally larger than in the limit, which can have serious consequences for estimation and evaluation of extreme risk. [1052] propose a new dynamic copula model where the parameter characterizing dependence follows an autoregressive process. As this model class includes the Gaussian copula with stochastic correlation process, it can be seen as a generalization of multivariate stochastic volatility models. The authors propose estimation in two steps using efficient importance sampling, discuss goodness-of-fit tests and ways to forecast the dependence parameter.

In [1101] a generalization of the Dynamic Conditional Correlation (DCC) model of Engle (2002) is suggested which allows for asset-specific correlation sensitivities and is useful in particular if one aims to summarize a large number of asset returns. Two estimation methods are proposed and an application to daily data on 39 UK stock returns in the FTSE is presented. It is now well recognized that the maximum likelihood estimator applied to the DCC model is severely biased in high dimensions and, in particular, in cases where the time series dimension is close to the sample size. [1053] propose to reduce the bias by using shrinkage for the sample covariance matrix to target methods, alternatively, using the identity matrix, a single factor model, and equicorrelation. The diffusion limit of DCC models is derived in [1050] which is degenerated in case of a standard DCC model, while for a modified version of the DCC model a non-degenerated diffusion limit can be obtained.

Generalizations of known results on asymptotic theory for GARCH models and their temporal aggregation to the multivariate case have been considered too. The effects of temporal aggregation on causality and forecasting in multivariate GARCH processes are analyzed in [1100]. Results are derived for the distribution of multivariate realized volatility if the high frequency process follows multivariate GARCH. In [1103] and [1105] the asymptotic theory is investigated for a factor GARCH model and a multivariate GARCH model in its general vector specification respectively. Sufficient conditions for strict stationarity, existence of certain moments, geometric ergodicity and beta-mixing with exponential decay rates are established. These conditions allow for volatility spill-over and integrated GARCH. The authors show further strong consistency and asymptotic normality of a quasi-maximum likelihood estimator. [1102] propose a fixed design wild bootstrap procedure to test parameter restrictions in vector autoregressive models, which is robust in cases of conditionally heteroscedastic error terms. The authors investigate the case of Granger causality in a bivariate system of inflation expectations in France and the United Kingdom.

Extreme values of a stationary, multivariate time series may exhibit dependence across coordinates and over time. In [1085], a new and potentially useful tool called tail process is introduced in order 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. A particularly useful application is that for certain time series with infinite second moments, the properly centered partial sum process converges to a stable Lévy process, the Lévy triplet depending among others on the distribution of clusters of extremes. The case of extremes of time series taking values in a Banach space is considered in [1068]. This rather general set-up allows the modeling of extremes of spatio-temporal phenomena.

In the context of multivariate time series the violation of stationarity and homogeneity assumptions are important problems. [1088] present a general approach of shrinkage (towards a multiple of the identity matrix) of spectral density matrix estimators that considerably decreases the mean-squared error and the conditioning number of resulting estimators. On the other hand, in [498] it is shown that in situations of time series panels which can be explained by a few common factors, shrinkage towards the first (few) factor(s) can considerably improve upon the mean-squared properties of periodogram-based spectral estimators. In [486] a dynamic factor model for large cross-section and time dimensions, simultaneously tending to infinity for consistent estimation, is fitted to a locally stationary time series panel with smoothly over time varying second-order structure. The common components of this factor model are estimated by the eigenvectors of a nonparametrically estimated spectral density matrix. In some sense, this work can be seen as a continuation of the previous approach of [1146] on static factor modeling of non-stationary panel data, where principal component regression has been performed on the time-varying covariance matrix of the multivariate data of high dimension. [1123] uses non-balanced Haar wavelets in order to develop a new comparison and classification scheme of time series which arise from a panel showing certain specific characteristics such as pronounced peaks of similar shape but different size and temporal localization. In time series discrimination [1126] combine the SLEX approach of fitting a piecewise stationary model to a second-order non-stationary time series with spectral shrinkage to improve the classification rate. Local stationarity in the context of multivariate volatility models and nonlinear dependence models are considered in [1618] and [1139] respectively. [1618] propose a new semiparametric multivariate volatility model which captures a slowly changing unconditional covariance matrix (low frequency volatility) in a nonparametric way, but also allows for dynamic evolution of the conditional covariance matrix (high frequency volatility) in a more standard fashion. Several estimation methods are proposed for low and high frequency volatility based on kernel methods combined with maximum likelihood. The authors establish semiparametric efficiency, and characterize the efficiency bound. In [1139] multiple time series are modeled based on univariate GARCH-type models for the marginals and copula models for the dependence which allows for time-varying dependence by specifying the copula parameters as deterministic functions of time.Asymptotic properties are shown of the two-step estimators and two types of goodness-of-fit tests are proposed. [1118] use a new impulse response methodology to analyze the effects of U.S. macroeconomic news announcements on the volatilities of three major exchange rates (Euro, Pound Sterling and Yen). Using a multivariate GARCH model with exogenous news effects the authors allow for different types of news and categories in the application.

2.2.4 Spatially related data

Another main objective of this workpackage is the analysis of spatial data.

In [1153] the authors study the asymptotic and finite sample properties of an estimator of a nonlinear regression function when errors are spatially correlated, and when the spatial dependence structure is unknown. The proposed method is based on a weighted nonlinear least squares approach, taking into account the spatial covariance.

There is a growing interest in improving the level of knowledge of spatial and spatio-temporal processes using spectral techniques. [1598] study the asymptotic properties of smoothed nonparametric kernel spectral density estimators for the spatial spectral density. They consider the case of continuous stationary spatial processes under a shrinking asymptotic framework. In [1588] properties of the multidimensional periodogram are studied, under both cases of tapering and no tapering, and the assumption of finite dimensionality of the regular lattice where the spatio-temporal process is observed. Detection and modeling the spatial correlation is an important issue in spatial data analysis, too. [1585] extend two different goodness-of-fit testing techniques for the spatial spectral density. The first approach is based on a smoothed version of the ratio between the periodogram and a parametric estimator of the spectral density. The second one is a generalized likelihood ratio test statistic, based on the log-periodogram representation as the response variable in a regression model. Spatio-temporal correlated data are quite common in many fields, such as ecology, meteorology or environmental sciences. In this situation, it is important to check the existence of spatio-temporal interactions or on the contrary, if the spatial and temporal dependence components can be modeled independently. In [1599] a testing technique is proposed for detecting separability in the spatio-temporal dependence structure. The approach is based on the representation of the log-periodogram as the response variable in a regression model. Within this context, separability can be interpreted as additivity in spatial and temporal frequency components.

Finally, in [1584] an application to real data is provided. The prediction of ozone concentration and the identification of episodes by modeling are fundamental for protecting and preventing the population and environment against the harmful effects of this species. Ozone concentration has been forecasted by spatial time series modeling. Descriptive and predictive models of ozone involving different parameters depending on the area considered have been developed. Satisfactory estimation of ozone concentration was obtained in the three cases with proved efficiency, since predictive values did not exceed the 95% confidence level.

2.3 Workpackage 3: Incomplete data

The contributing partners

All Belgian and international partners have contributed to WP3:

- Belgian partners: UCL, KUL-1, KUL-2, UG, UH
- International partners: UJF, UU, USC, LSHTM

Many strands of research output involve two or more partners, Belgian and/or international. Hence, WP3 has strongly contributed to the scientific integration of the various partners.

A summary description of the objectives

WP3's objectives are summarized as follows:

> "Incomplete data occur in time-to-events as censoring, and in longitudinal and multivariate studies as missing data. Correct (causal) inferences are needed. Further forms of incompleteness are misclassified data and measurement with error. Models for these

> types of incompleteness will be aimed at, and they will be corrected for varying intervention regimes. Models will be proposed for describing response shifts in test data. Such models typically need strong assumptions and thus appropriate sensitivity analysis tools will be proposed. For survival data, various types of censoring (left, right, interval) will be studied, with emphasis on informative and dependent censoring, and coupled to sensitivity analysis. Focus will be on nonparametric and parametric methods, in particular on frailty and copula-based models."

These are amplified in what follows.

- *Substantive applications/problems.*
  Censored (correlated) survival data occur in the medical and epidemiological fields, e.g., when the validity of prognostic indices is studied, when the heterogeneity in outcome between participating hospitals in multi-center studies is of scientific interest, and when the presence of a random treatment by center interaction needs to be investigated. Incomplete longitudinal studies, combined with survival outcomes, are commonly encountered in the areas of dental epidemiology, HIV studies, psychiatric trials, quality of life in oncology, and mental health epidemiological studies, to name a few. Especially in a regulatory context, flexible sensitivity analysis tools are needed. When a test is administered with a fixed time limit, some students may not have enough time to answer all questions, so-called test speededness, the effects of which are detrimental to the intended functioning of the test. Examinees affected by test speededness may hurry through, randomly guess or even fail to complete items, usually at the end of the test. The concept needs to be part of the model to avoid biased inferences. Causal impact of observed exposure is of prime importance in therapeutic treatments and marketing interventions. Individualized treatments are gaining popularity, also in view of genetic breakthroughs (Prentice et al., 2005). Inferences are hampered, though, by the very relationship between the treatment sequence and the outcome studied. Currently, large data streams are being recorded, forming a rich resource for evidence based marketing. Data mining techniques can help find association patterns (Cho et al., 2002). Ideally, sequentially randomised experiments provide a basis for unbiased estimation of the causal impact of evolving interventions based on a treatment and response history.

- *Models and primary objectives.*

  1. For censored survival data, copula modeling will be explored further, including model formulation, assessing goodness-of-fit, and bootstrap-based precision estimators. Various types of informative censoring will be combined into a single model, extending the copula-graphic estimator of Braekers and Veraverbeke (2005). The Koziol-Green model will be extended to handle dependent censoring and a mixture distribution for the censoring time will be considered. Regarding estimation in semiparametric regression with censored data, a first problem is that the mean function generally cannot be consistently estimated with censoring. Valid estimation procedures are in demand. These problems are currently under study for the single index model, but other semiparametric models need study too. Frailty models for right censored data received a lot of attention,

but questions on model diagnostics, hierarchical modeling, and methodological issues (e.g., likelihood ratio testing) are still open. Formulating frailty models for complex censoring are important but untouched. The link between copulas and frailties will be studied. Another research topic that will be studied is the analysis of recurrent events data by making use of transfer of tail information in order to improve the estimation of the so-called transition probabilities. This 'transfer of tail information' will be made possible by imposing a location-scale type regression model on the gap times between consecutive events.

2. Flexible models for incomplete longitudinal data, allowing for various types of sensitivity analysis, will be formulated within the selection model, pattern-mixture model, and shared-parameter model paradigms. Combined with mixture modeling ideas, hybrids between the various model families will be formulated, thereby allowing for misclassification. Sensitivity analysis tools, based on (semi-)parametric models, will be considered. Models for test speededness and/or shifts in test strategies, allowing for random guessing and omission of items, will be formulated. Subject-specific shifts and speeds will be allowed for. Sensitivity of inference will be assessed. The potential outcome formulation of causal parameters has connected causal inference and incomplete data methods. The focus will be on efficient semiparametric models for potential outcomes following different exposures, relying on instrumental variables or on the assumption of no unmeasured (time-varying) confounders. Focus will also be on non-linear effects, heterogeneity between subjects, and measurement error on exposure, extending linear model methods.

The main achievements in relation to the initial objectives

Table 1 provides a broad classification of main result areas, as presented in the proposal and objectives (left hand column) and themes for which research results were reached (right hand column). The numbering system in the right hand column refers to the sequential paragraph numbers.

Comments in case of deviations from the initial project workprogram

The deviation from the initial workprogram in WP3, as provided by way of overview in Table 1, is a beneficial one, in the sense that all areas proposed to be covered were actually covered, but then a number of additional themes were tackled, too. This is clear from the 'additional results' categories for both incomplete data and time-to-event data.

The work on incomplete data has been organized under the headings: (1) complex modeling approaches for missing data; (2) sensitivity analysis tools; (3) censored survival data; and (4) general incomplete data structures.

2.3.1 Complex modeling approaches for missing data

We will set out with high-profile references, then continue with efforts dealing with the shift from simplistic to more appropriate models as well as the proposition of such complex methods. We conclude this section by sensitivity analysis.

Table 1: Comparison between goals and results achieved for WP3.

Proposal goal	**Results actually achieved**
Incomplete data and general structures	
Flexible models	3. Proper analysis tools
Semiparametric modeling	6. Inverse probability weighting and double robustness
Selection models	Selection models are present throughout, given that it is the most commonly encountered framework in incomplete data
Pattern-mixture models	4. Pattern-mixture models
Shared-parameter models	8. Shared-parameter models
Sensitivity analysis	7. Sensitivity analysis
Speededness	Speededness; Item-response theory
Causal inference	Causal inference
Additional results	1. High-profile references; 2. Dealing with simple methods; 5. Multiple imputation
Time-to-event data	
Copula models	1. Copula models
Frailty models	2. Frailty models
Censoring	7. Correlated interval-censored data; 9. Left-censored data
Copula-frailty comparison	3. Copula–frailty comparison
Semiparametric approaches	4. Semiparametric and nonparametric methods
Measurement error	8. Measurement error
Recurrent events	7. Correlated interval-censored data
Additional results	5. Missing failure time indicators; 6. Regression methods; 10. Evaluation of treatment effects; 12. Bayesian methods

First, a number of **high-profile references** have been made. 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. It 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. In the book by [1713], several chapters, written by world renowned experts, are devoted to incomplete data and sensitivity analysis. [1551] devote a lot of attention in their book to missing data issues in clinical trials. [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. [1620] is a handbook on non-response in longitudinal studies. In a related fashion, 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. [1664] contributed a chapter to the Handbook of Epidemiology on incomplete data. 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. To add to these efforts [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. [1698] wrote an invited review for the Japanese statistical journal. Similarly, [1469] and [1470] wrote an invited contribution about incomplete data in clinical trials, with discussion. Conversely, [1696] wrote a discussion contribution to work by Lee and Nelder on unobservables, including missing data. [1453] prepared an invited review about analysis and sensitivity analysis of incomplete data. [1682] discuss the Points to Consider Document on Missing Data, as adopted by the Committee of Health and Medicinal Products (CHMP). In 2007, the CHMP issued a recommendation to review the document, with particular emphasis on summarizing and critically appraising the pattern of drop-outs, explaining the role and limitations of the *last observation carried forward* method relative to the use of mixed models. The initiative was taken by a working party of the organization, Statisticians in the Pharmaceutical Industry. [1627] offer a tutorial on handling incomplete data in the context of survival analysis.

Second, contributions have been made regarding avoiding the use of **less-than-optimal methods**. A still commonly used method is last observation carried forward. [1707] demonstrate that this approach always leads to implicit, highly implausible assumptions about the missing data mechanism. A general perspective on the issues arising is offered by [970].

Third, contributions have been made towards advocating **proper analysis tools**, for general use in a variety of applied settings. [425] show how hierarchical likelihood provides a convenient framework for fitting certain missing value models. 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. Also, 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.

Fourth, advanced modeling tools have been developed. When fitting **pattern-mixture models**, it may not always be straightforward how to derive marginal covariate effects, an issue studied by [1019]. In the same vein, [1188] propose a flexible marginal modeling strategy for non-monotone missing data.

Multiple imputation. [418] show, through re-weighing 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. [1626] developed a multiple imputation method for incomplete, multivariate, longitudinal nutritional data. The method is doubly iterative and makes use of quantile normal scores to transform to normality, performs imputations, and transforms back to the original scale. It operates well with high-dimensional, continuous or mixed-continuous, non-normal data. Binary or categorical variables can be accommodated. [1622] build upon the existing literature to formulate a class of hierarchical models for multivariate mixtures of Gaussian, ordered or unordered categorical responses and continuous distributions that are not Gaussian. Using MCMC and multiple imputation, it handles partially observed and incomplete data and missing data in generalized linear modeling. [1628] review the reasons why missing data may lead to bias and loss of information in epidemiological and clinical research. They discuss the circumstances in which multiple imputation may help by reducing bias or increasing precision, as well as describing potential pitfalls in its application. Finally, they describe the recent use and reporting of analyses using multiple imputation in general medical journals, and suggest guidelines for the conduct and reporting of such analyses.

Inverse probability weighting. [731] develop inverse probability weighting estimators, doubly robust estimators, and enriched doubly-robust estimators for controlled direct effects. [1402] develop marginal structural models for the impact of an intervention that takes place during a specified number of days (and may or may not persist afterwards), along with accompanying inference. The methodology is then used to assess the impact of an intensive care unit (ICU) patient acquiring infection on a given day since ICU-admission, versus not. [758] shows how IV-estimators for causal effects can be adjusted for systematic measurement error when an additional instrument for the measurement error is available. Grouped data are a special case of incomplete data. It is of great interest to be able to estimate a density reliably from (very) coarsely grouped data. A completely automatic Bayesian procedure is provided in [1012]. Apart from generalized estimating equations, also pseudo-likelihood has been considered. In its unweighted form, but allowing for MNAR, a practicable contribution has been developed by [1681]. [1697] develop a theory, based on single and double robustness, to apply pseudo-likelihood in a valid way under MAR.

Comparison between multiple imputation and inverse probability weighting. Such a comparison was undertaken by [777]. The LSHTM team has devoted additional research efforts to both multiple imputation, with a paper by [971] on the one hand, studying the use of multiple imputation in clinical and epidemiological studies, and a paper by [1675] on the other, investigating doubly-robust inverse probability weighting methods. [1398] offers a discussion on the importance of missing data assumptions that impose no testable restrictions on the observed data law. As an alternative to existing but often computationally complex estimation methods, [985] developed MCMC-based methodology. Two methods are commonly used to report on evidence carried by forensic DNA profiles: the so-called *Random Man Not Excluded* (RMNE) approach and the likelihood ratio (LR) approach. It is often claimed a major advantage of the LR method that drop-out can be assessed probabilistically. In [747], a new RMNE measure is proposed that likewise accounts for allelic drop-out in an observed forensic DNA profile. Also in [969], MCMC methodology is used. Indeed, this paper describes a model for multilevel mixed response data, together with a MCMC algorithm for fitting it. Its application to multilevel multiple imputation is then demonstrated.

[1638] replace the response distribution and logistic dropout model of the Diggle-Kenward model for incomplete longitudinal data by robust alternatives and fit using hierarchical likelihood.

2.3.2 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 analyses to explore the impact of potential deviations of this assumption in the direction of missingness at random.

First, the **dependence on assumptions** of models and their conclusions need to be carefully investigated. 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. 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. It is clear that such investigations are not a routine exercise, as further demonstrated in [1000], who point out the dangers of using standard methodology and offer formal and informal sensitivity analysis tools.

The intrinsic nature of sensitivity, owing to non-identifiable model parts, is studied in [995], who established that every missing not at random model has got a missing at random counterpart with equal fit.

It is therefore important to propose proper, carefully crafted tools to allow for exploration of such sensitivities and their impact on conclusions from complex models.

Second, the **shared-parameter models** framework has received considerable attention. [688] develop a new shared-parameter model, in which they avoid making parametric assumptions for the random effects distribution and leave it completely unspecified. For the estimation of the proposed model, a semiparametric method is used. [987] formulate a so-called latent-class mixture model that unites aspects of selection, pattern-mixture, and shared-parameter models. It offers an alternative basis for sensitivity analysis tools. [1684] extend the class of shared-parameter models, allowing to incorporate the missing at random mechanism within it, a feat previously considered impossible. They then explicitly develop sensitivity analysis tools for this framework [1232].

Further in the shared-parameter framework, [1304] and [1359] propose a class of non-ignorable models for handling non-monotone missingness in categorical longitudinal responses. This class of models includes the traditional selection models and shared parameter models. This allows us to perform a broader than usual, sensitivity analysis. In particular, instead of considering variations to a chosen non-ignorable model, they study sensitivity between different missing data frameworks. An appealing feature of the developed class is that parameters with a marginal interpretation are obtained, while algebraically simple models are considered. Specifically, marginalized mixed effects models are used for the longitudinal process that model separately the marginal mean and the correlation structure. For the correlation structure, random effects are introduced and their distribution is modeled either parametrically or nonparametrically to avoid potential misspecification.

Third, **a suite of sensitivity analysis tools** has been developed. [1648] provide an integrated sensitivity analysis, using a suite of tools, to the Slovenian plebiscite data, an often-studied set of data in the context of incompleteness. 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. [1686] 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. In ability testing, often a set of items is given to a sample of examinees with the instruction to solve them. Almost always there is a time limit, and this induces time pressure. As a result, towards the end of the test examinees will typically start to drop items, giving rise to missing data. [600] developed a probabilistic model for test speededness, and have investigated whether it is possible to detect persons that are particularly sensitive to time pressure using local influence diagnostics [1686] (see also WP2).

[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 trial to improve the quality of peer review for the *British Medical Journal.*

[289] study issues surrounding missingness, specific to the context of quality of life data from HIV clinical trials. [287] show 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.

2.3.3 Censored survival data

First, a lot of work has been done in the area of **copulas**. [986] prove 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] study 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] solve 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. [782] extend the Koziol-Green model to dependent censoring. Weak convergence is shown of a nonparametric estimator of the conditional distribution functional and its efficiency over the general copula-graphic estimator is established. [845] give further consideration and prove the weak convergence of the conditional copula-graphic estimator. Furthermore, they develop an asymptotic confidence band for this estimator. [829] explore nonparametric estimators for several functionals of the distribution function when the observations are subject to right censoring and when covariate information is taken into account. New results are obtained for the mean and median of the residual lifetime in a regression context. In nonparametric regression, it is common to assume independence between the error and the covariate. [686] generalize this by assuming a family of copula functions instead of independence. They establish the asymptotic normality of the parameters of the copula family and the weak convergence of the conditional distribution of the response. [1015] obtain a very general result on empirical likelihood with non-smooth criterion functions in which copula estimation appears as an important particular example. A full study of kernel estimation of copulas is done in [1016]. Weak convergence is the main result and goodness-of-fit testing is one of the applications. [1662] consider a semiparametric and a nonparametric 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.

Second, a very popular way of incorporating heterogeneity, association, and extra dispersion in time-to-event data is through **frailties**, a particular form of random effect. [1025] further publish the book entitled 'The Frailty Model' with Springer, a contemporary treatise on the frailty model and its extensions. 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]. [807] propose an alternative approach in which the original problem of ‘fitting a frailty model’ is reformulated into the problem of ‘fitting a linear mixed model’ using model transformation. Based on a simulation study, they show that the proposed method performs well for data sets with moderate to large samples sizes within covariate level subgroups in the cluster. This transformation idea also works for multivariate proportional odds models and multivariate additive risks models. [1023] develop and compare different techniques to fit frailty models with complex correlation structure, i.e., correlated frailty terms or hierarchical models. New frailty model methodology is also developed by [1659] to study the heterogeneity of prognostic indices in cancer over different centers. An important aspect of a prognostic index is its homogeneity over different centers, which makes that the prognostic index is generalizable. Frailty models with random center and center by prognostic index effects are fitted for this purpose. [1653] consider methods to model interval censored data with clustering, occurring in infection times of udder quarters of cows. They show that frailties can still be integrated out analytically, in the case of the gamma distribution, although it leads to quite complex likelihood expressions. Proportional hazards (PH) models with multivariate random effects (frailties) acting multiplicatively on the baseline hazard are a topic of intensive research. When analyzing multi-center clinical trials, such a model, including a random center effect and a random treatment by center interaction can be used to investigate heterogeneity in outcome and in treatment effect over centers. Several estimation procedures are proposed to deal with this type of models. [447] propose another approach to fit such a model generalizing a random effects accelerated failure time model. The generalization consists in using a penalized Gaussian mixture as an error distribution on top of multivariate random effects that are assumed to follow a normal distribution. The proposed method is illustrated on the same data base as before. In [1659], the same model is used in a different context with the objective now to investigate a further insight in validation of prognostic index (PI). Indeed, a major issue when proposing a new PI is its generalizability to daily clinical practice. While most validation techniques assess whether 'on average' the results obtained by the prognostic index in classifying patients in a new sample of patients are similar to the results obtained in the construction set, the heterogeneity of the prognostic index risk group hazard ratios is now considered over different centers.

Third, **copula and frailty models are contrasted**. [979] study the similarities and differences between frailty models and copulas. It is often stated in literature that frailty models with a particular density function for the frailties characterized by its Laplace transform correspond to Archimedean copulas. The authors demonstrate in their paper that this statement is only true for the particular case of Weibull marginal distributions with the positive stable density or copula.

Fourth, a lot of work has been devoted to **semiparametric and nonparametric methods**. For censored data, nonparametric estimation of a density function, using ideas of pre-smoothing has been studied in [605]. Pre-smoothing of the Kaplan-Meier estimator has been shown to be more efficient. In [605], local linear estimation techniques are used in the pre-smoothing step, either directly or via a local logistic approach. This type of pre-smoothing is compared to a Nadaraya-Watson type of pre-smoothing that had been used before. An extensive simulation study revealed that the use of a local logistic pre-smoothing leads to very good performance. [898] propose 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. They propose maximal partial likelihood estimators, establish the large sample properties of the estimators, and illustrate the results by a short simulation study and an application to a real data set in demography. In [920], under a nonparametric regression model, [1583] introduce two families of robust procedures to estimate the regression function when missing data occur in the response. The first proposal is based on a local M-functional applied to the conditional distribution function estimate adapted to the presence of missing data. The second proposal imputes the missing responses using the local M-smoother based on the observed sample and then estimates the regression function with the completed sample. They show that the robust procedures considered are consistent and asymptotically normally distributed. A robust procedure to select the smoothing parameter is also discussed. [922] study the problem of specification tests for conditional models when the data are subject to left truncation and right censoring. A general method is applied to derive tests for the polynomial regression, the proportional hazards, the additive risks and the proportional odds models. Bootstrap versions to approximate the critical values of the test are introduced and proved to work both from a theoretical viewpoint as well as in a small simulation study. In longitudinal studies of disease, patients can experience several events through a follow-up period. In these studies, the sequentially ordered events (gap times) are often of interest. The events of concern may be of the same nature (e.g., cancer patients may experience recurrent disease episodes) or represent different states in the disease process (e.g. alive and disease-free, alive with recurrence and dead). If the events are of the same nature this is usually referred to as recurrent event, whereas if they represent different states (i.e., multi-state models) they are usually modeled through their intensity functions. [934] present nonparametric estimators for several quantities in a progressive three state model. They derive a simple estimator for the bivariate distribution function for censored gap times and estimators for the transition probabilities. The main objective of [957] is the nonparametric estimation of the regression function with correlated errors when observations are missing in the response variable. Two nonparametric estimators of the regression function are proposed. The asymptotic properties of these estimators are studied; expressions for the bias and the variance are obtained and the joint asymptotic normality is established. A more complex situation in survival analysis occurs when for censored data some failure indicators are missing (under different mechanism of missingness). 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. A key issue in [1572] is to study the estimation problem for Cox's proportional hazards regression models using a framework that extends the theory, the computational advantages and the optimal asymptotic rate properties of the so-called Dantzig selector to the class of Cox's proportional hazards under appropriate sparsity scenarios. The Dantzig selector is a recent approach to estimation in high-dimensional linear regression models with a large number of explanatory variables and a relatively small number of observations. As in the least absolute shrinkage and selection operator (LASSO), this approach sets certain regression coefficients exactly to zero, thus performing variable selection. The paper [1572] is the first to use the Dantzig selector in regression models for survival data with censoring. Some overview papers related to this workpackage have been written. An overview of the famous Kaplan-Mayor estimator can be found in [1237]. A more extensive overview work on censored data, including a discussion on various censoring schemes, methods for estimating a survival and a hazard function, and an overview on the most popular parametric regression models is in [1236]. Apart from estimation of the cumulative distribution function or the hazard function, it is also of interest to estimate quantiles (e.g. the median). Censored depth quantiles were studied in [593]. The above procedures naturally involve the development of goodness-of-fit tests: in [980] a first set of procedures (achieved in collaboration with the European partner of Santiago de Compostela) is based on empirical processes, and tests parametric conditional means and variances, while second a method based on the difference between two residual distributions is proposed in [1045] and [1141]. Consider a nonparametric location-scale regression model, where the error is independent of the covariate. In [1117] and [457] the authors construct tests for the hypothesis that the regression function belongs to some parametric family. The former paper assumes that the response and the covariate are subject to selection bias, whereas the latter paper allows the response to be subject to random right censoring. The proposed tests compare the nonparametric MLE (NMLE) based on the residuals obtained under the assumed parametric model, with the NPMLE based on nonparametric residuals. In the setting of a two-sample problem with data subject to left truncation and right censoring, [1070] develop an empirical likelihood method to do inference for the relative distribution. They obtain a nonparametric generalization of Wilks' theorem and construct nonparametric pointwise confidence intervals for the relative distribution.

Fifth, a more complex situation in survival analysis occurs for censored data with **missing failure time indicators.** 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.

Sixth, in particular UCL has been very active in the area of **regression approaches** with incomplete, mostly censored, data. Hence, there is a close connection with WP1. [1145] address estimating nonparametrically a regression function, when the response variable is subject to various filtering schemes. Both the mean and the median regression case are considered. The authors present a completely nonparametric estimation methodology, assume that the error is independent of the covariate, and show how to construct a more efficient estimator that takes account of the common shape. In [484], the authors first propose a new estimator of the joint distribution of a d-dimensional covariate X and a univariate response Y, that is subject to random right censoring. The estimator overcomes the common curse-of-dimensionality problem, by using a new dimension reduction technique. Second, they assume that the relation between X and Y is given by a single-index model, and propose a new estimator of the parameters in this model. Further, consider a nonparametric location-scale regression model, where the error is independent of the covariate, and the response and the covariate are subject to selection bias. [1117] construct tests for the hypothesis that the regression function belongs to some parametric family. The proposed tests compare the nonparametric MLE (NPMLE) based on the residuals obtained under the assumed parametric model, with the NPMLE based on nonparametric residuals.

In [1153] and [1152], a semiparametric time-varying coefficients regression model is considered, which reduces for special choices of the link function to e.g., the additive hazards model or the Cox proportional hazards model with time dependent coefficients. The response is supposed to be subject to left truncation and right censoring. The authors construct an omnibus goodness-of-fit test for this model, and they develop the asymptotic distribution of the proposed test. Finally, consider a pair of gap times corresponding to two consecutive events, which are observed subject to random right censoring, and suppose the gap times satisfy a nonparametric location-scale regression model. Under this model, [495] propose a nonparametric estimator of the distribution of the error variable. They then use the proposed estimator to introduce nonparametric estimators for important targets such as: (a) the conditional distribution of the second gap time given the first; (b) the bivariate distribution of the gap times; and (c) the so-called transition probabilities. Further, [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 nonparametric kernel estimation of the regression function. In [64], this estimator is extended to the case of any (nonlinear) parametric regression model. In addition, [63] consider the estimation of the regression function under a nonparametric 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 nonparametric estimator of the regression function, which usually behaves quite poorly when the censoring proportion is large. When the conditional location function is not supposed to have any parametric form, new procedures to estimate general L-functionals (conditional mean, trimmed mean, quantiles...) are provided in [1140] and [63]. All five papers are closely connected to WP1.

Seventh, [171] provide 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] study the AFT setting from a Bayesian angle.

Eighth, 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 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. [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 models 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.

Ninth, 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.

Tenth, the **evaluation of treatment effects** still poses fundamental statistical challenges in terms of both analysis and design. Several such prevalent problems are studied. First the evaluation of cause-specific failure time outcomes is examined when the cause of failure assessment is prone to error. This problem is well known to occur in many data registers for instance, and is perhaps at its worst in a developing world setting where cause of death is obtained through verbal autopsy. Inspired by the partial likelihood approach an optimal test is derived for the treatment effect when the type 1 and type 2 error are known and the alternative treatment effect is confined to the cause-specific hazard. For the envisaged alternative, the test can gain substantial power. Under contiguous alternatives, sample size calculations for this test are derived as well as for the corresponding all-cause mortality and naive cause-specific log-rank test which ignores the misclassification. It is found that this can lead to substantial savings in sample size and may lead to new options at the design stage which are also discussed. [1420] build on this to derive more precise and asymptotically unbiased estimators for the relative cause-specific hazards, allowing for a non-constrained relationship between the hazards for different causes ([1420]).

Eleventh, various **specific applications** are considered. [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 nonparametric maximum likelihood estimator of the survival function is derived. [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].

Twelfth, a Bayesian perspective is taken. [1029] discuss the use of Bayesian P-splines and of the composite link model to estimate survival functions and hazard ratios from interval-censored data. Under the assumption of proportional hazards, the proposed strategy provides a smoothed estimate of the baseline hazard along with estimates of global covariate effects. The frequentist properties of the Bayesian estimators are assessed by an extensive simulation study and further illustrated by three examples. [1061] show how penalized B-splines combined with the composite link model can be used to estimate a bivariate density from histogram data. Two strategies are proposed: the first one is semiparametric with flexible margins modeled using B-splines and a parametric copula for the dependence structure; the second one is nonparametric and is based on Kronecker products of the marginal B-splines bases. Frequentist and Bayesian estimations are described. In the context of grouped data, [1658] propose a combination of the composite link model with roughness penalties to estimate smooth densities in a Bayesian framework. A simulation study is used to evaluate the performances of the strategy in the estimation of a density, of its quantiles and first moments.

2.3.4 General incomplete data structures

Unobserved data structures can take forms beyond missing data and censored survival data. An important latent structure is random effects, often in conjunction with incomplete data. An encompassing name, coined by [1021], is **data enrichment**. These authors, in line with [995], demonstrate that the predictive distribution of unobserved outcomes in the missing data context, but also the posterior random-effects distribution, are unobservable from the data alone and, therefore should be approached with the utmost caution. Such findings are important when pursuing model assessment with latent structures, such as done in [775, 804] for the random-effects distribution in the generalized linear mixed model. The case of latent classes, another type of unobservables, has received attention in [1681].

[600] propose an **item response model** for dealing with so-called test speededness. The model consists of two random processes, a problem solving process and a random guessing process. [1686] apply local-influence diagnostics for the detection of test speededness in a model describing non-response in test data. This approach is compared to the optimal person fit index (Drasgow and Levine, 1986).

[968] provide empirical evidence that the so-called **Copas selection model** can usefully be used in a large fraction of meta-analyses in clinical studies.

[1378] adjust for noncompliance in the analysis of randomized clinical trials using an **instrumental variables approach** based on linear and log-linear Structural Mean Models. They further propose formal goodness-of-fit tests for model checking. Because compliance measurements are often subject to systematic measurement error, [1399] develop techniques to adjust for systematic measurement error when information is available on an instrumental variable for the measurement error which is associated with the exposure, but known not to modify the treatment effect of interest. The impact of measurement error on other estimators of causal effects in point-treatment studies is investigated in [1371], who find that different methods of confounding adjustment (regression adjustment, propensity score adjustment and inverse weighting) are typically equally affected by measurement error in the exposure.

Also in the HIV area, E. Goetghebeur (UG) is co-author of a report ([763]) 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. Sequentially randomized trials [1409] enable unbiased evaluation of how to best adapt treatment interventions to a history of outcomes with the goal to optimize future treatment adherence and/or response. To estimate the treatment effect one must then allow for time-varying covariates predicting possibly treatment as well as outcome. Doubly robust estimators are developed for this purpose. Furthermore, HIV prevention trials (see [763]) generally, and microbicide trials more specifically, again suffer great power losses in clinical trials, among other reasons due to substantial treatment non-adherence (see [1418]). This has repeatedly resulted in very expensive trials ending with inconclusive results. In several manuscripts we have sought to find statistical solutions for this and related problems. The authors have examined for instance the role of reduced hazard ratios over time due to risk heterogeneity in the dataset and factorial designs to allow for testing of different adherence enhancing interventions.

2.4 Workpackage 4: Data with latent heterogeneity

The contributing partners

All Belgian partners and most international partners (UU, LSHTM) have contributed to WP4.

A summary description of the objectives

WP4's objectives are summarized as follows:

> "Unobserved heterogeneity can be modeled in different ways. A natural and common way to model this heterogeneity is by means of mixed models, which have been extensively studied during Phase V. The gained expertise opens the door to study in particular mixed models with partially specified residual dependencies conditional on the values of the random effects, and generalized linear mixed models. For the latter model, flexible models for the random effects distribution will be investigated, like mixtures of normals to approximate a B-spline basis."

The main achievements in relation to the initial objectives

All initial objectives have been met. First, classical models with latent structures have been extended in various ways. Mixed models with flexible random-effects distributions have been proposed by several partners, and publicly available software has been developed, allowing users to apply many of the proposed models. Second, various proposals have been made to incorporate residual dependencies in the model, through, e.g., copula models. A further extension of the models is towards modeling multiple outcomes simultaneously, e.g., multivariate longitudinal profiles, multivariate survival data (link with WP3), or longitudinal and time-to-event data (link with WP3). Again, not only novel models have been proposed, also software has been developed. Finally, a new research line has been started in which random effects are introduced at various places in mixed models, leading to much more flexible models than the ones traditionally used in applied statistics.

Comments in case of deviations from the initial project workprogram

As indicated, the main achievements have been met. Moreover, the summary clearly shows that several other new research lines have been initiated, by individual partners, but also jointly by several partners of the network. While these new research lines still need to be developed further, many results are available yet, and have been published in the international literature.

Many statistical models in which latent structures are used to explain observed heterogeneity are developed and investigated. The classical linear, generalized linear and non-linear mixed effects models are 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 are 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 is also 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 are developed as well as score tests for the random effects structure and techniques for prediction. A rather new development is the use of random effects to build models for multivariate longitudinal data, i.e., data sets in which multiple outcomes are measured repeatedly over time. Data from various disciplines (medical, psychology, etc.) are taken to illustrate the performance of the various proposals.

In many situations, statistical models are used that assume the presence of latent, unobservable, structures to explain the variability observed in the data. Exactly the fact that those structures, by definition, can never be observed poses particular problems with respect to identifiability, as illustrated in [1021]. Nevertheless, those models have proven to be very useful in various contexts, and some illustrations of this are summarized in Section 2.4.1. Overview works are [1713], an edited handbook on longitudinal data analysis, [1714], the new edition of Verbeke and Molenberghs (2000) on linear mixed models for longitudinal data, and the book by P. Kroonenberg on multiway data, a review of which has been presented in [874].

In a number of applications, latent structures are needed at various levels, implying models with so-called crossed random effects. Some examples are discussed in Section 2.4.2. Many statistical models employ latent structures to generate association structures in correlated data settings. One particular application is the construction of models for the joint analysis of multiple outcomes. A number of examples are discussed in Section 2.4.3. Many standard statistical models are based on very strict underlying assumptions. So are many mixed models. In specific applications, model assumptions need to be relaxed in order for the models to be realistic for specific data sets at hand. Examples are presented in Section 2.4.4. Although the models studied in this workpackage have been around for several decades, they still pose specific problems with respect to model fitting, and there is often lack of software to fit the particular models needed to answer subject-matter research questions. Contributions on this are discussed in Section 2.4.5.

2.4.1 Applications of models with latent structures

When the latent structures are assumed to be of a continuous nature, one traditionally refers to the models as mixed-effects models. Often the models are linear, generalized linear, or nonlinear regression models in which some parameters are assumed to be sampled from a continuous distribution, often assumed (multivariate) normal. Those random effects can have various interpretations such as levels of general practitioners in a multicenter study ([663, 1261, 1324]), levels of hospitals in a study comparing various hospitals with respect to patient safety ([679, 1305, 1322]), patient effects in longitudinal studies ([667]), study effects in a meta-analysis ([1261]), or gene effects in the context of transgenic improvement of a commercial crop ([1685]). Sometimes, the latent variables are assumed to be discrete. The models are then referred to as mixture models. Sometimes, mixture models are used because they are very flexible, often posing less strict assumptions than other parametric models. An example is given by [987] for the analysis of longitudinal data subject to dropout. Another advantage of mixture models is that they provide a natural way of detecting clusters and/or of classifying observations in mixture components. For example, [705] apply mixture models to detect chromosomal and segmental aneuploidies in single cells, based on smoothed clone-specific posterior probabilities for the presence of deletions or duplications.

Another area where latent structures are applied is in modeling differential item functioning (DIF), referring to the phenomenon that questions if a test may mean something different to different groups. It is important to detect DIF in order to avoid test bias (and potential law suits).

On the one hand, novel models are constructed that incorporate DIF, see [568], and [1249]. On the other hand, new DIF screening tests are proposed as well, see [1177] and [1178]; those tests are based on robust statistics and outlier detection. In a strand of research related to DIF, methods are developed that allow to investigate whether latent heterogeneity is due in part to qualitative or group differences, see [583] and [143].

Often, random effects are used to implicitly model correlation structures in the data. A typical example is the analysis of repeated measurements, where the correlation between repeated measurements from the same subject is modeled through random subject-specific coefficients, shared by those repeated measurements. Examples can be found in [575, 699]. In the latter paper, three different mixed models are compared for the analysis of scores that are constrained within an interval and measured repeatedly over time. A book chapter on popular models for longitudinal data, including mixed models is [1312].

A large amount of work is done in the context of the evaluation of surrogate markers in clinical studies ([253, 254, 255, 256, 257, 268, 270, 274, 296, 811, 785, 811, 785, 784, 786, 1670, 1695, 1680, 1704, 1670, 1244, 1680, 1670, 1695]). 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, 825, 852, 853, 884]). 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. Both continuous and non-continuous data can be handled in this way. The Rasch model in psychometric research, too ([600, 567, 595, 635]), has received a lot of attention. Other examples of mixed models in the psychometric literature can be found in [619, 941, 1215, 1250].

Mixed models are also very useful 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 a 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 are applied to cDNA data in [284].

In [1334], mixed models are used to classify subjects into two or more prognostic groups using longitudinally observed values of markers related to the prognosis. The proposed method proceeds in two steps as described earlier in the literature. First, multivariate linear mixed models are fitted in each prognostic group from a training data set to model the dependence of markers on time and on possibly other covariates. Second, the fitted mixed models are used to develop a discrimination rule for future subjects. The method improves upon existing approaches by relaxing the normality assumption of random effects in the underlying mixed models. Namely, a heteroscedastic multivariate normal mixture is assumed and inference is performed within the Bayesian framework using the Markov chain Monte Carlo methodology. Software is written for the proposed method and it is freely available. The methodology is applied to data from the Dutch Primary Biliary Cirrhosis Study.

An area in which mixed models are also applied is the analysis of data from microarray experiments. For example, in [1300], a calibration method is proposed for preprocessing spiked-in microarray experiments based on nonlinear mixed-effects models. This method uses a spike in calibration curve to estimate normalized absolute expression values and asymptotic confidence intervals for the estimated expression values are constructed. Other contexts in which non-linear mixed models are applied are in modeling regeneration of blood vessels, see [383], or in the modeling of drug concentrations in plasma levels in pharmacokinetic studies (see [1109, 481]). Related to the latter context is the use in [482] of P-splines in a varying-coefficient regression model for estimation of the association between receptor occupancy and drug concentration in plasma measurements.

Another project in which models with latent structures are applied many times is the Signal-Tandmobiel study data ([210, 183, 187]). For example, the variability of permanent tooth emergence is examined in [193].

For the analysis of discrete bounded-outcome score data, [195] suggest 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 [685].

In [586], mixed models are used to model claims reserving and credibility in the context of insurance and actuarial sciences.

2.4.2 Crossed random effects

Crossed random effects models are an exciting tool to deal with cases in which latent heterogeneity is present in two modes of the data. As an example, one may think of person by item data with latent heterogeneity in both the person and the item mode. Linking up with this example, a crossed random effects approach is shown to be very fruitful in the context of item response models, see [459] and [1248], which includes a latent regression model for the item difficulties. Beyond the case of person by item data, we have extended the crossed random effects approach to structural equation models for three-way data in [941], which allows to investigate structural relations among units of two of the three modes. Crossed random effects can also be generalized to analyze data from complex experimental designs as shown in [573] and [615], the analysis of multivariate network data, see [1201], and the modeling of repeated decision trials, see [1254] and [1181]. The algorithmic approximation work reported in [614], based on variational methods, is also shown to be promising in the context of crossed random effects models.

2.4.3 Latent structures for joint modeling

Many subject-matter research questions can only be answered by modeling several outcomes jointly. Assuming a mixed or mixture model for each outcome separately and allowing the latent structures to be common or correlated for the various outcomes, implies joint models in a very straightforward way. For example [669] uses this approach to jointly model 4 different markers, measured longitudinally in patients who have undergone renal transplantation. The aim is to anticipate a failure of the renal graft. It is shown that this joint model allows for a much better prognosis than what was available before based on univariate models. In [683, 688] latent variables are used to jointly model a survival outcome with a longitudinally measured binary outcome. In [559], a latent variable framework is proposed to model multivariate binary anomaly data as provided in some teratology studies. The usual conditional independence assumption, i.e., the assumption that the outcomes are independent conditionally on the latent variables, is relaxed by means of copula functions and the distribution of the latent variable is kept flexible through a finite mixture of normals. Another example of the use of random-effect models for the construction of joint models is [1303], in the context of longitudinal data subject to missingness. In [1366], latent variables are used to derive the joint distribution of bivariate ordinal outcomes, measured longitudinally. Specifically, the ordinal outcomes are related to a bivariate latent variable, which is modeled over time using a linear mixed model. Random effects terms are used to tie together all repeated observations from the same subject. The cross-sectional association between the two outcomes is modeled through the correlation coefficient of the bivariate latent variable, conditional on random effects.

2.4.4 Model extensions and/or flexible models

Many standard mixed and mixture models rely on very strict assumptions. A new, though important area of research is goodness-of-fit for assessing random-effects models, and the development of model extensions which pose less strict assumptions than the traditional models. Several partners have contributed to this topic. The assumption of a normal random effects distribution is extended by Ghidey 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 [670] it is shown that the approach compares favorably to three competing approaches. Further, this technique is 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 is used to generalize the distribution of the baseline risk, see [184]. In [673], the random effects distribution of a generalized linear mixed effects model is assumed of this type. The same approach is applied in a variety of survival applications using a Bayesian approach, see [188, 447, 190, 439, 1659] for applications of this approach in accelerated failure time models and frailty models whereby the random effects and/or the measurement error distribution is assumed to be of the above mixture type. A similar model is used in [662] for the estimation of the association between two interval-censored outcomes. They show that the approach works well in many realistic situations and that it is superior to other existing approaches. In [167] the approach is used 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 [688, 175] 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. In [375] one therefore combines smoothness and log-concave shapes to provide effective nonparametric models. In [657], 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. Nonparametric 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 is applied in [656] to the same dental problem as in [190]. A two-part shared parameter model is suggested in [683] 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.

Models with flexible assumptions about latent structures are applied in many different contexts. For example, [910] compare mortality data from different countries (at the same age and in the same year) using models with a smooth, nonparametric, mixing distribution of the log mortality rates. In [1323], some recently suggested approaches to flexibly estimate the random effects distribution in linear mixed models are compared. In a Bayesian context, [1332] relaxing generalized linear mixed models by allowing random effects to follow a mixture of multivariate Polya trees, center on the commonly used parametric normal family. This work provides a robust alternative to the standard models. Issues on implementation and computational challenges are discussed as well. In a dental context, [682] model zero-inflated counts observed for the dmft-index which is the sum of milk teeth with caries experience. Zero-inflation occurs because of correlation among caries experience on teeth in the same mouth, but this also implies the overdispersed distribution of non-zero counts (negative binomial distribution). Since scoring caries is not always straightforward, an appropriate analysis needs to correct for misclassification. In [580], the authors combine the Wiener diffusion process for choice response times with techniques from psychometrics in order to construct a hierarchical diffusion model. This leads to a modeling framework that is highly flexible and easy to work with. The same authors also consider the Ratcliff diffusion model, a well developed process account of the time course of human decision-making in two-choice tasks, and suggest adopting Bayesian estimation methods of the parameters. This method promises to broaden the scope of psychological problems the models can address ([618]). An application of those Bayesian tools is described in [1197], in which practice effects on the diffusion model parameters are investigated. In another project, the model is extended to allow also for latent between-person heterogeneity ([1181]).

Other work in the context of formulating models with both normal and non-normal (conjugate) random effects for longitudinal, clustered, and otherwise hierarchical data can be found in [454, 1065]. These models have proven very useful, e.g., in psychometric validation, surrogate markers, and whenever association functions are of interest ([1706]). While mixed models are particularly useful to model mean and/or dispersion functions flexibly, allowing for overdispersion (see, e.g., [1342]), alternatives are studied in [1239], who also compare their approach to mixed models.

In some situations, it is considered a drawback that the parameters in mixed models have a cluster-specific interpretation (i.e., conditional on the latent structures), rather than a population average interpretation. In [208] a latent-variable model is suggested for longitudinal bivariate ordinal data, whereby the fixed-effects parameters retain their population averaged interpretation when properly scaled.

Finally, mixed models often are based on the so-called assumption of conditional independence. The impact of deviations from this assumption is studied by [680], in the context of diagnostic testing. In the context of item response models, this is also investigated in [1206] via a new family of item response models, and in [127] via copulas. The copula-based approach, which allows for quite a flexible modeling of residual dependencies, is also successfully applied in a biostatistical context with a study of teratology, see [1188]; otherwise, in this study copulas are combined with a mixture model to allow for qualitative latent heterogeneity. In addition, a copula-based approach is also developed to capture residual dependencies in structural equation modeling for categorical data (where coupled error terms often lead to a high computational burden, for which copulas present a neat way out), see [1159].

2.4.5 Development of estimation methods, inference tools, and software

Since model extensions are often required to address subject-matter research questions adequately, standard estimation methods, inferential procedures, and/or standard software tools often do not suffice to fit the models at hand or to answer the research questions of interest. In this workpackage, several teams have contributed to the development of new estimation theory, fitting algorithms and/or software.

For the fitting of mixed models, with continuous latent variables, standard software relies on approximation methods of the model or the data. Methods such as marginal quasi-likelihood (MQL) and penalized quasi-likelihood (PQL) are easy to implement, but often lead to biased parameter estimates. Alternatively, quadrature methods can be used, but they are often very unstable and time-consuming, especially in models with many random effects. [670] compare various methods for the estimation of the latent variable distribution, in the context of linear mixed models. In [1645, 658] a fully exponential Laplace approximations is proposed, which is shown to be fast and stable. An alternative is presented by the authors of [614], who propose variational estimation, not requiring numerical evaluation of integrals, based on a lower bound approximation of the logistic model. In the context of frailty models for survival data, [807] propose a novel estimation method based on a cumulative hazard transformation of the data such that a linear mixed model can be used for parameter estimation and inference. The properties of the method are studied by using a simulation study and a real life case study. In [651, 1257], the statistical properties of the Expectancy Valence model of the Iowa gambling task are studied. A Bayesian hierarchical estimation procedure is implemented and applied to data from an experimental study. Finally, [574] propose an iterative maximum a posteriori estimator in item response theory as an enhanced technique for estimating proficiency levels. This new estimator is compared to classical methods by means of simulations. Also in a Bayesian framework, [969] propose a MCMC algorithm for fitting a multilevel mixed model and apply this for multilevel multiple imputation. As discussed in Section 2.4.3, mixed models are often used to generate joint models for multiple outcomes.

This often leads to high-dimensional latent structures, implying model fitting to be complex. In [180], [443] and [211], a computational procedure is developed to obtain parameter estimates for a multivariate mixed 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 presented as well. In [647, 107] 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 the same context, various types of multivariate multilevel continuous time models are developed, and associated Bayesian inferential techniques are proposed, see [575], [1179], and [1214]. 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. [373] propose a novel way to estimate a nonparametric two-dimensional latent distributions in meta-analysis with case-control data. [439] examine various estimation approaches for frailty models and compare these by using a simulation study. None of the approaches proves to be uniformly better than another.

It has been shown that inference in mixed models often does not follow standard statistical theory. One such situation is the testing of variance components. [452] and [461] study the non-standard behavior of the score, Wald, and likelihood-ratio tests for variance components, in the case the null hypothesis is on the boundary of the parameter space.

As far as software development is concerned, [561] and [106] propose estimation methods, possible with easily available software, for the fitting of a one/two-dimensional mixture model for DIF and proficiency levels, respectively. [617, 158] present a software tool freely downloadable, intended to make the Ratcliff diffusion model (see also [114]) for reaction time and accuracy data more accessible to experimental psychologists. An R package to apply a variety of DIF methods is presented in [1176]. In [560] a MATLAB IRT modeling (IRTm) toolbox is developed, which can fit a large variety of IRT models, with inclusion of recent developed copula IRT models to handle local item dependencies. Within the Bayesian context, [186] propose a freely downloadable R package, DPpackage, allowing models with Dirichlet Processes, Polya Trees, mixtures of Triangular distributions, and Random Bernstein polynomials priors. Finally, a copula approach to accommodate deviations from the conditional independence assumption is included in a Matlab toolbox which incorporates a large variety of IRT models, see [1189].

2.5 Workpackage 5: High-dimensional and compound data

The contributing partners

The contributing partners to this workpackage consist of UG, KUL-1, UCL, UH and the European partners UJF, UU and USC. Joint work has taken place mainly between UG and KUL-1 and also with the European partners.

A summary description of the objectives

High-dimensional and compound data require the development of specific analysis tools. To reveal and model association or dependence in real-valued object-by-variable data involving a high number of variables (e.g. micro-array data) novel models based on dimension-reduction techniques will be developed. New multiple testing procedures are proposed to promote reproducibility in the presence of such high numbers of variables. Additional information will be incorporated, in the form of one (or more) added dimensions to the object-by-variable data (thus creating multi-way data) or through extra high-dimensional blocks of data from different sources pertaining to the same problem (compound data) and new statistical techniques will be developed for such multi-way and compound data.

The main achievements in relation to the initial objectives

Workpackage 5 has three main initial objectives; for each of these objectives major contributions are made in the different disciplines in which researchers are active. We now briefly discuss some of the important achievements for these three initial objectives:

- *Dimension-reduction*: Different new dimension-reduction techniques are developed, for instance for micro-array data. Also state of the art robust techniques are developed in a joint research effort between UG and KUL-1, with important applications in data mining.
- *Multiple testing and model selection/validation*: The multiple testing problem is, for instance, tackled in the context of gene screening with many possible candidates. New techniques for model selection and validation are developed both in the context of bioinformatics, data mining and psychometrics.
- *Analysis for multi-way and multiblock/multiset data*: General methods are proposed for the analysis of these types of data. Such methods can be applied in any of the three different disciplines.

Comments in case of deviations from the initial workprogram

This workpackage has made major contributions to any of the initially stated research problems, and is therefore not deviating from its initial workprogram.

High-dimensional data and compound data occur in different disciplines. Most of the methods developed in this workpackage are concerned with a specific discipline; the research output in this workpackage is therefore split up into the three disciplines bioinformatics, data mining and psychometrics. Obviously many of the methods developed in one discipline might be highly relevant for another discipline. We collect techniques that are developed in more than one discipline in the fourth section, 'Across different disciplines'.

2.5.1 Bioinformatics

A first and important research topic in bioinformatics concerns the development of tests for statistical and sufficient cause interaction, with a focus on testing for gene-environment interaction, on mediation analysis (i.e., disentangling direct and indirect exposure effects) and on the adjustment for time-varying confounding. Each of these are discussed in turn below.

[748] propose tests for gene-environment interaction on the basis of complex traits in family studies. Desirable features of the approach are that it enables the estimation of genetic effects in

family studies (whereas the previous focus has been on testing), that it adjusts for unmeasured confounding due to population admixture, and that it only requires modeling of the main genetic effect. The approach is extended to case-only data in [1417] and to discordant sibships in [1383]. General methodological developments on interaction tests are made in [748], who propose multiple robust statistical interaction tests that remain unbiased under model misspecification, provided information on the exposure distribution is available, and are therefore applicable even when adjustment for a high-dimensional covariate is needed. [1396] propose tests for so-called sufficient cause interactions that signal the presence of a mechanistic interaction between 2 exposures. Finally, [1384] propose a stepwise approach for family-based designs for selecting a set of markers in a gene that are independently associated with the disease. The approach is based on testing the effect of a set of markers conditional on another set of markers.

[712] propose evolutionary substitution models allowing for nearest neighbor dependencies. On the basis of nucleotide information for a number of species for whom the phylogenetic tree is known, they develop inference for evolutionary parameters based on data augmentation and find strong support for the presence of neighbor dependencies. Because of the high-dimensionality of the models and the computational complexity, [1372] propose heuristic model selection strategies based on posterior distribution for use when Bayes Factors are computationally too intensive.

[731] develop a general theory for the assessment of direct effects in the presence of high-dimensional, intermediate confounding (i.e., when there are common causes of mediator and outcome that are affected by the exposure of interest). One of the proposed estimators is studied in detail in the context of genetic association studies with pleiotropy in [748] and extended to case-control studies in [1397]. A review of methods for mediation analysis is given in [1395]. To relax the typical assumption that the association between mediator and outcome is only confounded by measured factors, a sensitivity analysis approach is developed in [1391].

Progress on adjustment techniques for high-dimensional time-varying confounding has been made in a series of papers by [1402], [1374] and [1373]. Here, the first paper deals with the problem that the exposure of interest is not observed until the primary event time (e.g. patient information is only recorded during a patient's stay in the intensive care unit, but hospital mortality is the outcome of interest). The second paper discusses how to deal with competing risks when assessing the effect of a time-varying exposure on an outcome. The third is a commentary, pointing out the problems of time-varying confounding adjustment in the intensive care literature.

As a final contribution, [730] propose semiparametric estimators for within-cluster effects when no model assumptions are imposed on between-cluster effects. Their approach extends conditional likelihood methods to enable general correlation structures and to avoid fully parametric distributional assumptions.

Another topic studied by [734] concerns the problem of false positive findings due to repeated testing when screening for genes with an effect on the trait. In this, and other settings of repeated testing, the need to control false positive results has lead to a much stricter control of type 1 errors with a drop in power as the consequence. When plant breeders screen for promising genes, they are however greatly concerned with type 2 errors, not wishing to loose sight too early of truly promising genes. To respond to these needs we have developed a balanced test for screening, maximizing a weighted average of the probability of drawing the correct conclusion under the null and the alternative hypothesis. A fundamentally different ranking of promising genes results from this approach with due regard for the alternative hypothesis. A corresponding measure of relative evidence in favor of the alternative versus the null is derived. This approach is further adapted to two-stage designs and several adaptations to specific data structures are developed. A similar problem of over optimism shows up when a regression model is fitted for prediction and [239] aim to assess validity of the fitted prediction model. In a regression approach that targets prediction directly on the scale of the ordinal data, using the L1 norm, we consider the absolute prediction error, the expected value of the absolute difference between the future and predicted responses, as the model evaluation criterion. [239] show that the distributions of the apparent error and its cross-validation counterparts are approximately normal even under a misspecified fitted model. When the prediction rule is 'unsmooth', the variance of the above normal distribution can be estimated well via a perturbation-resampling method. We also show how to approximate the distribution of the difference of the estimated prediction errors from two competing models. With two real examples, we demonstrate that the resulting interval estimates for prediction errors provide much more information about model adequacy than the point estimates alone.

A further topic studied by [1403] in forensic research is related to the matching of dna profiles obtained from two sources. Two methods are commonly used to report on evidence carried by forensic DNA profiles: the 'Random Man Not Excluded' (RMNE) approach and the likelihood ratio (LR) approach. It is often claimed a major advantage of the LR method that dropout can be assessed probabilistically. A new RMNE measure is proposed that like-wise accounts for allelic dropout in an observed forensic DNA profile. We discuss the necessary calculations, underline their simplicity and provide a tool for performing the calculations.

In plant breeding, [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. As an extension to this work, design aspects of such trials aiming at evaluating a specific gene effect and producing a productive mutant are studied by [1685].

Further, 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 apply 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. [828] further performed a simulation study to investigate the performance of several gene selection methods in combination with several classification techniques in the microarray context. The stability of the methods with respect to distributional assumptions is examined by also considering data simulated from a symmetric and asymmetric Laplace distribution in addition to normally distributed microarray data. [803] investigate, by using a simulation study and a case study, the performance of SAM in the many to one microarray experiments. In particular, the influence of the presence of genes with a small variability on operational characteristics of SAM is evaluated.

[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. On the level of model selection/hypothesis testing, [1671] look for an appropriate statistical testing procedure to find genes that are over expressed in a single tissue as compared to all members of a panel of other tissues (problem of hypothesis testing with an alternative hypothesis that implies an intersection). For this purpose, [1671] successfully develop a Bayesian method that is shown to outperform the current best solution to this problem (viz., a so-called intersection-union test).

[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. High-dimensional data are encountered frequently in modern genetics. One problem, tackled by [912], is to estimate genotypes of SNPs reliably from fluorescence data. Remarkably stable patterns can be found in these signals, offering good opportunities for improved accuracy of SNP genotyping. Once SNPs have been genotyped, the haplotype problem comes into play: estimating and separating the states of two binary strings (representing two chromosomes) from their sum. A penalized composite link model is proposed and applied by [914].

[390, 1007] revise and comment on the work by Tyekucheva and Chiaromonte (2008). They revise the different ways of solving the problem caused by the singularity of the estimated covariance matrix, starting from the estimation of a pseudo-inverse and also considering bagging techniques for generating averages of pseudo-inverse covariance matrices. The authors propose a new estimator of the covariance matrix and its inverse is obtained by the augmented Bootstrap. The performance of this estimator is checked by a simulation study and some real data applications.

[1554] use the independent factor analysis (IFA) model as a dimensionality reduction technique to perform multivariate nonparametric density estimation. The proposed estimator of the density is optimal in a minimax sense (up to a logarithmic factor), and can be approximated in a computationally feasible way. An interesting property of the estimator is that it automatically determines the number of necessary factors in a principled way. Besides, the plug-in classifier resulting from the estimator is also proved to be minimax optimal (up to a logarithmic factor) for classification. The estimator is tested on an artificial density estimation task and on a classification task related to meteorological satellite images. Good quality statistical analysis of gene expression data is an important issue. [1556] present some statistical methods for generating sparse models in the gene expression field, where the number of covariates (genes) under study moves about the thousands, while the sample sizes seldom reach a hundred of individuals. For phenotype classification, the authors propose different lasso logistic regression approaches with specific penalizations for each gene.

2.5.2 Data mining

The development of robust techniques in data mining is partially a joint research effort of researchers of UG and KUL-1. [446, 993] 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. [738] further describe fast and robust bootstrap methods in general and consider its application to perform inference based on robust estimators for linear regression and multivariate models. Confidence and prediction intervals and tests of hypotheses for linear regression models are studied as well as inference for location-scatter parameters, principal components, and classification error estimation for 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.

Furthermore, general techniques are developed for a typical setting in data mining with a lot of variables observed and the requirement to select the most important variables. Variable selection in an additive model structure is accomplished in [1640] using a combination of the nonnegative garotte technique and P-splines estimation. This results into a powerful variable selection method, for which the performance in comparison with other recent variable selection procedures is illustrated.

For high-dimensional data, dimension reduction will often be applied by means of a principal component analysis (PCA) method. In [1194] theoretical properties of a robust PCA-method, called the ROBPCA method, are studied by deriving its influence function. Further, a robust PCA method for skewed unimodal data and appropriate outlier maps to visualize the possible outliers is developed by [1205]. Also a new kernel PCA method and diagnostic tools to detect influential data points are proposed in [1233]. Support-vector machine techniques are among the methods used for high-dimensional data. Such kernel-based regression are studied in [1193] and [1264]. In the context of data mining, robust learning from bites is discussed in [132].

[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. [826] propose an approach to construct a diagnostic classification rule based on proteomic mass spectrometry. The rule is developed on a test dataset and evaluated on a validation data set of mass spectra obtained from breast cancer patients. [826] further develop a model that allows to predict the isotopic distribution of peptides. The predicted distribution is then used to discover peptide related peaks in mass spectra. Finally, [875] review approaches, which use the concept of the isotopic distribution to discovering peptide related peaks in mass spectra.

Data mining in chemometrics is also considered in the network. One of the major difficulties within the context of the development of chromatographic methods consists in the automated detection of the peaks coming from complex chemical matrices. [471] propose an integrated methodology based on Independent Components Analysis (ICA) and clustering to solve this problem and apply ICA successively to HPLC-UV-DAD chromatograms. A new method for modeling chromatographic responses is presented by [514] as a critical piece for the achievement of automated development of analytical methods. This methodology combines experimental design, multivariate response surface modeling, multi-criteria optimization and uncertainty estimation by Monte Carlo simulations to derive optimal analytical parameters in chromatographic frameworks. 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. [520] explore the respective effectiveness of six multivariate methods in this context : multiple hypotheses testing, supervised extensions of principal (PCA) and independent components analysis (ICA), discriminant partial least squares, linear logistic regression and classification trees. On this basis, [1074] develop also an integrated method for biomarker identification combining ICA and mixed linear statistical modeling. It improves classical PCA chemometricians approaches in providing candidate spectral biomarkers which are more easy to interpret, less sensitive to biological and analytical noises and with a statistical confirmation of significance of their effect. The computer screening or large molecule libraries is also a context where chemists or pharmacists have the opportunity to discover new products of interest for the industry. [1125] design a genetic algorithm to explore potential libraries of candidate molecules built on the basis of high-throughput experiments. They apply it successfully in the performance optimization of catalysts for asymmetric transfer hydrogenation.

Finally, [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 are compared. [906] further advice on how the construction of biplots can be improved. One way to summarize and display high-dimensional data is projection on principal components. Both the observations and the variables can be represented in biplots. Unfortunately most biplots are much less informative than they could be.

2.5.3 Psychometrics

High-dimensional and compound data also occur frequently in the psychological literature. Different new techniques are developed to cope with these data complexities.

To deal with vast data sizes (of high-dimensional object by variable data type), the network has been working on a broad range of methods that imply a reduction of the variables (and possibly also of the objects), with this reduction being either categorical or dimensional in nature (clustering resp. dimension reduction). Contributions in this area include the development of novel models, such as a model that clusters the objects in overlapping clusters and simultaneously reduces the variable space ([1167]), and a model that simultaneously groups both objects and variables in overlapping clusters ([164]); the latter model goes with a particularly useful extension that allows to unveil the nature of interactions as present in the data ([1180]). The methods that are developed are both deterministic and stochastic in nature ([610]) (with a generic method being proposed to expand deterministic models into their immediate stochastic counterparts ([1235])). The development of novel models goes along with a study of their identifiability ([1166], [578]), and with the development of suitable associated algorithms (and the study of their performance) ([1216], [596], [1165], [131] and [161]).

To deal with multiway data, models are developed to capture the structure of three-way data making use of random effects ([941], [1201]); this represents an obvious cross-link with WP4. The relation between the two previously developed simultaneous multiway clustering models was further investigated ([1212]). Furthermore, beyond specific models, a unifying model is developed that implies a categorical and/or dimensional reduction of one or several modes of a multiway data set ([165]); this encompasses a broad range of existing reduction models as special cases and may act as a powerful tool for the creation of novel models.

Apart from some algorithmic work ([1168]), two important data-analytic issues are dealt with. The first of these pertains to model selection. Within the context of models for multiway data, this typically is a multidimensional problem, and therefore a more complex one than in the two-way case. Several model selection procedures are compared and a best deal for model selection in the case of multi-mode partitioning models [616] is found. The second issue pertains to estimation uncertainty. In this regard a novel procedure is successfully developed to discriminate between strong and weak structural aspects of three-way component solutions ([1190]).

A particularly intriguing subtype of coupled data pertains to information on two or more links of a sequential process (with information on each set of variables as included as a node in the sequential chain constituting a separate data block). As an example, one may think of psychological data on behaviors as displayed by persons in several situations (one data block), with the occurrence or intensity of behaviors being mediated by the thoughts and feelings that those persons experience in the situations in question (second data block, linked to the first through both persons and situations). A major challenge when analyzing such sequential process data is that each link in the chain may be subject to heterogeneity (cross-link with WP4). For instance, in the psychological chain 'situation - thought/feeling - behavior', both the 'situation - thought/feeling' link and the 'thought/feeling - behavior' link may be subject to individual differences. A contribution is made to the development of novel tools to capture these sources of heterogeneity, both in terms of a variance decomposition (based on a conjunction of a multivariate analysis of variance and a linear mixed model) and in terms of a representation of the contents of the heterogeneity (by means of a novel family of sequential clusterwise regression models), see [609], [592], [1161].

With respect to the development of numerical tools, a method is proposed to overcome the computational burden associated with latent Markov models applied to panel data; this method is based on transformations applied to a directed acyclic graph associated with the models in question ([613], [613]). In the same context, a user-friendly software to estimate a specific model for data fusion is developed ([1227]). For model selection and the estimation of uncertainty, a numerical convex hull based procedure for selecting among several multilevel simultaneous component solutions is developed ([1162]). Furthermore, [1217] also develop different bootstrap strategies for estimating confidence intervals of the parameters of models from the same family.

2.5.4 Across different disciplines

The problem of the analysis of data sets that include multiple sorts of information regarding the same system is also addressed in the network. Such data sets can be denoted by the terms coupled, linked, or multiset data, and the associated data analysis can be denoted by the term data fusion. As an example one may think of data that are obtained on the same organisms in terms of gene expression, ChIP-on-Chip, alternative splicing (Exon arrays), copy-number measurements (CGH arrays) and polymorphism genotyping (SNP arrays).

A broad range of novel models are developed for data fusion problems (see, e.g., [1169], [620], [1252]). Moreover, those models as well as related ones, appear to yield most insightful results, when applied to transcriptomics and metabolomics data from systems biology ([1182], [1220]), as well as to data from cross-cultural psychology ([1251]).

In addition to the development of particular novel models for data fusion, relations between data fusion models are also studied. At this point, an encompassing framework is built for a broad range of simultaneous component methods as developed in various research domains (including chemometrics and psychometrics), which provides a clear insight into both the common core of the methods and the specific elements with regard to which they differ; this implies most useful clues for a suitable choice between the methods in question in data-analytic practice ([1223]). The further study of model interrelations culminates in an overall conceptual framework for data fusion, along with a generic model that subsumes an important subset of specific (existing as well as to be developed) data fusion approaches.

In this context, the problem of weighting is also considered. When dealing with the estimation of a global model to coupled data at hand, an important question reads how the information from the different blocks is to be integrated into an overall loss or objective function. This question becomes especially important if the data blocks differ rather strongly in size and/or if they are subject to rather different amounts of noise. To deal with this issue, different possible weighting schemes are studied and likelihood-based approaches, making use of both synthetic data and data from a study on functional genomics ([1221]). This makes clear that: (a) in a homogeneous error setting, it is advisable not to correct for block size by down weighting larger blocks ([1185]), and (b) when data blocks differ considerably both in size and presumed noise level, one may benefit from using a likelihood-based approach ([1185]).

3 The networking

3.1 Major joint activities

The scientific collaborations between various groups of the network were organized at six different levels:

- **Workshops.** Three workshops took place. They play a key role in the organization of the project and join together all members of the IAP network.
- **Meetings.** A set of meetings was organized related to the topics of the network. These meetings were attended by researchers of the IAP network but were also open to other researchers.
- **Research seminars.** In the institutional seminar series of each group of the network, we organized seminars on topics of direct interest for the research network. In addition, a number of seminars have been organized by the network itself.
- **Working groups.** Their goal is to facilitate the interaction between a limited number of members from the network on a specific topic. They often lead to specific collaborations between members of the network.
- **Training.** Intensive courses on various topics of the network were organized by the nine partners. These courses were taught by members of the network or external experts. Special attention was paid to encourage doctoral and post-doctoral students to attend these courses. However the audience was not limited to the members of the network and very often external students attended these intensive courses too.
- **Mobility of researchers.** Some exchanges of researchers were organized within the network.

These six levels of joint scientific activities are now described in detail. Overall, from 2007 on, the number of scientific collaborations within the network growed continuously, as can be seen from the list of joint activities below.

3.1.1 Workshops

The annual workshops were organized by different partners of the network and play an essential role in the organization of the network. The scheduling over time links up with the advancement of the whole project.

The annual workshop of 2007 entitled ‘Statistical Inference for Dependent Data’ was organized by the University of Hasselt on April 26-27, 2007. The local organizer was Noël Veraverbeke. The central theme of the workshop (‘How to deal with dependent data ?’) was considered from different angles and approaches, and brought together researchers from the network working in diverse areas of statistics. 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 and KUL-1). There were also 11 contributed talks.

The annual workshop of 2008 entitled ‘Missing Information in Survival Data Beyond Right Censoring’ was organized by the University of Ghent on 18-19 September 2008. The local organizer was Luc Duchateau. The workshop focused on the cornerstone ‘Data’, which is one of the three cornerstones on which the IAP project is based (the two others being ‘Models’ and ‘Methods’, see Section 1.3 for more details). More precisely, the workshop concentrated on incomplete survival data. Examples of incomplete data are missing, right censored and interval-censored data, and common survival models for this type of data are competing and semicompeting risks models and frailty models. In this workshop new methodological developments for these data and models were considered. Three keynote speakers were invited: Rick Chappell (University of Wisconsin, USA), Jianguo Sun (University of Missouri, USA) and Jerry Lawless (University of Waterloo, Canada). Furthermore, there were contributed sessions organized around the five workpackages of the network. Many contributed papers from the whole IAP network were presented. A total of 75 IAP members participated to this workshop. A social activity (guided tour through Ghent, followed by an informal lunch) was organized on September 20, in order to stimulate interactions (both scientifically and socially) in the network.

The next annual workshop was organized by the KUL-2 group. It took place on 19-20 November 2009, and the local organizers were Geert Verbeke and Geert Molenberghs. The focus was on the cornerstone ‘Methods’, and more precisely on nonparametric and Bayesian methods. The keynote speakers of this workshop were Gerda Claeskens (KULeuven), who talked about ‘Model testing and selection’ and Timothy Hanson (University of Minnesota, U.S.A.), whose talk was on ‘Bayesian survival analysis: An overview of models and methods’. There were 6 contributed papers sessions: one for each workpackage, plus one additional session for papers that did not clearly belong to any of the workpackages. A total of 96 people attended this workshop. On 17-18 November Timothy Hanson together with Alejandro Jara Vallejos (Universidad de Concepción, Chile), also gave a short course on ‘Nonparametric Bayesian statistics’, which was attended by a large number of PhD students in the network.

The annual workshop of 2010 will be organized by the KUL-1 partner. It will take place in October 2010 and will be focused on the cornerstone ‘Models’. The workshop will focus on cross-links between the different workpackages and on recurrent research themes that are being studied by several partners from different angles. Modeling approaches that are shared by many workpackages are e.g. copula models (as a key tool to deal with dependencies), and variable selection and sparseness (to deal with high-dimensional data).

The final workshop in 2011 will be organized by the UCL partner. The workshop will focus on the interaction and synthesis between the 3 cornerstones of the network. A possible approach to reach this goal could be the application of central methods and modeling approaches to a number of benchmarking datasets with interlocking complexities.

3.1.2 Meetings

Apart from the above workshops, several meetings were organized in relation to the IAP network.

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`

- Meeting on 'Mixture modeling' with special interest to applications in educational measurement and bioinformatics, 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.

The following meetings were organized by the network in 2008:

- On September 10, 2008, Norman Breslow received an honorary doctorate from UH and KUL-2 combined, at the occasion of the start of the Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat), a joint venture of UH and KUL-2. It is fair to say that this well advanced level of collaboration has been catalyzed by the IAP-Network in Statistics. Incidentally, Norman Breslow has published extensively in the fields of models with latent structures, such as the overdispersed Poisson model and the generalized linear mixed model. He is considered one of the founding fathers of the latter.
- On October 16-17, 2008, the 16th Annual Meeting of the Belgian Statistical Society (BSS) took place in Wépion-Namur. The IAP network took part in the organization of this meeting, and organized one session, with Vladimir Spokoiny (Weierstrass Institute, Berlin) as invited speaker. The IAP network sponsored this meeting for the first time, and agreed with the BSS to do this on an annual basis, to increase the visibility of the network within the Belgian statistical community.
- On November 5-7, 2008, the 'International Seminar on Nonparametric Inference' (ISNI2008) took place in Vigo, Spain. The meeting was co-organized by the European partner of Santiago de Compostela under the heading of the IAP network. Three members of the IAP network, as well as A. Davison (member of the follow-up committee) gave invited lectures on this meeting. Many members of the network participated to this meeting (especially from UCL, UH and UJF).
- On November 12-14, 2008, an international workshop on 'Flexible modeling: smoothness and robustness' was held at KUL-1 (main organizer: I. Gijbels).Two of the ten invited speakers and several contributed speakers and presenters of posters were members of the IAP-network. The international workshop was attended by approximately hundred participants including several members of the IAP-network.

The following meetings were organized in 2009:

- Tom Snijders (member of the follow-up committee) has been awarded the prestigious Francqui Chair 2008-2009 upon a nomination by the KUL-1 group on 16 February 2009. On this occasion, he has given lectures on ‘Statistical models for social network data’ on 17 and 18 February and on 23 and 24 March, 2009. They were co-organized by the IAP network.
- An international conference on ‘Exploring research frontiers in contemporary statistics and econometrics’ was organized by the UCL-partner in honor of the main coordinator of the network, Léopold Simar (UCL), on the occasion of his retirement. The conference took place on May 14-15, 2009 at the UCL, and was co-organized by the network.

The following meetings will be held in 2010:

- The European partner of Santiago de Compostela (Spain) will organize the METMAV Conference from 30 June to 2 July 2010. The purpose of this conference is to promote the development and application of spatio-temporal statistical methods in different fields related to environmental sciences.

  Other conferences will follow in the near future.

3.1.3 Research 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 these 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 (*).

We took the opportunity of these seminars to facilitate the transmission of our research results within the network. In particular, each group invited on a regular basis members of the partner teams to present their research. A list of these seminars is presented below. We start with 2007:

- *January 25, 2007: Philippe Lambert (UCL), ‘Bayesian multi-dimensional density estimation with P-splines’, at KUL-1
- February 5, 2007: Juan Carlos Pardo-Fernández (University of Vigo), ‘Tests in nonparametric regression based on the error distribution’, at UCL
- *February 5, 2007: Wenceslao González-Manteiga (USC), ‘Comparing spatial dependence structures’, at UCL
- February 23, 2007: Jérémie Bigot (Université Toulouse III), ‘Homeomorphic smoothing splines: a new tool for monotonizing a non-constrained estimator in nonparametric regression’, at UCL
- February 23, 2007: Geurt Jongbloed (Free University of Amsterdam), ‘Density estimation in the current status model’, at UCL

- *April 20, 2007: Paul Eilers (UU), 'Statistical classification for reliable high-volume genetic measurements', at UCL
- *April 20, 2007: Kristel Van Steen (UG), 'Perspectives on genetic association screening, at UCL
- *June 12, 2007: Ingrid Van Keilegom (UCL), 'Estimation of a semiparametric transformation model', at USC
- *October 4, 2007: Cédric Heuchenne (UCL), 'Testing parametric forms in location-scale models with censored data', at KUL-1
- October 12, 2007: Mark van der Laan (University of California, Berkeley University), 'Targeted maximum likelihood learning: Applications in discovery and randomized trials', at KUL-1
- *October 25, 2007: Marek Omelka (IAP-postdoc, KUL-1 and UH), 'Uniform asymptotics for S- and MM-regression estimators', at KUL-1
- *November 8, 2007: Noël Veraverbeke (UH), 'Copulas in models for survival data', at KUL-1
- November 29, 2007: Anthony Davison (EPFL, Lausanne), 'Likelihood inference for a problem in particle physics', at UCL
- November 29, 2007: Ernesto San Martin (PUC Santiago, Chili), 'On the statistical meaning of the parameters of skew-normal experiments', at UCL
- *December 6, 2007: Stefan Van Aelst (UG), 'Feasible robust variable selection', at KUL-1

The two seminars that took place on November 29 were organized on the occasion of the administrative meeting of the network, that took place in the morning. 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). In 2008 the following seminars took place:

- *February 8, 2008: Wenceslao González-Manteiga (USC), 'Using SiZer map for testing problems', at UCL
- February 8, 2008: John Einmahl (University of Tilburg), 'Asymptotics for the Hirsch index', at UCL
- May 13, 2008: Pieter Kroonenberg (Leiden University), 'Analyzing two cases of three-mode data', at KUL-1
- *May 16, 2008: Catherine Legrand (UCL), 'A Bayesian approach to jointly estimate a random effect and a random interaction in a proportional hazards model. Application to the validation of a prognostic index in bladder cancer patients', at UCL
- *May 16, 2008: Noël Veraverbeke (UH),'Copulas in survival models', at UCL

- *October 2, 2008: Goele Massonnet (UH), ‘Contributions to frailty and copula modeling with applications to clinical trials and dairy cows data’, at KUL-1
- *October 31, 2008: Noël Veraverbeke (UH), ‘Half a century of Kaplan-Meier estimation in survival analysis (1958 - 2008)’, at KUL-1
- November 14, 2008: Dimitris Rizopoulos (Erasmus University, Rotterdam), ‘Laplace approximations for joint models’, at UCL
- *November 14, 2008: Goele Massonnet (UH), ‘Contributions to frailty and copula modeling with applications to clinical trials and dairy cows data’, at UCL
- December 9, 2008: Herbert Hoijtink (University Utrecht), ‘Learning more from empirical data using prior knowledge’, at KUL-1

We continue with summarizing the seminars that took place in 2009:

- January 6, 2009: Antonio Forcina (University of Perugia), ‘ Likelihood inference on equality and inequality constraints for contingency tables’, at UG
- February 12, 2009: Thomas Lumley (University of Washington), ‘The case-only design allowing for interactions in a multiplicative model for rare events when genetic and environmental exposures are independent’, at UG
- *March 26, 2009: Léopold Simar (UCL), ‘Frontier estimation and extreme value theory’, at UCL
- *March 26, 2009: Geert Molenberghs (UH and KUL-2): ‘Every missing not at random model for incomplete data has got a missing at random counterpart with equal fit’, at UCL
- *May 4, 2009: Fentaw Abegaz (currently IAP-postdoc KUL-1), ‘Statistical dependence modeling and inference using copulas’, at KUL-1
- May 27, 2009: Jack Bowden (University of Cambridge), ‘ Unbiased estimation of selected treatment means in two-stage trials’, at UG
- *October 1, 2009: Tetyana Kadankova (UH), ‘Two-sided exit problems for several classes of stochastic processes’, at KUL-1
- November 16, 2009: Tom Palmer, (University of Bristol), ‘Including multiple instrumental variables in mendelian randomization analyses’, at UG
- *December 3, 2009: Anouar El Ghouch (UCL), ‘Measuring the discrepancy of a parametric model via local polynomial smoothing’, at KUL-1

The two seminars of 26 March 2009 were organized in the context of an administrative meeting with all promoters of the network that took place at UCL after the two seminar talks. In 2010 the following seminars took place or are planned. More seminars will be organized in the near future.

- 8 January, 2010: Yi Li (Harvard University), 'The Dantzig selector for censored linear regression models: identifying predictive genes for myeloma disease progression', at UG
- *January 22, 2010: Auguste Gaddah (UH), 'Flexible modeling in the generalized conditional Koziol-Green model by a copula function', at UCL
- *February 11, 2010: Jan Johannes, (UCL), 'Adaptive circular deconvolution by model selection under unknown error distribution, at KUL-1
- *March 18, 2010: Auguste Gaddah (UH), 'Flexible modeling in the generalized conditional Koziol-Green model by a copula function', at KUL-1
- *March 18, 2010: Fentaw Abegaz (postdoc IAP, joint at KUL-1 and UH), 'Semiparametric estimation of conditional copulas', at KUL-1. This seminar and the previous one were organized together around the common theme 'modeling dependencies in complex data using copulas'.
- May 21, 2010: Uschi Müller-Harknett (Texas A&M University), title to be fixed, at UCL
- May 21, 2010: Melanie Birke (University of Bochum), title to be fixed, at UCL

Finally, we like to mention that Anastasios Tsiatis (North Carolina State University) has been awarded the 2010-2011 Princess Lilian Visiting Professorship. The promoters (hosts) are Geert Molenberghs (UH and KUL-2), Geert Verbeke (KUL-2) and Marc Aerts (UH). During his visit, which is planned in May-June 2011, Anastasios Tsiatis will teach a doctoral-training course, he will provide guidance to the doctoral students of KUL-2 and UH, and he will give a sequence of research seminars at UH, KUL-2 and UCL.

3.1.4 Working groups

The IAP network is working on a broad range of research topics in statistics. The interactions and collaborations between network partners organized as formal and informal working groups, are manifold. This can also be seen from the list of technical reports and publications.

The interaction between researchers of the network is facilitated through the working groups. They allow an intensive discussion on specific topics between a limited number of researchers. 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.

- *Sensitivity analysis*

  In the context of incomplete data, the sensitivity analysis working group has existed as a collaboration between UH and KUL-2, with contributions from LSHTM. This sensitivity analysis work is to be situated within the context of mixed models or other modeling contexts for hierarchical data, thus placing itself at the crossroads of WP3 and WP4. It is here that the work of [964] and by [967] are aimed at. By way of another example, [1700] studied the behavior of shared-parameter models under random-effects misspecification. This work connects various themes within WP3 (sensitivity analysis, shared-parameter models, and other latent structures) and within WP4 (random effects structures).

- *Frailty models*
  UH, UG, UCL, and KUL-2 have established and maintained a working group on frailty models for survival analysis. This has led to a suite of high-profile references, in particular [1025] is a state-of-the-art treatise on the frailty model and extensions thereof. The intensive collaboration started in 2002 and continued up to now. The working group concentrates on research issues under workpackage 3. Since 2002, the group met in a regular way to discuss research progress and for a series of seminars mainly to introduce relevant methodology to new PhD students. Participants in 2007-2009 were: B. Ampe (UG), R. Braekers (UH), T. Burzykowski (UH), J. Cortiñas (UH), L. Duchateau (UG), A. Gaddah (UH), E. Goetghebeur (UG), K. Goethals (UG),P. Janssen (UH), Komárek (KUL-2), P. Lambert (UCL), C. Legrand (UCL), Lesaffre (KUL-2), G. Massonnet (UH), R. Sylvester (EORTC), B. Van Rompaye (UG), I. Van Keilegom (UCL) and N. Veraverbeke (UH). The main research topics discussed in this working group are related to competing risks and to transformation models for survival data (where the transformed cumulative hazard is modeled as a function of covariates in a linear way and where such models are extended to time-varying covariates and frailties).

- *Surrogate markers*
  The work on the evaluation of surrogate markers from clinical studies, is located at the interaction between WP3, WP4, and WP5. There has been an informal working group in existence ever since the beginning of the network, encompassing colleagues from UH, KUL-2, and various external partners, from Belgium and abroad.

- *Copulas*
  The working group on ‘Modeling dependencies and inference based on copulas’ is very active, and met approximately every 6 weeks in 2007–2010. The close collaboration of members of KUL-1 and UH within the context of this working group led already to several joint publications of members of these two universities. Two IAP-postdocs (in the period 2007–2010) M. Omelka and F. Abegaz participated to this working group.

- *Variable selection and sparseness*
  Members of the KUL-1 (in particular I. Gijbels and A. Verhasselt) and of UJF (A. 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. These collaborations are situated in WP1 and on the interface between WP1 and WP4. These collaborations resulted already in several joint papers. A. Verhasselt spent in spring 2008 more than 2 months at UJF, working on a joint project between KUL-1 and UJF.

- *Longitudinal data*
  There is a working group on longitudinal data, focusing on developments originating from [454], and rooted in UH, KUL-2, and LSHTM. The purpose is to develop flexible methodology to allow for complex hierarchical models for designs that encompass repeated measures,

multi-level structures, and to correct for overdispersion when data are non-Gaussian. Applications are situated in surrogate marker evaluation in clinical trials, psychometric validation, incomplete longitudinal data, etc.
Given that latent structures can be conceived as a form of incompleteness, there are strong and natural links between WP3 and WP4. Incidentally, various book projects are situated within WP3 and on this particular bridge between the two packages. These are oftentimes outgrowths of informal inter-university working groups. Much of the work on longitudinal data (WP4) faces issues of incompleteness (WP3), as well. G. Verbeke (KUL-2) and G. Molenberghs (UH) are co-editors, jointly with G. Fitzmaurice (Harvard School of Public Health, Boston, MA) and M. Davidian (North Carolina State University, Raleigh, NC), on the state-of-the art volume on recent advances in longitudinal data analysis (WP4), with an important focus on incomplete data and sensitivity analysis (WP3) ([1026]).

- *Goodness-of-fit tests*
  C. Heuchenne and I. Van Keilegom (UCL) has extensive collaborations with members of the USC partner (W. González-Manteiga and R. Crujeiras Casais) 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 inference for censored spatial data. The collaborations have lead already to many publications, mainly in WP1 but also in WP2 and WP3. R. Crujeiras-Casais worked as an IAP postdoc at UCL in 2008 on these and other topics related to the network. W. González-Manteiga visited the UCL for one week in February 2007, for three weeks in February 2008 and for one week in May 2010. I. Van Keilegom and I. Heuchenne visited USC on average once or twice a year, each time for about one week to finish ongoing work and discuss new research directions.

- *Bioinformatics*
  I. Van Mechelen (KUL-1) and G. Verbeke (KUL-2) are both partners in a KUL Center of Excellence for computational systems biology. Their teams have made several methodological contributions (of immediate relevance for WP5) in the context of statistical bioinformatics, which have lead to new biological and medical insights.

3.1.5 Training

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 is given below.

- January 30-31 and February 1, 2007: Short course on ‘Modeling dependence with copulas’, by Christian Genest (Université Laval, Canada), at UCL, jointly organized by the FNRS Graduate School in Statistics and the IAP network.

- May 21-24, 2007: Summer school on ‘Advanced nonparametric statistics’, including three short courses given by Jean Opsomer (Iowa State University, USA), Marten Wegkamp (Florida State University, USA), and Oliver Linton (London School of Economics, UK), at UCL, jointly organized by the FNRS Graduate School in Statistics and the IAP network. The summer school also attracted a large number of PhD students, coming from several European countries, including a number of PhD students from the European partner of USC.
- September 4, 2007: Short course on ‘The statistical evaluation of surrogate endpoints in clinical trials’, by Geert Molenberghs (UH and KUL-2), on the premises of the UCB company in Braine-l’Alleud, jointly organized by the Biostatistics Section of the Belgian Statistical Society and the IAP Network.
- November 22-23, 2007: Short course on ‘Mixed models and incomplete data’, by Geert Molenberghs (UH and KUL-2) and Geert Verbeke (KUL-2), at KUL, organized by KUL-1 on behalf of the ‘Interuniversity Graduate School of Psychometrics and Sociometrics’ and the IAP Network.
- January 21-22, 2008: Short course on ‘Statistical inference for curve estimation and classification’, by Lutz Dümbgen (University of Bern, Switzerland), at UCL.
- February 12-15, 2008: Short course on ‘Goodness-of-fit testing in regression’, by Wenceslao González-Manteiga (USC), at UCL.
- March 3-4-5, 2008: Short course on ‘Local parametric estimation with applications in volatility estimation and risk management’, by Vladimir Spokoiny (Weierstrass Institute, Berlin, Germany), at UCL.
- April 28-29, 2008: Short course on ‘Introduction to numerical techniques for statisticians’, by Geert Molenberghs (UH and KUL-2), Francis Tuerlinckx (KUL-1) and Geert Verbeke (KUL-2), at KUL, organized by KUL-1, KUL-2, UH.
- May 20-21, 2008: Short course on ‘Modeling with heavy tails and extremes’, by Sydney Resnik (Cornell University, USA), at UCL.
- September 16-18, 2008: Short course on ‘On the bridge between genetics and statistics’, by Amy Anderson (Western Washington University, USA), at UCL.
- October 20-21, 2008: Short course on ‘Financial time series’, by Timo Terasvirta (University of Aarhus, Denmark), at UCL, jointly organized by the FNRS Graduate School in Statistics and the IAP network.
- November 3 and 5, 2008: Short course on ‘Statistical learning theory’, by Sarel Steel (University of Stellenbosch, South Africa), at KUL-1.
- November 14, 2008: Short course on ‘Penalized spline smoothing and geoadditive regression’, by Thomas Kneib, Ludwig-Maximilians-University Münich, Germany, at KUL-1.

- January 29-30 and February 3-4, 2009: Short course on ‘Empirical log optimum portfolio selection’, by László Györfi and András Urbán (Budapest University of Technology and Economics, Hungary), at KUL-1.
- February-March, 2009: Belgian Francqui Chair 2008-2009 awarded to Tom A.B. Snijders (University of Oxford and University of Groningen) and member of the IAP follow-up committee. The chair, organized by KUL-1, comprised an inauguration and a lecture series on the topic of ‘Statistical models for social network data’.
- November 3-6, 2009: IERI Autumn Academy on ‘A nonlinear mixed model framework for item response theory’, at KUL-1.
- January 25-26, 2010: Short course on ‘An introduction to statistical inverse problems’, by Axel Munk (University of Goettingen, Germany), at UCL, jointly organized by the FNRS Graduate School in Statistics and the IAP network.
- February 17, 2010: Short course on ‘Bayesian data analysis’, by Andrew Gelman (Columbia University, New York, USA), at KUL-1. More than 180 participants registered for this course, including many PhD students and postdocs from all IAP partners. This activity can also be considered to have contributed significantly to the visibility of our network.
- June 1 and 3, 2010: Short course on ‘Regression estimation under shape constraints - an overview’, by Melanie Birke (Ruhr Universität Bochum, Germany), at UCL.

We also like to mention that in the framework of the Princess Lilian Visiting Professorship that has been awarded by KUL-2 and UH to Anastasios Tsiatis (North Carolina State University) in 2010-2011 (see Section 3.1.3 for more details), Anastasios Tsiatis will, among other activities, teach a doctoral-training course based on his state-of-the-art Springer book on semiparametric theory and incomplete data.

3.1.6 Mobility of researchers

Mobility of researchers is a crucial issue in the network. In particular with respect to training of PhD students, the network really offers a great opportunity. Apart from mobility in terms of training for PhD students it is also of vital importance that researchers of the network visit other network members for collaboration.

The KUL-1 partner member A. Verhasselt, PhD student at the KUL-1 group, spent twice a study leave (of 2.5 months and one month) in 2008-2009 at UJF. These study leaves and the resulting collaborations between the KUL-1 team and the UJF-team resulted in three papers, of which one is submitted, and two are currently prepared for submission. Moreover, it is planned that a graduate student of W. González-Manteiga (USC) will spend in 2011 several months at UCL to work with I. Van Keilegom.

In addition, a lot of mutual visits took place between academic members of the network, usually for short periods. For instance, faculty members of KUL-1 (I. Gijbels) and of UJF (A. Antoniadis and S. Lambert-Lacroix) made several mutual visits. The aim of these visits is an intensive working on the joint research projects in order to make substantial progress on them. Also, several visits per year took place between the European partner USC (W. González-Manteiga) and the UCL partner (C. Heuchenne and I. Van Keilegom), and between LSHTM (M. Kenward) and the groups of G. Verbeke and G. Molenberghs (KUL-2 and UH).

Below are a few examples of joint appointments of postdoctoral positions financed by the network. The postdocs financed by the IAP network have in most cases been hired after a joint call, collecting all the positions offered by groups of the network in one single announcement, that was sent out on an international scale to many universities in the whole scientific community.

- Marek Omelka was an IAP postdoc coming from Prague University. He worked 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.
- Rosa Crujeiras-Casais, who did her PhD at USC, was an IAP postdoc at UCL from August 2008 till December 2008. On January 1, 2009, she obtained an academic position at USC.
- Fentaw Abegaz is an IAP postdoc, who currently works for 6 months at KUL-1 (October 2009-April 2010) followed by 6 months at UH (April 2010 - October 2010).
- Auguste Gaddah will finish his PhD in September 2010 at UH, and will start working as a postdoc at UCL from October on.

These types of postdoc positions are very interesting for the network, as it stimulates interactions among the different groups.

Several members of the network are members of PhD guidance committees of PhD students at other universities in the network. A list of PhD theses currently in preparation in the network can be found on the website:

`http://www.stat.ucl.ac.be/IAP/PhaseVI/research_theses.html`

The list mentions (among others) which members of partner universities of the network take part in the PhD committee. This participation is a very useful way to get familiar with the research carried out at other groups of the network. The website also contains a list of defended theses in the network since 2007.

3.2 Added value gained through the network

The whole overview report reflects the high added value gained through the network. The researchers of the network are sometimes faced to similar problems but in different fields. An example is given by the use of copulas as a key tool to deal with dependencies, and variable selection / sparseness methods to deal with high-dimensional data. They are proved to be useful and popular in many workpackages. In this respect, each partner clearly benefits from the expertise of the whole network.

The list of working groups (Section 3.1.4) illustrates other common scientific concerns between the members of the network. Overall, the network is in a unique position since it brings together researchers and hence methods from biostatistics, mathematical statistics, econometrics and psychometrics. This is testified, for example, by the frailty working group (biostatistics and mathematical statistics, UG, UH, KUL-2, UCL, ...), by the work in WP2 on time series, which brings together mathematical statisticians and econometricians, and by the mixtures, mixed, and missing-data related interactions (WP3 and WP4), where psychometrics and biostatistics have come together.

The network also leads to a significant increase of the research activities in statistics in Belgium. To support this point, a large number of high level courses offered in Belgium were organized thanks to the network. Students from all universities of Belgium (and sometimes from abroad) participated to these courses. We can say that the network pushed the Belgian statistical community to a more active and a more visible situation.

KUL-2 and UH have a strong, integrated collaboration in the field of biostatistics and statistical bioinformatics. It has led to the foundation of a joint research institute, termed Interuniversity Institute for Biostatistics and statistical Bioinformatics (I-BioStat). This is, in part, an outgrowth of strong joint research in the context of WP3 and WP4. Without exaggeration, it can be stated that the network has strongly facilitated such an integrated structure.

KUL-1 acquired a new GOA with a major emphasis on a broad range of statistical models with a tight relation to WP2, 4 and 5 (2010-2014, 1.4 million €). Moreover, KUL-1 and KUL-2 are partners in a project for Strategic Basic Research on an algorithmic framework for integrative modeling in systems biology (2007-2011) that is of immediate relevance for WP5.

Members of the KUL-1 team of the Mathematics Department have become partners in an IWT-innovation project, in which their expertise in flexible and robust regression modeling and variable selection is sought for. This expertise is closely related to many topics in workpackages WP1 and WP5.

Another important added value gained through the network are the new interuniversity collaborations that are set up due to the existence of the network. A most rewarding situation is to have IAP-researchers spending part of their time at one university and another part at another university of the network, while working on one joint project for the whole period. Experiences of this kind have been realized by the KUL-1 partner and the UH partner who worked on joint research projects with joined postdoc M. Omelka and F. Abegaz. These collaborations so far resulted in one published paper (in the Top 10 journal *The Annals of Statistics*), two submitted papers (technical reports), one paper that is to-be-submitted soon, and one manuscript in preparation.

3.3 Circulation of information in the network

The above joint scientific activities constitute an important way to facilitate the circulation of information in the network. Apart from these activities, the network also developed a set of tools for this task. These are listed below.

- *Website*

  All activities of the IAP-statistics network can be followed very closely from our website which was created in January 2002, during phase V of the network. The address of the website is

  `http://www.stat.ucl.ac.be/IAP/PhaseVI`.

The website contains e.g. the following information :

- Our logo
- Call for applications
- Description of the project
- List of scientific personnel working under the IAP project
- List of IAP members and their email addresses
- List of visitors
- Research activities (workshops, meetings, seminars, short courses,...)
- Defended theses and theses in preparation
- Downloadable technical reports, list of publications and books written my members of the network
- Annual reports and reports of scientific meetings
- Contact details
- Link to the website of Phase V of the network

- *Technical Reports and Publications Series*
  Two series, available via the website

  `http://www.stat.ucl.ac.be/IAP/PhaseVI/publication.html`,

  report on the scientific results obtained within the IAP-statistics network: the Technical Report Series and the Publications Series.

  The Technical Reports Series groups all papers written under the IAP-statistics network. For each paper the title and authors are given and the pdf file of the paper is provided (in most cases). Each paper in this series has been submitted for publication in an international journal.

  Once a paper has been accepted for publication in an international journal and has been printed, it is listed in the IAP-statistics Publications Series, which mentions for each paper the title, authors and precise reference.

  The following table reports the number of written papers for each publication series :

	2007	2008	2009-2010*
Technical Reports	85	77	109
Publications	222	242	253**

(* until April, 30th)
(** the exact number of publications will probably be somewhat higher, since for a number of papers the date of publication is not known yet).

It is worth mentioning that the number of joint publications is increasing year after year. Starting with 10 joint papers in 2002 (under Phase V of the network), the network produced a total of 37 joint papers in 2007 (all types of papers confounded), 52 in 2008 and 75 in 2009 (without counting the papers that appeared so far in 2010). This evolution is visually represented in Figure 2.

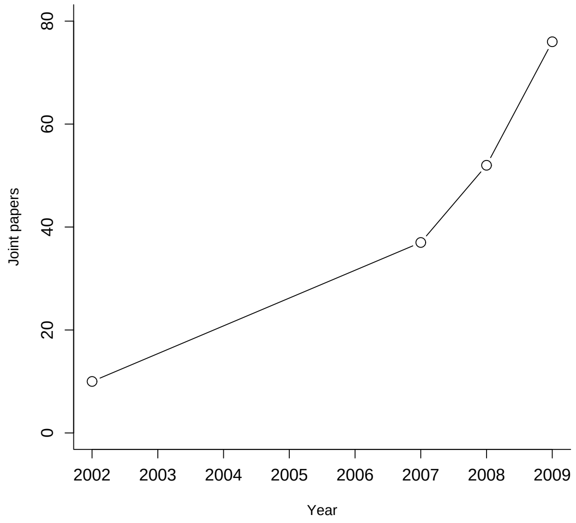

Figure 2: *Evolution of the number of joint papers written in the network.*

- *B-Stat News*

  This journal is the official journal of the Belgian Statistical Society. The major events organized within the IAP network are announced in this journal.

- *Mailing list*

  An announcement of the events organized by the IAP are sent by e-mail to all members of the IAP network.

3.4 Functioning of the follow-up committee

The follow-up committee is composed of Tom Snijders (University of Groningen and Oxford University), Anthony Davison (EPFL, Lausanne) and Ursula Müller (University of Dortmund).

The follow-up committee stays informed of the activities of the network by means of the annual reports of the network. It also plays an important role during the administrative meetings. These meetings took place on

- 29 November 2007 at UCL
- 19 September 2008 in Ghent during the annual workshop organized by UG
- 26 March 2009 at UCL
- 19 November 2009 in Leuven during the annual workshop organized by KUL-2.

During these meetings the committee has been informed about the latest research developments and the administrative evolution of the IAP network. Members of the follow-up committee made several invaluable suggestions for a further strengthening of the network, including the suggestion to organize a few more short courses. This suggestion appeared to be most useful, taking into account the success of, for instance, the most recent short course on Bayesian data analysis.

4 Position of the IAP network

4.1 Cutting-edge research

4.1.1 Scientific highlights of the network

The network has been very active in making important scientific contributions in many areas of statistics : longitudinal data, mixed models, frontier models, censored data, classification models, nonparametric regression, time series, semiparametric models, etc. We list below a selection of important examples of these scientific highlights.

- *Incomplete data and sensitivity analysis*
  WP3 has contributed to number of high-profile references. 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. It 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. In the book by [1713], several chapters, written by world renowned experts, are devoted to incomplete data and sensitivity analysis. [1551] devote a lot of attention in their book to missing data issues in clinical trials. [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. [1620] is a handbook on non-response in longitudinal studies. In a related fashion, 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. [1664] contributed a chapter to the Handbook of Epidemiology on incomplete data. 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. To add to these efforts [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. [1698] wrote an invited review for the Japanese statistical journal. Similarly, [1469] and [1470] wrote an invited contribution about incomplete data in clinical trials, with discussion. Conversely, [1696] wrote a discussion contribution to a work by Lee and Nelder on unobservables, including missing data. [1453] prepared an invited review about analysis and sensitivity analysis of incomplete data. [1682] discuss the Points to Consider Document on Missing Data, as adopted by the Committee of Health and Medicinal Products (CHMP). In 2007, the CHMP issued a recommendation to review the document, with particular emphasis on summarizing and critically appraising the pattern of drop-outs, explaining the role and limitations of the *last observation carried forward* method relative to the use of mixed models. The initiative was taken by a working party of the organization, Statisticians in the Pharmaceutical Industry. [1627] offer a tutorial on handling incomplete data in the context of survival analysis.

- *Frailty models*

  The research undertaken by the frailty model workgroup, with contributions of most of the partners of this IAP network, has led to the development of new techniques (with papers in, for instance, Biometrics, JABES, JSPI, Am. Statistician, and JRSS C), which makes that the group is considered to be one of the major research groups on frailty models internationally. This is emphasized further by the publication of the standard work on frailty models, entitled 'The frailty model', published by Springer Verlag, and authored by L. Duchateau and P. Janssen.

- *Modeling dependencies through copulas*

  Regarding modeling dependencies via copulas there is a large and very important output in the network. Fundamental theoretical contributions were obtained by several members of the network, including members from KUL-1, UHasselt and UCL, and published in top journals. An example is the joint paper by members I. Gijbels (KUL-1), N. Veraverbeke (UH) and M. Omelka (postdoc IAP, KUL-1 and UH) published in *The Annals of Statistics*. Copula functions are also used in many application areas covered by the network. The research on copulas is on the forefront of research in that field worldwide.

- *Flexible modeling and regularization techniques.*

  When analyzing complex data, flexible modeling is a key issue. An interesting method that allows for such a flexible modeling are P-splines. Several members of the network (e.g. KUL-1, UJF, Utrecht) have made seminal contributions to the research area, including for example data-driven choices of the penalty function, flexible modeling of mean and variance/dispersion functions, and variable selection methods. The methods developed on estimation of mean and dispersion function, in an extended generalized linear models setup are among the first to be able to deal with over dispersed as well as under dispersed data, and even crossovers between the two situations. The research currently carried out on variable selection and additive modeling opens very interesting new research perspectives.

- *Frontier estimation*

  The use of nonparametric frontier models remains a hot research topic in efficiency and productivity analysis because they rely on very few behavioral assumptions. Statistical inference covers now most of the underlying economic models (assumptions on returns to scale, etc.) and the theory behind the bootstrap is now translated in efficient and easy to implement algorithms. Extensions have been proposed which allow to handle noise on the data, to condition on environmental variables and to be robust to extremes and outliers. The links with extreme value theory allowed to extend previous asymptotic results on nonparametric frontier estimators but also provided new estimators of the frontier function having asymptotic normal distributions. The asymptotics for conditional efficiency scores has also been provided, with appropriate bandwidth selection procedures. Many papers were published in the best journals ranging from theoretical statistics, theoretical econometrics but also to more applied papers in the field (*Journal of Econometrics* (5), *Annals of Statistics* (1), *Bernoulli* (1), *Journal of Statistical Planning and Inference* (1), *Econometric Theory* (2), *Econometric*

Review (2), *Journal of Applied Econometrics* (1), *Journal of Productivity Analysis* (5), *Annals of Operations Research* (1) and *European Journal of Operational Research* (1)). Some of these methods were also presented in a book published at Springer, New-York (Daraio and Simar, 2007). L. Simar was also invited, as Quality Manager, to join an European project (EUMIDA) to build micro-data on the activities of European Universities. The idea will be in a second phase to implement these methods to facilitate the relative positioning of the European universities.

- *Non- and semiparametric regression with censored data*
  Lots of research has been done by the UCL and UH partner in the context of semi- and nonparametric regression with censored data. This is considered as an important research area, as in many practical situations the classical Cox and accelerated failure type models are too restrictive. In particular, research has been carried out in the context of generalized conditional linear models with censored data, nonlinear regression with censored data, semiparametric transformation models, nonparametric regression with dependent censored data, goodness-of-fit tests for parametric and semiparametric models in multiresponse regression, single index regression models in the presence of censoring, etc.

- *Inference for semiparametric Z-estimators*
  The UCL partner has made lots of progress on inference and asymptotics for semiparametric estimation problems, in particular for estimators that can be written as approximate Z-estimators. The UCL has also collaborated with the USC and UH partner on this type of problems. The methodology for this type of estimators has been applied to a number of novel research ideas, coming from a broad range of areas in statistics (copula estimation in nonparametric regression, semiparametric regression with missing data, semiparametric transformation models, comparison of semiparametric backfitting and profiling procedures, semiparametric location-scale regression models, ...). The asymptotics for this type of estimators is often highly complex, especially when the criterion function is non smooth in the nonparametric component of the model.

- *Nonparametric location-scale models*
  The study of nonparametric location-scale models remains an important research area in the network, especially for the UCL, USC and UH partners. During Phase VI of the network, a wide range of testing problems has been studied in this model, like (among others) goodness-of-fit tests for the mean and variance function based on the comparison of two estimators of the error distribution (one obtained under the null hypothesis and one under the alternative), tests for the independence between the error and the covariate in this model, change-point tests and the comparison of regression curves. Moreover, the nonparametric location-scale models have also been studied in the presence of censored successive survival times and when the response variable is subject to selection bias. Finally, the estimation of conditional ROC curves has been investigated, copulas have been used to model the (possible) relation between the error and the covariate in the model, and the important extension to multiple regression has been developed. The latter result offers a wide range of possibilities for further research,

since so far all results developed in the literature were restricted to one-dimensional covariates, the case of more-dimensional covariates being in fact substantially harder to analyze theoretically.

- *Robust model selection*
  Large datasets are often of unequal quality which leads to outliers and other data anomalies. Building reliable models from such data sets is a challenging task of great importance. The problem of robust model selection appears in all multivariate analysis settings, such as regression, classification and clustering. The selected models have to be stable on the one hand and obtained in a reasonable amount of time on the other hand. The UG team (S. Van Aelst) together with international collaborators have developed and studied several robust, computationally efficient procedures to select stable models in regression and clustering contexts. This research has led to publications in high quality journals such as the Journal of the American Statistical Association, the Journal of the Royal Statistical Society-Series B, and Computational Statistics and Data Analysis.

- *Robustness in high-dimensional data*
  The standard robustness model of regular (clean) and contaminated observations becomes less realistic and useful in high-dimensional data settings. In such settings there are too few observations to discard whole observations when only few of their components are contaminated. On the other hand, one can not assume that there is a majority of completely regular observations without anomaly in any of its components. The UG team (S. Van Aelst) together with international collaborators have proposed new contamination models for high-dimensional data and shown the lack of robustness of the current high-breakdown robust methods under this new framework (published in Annals of Statistics). This shows the need to develop new methods that yield robust results when analyzing high-dimensional data.

- *Data fusion*
  As a consequence of our information society, not only more and larger data sets become available, but also data sets that include multiple sorts of information regarding the same system. Such data sets can be denoted by the terms coupled, linked, or multiset data, and the associated data analysis can be denoted by the term data fusion. The KUL-1 team (along with international partners) has contributed significantly to this challenging area through the development of novel models, through the study of model interrelations (which are of utmost relevance in model selection), and through the study of various strategies for optimally integrating information from various linked data blocks (that may differ in size, amount of noise, redundancy, relatedness to other blocks etc.). This work, which resulted in several top journal papers, meets a broad need as nowadays problems of data fusion are ubiquitous in many research disciplines.

4.1.2 Perspectives

Each of the results described above gives a (partial) answer to an important research question. In many cases, there are however still a lot of unsolved related questions, which definitely deserve and require further close attention in the near future. The following list of research topics is meant to give an idea of the research questions the network members plan to work on during the coming years.

- *Frailty models*
  There are still quite a few open problems related to multivariate survival data. First, there is still the need for diagnostic tools (aptness of the model), very few such tools are available. Secondly, many of the existing statistics to summarize data, such as Kendall's tau and the cross-correlation coefficient, are derived from bivariate survival data, mostly from a point of view of the copula model. This warrants extension to a multivariate setting. The correlated frailty model -as compared to the shared frailty model- is extensively used in the genetic analysis of twin data. These models should be extended to more dimensions than the bivariate case, as also in higher dimensions the need arises to have different correlations between the different members in a cluster. Finally, also new correlation structures need to be adopted in the frailty model context to describe and model spatial survival data. Only a limited amount of research has been done up to now on spatial survival data analysis in the context of the frailty model.

- *Modeling dependencies through copulas*
  Semiparametric modeling of conditional copulas is a very interesting open research avenue. Through such modeling the influence of covariates on the dependence structure between two variables can be modeled in a flexible way, allowing to capture many interesting dependencies in applications, but at the same time referring to important classes of copula functions. In addition, this framework allows to testing for the null hypothesis that the dependence structure between the two variables does not depend on the covariates. This issue is of importance in many applications. Tackling such testing problems is an unchartered area.

- *Flexible modeling and regularization techniques*
  A lot of experience has been gained in the last year with using P-splines regression techniques in powerful variable selection procedures. Such procedures aim not only at selecting the significant variables among a large set of variables, but at the same time come up with a nonparametric estimation of the impact that these selected variables have on the variable of interest. In other words, the statistical procedures should be estimation consistent as well as selection consistent. Crucial is thus to make appropriate selection of the regularization parameters such that both tasks – estimation and selection – are done properly. This requires a selection criterion that in fact focuses on both goals – estimation and selection – simultaneously. This is a challenging open question. Another challenging issue is how to robustify the developed variable selection procedures, so that the influence of aberrant observations can be diminished.

- *Frontier estimation*
  In nonparametric frontier estimation the bootstrap is the only way to implement statistical inference in practice. Subsampling techniques provide a very appealing approach where the subsample size can be determined by data driven techniques. This has now to be implemented

for most of the testing problems that are of interest for the economists (convexity of the production set, returns to scale, separability, technical changes, etc.). The link with extreme value theory opens new perspectives of deriving new (regularized) frontier estimators being robust to extreme and outliers and having simple asymptotic normal distribution. The pointwise results obtained so far should be extended to cases where regularity of the frontier function should help in deriving estimators with even better properties. Introducing noise in the nonparametric frontier models remains a challenging issue: this is linked to deconvolutions problems. We will investigate new methods allowing to identify, under a minimal set of assumptions, noise from the signal (inefficiencies). Again here, partial frontiers (quantile and order-m frontiers) could be very useful because they rely on statistical objects (survival functions) which are easier to estimate than the boundary of the support of the observed variable. Then in a second stage the inherent bias introduced by these partial frontiers could be removed. The use of these partial frontiers should also be extended in order to define more general directional distance functions (and not only radial distances).

- *Inference for semiparametric M- and Z-estimators*
  There are still many open problems in the area of semiparametric M- and Z-estimation. The theory for M-estimators is useful whenever the criterion function is not differentiable, in which case an M-estimation problem cannot be reduced to a Z-estimation problem by taking the derivative of the criterion function. It would therefore be very useful to develop a general theory for this class of estimators. This theory will open new horizons for a wide variety of problems in semiparametric statistics. There are other related problems on which only limited or no research has been done so far, and which should be able to tackle with the expertise built up over the last years. These include inference for semiparametric conditional Z- and M-estimation problems, the use of sieves to estimate jointly the parametric and the nonparametric component in a general semiparametric model, the case where the dimension of the parameter vector in a semiparametric model grows with the sample size, inference for semiparametric models by means of empirical likelihood, etc.

- *Nonparametric location-scale models*
  As mentioned in Section 4.1.1 lots of work on nonparametric location-scale models remains to be done. The recently obtained extension to the case of multiple regression opens lots of perspectives for further research. For instance the study of semiparametric location-scale models, based on e.g. a partial linear or single index structure for the regression and/or scale function are now possible. Another topic of further interest to the network is the use of nonparametric location-scale models to estimate the frontier function in a more accurate way. Finally, it is planned to study the location-scale models in the context of mixed effects model, opening in this way an interesting bridge between WP1 and WP4.

- *Robust model selection*
  For only a few multivariate models computationally efficient and robust model selection procedures have been developed. There still remains a lot of work to develop such methods for other multivariate models, such as classification models. Moreover, the current methods are not yet optimal. There is still a lot o room for improvement.

- *Robustness in high-dimensional data*
  This important area of research has just emerged by showing that the standard robust methods fail in realistic high-dimensional contamination models. The next step is to develop methods that are robust according to these high-dimensional contamination models.

- *Dimension reduction in high-dimensional data and data fusion*
  Classical dimension reduction techniques, along with their typically very insightful graphical representations, often break down in situations with very high-dimensional data (which nowadays show up often in statistical practice). This implies the intriguing challenge of developing novel types of custom-made reduction approaches, such as methods that imply multiple nested types of (categorical and dimensional) reductions, and variable selection/sparseness techniques applied to new statistical areas such as that of data fusion.

4.1.3 Recognition/critical mass

The IAP network has built up over the last years an important recognition at national and international level. This recognition is the result of (1) the fact that the network is getting well known in Belgium and abroad; (2) successful collaborations between members of the network; and (3) important contributions/achievements of individual members of the network. The following list demonstrates this recognition by means of a number of examples.

1. *Actual position of former PhD students.*
   Many of the PhD students in the network have obtained after their PhD postdoctoral or academic positions. The following list gives an impression of their present positions.

   - Carlos Almeida (UCL): postdoctoral researcher at the University of Münich
   - Katrien Antonio (KUL-1): assistant professor, University of Amsterdam
   - Johan Braeken (KUL-1): assistant professor at Tilburg University
   - Frank Coopman (UG): assistant professor at Artevelde Hogeschool
   - Michiel Debruyne (KUL-1): lecturer, Antwerp University
   - Anouar El Ghouch (UCL): assistant professor at UCL
   - Gery Geenens (UCL): assistant professor at University of New South Wales (Australia)
   - Giovanni Motta (UCL): postdoctoral researcher at Maastricht University
   - Ella Roelant (UG): part-time lecturer at University of Newcastle
   - Jan Schepers (KUL-1): assistant professor at Maastricht University

2. *Actual position of former postdocs.*
   The postdocs who have worked under the IAP-network, often found academic positions afterwards:

   - Isabel Casas (KUL-1): postdoctoral researcher at Aarhus University
   - Eva Ceulemans (KUL-1): assistant professor at KUL-1

- Rosa Crujeiras-Casais (UCL): assistant professor at USC
- Tijl De Bie (KUL-1): lecturer at University of Bristol
- Laurent Delsol (UCL): assistant professor at Université d'Orléans
- Beatrijs Moerkerke (UG): assistant professor at UG
- Marek Omelka (KUL-1 and UH): assistant professor at Charles University Prague

3. *Important points of recognition of network members.*
   A number of members of the network have obtained important prizes, medals, elected fellowships, lectureships, etc. These points of recognition contributed in a large extent to the building of an important recognition of the network.

   - The network has been involved in two Belgian Francqui chairs for the academic year 2008–2009. T. Snijders (member of the follow-up committee) is recipient of a chair awarded to KUL-1. G. Molenberghs (UH) is recipient of a chair awarded to the University of Antwerp.
   - L. Simar (UCL) has held the 'Chaire d'Excellence Pierre de Fermat', Université Toulouse I (2008-09), which is a prestigious research chair financed by the 'Conseil Régional de Midi-Pyrénées'.
   - I. Van Keilegom (UCL) obtained in 2008 an important five-year grant from the ERC (European Research Council) under the European Community's Seventh Framework Programme. The grant (750,000 €), which is on 'M- and Z-estimation in semiparametric statistics: applications in various fields', was obtained after a competitive selection (success rate over the whole of Europe: 2.5 %). In 2008, she also received Elected Fellowship of the IMS (Institute of Mathematical Statistics).
   - R. von Sachs (UCL) received Elected Fellowship of the IMS (Institute of Mathematical Statistics) in 2009.
   - E. Lesaffre (KUL-2) received Elected Fellowship of the American Statistical Association, for his work on hierarchical, complex data, approached from a Bayesian angle.
   - G. Molenberghs (UH) is Member of the National Academy of Sciences (U.S.) Oversight Committee on the 'Handling of Incomplete Data in Clinical Trials', an initiative taken by the U.S. Food and Drug Administration. He is also the Belgian representative of the Steering Committee for the EU sponsored program 'Quantitative Methods in the Social Sciences', thereby building a bridge between various working packages.
   - G. Verbeke (KUL-2) is International Representative on the Board of Directors of the American Statistical Association.
   - G. Molenberghs (UH) and G. Verbeke (KUL-2) received, for the third time, the Excellence-In-Continuing-Education Award for 2008, for their joint short course 'Models for discrete repeated measures,' taught on August 5, 2008, at the Joint Statistical Meetings in Denver, U.S.A.

- I. Van Mechelen (KUL-1) has been elected President of the International Federation of Classification Societies (IFCS), the worldwide federation of national, regional, and linguistically-based societies for research on (unsupervised and supervised) classification methods (2010-2011 President-elect, 2012-2013 President, 2014-2015 Past President).

4. *Editorships/associate editorships of network members.*

   A number of members of the IAP network have contributed to the international recognition of the network, by accepting editorships/associate editorships of important journals in statistics.

   - G. Verbeke (KUL-2) is former editor of the Journal of the Royal Statistical Society, Series A and current editor of Biometrics.
   - G. Molenberghs (UH) is former editor of Biometrics and current editor of Biostatistics.
   - S. Van Aelst (UG) is associate editor of Journal of Statistical Planning and Inference (2006-) and Computational Statistics and Data Analysis (2007-). For the latter journal he was guest editor for a special issue on Machine Learning and Robust Data Mining in 2007 and for a special issue on Variable Selection and Robust Procedures that will appear in 2010.
   - I. Van Keilegom (UCL) is associate editor of Annals of Statistics (2007-), Scandinavian Journal of Statistics (2004-), Annals of the Institute of Statistical Mathematics (2006-), Statistics and Probability Letters (2007-) and International Journal of Biostatistics (2008-).
   - L. Simar (UCL) is associate editor of Journal of Productivity Analysis.
   - J. Segers (UCL) is/was associate editor of Journal of the Royal Statistical Society - Series B (2005-2009), Bernoulli, Stochastic Processes and Their Applications, Advances in Applied Probability, and Journal of Applied Probability.
   - C. Legrand (UCL) is associate editor of Biometrics.
   - I. Gijbels (KUL-1) is/was associate editor of Annals of Statistics (2004-2009), Journal of Computational and Graphical Statistics (2000-) and Journal of Nonparametric Statistics (2010-).
   - M. Hubert (KUL-1) is associate editor of Computational Statistics & Data Analysis and Technometrics.

4.2 International role

4.2.1 Collaboration with the European partners

The added value of having the European partners in the network is clear and important. The European partners have been chosen for their international recognized expertise in the topics studied in the network. They have clearly demonstrated and shared their expertise with the network members through scientific collaborations, organization and participation to workshops and meetings of the network, mutual visits between the European and the Belgian partners, etc.

Thanks to the network, the contact and collaboration with the European partners has certainly risen to a much higher level than before.

The scientific collaboration with the European partners has been very diverse and intensive :

- KUL-1 and UJF have a longstanding history of successful collaboration. Within Phase VI this collaboration has been extended to several new topics, including regularization techniques for varying coefficient models, variable selection within the context of mixture models, and sparseness techniques within the context of component and unfolding models. Moreover, the previous collaborations have also been extended to several new members of both universities (e.g., S. Lambert-Lacroix, K. Van Deun, and A. Verhasselt) .
- KUL-2, UCL and UU have collaborated on B-spline models and Bayesian inference procedures motivated by and applied to a variety of problems coming from medicine and dental research.
- UCL and USC have extensively collaborated on topics at the heart of WP1 and WP3, like e.g. on empirical likelihood procedures, goodness-of-fit tests, nonparametric location-scale models, and ROC curves in the presence or absence of censored data. Recently, collaborations started on semiparametric random effects regression models, which have a strong link with WP4.
- UH, KUL-2, UG and LSHTM have collaborated extensively on incomplete data and sensitivity analysis, situated within the theme of WP3. The same partners have contributed to work on hierarchical data, through mixed models and beyond, the theme of WP4.

More details about the collaborations with and the visits to the European partners can be found in Sections 3.1.4 and 3.1.6.

4.2.2 International activities

Many members of the network have been very active on international level, by participating in European and international research networks, giving important invited presentations at conferences, giving short courses, collaborating with researchers from all over the world, organizing international symposia, etc.

1. *Participation in European and international research networks*
    - The European partner USC (W. González Manteiga) is principal investigator and UCL (I. Van Keilegom) is co-investigator of the grant 'Metodologia y aplicaciones en estadistica semiparametrica, funcional y espacio temporal' from the Spanish Ministerio de Educación y Ciencia.
    - L. Simar (UCL) is Quality Manager for the EUMIDA (European Universities Micro Data) Consortium. This consortium, which is led by the University of Pisa, has been appointed by the European Commission (DG Research, DG Education and Culture and EUROSTAT) to explore the feasibility of building a consistent and transparent European statistical infrastructure at the level of individual higher education institutions. The goal is to provide these institutions and policy makers with relevant information

for the benchmarking and monitoring of trends for modernization in higher education institutions. EUMIDA will propose a sustainable infrastructure for collecting data on a regular basis, develop the methodological components and build up a complete census of EU higher education institutions.

- S. Van Aelst (UG) is member of the 'European Research Consortium for Informatics and Mathematics' (ERCIM) working group on Computing & Statistics and co-chair of the track on Robust Analysis of Complex Data Sets.
- S. Van Aelst (UG) is research member of the 'European Science Foundation' (ESF) COST project on 'Combining Soft Computing Techniques and Statistical Methods to Improve Data Analysis Solutions' (Action IC0702 - SoftStat).

2. *Important invited presentations at conferences*
   The following list contains a number of examples of important invited presentations given by members of the network.

   - G. Molenberghs (UH) has given the following important keynote and inaugural lectures:
     - Inaugural lecture on 'The applied statistical scientist in a high-profile academic environment' at the Faculté Universitaire de Gembloux on April 18, 2007.
     - Keynote lecture on 'The meta-analytic framework for the evaluation of surrogate endpoints in clinical trials' at the Pharma Statistics Conference, Johnson & Johnson Pharmaceutical Research and Development, Titusville, New Jersey on September 26, 2007.
     - Closing lecture on 'The applied statistical scientist in a high-profile academic environment' at the IV Meeting of the Central American and Caribbean Region of the International Society and IV Colombian-Venezuelan Statistical Meeting, Margarita Island (Venezuela) on October 5, 2007.
     - Inaugural lecture on 'The applied statistical scientist in a high-profile academic environment' on the occasion of the reception of the Belgian Francqui Chair, University of Antwerp on February 23, 2009.
     - Series of lectures, on the occasion of the reception of the Belgian Francqui Chair, University of Antwerp on March 5, 12, 26 and 30, 2009.
     - Invited lecture on 'Longitudinal and incomplete data in clinical studies' at the Biometric Seminar, Osaka (Japan) on July 8, 2009.
   - G. Verbeke (KUL-2) has given an invited lecture in the Biometrics Showcase Session at the International Biometric Conference, Dublin, Ireland, July 13-18, 2008. In addition, he has given important invited presentations at the Joint meeting of the Western North American Region of The International Biometric Society and the Institute of Mathematical Statistics, Portland (June 14-17, 2009), at the Third Annual Clinical Forum, Nice (October 19-21, 2009) and at the Spring meeting of the Eastern North American Region (ENAR) of the International Biometric Society, New Orleans (March 21-24, 2010).
   - Between 2007 and 2010, I. Gijbels (KUL-1) was invited speaker at several international conferences and presented seminars on invitation in various countries.

- I. Van Mechelen (KUL-1) has given invited lectures at various international conferences including COMPSTAT, meetings of the Gesellschaft für Klassifikation, and the Convention of the Association for Psychological Science.
- L. Simar (UCL) was keynote speaker at the 15th Annual Meeting of The Belgian Statistical Society, which took place on 18-20 October 2007 at the University of Antwerp, He was also keynote speaker and member of the scientific committee of the 11th European Workshop on Efficiency and Productivity Analysis, which took place on 26-29 June 2009 in Pisa, Italy.
- R. von Sachs (UCL) has given an invited talk at the symposium 'Wavelets in the presence and in the future; in the honor of Ingrid Daubechies', May 27, 2008, University of Hasselt.
- I. Van Keilegom (UCL) will be keynote speaker at the 'Stochastics Meeting Lunteren'. the annual meeting of Dutch mathematical statisticians and probabilists, which will take place on 15-17 November 2010. She has also been an invited speaker at 12 international conferences in 2007, 2008 and 2009, and presented 19 seminars on invitation in various countries during this period.

3. *Short courses*

- Centered around the themes of WP3 and WP4, G. Verbeke (KUL-2) and G. Molenberghs (UH) have given about *50 short courses*, around the world (Belgium, the Netherlands, France, Switzerland, Denmark, the United Kingdom, Spain, the United States of America, Cuba, Venezuela, Uruguay, Argentina, Brazil, Mozambique, South Africa, and Japan). Some examples:
    - 'Models for repeated discrete data', International Workshop on Statistical Modeling, Barcelona (Spain), July 1, 2007
    - 'Models for longitudinal and incomplete data', 73rd Annual meeting of the Psychometric Society, Durham (U.S.A.), June 29, 2008
    - 'Models for discrete repeated measures', Joint Statistical Meetings, Denver (U.S.A.), August 4, 2008
    - 'Models for discrete repeated measures', Joint Statistical Meetings, Washington DC (U.S.A.), August 2, 2009
- Members of the KUL-1 group organized and taught a short course on 'A nonlinear mixed models approach to IRT' at the National Council on Measurement in Education (NCME), San Diego (April 12, 2009) and New York (March 24, 2008).
- L. Simar (UCL) has given the following short courses since 2007 :
    - short course on 'Efficiency Analysis: the econometric approach' at the University of Bologna (Italy) on 24-29 April 2008
    - summer school on 'Advanced Quantitative Methods: Applied Multivariate Data Analysis, Bootstrapping and Productivity/Efficiency Measurements' at the Università di Pisa (Italy) on 6-12 July 2008

  - course on ‘Introduction to the Bootstrap with Applications in Econometrics’ at University of Bologna (Italy) on 14-16 April 2009
  - course on ‘Bootstrap, Panacea for Statistical Inference?’ at Università di Verona (Italy) on 20-23 April 2009
- I. Van Keilegom (UCL) has given a short course on ‘Inference for semiparametric Z-estimators’ at Indiana University (February 2010) and at the Toulouse School of Economics (April 2010). She will also give this short course in August 2010 during the European Meeting for Statisticians in Piraeus (Greece).
- L. Duchateau (UG) and P. Janssen (UH) gave a short course on frailty models at the Erasmus University Medical Center Rotterdam in October 2009.
- In the COMISEF - COST (Action IC0702 - SoftStat) Tutorial on Statistical Model Selection, organized in Cyprus, 26-28 October, 2009, S. Van Aelst (UG) gave the part on ‘Robust strategies and model selection (including R hands-on session)’.

4. *Important international collaborations*

   Many network members are very active on international level. Some example of the most important international collaborations are mentioned below.

   - WP1 : Research projects on WP1 have been carried out in collaboration with R. Carroll (Texas A & M University), J.-P. Florens (University of Toulouse), P. Hall (University of Melbourne), J. Einmahl (University of Tilburg), P. Wilson (Clemson University), B. Park (University of Seoul), A. Daouia (University of Toulouse), C. Genest (Université Laval), R. Zamar (University of British Columbia), A. Delaigle (University of Bristol and University of Melbourne), L.-A. Garcia-Escudero (University of Valladolid) (among others)
   - WP2 : Research activities for this workpackage have been conducted in collaboration with (among others): O. Linton (London School of Economics), J. C. Escanciano (Indiana University), R. Dahlhaus (University of Heidelberg), E. Mammen (University of Mannheim), J. Gao (University of Adelaide), B. Basrak (University of Zagreb), M. Lee (University of California), J. Rouder (University of Missouri)
   - WP3 : Research cooperations have been conducted with R. Cao (University of La Coruña), N. Hjort (University of Oslo), S. Chen (Iowa State University), L. Wang (University of Minnesota), V. Ducrocq (INRA, France), A. Wienke (Halle University) and D. Harrington (Harvard School of Public Health), I. McKeague (Columbia University) (among others)
   - WP4 : Research projects on WP4 have been carried out in collaboration with G. Fitzmaurice (Harvard School of Public Health), M. Davidian (North Carolina State University), C. Demetrio (Escola Superior de Agricultura, Priacicaba, Brazil) (among others).
   - WP5 : Research activities for this workpackage have been conducted in collaboration with (among others): M. Salibian-barrera (University of British Columbia), V. Yohai (University of Buenos Aires), R. Zamar (University of British Columbia), A. Gelman

(Columbia University), H. Kiers (University of Groningen), A. Smilde (University of Amsterdam)

5. *Organization of international symposia*

   Apart from the international workshops the network organizes on a yearly basis (see Section 3.1.1 for more details), members of the network had important responsibilities in the organization of some international symposia. We list here some examples.

   - KUL-1 (I. Gijbels) and G. Claeskens organized a Leuven Statistics Day on 'Semi- and Nonparametric Statistical Analysis and Smoothing Techniques', KULeuven, May 25, 2007.
   - R. von Sachs (UCL) co-organized the International Workshop on 'Recent Advances in Time Series Analysis', which took place on June 8-11, 2008 in Protaras (Cyprus).
   - S. Van Aelst is/was co-chair of the First, Second and Third Workshop of the ERCIM Working Group on Computing & Statistics, which took (will take) place in Neuchâtel (Switzerland) on June 19-21, 2008, in Limassol (Cyprus) on October 29-31, 2009 and in London on December 11-13, 2010 (respectively).
   - On November 5-7, 2008, the 'International Seminar on Nonparametric Inference' (ISNI2008) took place in Vigo, Spain. The meeting was co-organized by the European partner of Santiago de Compostela under the heading of the IAP network. Three members of the IAP network, as well as A. Davison (member of the follow-up committee) gave invited lectures on this meeting. Many members of the network participated to this meeting (especially from UCL, UH and UJF).
   - The KUL-1 team of I. Gijbels was main organizer of an international conference on 'Flexible Modeling: Smoothing and Robustness' that took place on November 12-14, 2008, at the KULeuven. Among the invited speakers were two European-partner members of the IAP-network (A. Antoniadis and P. Eilers). The international workshop was attended by approximately hundred participants including several members of the IAP-network.
   - The UCL-partner (I. Van Keilegom) organized an international conference on 'Exploring research frontiers in contemporary statistics and econometrics' in honor of the main coordinator of the network, Léopold Simar (UCL), on the occasion of his retirement. The conference took take place on May 14-15, 2009 at the UCL, and was co-organized by the network.
   - R. von Sachs (UCL) has organized a Miniworkshop on 'Semiparametric modeling for multivariate time series in econometrics and finance', which took place in Oberwolfach (Germany) on January 17-23, 2010.
   - The European partner of Santiago de Compostela (Spain) will organize the METMAV Conference from 30 June to 2 July 2010. The purpose of this conference is to promote the development and application of spatio-temporal statistical methods in different fields related to environmental sciences.

4.3 Durability of the IAP

The need for a continuation of the existing research network can be motivated by means of many arguments :

- As can be seen from Figure 2 in Section 3.3 the efforts to work together have resulted in a vast increase in the number of joint publications with two or more partners in the network. It is clear that it would therefore be highly desirable to continue the network activities with the common expertise built up over the last 10 years acting as a major asset and as a sound basis to launch top-level future joint work.
- The network has further ensured that Belgium is on the map from a statistical point of view. The network partners have intensive collaborations, and bring together theoretical and applied expertise across a variety of areas. This is a very strong asset worth of further support.
- Statisticians, due to the interdisciplinary nature of their expertise and teaching duties, are typically scattered among faculties, schools, departments, etc. More than any other field, they need help in order to meet and share their knowledge and experience. This network is a vital opportunity for the main statistical groups of Belgian universities to exist as a community.
- The methodology used in different areas of statistics (biostatistics, psychometrics, econometrics, etc.) is often similar in nature. It is therefore important to have a network in order to combine forces and to look at statistical problems from different angles.
- In practical settings problems come together. What typically needs to be done is the integration of different methodologies to tackle practical complex research questions. The existence of a research network in which expertise on these various research topics is present, is therefore desirable.
- The network has proven to be successful in its research output (papers, books, proceedings, invited and contributed presentations, seminars). Further, it is very internationally active, in a large variety of ways, including but not restricted to, inviting scholars and paying scientific visits, giving courses abroad at universities, companies, and governmental agencies, etc. Its international reputation can also be testified by e.g. the high number of papers that are published in top journals in their field, the fact that PhD students are often offered academic positions after graduating, etc. As such, it is recognized as a strong international partner, for a number of areas even as a world leader.

Overall, the network has been very successful so far and one would like to continue with the same general architecture. All Belgian and European partners have been very active and working together a lot. The sharing of knowledge and information in the network is working very well, there is a very good team spirit and a real intention to push the network as far as possible on the international level. The network would however like to improve the current linguistic imbalance in the next phase : at present there is only one partner in the French speaking part (UCL, who is the main coordinator) against four partners in the Flemish speaking part (KUL-1, KUL-2, UG and UH). The network would like to add a group at the Université Libre de Bruxelles and/or the Université de Liège to the network, in order to arrive at a more balanced situation. At both universities research is done on the themes of the network, and they could both be an important scientific asset for the network.

5 Output

5.1 Most relevant publications

The following list contains a selection of the most relevant publications, written by members of the IAP-network.

- Antoniadis, A., Gijbels, I. and Poggi, J.-M. (2009). Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica, Special issue: Multiscale Methods and Statistics: A Productive Marriage*, **19**, 1371-1387.
- Braeken, J. and Tuerlinckx, F. (2009). A mixed model framework for teratology studies. *Biostatistics*, **10**, 744-755.
- Duchateau, L. and Janssen, P. (2008). *The Frailty Model.* Springer, New York.
- Goetgeluk, S., Vansteelandt, S. and Goetghebeur, E. (2008). Estimation of controlled direct effects. *Journal of the Royal Statistical Society - Series B*, **70**, 1049-1066.
- Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Annals of Statistics*, **37**, 1079-1111.
- Molenberghs, G., Beunckens, C., Sotto, C., and Kenward, M.G. (2008). 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**, 371-388.
- Omelka, M., Gijbels, I. and Veraverbeke, N. (2009). Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *Annals of Statistics*, **37**, 3023-3058.
- Rizopoulos, D., Verbeke, G. and Molenberghs, G. (2008). Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63-74.
- Simar, L and Wilson, P. (2007), Estimation and Inference in two-stage, semiparametric models of production processes, *Journal of Econometrics*, **136**, 31-64.
- Verbeke, G. and Fieuws, S. (2007). The effect of misspecified baseline characteristics on inference for longitudinal trends in linear models. *Biostatistics*, **8**, 772-783.

5.2 Appeal of the IAP

In order to make the IAP network well visible to the scientific community and in order to be considered as an active, productive and stimulating research network, we have taken the following initiatives:

- *Logo.* The logo has been created at the start of Phase V of the network. It can be found on the front page of this report, as well as on the web page of the network (`http://www.stat.ucl.ac.be/IAP/PhaseVI`). It has been designed by the young researchers of the UCL-team and includes a number of hidden statistical concepts.

- *Website.* An extensive website has been set up, which contains information about e.g. call for applications, description of the project, list of personnel working under the IAP project, member lists,... See Section 3.3 for more information.
- *Technical Reports and Publications Series.* Two series, available via the website, report on scientific results obtained within the IAP-statistics network: the Technical Report Series and the Publications Series. See Section 3.3 for more details.
- *Email lists.* A regularly updated list containing all email addresses of members of the network is used to disseminate information about short courses, visitors, calls for applications, ... In addition, this information can be found on the web page.
- *Other ways to make the IAP visible.* Announcements of pre/post-doc positions of the IAP network are widely spread via numerous email lists all over the world. The activities of the network are also mentioned in B-Stat News, the official journal of the Belgian Statistical Society.

The visibility of the IAP network can also be seen from the high number of foreign PhD and postdoctoral researchers that come from their country to one of the partner universities of the network to develop their scientific career. The wide variety of countries that are represented in the network, also strengthens on its turn even more the international reputation and visibility.

5.3 PhD's and postdoc training

The following table summarizes the number of PhD students and postdocs that were trained in the network. The numbers in the table represent the average number **per year**, and are based on the number of researchers mentioned in the annual administrative reports. For simplicity, no distinction is made between part-time and full-time positions.

Partner	PhD		Postdoc	
	IAP	not IAP	IAP	not IAP
UCL	2	18	1	0
KUL-1	3	18	1	9
KUL-2	1	7	0	1
UG	3	9	0	2
UH	2	25	1	9
UJF	0	2	0	0
UU	0	0	0	0
USC	0	1	0	2
LSHTM	0	0	0	1
Total	11	80	3	24

The added value the network offers to the careers of PhD and postdocs can be described as follows :

- We witness in general a lack of statisticians in many research domains; through the training as organized by the IAP-network and the openness for different research domains as being stimulated by exchanges in the network, the network yielded well-trained, open-minded young statisticians that may subsequently contribute to many disciplines other than statistics. (One element to support this is the easiness with which young statisticians stemming from the network find jobs in all kinds of research domains.)
- Thanks to the IAP network, there is a better exchange of ideas between students of the different partners. This can be explained by the fact that they meet each other more regularly than before the existence of the network (at workshops, meetings, joint seminars, short courses,...). Fruitful scientific interactions are the result of this.
- The short courses offer a unique chance to the young researchers to learn from specialists in the field about areas they are less familiar with. It is an efficient way to broaden their view on statistics.
- The network has organized or co-organized a high number of short courses often given by experts in the field. See Section 3.1.5 for a detailed list of the organized short courses. These courses are a very valuable asset for the PhD students of the network to get familiar with other research areas, and to get to learn other Belgian and international researchers. Among the short courses organized by the network, the following ones are examples of short courses that brought together many people from the network and from outside, and that were characterized by lots of discussions and social contacts among the participants:
  - Summer school on ‘Advanced nonparametric statistics’ organized by the UCL in May 2007, including a total of three short courses on related topics. Lots of international advertisement was made for this summer school, which was solely organized by the IAP, and which resulted in an important international visibility of the network.
  - Short course on ‘Introduction to numerical techniques for statisticians’, given by Geert Molenberghs (UH), Francis Tuerlinckx (KUL-1) and Geert Verbeke (KUL-2), at KUL, organized by KUL-1, KUL-2, UH in April 2008. The short course was a big success, thanks to the high quality of the presentations, the international reputation of the speakers, and the fact that statisticians throughout all subdisciplines are confronted with numerical problems when implementing their research work via statistical computer packages.
  - Short course on ‘Bayesian data analysis’, by Andrew Gelman (Columbia University, New York), organized by KUL-1 in February 2010. More than 180 participants registered for this course, including many PhD students and postdocs from all IAP partners. This activity also contributed significantly to the visibility of our network.

5.4 New research teams

With respect to Phase V of the network, the following groups have joined the network at the start of Phase VI : the group of the University of Ghent, and the European partners UU, USC and LSHTM. Thanks to the network, the group of UG (L. Duchateau) has developed important research contacts with the partners of UCL (in particular with C. Legrand), the colleagues of KUL-1 (M. Hubert), the partners at KUL-2 (with G. Verbeke and colleagues), the group at UH (with P. Janssen and his students) and also with the European partner LSHTM in London (with J. Carpenter and colleagues). The UG group has organized the annual workshop of the network in 2008. At that occasion they also organized a social event (tourist tour in Ghent) in order to get to know the members of the network better. They have also set up and lead a number of working groups, the most important one being the one on frailty models which brought together researchers from UG, UH, KUL-2 and UCL to discuss common research problems related to clustered data and competing risks in survival analysis. The UG group (and in particular L. Duchateau) has played and still plays a very dynamic role in this working group. So, one can fairly say that there have been considerable efforts to integrate into the network, and that this integration has been a success story.

A close scientific collaboration between the European and Belgian partners has been developed over the years. It's clear that the scientific expertise of the European partners in the research topics of the network has been very useful for the network, and has resulted in a considerable number of publications in high level and even top international journals.

Appendix:

Publications of the IAP network

during the period

January 2007 - April 2010

A Organization of the Appendix

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, in 2008 and in 2009 (plus the four first months of 2010). Different types of papers are considered. For the year 2007 we used only two categories (namely 'technical reports' and 'publications'). From 2008 on, we refined our reference system. Below we explain the different types and the corresponding reference numbers used to distinguish between them. We restrict attention to 2008, but the same system also applies to 2009 and 2010.

- Technical reports: These are manuscripts that have been written in 2008, and have been submitted for publication to an international journal. The reports are also available on our web site:

  `http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_tr.html`.

  Each Technical Report has a number of the form TR08xxx, and we mention these reference numbers below. The website also contains the pdf-file of many of the Technical Reports.

- Refereed publications: We list all published papers in international journals in 2008 (with refereeing system). We make the distinction between published papers and papers in press. See also the IAP-Statistics Reprints Series on our web site:

  `http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_reprint.html`,

  for the published papers (reference numbers are of the form R08xxx). The papers in press have a label of the form RP08xxx.

- Non-refereed publications: We also include (an incomplete list of) papers that have been published without undergoing a peer review. The reference numbers are of the form NR08xxx (for the published ones) and NRP08xxx (for the ones in press).

- Books: These are books written by members of the network, that are published by international editors. They can also be found on the webpage

  `http://www.stat.ucl.ac.be/IAP/PhaseVI/publication_books.html`

  (reference numbers are of the form B08xxx and BP08xxx).

Below we list the research output of the IAP-network for each of the categories described above. We start with separate lists for each partner in the network, followed by a list of the technical reports and publications that are co-signed by researchers from at least two different groups from the network.

Important remark

It is important to note that certain papers appear two or three times in the Appendix. In fact, given that we include in the reference list not only published papers but also technical reports and papers in press, it is possible that a paper is mentioned e.g. as technical report in 2007, as paper in press in 2008 and as published paper in 2009. Therefore, although the total number of manuscripts mentioned in this Appendix is given by 1714, the number of distinct papers is estimated to be more or less equal to 1000.

B Publications of the network in 2007

B.1 Université catholique de Louvain, UCL

B.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 Tilleghem, 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 succesive 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.

B.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., Räty, T. and L. Simar, Efficiency and university size: disciplinewize evidence from the European universities. *Hyvinvointipalvelujen Tuottavuus: Tuloksia opintien varrelta*, VATT Publications, Helsinki, Finland, 2007. R07109.

[40] Bonaccorsi, A., Daraio, C. and L. Simar, Efficiency and productivity in European universities. 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 University System*, Edward Elgar PRIME Collection, 2007. R07108.

[41] Boulanger, B., Dewé, W., Gilbert, A., Govaerts, B. and M. Maumy-Bertrand, Risk management 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 Archimedian copulas: characterizations and pitfalls. *Insurance: Mathematics and Economics*, 40, 525-532, 2007. R07074.

[43] Charpentier, A. and J. Segers, Convergence of Archimedean copulas. *Statistics and Probability 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 estimation for dependent observations. *Bernouilli*, 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 quantile approach. *Journal of Econometrics*, 140, 375-400, 2007. R07113.

[48] Daraio, C. and L. Simar, *Advanced Robust and Nonparametric Methods in Efficiency Analysis: 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 Levell and Shelton Schmidt editors, Oxford University Press, 2008. R08002.

[78] Simar, L. and V. Zelenyuk, Statistical inference for aggregates of Farell-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.

B.2 Katholieke Universiteit Leuven, KUL-1

B.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 muturity 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. Raîche, 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.

B.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., Ceulemens, 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 Behavorial 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.

B.3 Katholieke Universiteit Leuven, KUL-2

B.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.

B.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 Weerdt, 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 Weerdt, 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.

B.4 Universiteit Gent, UG

B.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.

B.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., Lefebre, 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 an 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.

B.5 Universiteit Hasselt, UH

B.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.

B.5.2 List of Publications

[252] Aelvoet, W., Windey, F., Molenberghs, G., Temmerman, M., Van Reepts, 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., Nabholtz, J.-M., Paridaens, R., Biganzoli, L., Jassem, J., Bontenbal, M., Bonneterre, 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., Caroll, 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 polynominals 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 *Longidutinal 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, Occurence 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. Bijnens, 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. Bijnens, 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. Bijnens, 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. Bijnens, 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. *Lifetime 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., Nabholtz, J.-M., Paridaens, R., Biganzoli, L., Jassem, J., Bontenbal, M., Bonneterre, 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] Robaeys, 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 aarabidopsis 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 chemonaive patients with squamous cell carcimona 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., Burzykowksi, 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 hemilaminecdtomy without fenestration in the treatment of thoracolumbar disc disease in chrondrodystophic 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] Valkenborg, 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 Spectronometry*, 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 intraducal 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., Bijnens, 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. *Ophtalmology*, 144, 450-453, 2007. R07198.

B.6 Université Joseph Fourier, UJF–LMC–IMAG

B.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. Tourneret, 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.

B.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.

B.7 Universiteit Utrecht, UU

B.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 spectronometry-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.

B.8 Universidad de Santiago de Compostela, USC

B.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 hetoroskedastic 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.

B.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.

B.9 London School of Hygiene and Tropical Medicine, LSHTM

B.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.

B.9.2 List of Publications

[418] Carpenter, J.R., Kenward, M.G. and I.R. White, Sensitiviy 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., 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.

[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.

B.10 List of joint publications

B.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.

B.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.

C Publications of the network in 2008

C.1 Université catholique de Louvain, UCL

C.1.1 Technical reports

[463] Badin, C., Daraio, C. and L. Simar, Optimal bandwith selection for conditional efficiency measures : a data-driven approach, 2008. TR08038.

[464] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, 2008. TR08035.

[465] Böhm, H. and R. von Sachs, Structural shrinkage of nonparametric spectral estimators for multivariate time series, 2008. TR08006.

[466] Bouezmarni, T., Rombouts, J. and A. Taamouti, Asymptotic properties of the Bernstein density copula for dependent data, 2008. TR08042.

[467] Charpentier, A. and J. Segers, Tails of multivariate Archimedean copulas, 2008. TR08013.

[468] Crujeiras, R. M. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends, 2008. TR08044.

[469] Daouia, A., Florens, J.P. and L. Simar, Frontier estimation and extreme values theory, 2008. TR08008.

[470] Daskovska, A., Simar, L. and S. Van Bellegem, Forecasting the Malmquist productivity index, 2008. TR08032.

[471] 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.

[472] Einmahl, J. and J. Segers, Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution, 2008. TR08028.

[473] Ferraty, F., Van Keilegom, I. and P. Vieu, On the validity of bootstrap in nonparametric functional regression, 2008. TR08027.

[474] Fils-Villetard, A., Guillou, A. and J. Segers, Projection estimators of Picklands dependence functions, 2008. TR08014.

[475] Geenens, G., Explicit formula for asymptotic higher moments of the Nadaraya-Watson estimator, 2008. TR08004.

[476] Geenens, G. and L. Simar, Nonparametric test for conditional independence in two-way contingency tables, 2008. TR08003.

[477] Geenens, G. and L. Simar, Single-index modelling of conditional probabilities in two-way contingency tables, 2008. TR08012.

[478] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Goodness-of-fit tests for censored regression based on artificial data points, 2008. TR08060.

[479] González-Manteiga, W., Pardo-Fernandez, J. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models, 2008. TR08045.

[480] 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.

[481] Jullion, A. and Ph. Lambert, Extensions of Bayesian P-splines models for fitting PK curves, 2008. TR08026.

[482] Jullion, A., Lambert, Ph. and F. Vandenhende, Estimation of receptor occupancy using varying coefficients models, 2008. TR08043.

[483] Linton, O., Mammen, E., Nielsen, J.P., and I. Van Keilegom, Nonparametric regression with filtered data, 2008. TR08041.

[484] Lopez, O., Patilea, V. and I. Van Keilegom, Single index regression models in the presence of censoring depending on the covariates, 2008. TR08039.

[485] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008. TR08023.

[486] Motta, G., Eichler, M. and R. von Sachs, Fitting dynamic factor models to nonstationary time series, 2008. TR08040.

[487] Neumeyer, N. and I. Van Keilegom, Estimating the error distribution in nonparametric multiple regression with applications to model testing, 2008. TR08031.

[488] Ojeda Cabrera, J.L. and I. Van Keilegom, Goodness-of-fit tests for parametric regression with selection based data, 2008. TR08011.

[489] Robert, Y.R., Segers, J. and C. Ferro, A sliding blocks estimator for the extremal index, 2008. TR08046.

[490] Russo, F., Wunsch, G. and M. Mouchart, Potential outcomes, counterfactuals, and structural modelling, 2008. TR08036.

[491] San Martin, E., Jara, A., Rolin, J.M. and M. Mouchart, On the analysis of Bayesian semiparametric IRT-type models, 2008. TR08029.

[492] Segers, J., Van Den Akker, R. and J.M. Werker, Improving upon the marginal empirical distribution functions when the copula is known, 2008. TR08030.

[493] Simar, L. and V. Zelenyuk, Stochastic FDH/DEA estimators for frontier analysis, 2008. TR08025.

[494] Teodorescu, B. and I. Van Keilegom, Goodness-of-fit test for generalized conditional linear models under left truncation and right censoring, 2008. TR08034.

[495] Van Keilegom, I., de Uña-Alvarez, J. and L. Meira-Machado, Nonparametric location-scale models for successive survival times under dependent censoring, 2008. TR08002.

[496] Van Keilegom, I. and L. Wang, Semiparametric modeling and estimation of the dispersion function in regression, 2008. TR08037.

C.1.2 Refereed publications (published)

[497] Beuthe, M., Bouffioux, C., Krier, C. and M. Mouchart, A comparison of conjoint, multi-criteria, conditional logic and neural network analyses for rank-ordered preference data. *Recent Developments in Transport Modelling: Lessons for the Freight Sector, Ben-Akiva, M., Meersman, H. and E. Van de Voorde (eds), Emerald Group Publishing Limited, UK*, 157-178, 2008. R08066.

[498] Böhm, H. and R. von Sachs, Structural shrinkage of nonparametric spectral estimators for multivariate time series. *Electronic Journal of Statistics*, **2**, 696-721, 2008. R08061.

[499] Crujeiras, R. and W. González-Manteiga, Comments on: Augmenting the bootstrap to analyze high dimensional genomic data. *TEST*, **17**, 40-42, 2008. R08075.

[500] 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.

[501] Einmahl, J.H.J., Krajina, A. and J. Segers, A method of moments estimator of tail dependence. *Bernoulli*, **14**, 1003-1026, 2008. R08123.

[502] Einmahl, J. and I. Van Keilegom, Tests for independence in nonparametric regression. *Statistica Sinica*, **18**, 601-616, 2008. R08115.

[503] Einmahl, J.H.J. and I. Van Keilegom, Specification tests in nonparametric regression. *Journal of Econometrics*, **143**, 88-102, 2008. R08001.

[504] El Ghouch, A. and I. Van Keilegom, Non-parametric regression with dependent censored data. *Scandinavian Journal of Statistics*, **35**, 228-247, 2008. R08008.

[505] Fils-Villetard, A., Guillou, A. and J. Segers, Projection estimators of Pickands dependence functions. *The Canadian Journal of Statistics*, **36**, 3, 369-382, 2008. R08119.

[506] Florens, J.P., Fougere, D. and M. Mouchart, Duration models and point processes. *The Econometrics of Panel Data, Màtyàs, L. and P. Sevestre (eds), Springer-Verlag Berlin Heidelberg*, 547-601, 2008. R08063.

[507] Fryzlewicz, P., Nason, G.P. and R. von Sachs, A wavelet-Fisz approach to spectrum estimation. *Journal of Time Series Analysis*, **29**, 5, 868-880, 2008. R08062.

[508] Gao, J., Gijbels, I. and S. Van Bellegem, Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics*, **143**, 123-142, 2008. R08005.

[509] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Parametric conditional mean and variance testing with censored data. *H. Skiadas, Christos, ed. Recent Advances in Applied Stochastic Models and Data Analysis. N.p.: World Scientific Publishing Co. Pte Ltd*, 2008. R08104.

[510] Govaerts, B., Dewé, W., Maumy, M. and B. Boulanger, Pre-study analytical method validation: comparison of four alternative approaches based on quality-level estimation and tolerance intervals. *Quality and Reliability Engineering International*, **24**, 667-680, 2008. R08130.

[511] Hafner, C., Temporal aggregation of multivariate GARCH models. *Journal of Econometrics*, **142**, 467-483, 2008. R08240.

[512] Heuchenne, C., Strong uniform consistency results of the weighted average of conditional artificial data points. *Journal of Statistical Planning and Inference*, **138**, 1496-1515, 2008. R08106.

[513] Kneip, A., Simar, L. and P. W. Wilson, Asymptotics and consistent bootstraps for DEA estimators in non-parametric frontier models. *Econometric Theory*, **24**, 1663-1697, 2008. R08092

[514] Lebrun, P., Govaerts, B., Debrus, B., Ceccato, A., Caliaro, G., Hubert, P. and B. Boulanger, Development of a new predictive modelling technique to find with confidence equivalence zone and design space of chromatographic analytical methods. *Chemometrics and Intelligent Laboratory Systems*, **91**, 4-16, 2008. R08118.

[515] Li, S.K., Simar, L., Wilson, P.W. and V. Zelenyuk, Introduction for the special issue on transitionning economics. *Journal of Productivity Analysis*, **29**, 77, 2008. R08097.

[516] Linton, O., Sperlich, S. and I. Van Keilegom, Estimation of a semiparametric transformation model. *Annals of Statistics*, **36**, 2, 686-718, 2008. R08012.

[517] Ombao, H. and S. Van Bellegem, Evolutionary coherence of nonstationary signals. *IEEE Transactions on Signal Processing*, **56**, 6, 2259-2266, 2008. R08054.

[518] Park, B., Simar, L. and V. Zelenyuk, Local likelihood estimation of truncated regression and its partial derivatives: theory and application. *Journal of Econometrics*, **146**, 185-198, 2008. R08096.

[519] Robert, C.Y. and J. Segers, Tails of random sums of a heavy-tailed number of light-tailed terms. *Insurance: Mathematics and Economics*, **43**, 85-92, 2008. R08125.

[520] Rousseau, R., Govaerts, B., Verleysen, M. and B. Boulanger, Comparison of some chemometric tools for metabonomics biomarker identification. *Chemometrics and Intelligent Laboratory Systems*, **91**, 54-66, 2008. R08126.

[521] Simar, L. and P.W. Wilson, Statistical inference in nonparametric frontier models: recent developments and perspectives. *The Measurements of Productive Efficiency*, 2nd Edition, Harold Fried, C. A. Knox Lovell and Shelton Schmidt, editors, Oxford University Press, 2008. R08002.

[522] Sougne, D., Heuchenne, C. and G. Hubner, The determinants of CDS prices: an industry-based investigation. *Wagner, Niklas, ed. Credit Risk: Models, Derivatives and Management. Empirical Studies and Analysis. Financial Mathematics Series, Volume 6. Boca Raton, London, New York: Chapman and Hall / CRC*, 85-96, 2008. R08238.

[523] Van Bellegem, S. and R. von Sachs, Locally adaptive estimation of sparse evolutionary wavelet spectra. *The Annals of Statistics*, **36**, 4, 1879-1924, 2008. R08060.

[524] Van Keilegom, I., González-Manteiga, W. and C. Sánchez Sellero, Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401-415, 2008. R08015.

[525] Van Keilegom, I., Sánchez Sellero, C. and W. González-Manteiga, Empirical likelihood based testing for regression. *Electronic Journal of Statistics*, **2**, 581-604, 2008. R08117.

[526] Wang, L., Akritas, M.G. and I. Van Keilegom, An ANOVA-type nonparametric diagnostic test for heteroscedastic regression models. *Journal of Nonparametric Statistics*, **20**, 5, 365-382, 2008. R08064.

C.1.3 Refereed publications (in press)

[527] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of spatially and temporally heterogeneous brain response images. *Journal of Computational and Graphical Statistics*, 2008. RP08001.

[528] Badin, L. and L. Simar, A bias corrected nonparametric envelopment estimator of frontiers. *Econometric Theory*, 2009. RP08148.

[529] Basrak, B. and J. Segers, Regularly varying multivariate time series. *Stochastic processes and their Applications*, 2009. RP08139.

[530] Beirlant, J., Joossens, E. and J. Segers, Second order refined Peaks-over-Threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 2009. RP08039.

[531] Böhm, H. and R. von Sachs, Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, 2009. RP08140.

[532] Braekers, R. and I. Van Keilegom, Flexible modeling base on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, 2008. RP08090.

[533] Charpentier, A. and J. Segers, Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis*, 2009. RP08141.

[534] Crujeiras, R., Fernández-Casal, R. and González-Manteiga, An L2-test for comparing spatial spectral densities. *Statistics and Probability Letters*, RP08011.

[535] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, Goodness-of-fit tests for the spatial spectral density. *Stochastic Environmental Research and Risk Assesment*, RP08015.

[536] Einmahl, J.H.J. and J. Segers, Maximum empirical likelihood estimation of the spectral measure of an extreme value distribution. *Annals of Statistics*, 2009. RP08142.

[537] El Ghouch, A. and I. Van Keilegom, Local linear quantile regression with dependent censored data. *Statistica Sinica*, 2009. RP08135.

[538] El Ghouch, A., Van Keilegom, I. and I. McKeague, Empirical likelihood confidence intervals for dependent duration data. *Econometric Theory*, 2009. RP08131.

[539] Genest, C. and J. Segers, Rank-based inference for bivariate extreme-value copulas. *Annals of Statistics*, 2009. RP08143.

[540] Hafner, C. and H. Herwartz, Analytical quasi maximum likelihood inference in multivariate volatility models. *Metrika*, 2009. RP08144.

[541] Hafner, C. and H. Herwartz, Testing causality in variance using multivariate GARCH models. *Annales d'Economie et de Statistique*, 2009. RP08145.

[542] Hafner, C. and A. Preminger, Asymptotic theory for a factor GARCH model. *Econometric Theory*, 2009. RP08146.

[543] Hall, P. and I. Van Keilegom, Nonparametric "regression" when errors are centred at endpoints. *Bernoulli*, 2009. RP08137.

[544] Heuchenne, C. and I. Van Keilegom, Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics*, 2009. RP08133.

[545] Hjort, N.L., McKeague, I.W. and I. Van Keilegom, Extending the scope of empirical likelihood. *Annals of Statistics*, 2009. RP08132.

[546] Jeong, S.O., Park, B.U. and L. Simar, Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, 2009. RP08147.

[547] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, 2008. RP08007.

[548] Legrand, C., Duchateau, L. Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis*, 2008/2009. RP08100.

[549] Molanes, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, 2008. RP08068.

[550] Neumeyer, N. and I. Van Keilegom, Change-point tests for the error distribution in nonparametric regression. *Scandinavian Journal of Statistics*, 2009. RP08136.

[551] Ojeda-Cabrera, J. and I. Van Keilegom, Goodness-of-fit tests for parametric regression with selection biased data. *Journal of Statistical Planning and Inference*, 2009. RP08138.

[552] Porcu, E., Crujeiras, R., Mateu, J. and W. González-Manteiga, On the second order properties of the multidimensional peridogram for regularly spaced data. *Theory of Probability and its Applications*, 2008. RP08008.

[553] Simar, L. and P.W. Wilson, Estimation and inference in cross-sectional stochastic frontier models. *Econometric Review*, 2009. RP08149.

[554] Teodorescu, B., Van Keilegom, I. and R. Cao, Generalized conditional linear models under left truncation and right censoring. *Annals of the Institute of Statistical Mathematics*, 2009. RP08134.

C.1.4 Books (published)

[555] Florens, J.P., Mouchart, M. and J.M. Rolin, *Elements of Bayesian Statistics.* Marcel Dekker, New York, 2008. B08003.

[556] Franke, J., Härdle, W.K. and C.M. Hafner, *Statistics of Financial Markets, An Introduction.* Springer, New York, 2008. B08004.

C.2 Katholieke Universiteit Leuven, KUL-1

C.2.1 Technical reports

[557] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties, 2008. TR08050.

[558] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, 2008. TR08035.

[559] Braeken, J., Tuerlinckx, F. and P. De Boeck, A latent variable framework for teratology studies, 2008. TR08061.

[560] Braeken, J., Tuerlinckx, F. and P. De Boeck, Fitting explanatory item response theory models: a Matlab IRTm toolbox. 2008. TR08062.

[561] De Boeck, P., Cho, S.-J. and M. Wilson, Secondary dimension modeling of DIF: A mixture approach to test speededness, 2008. TR08063.

[562] Debruyne, M., Hubert, M. and J. Van Horebeek, Detecting influential observations in kernel PCA, 2008. TR08001.

[563] Depril, D. and I. Van Mechelen, Low dimensional overlapping clustering, 2008. TR08064.

[564] Depril, D. and I. Van Mechelen, New ALS approaches for fitting the INDCLUS, 2008. TR08065.

[565] Desmet, L. and I. Gijbels, Improved spectral density estimation at peaks, 2008. TR08052.

[566] Desmet, L. and I. Gijbels, Local linear fitting and improved estimation near peaks, 2008. TR08051.

[567] Dutilh, G., Vandekerckhove, J., Tuerlinckx, F. and E.J. Wagenmakers, A diffusion model decomposition of the practice effect, 2008. TR08066.

[568] Frederickx, S., Tuerlinckx, T., De Boeck, P. and D. Magis, An item mixture model to detect differential item functioning, 2008. TR08067.

[569] Frederickx, S., Kuppens, P., Tuerlinckx, F. and I. Van Mechelen, Desequens: an R-package for the variance decomposition of sequential processes, 2008. TR08068.

[570] Gijbels, I., Prosdocimi, I. and G. Claeskens, Nonparametric estimation of mean and dispersion functions inextended generalized linear models, 2008. TR08053.

[571] Gijbels, I. and A. Verhasselt, P-splines regression and difference type of penalty, 2008. TR08054.

[572] Gijbels, I. and A. Verhasselt, Regularization and P-splines in generalized linear models, 2008. TR08055.

[573] González, J., Tuerlinckx, F., and P. De Boeck, Crossed random effects models for the analysis of complex designs, 2008. TR08019.

[574] Magis, D. and G. Raîche, An interactive maximum a posteriori estimation of proficiency level to detect multiple local likelihood maxima, 2008. TR08069.

[575] Oravecz, Z., Tuerlinckx, F. and J. Vandekerckhove, A hierarchical latent stochastic differential equation model for affective dynamics, 2008. TR08070.

[576] Schepers, J. and I. Van Mechelen, Capturing the nature of two-way interactions by means of a two-mode clustering method, 2008. TR08021.

[577] Schepers, J., Van Mechelen, I. and E. Ceulemans, The real-valued model of hierarchical classes, 2008.TR08072.

[578] Schepers, J. and I. Van Mechelen, Uniqueness of real-valued hierarchical classes models, 2008. TR08071.

[579] Timmerman, M.E., Ceulemans, E. and H.A.L. Kiers, Factorial K-means analysis reconsidered, 2008. TR08073.

[580] Vandekerckhove, J., Tuerlinckx, F. and M.D. Lee, Hierarchical diffusion models for two-choice response times, 2008. TR08074.

[581] Vandekerckhove, J., Verheyen, S. and F. Tuerlinckx, A crossed random effects diffusion model for speeded semantic categorization decisions, 2008. TR08075.

[582] Van Deun, K., Smilde, A., van der Werf, M.J., Kiers, H.A.L. and I. Van Mechelen, A general framework for simultaneous component methods, 2008. TR08076.

[583] Van Nijlen, D. and R. Janssen, Modeling quantitative and qualitative differences in spelling ability, 2008. TR08077.

[584] Veraverbeke, N., Omelka, M. and I. Gijbels, Weak convergence of the empirical copula process with covariates, 2008. TR08056.

C.2.2 Refereed publications (published)

[585] Albrecher, H., Mayer, Ph. and W. Schoutens, General lower bounds for arithmetic Asian option prices. *Applied Mathematical Finance*, **15**, 123-149, 2008. R08116.

[586] Antonio, K. and J. Beirlant, Issues in claims reserving and credibility: A semiparametric approach with mixed models. *Journal of Risk and Insurance*, **75**, 643-676, 2008. R08010.

[587] Beirlant J., Berlinet A. and G. Biau, Higher order estimation at Lebesgue points. *The Annals of the Institute of Statistical Mathematics*, **60**, 651-677, 2008. R08009.

[588] Beirlant, J., de Waal, D.J. and G. Dierckx, Predicting high quantiles through the Dirichlet process on extreme modelling. *South African Statistical Journal*, **42**, 101-124, 2008. R08120.

[589] Beirlant J., Figueiredo F., Gomes M.I. and B. Vandewalle, Improved reduced-bias tail index and quantile estimators. *Journal of Statistical Planning and Inference*, **138**, 1851-1870, 2008. R08011.

[590] Brys, G., Hubert, M. and A. Struyf, Goodness-of-fit tests based on a robust measure of skewness. *Computational Statistics*, **23**, 429-442, 2008. R08121.

[591] Carr, P. and W. Schoutens, Hedging under the Heston model with jump-to-default. *International Journal of Theoretical and Applied Finance*, **11**, 403-414, 2008. R08122.

[592] Ceulemans, E. and I. Van Mechelen, CLASSI: a classification model for the study of sequential processes and individual differences therein. *Psychometrika*, **73**, 107-124, 2008. R08185.

[593] Debruyne, M., Hubert, M., Portnoy, S. and K. Vanden Branden, Censored depth quantiles. *Computational Statistics and Data Analysis*, **52**, 1604-1614. 2008, R08016.

[594] Debruyne, M., Hubert, M. and J. Suykens, Model selection for kernel regression using the influence function. *Journal of Machine Learning Research*, **9**, 2377-2400, 2008. R08124.

[595] del Pino, G., San Martín, E., González, J. and P. De Boeck, On the relationships between sum score based estimation and joint maximum likelihood estimation. *Psychometrika*, **73**, 145-151, 2008. R08186.

[596] Depril, D., Van Mechelen, I. and B.G. Mirkin, Algorithms for additive clustering of rectangular data tables. *Computational Statistics and Data Analysis*, **52**, 4923-4938, 2008. R08187.

[597] 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.

[598] Garcia, J., Goossens, S. and W. Schoutens, Let's jump together: pricing credit derivatives. *Risk Magazine*, 2008. R08177.

[599] Gijbels, I., Smoothing and preservation of irregularities using local linear fitting. *Applications of Mathematics*, **53**, 177-194, 2008. R08017.

[600] Goegebeur, Y., De Boeck, P., Wollack, J.A. and A.S. Cohen, A speeded item response model with gradual process change. *Psychometrika*, **73**, 65-87, 2008. R08188.

[601] González, J., De Boeck, P. and F. Tuerlinckx, A double-structure structural equation model for three-mode data. *Psychological Methods*, **13**, 337-353, 2008. R08074.

[602] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, **23**, 92-119, 2008. R08203.

[603] Hubert, M. and S. Van der Veeken, Outlier detection for skewed data, *Journal of Chemometrics*, **22**, 235-246, 2008. R08128.

[604] Hubert, M. and E. Vandervieren, An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, **52** , 5186-5201, 2008. R08129.

[605] Jácome, A., Gijbels, I. and R. Cao, Comparison of the Nadaraya-Watson and local linear methods in presmoothed density estimation under censoring. *Computational Statistics*, **23**, 381-406, 2008. R08018.

[606] Jansen, M. Nason, G. and B. Silverman, Multiscale methods for data on graphs and irregular multidimensional situations. *Journal of the Royal Statistical Society, series B*, **71**, 97-125, 2009. R08131.

[607] Jansen, M., Refinement independent wavelets for use in adaptive multiresolution schemes. *International Journal of Wavelets, Multiresolution and Information Processing (IJWMIP)*, **6**, 521-539, 2008. R08132.

[608] Jansen, M., Stable edge-adaptive multiscale decompositions using updated normal offsets. *IEEE Transactions on Signal Processing*, **56**, 2718-2727, 2008. R08133.

[609] Kuppens, P., Van Mechelen, I. and F. Rijmen, Towards disentangling sources of individual differences in appraisal and anger. *Journal of Personality*, **76**, 969-1000, 2008. R08190.

[610] Leenen, I., Van Mechelen, I., Gelman, A. and S. De Knop, Bayesian hierarchical classes analysis. *Psychometrika*, **73**, 39-64, 2008. R08191.

[611] Leoni, P. and W. Schoutens, Multivariate smiling. *Wilmott Magazine*, 2008. R08182.

[612] Madan, D.B. and W. Schoutens, Break on through to the single side. *Journal of Credit Risk*, 2008. R08181.

[613] Rijmen, F., Vansteelandt, K. and P. De Boeck, Latent class models for diary method data: parameter estimation by local computations. *Psychometrika*, **73**, 167-182, 2008. R08192.

[614] Rijmen, F. and J. Vomlel, Assessing the performance of variational methods for mixed logistic regression models. *Journal of Statistical Computation and Simulation*, **78**, 765-779, 2008. R08193.

[615] Rouder, J., Tuerlinckx, F., Speckman, P., Lu, J. and P. Gomez, A hierarchical approach for fitting curves to response time measurements. *Psychonomic Bulletin and Review*, **15**, 1201-1208, 2008. R08194.

[616] Schepers, J., Ceulemans, E. and I. Van Mechelen, Selecting among multi-mode partitioning models of different complexities: A comparison of four model selection criteria. *Journal of Classification*, **25**, 67-85, 2008. R08195.

[617] Vandekerckhove, J. and F. Tuerlinckx, Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, **40**, 61-72, 2008. R08196.

[618] Vandekerckhove, J., Tuerlinckx, F. and M.D. Lee, A Bayesian approach to diffusion process models of decision-making. *V. Sloutsky, B. Love and K. McRae (Eds.), Proceedings of the 30th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society*, 1429-1434, 2008. R08197.

[619] Van Nijlen, D. and R. Janssen, Modeling judgments in the Angoff and contrasting-groups method of standard setting. *Journal of Educational Measurement*, **45**, 45-63, 2008. R08198.

[620] Wilderjans, T., Ceulemans, E. and I. Van Mechelen, The chic model: a global model for coupled binary data. *Psychometrika*, **73**, 729-751, 2008. R08199.

[621] Willems, P., Beirlant, J. and A. Guillou, Bias correction in hydrologic GPD based extreme value analysis by means of a slowly varying function. *Journal of Hydrology*, **338**, 221-236, 2008. R08134.

C.2.3 Refereed publications (in press)

[622] Antoniadis, A., Gijbels, I. and J.M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, 2008. RP08038.

[623] Beirlant, J., Dierckx, G., Onghena, P. and P. Luwel, Analysis of extreme values in education. *The International Encyclopedia of Education. 3rd Edition*, 2009. BP08002.

[624] Beirlant, J., Joossens, E. and J. Segers, Second order refined Peaks-over-Threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 2009. RP08039.

[625] Cariboni, J. and W. Schoutens, Jumps in intensity models: investigating the performance of Ornstein-Uhlenbeck processes. *Metrika*, 2008. RP08040.

[626] Ceulemans, E. and H.A.L. Kiers, Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical and Statistical Psychology*. RP08109.

[627] Debruyne, M. and M. Hubert, The influence function of the Stahel-Donoho covariance estimator of smallest outlyingness. *Statistics and Probability Letters*, 2008. RP08041.

[628] Debruyne, M., Hubert, M. and J. Suykens, Model selection for kernel regression using the influence function, chapter in book *Hand-on Pattern Recognition edited by Isabelle Guyon, Gavin Cawley, Gideon Dror and Amir Saffari*, 2008. BP08003.

[629] Dierckx, G., Beirlant, J., De Waal, J. D. and A. Guillou, A new estimation method for Weibull-type tails based on the mean excess function. *Journal of Statistical Planning and Inference*, 2009. RP08042.

[630] Gao, J. and I. Gijbels, Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, **484**, 2008. RP08043.

[631] Gijbels, I. and M. Hubert, Robust and nonparametric statistical methods, *Encyclopedia Comprehensive Chemometrics, edited by B. Walczak, R. Tauler Ferré, S. Brown, Elsevier*, 2008. BP08004.

[632] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology*. RP08112.

[633] Hubert, M., Linear regression modeling: robust calibration, *Encyclopedia Comprehensive Chemometrics, edited by B. Walczak, R. Tauler Ferré, S. Brown, Elsevier*, 2008. BP08005.

[634] Hubert, M., Rousseeuw, P.J. and T. Verdonck, Robust PCA for skewed data and its outlier map. *Computational Statistics and Data Analysis*, 2008. RP08044.

[635] Ip, E.H., Smits, D.J.M. and P. De Boeck, Locally dependent linear logistic test model with person covariates. *Applied Psychological Measurement*. RP08113.

[636] Janssen, R., Modeling the effect of item designs within the Rasch model. *New directions in psychological measurement with model-based approaches. S. Embretson and J. Roberts (Eds.), American Psychological Association. Washington, DC*. BP08008.

[637] Jönsson, H. and W. Schoutens, Single name credit default swaptions meet single sided jump models. *Review of Derivative Research*, 2008. RP08045.

[638] Jönsson, H., Schoutens, W. and G. Van Damme, Modeling default and prepayment using Lévy processes: an application to asset backed securities. *Radon Series Comp. Appl. Math*, **8**, 2009. RP08046.

[639] Joossens, E. and W. Schoutens, Portfolio insurances, CPPI and CPDO, truth or illusion? *Alternative Investments and Strategies, World Scientific Publishing. Kiesel. R. And Scherer, M. (eds)*, 2009. BP08006.

[640] Leenen, I. and E. Ceulemans, Three-way hierarchical classes: a comparison of INDCLAS and Tucker3-HICLAS models. *Applied Multivariate Research*. RP08115.

[641] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, 2008. RP08047.

[642] Oravecz, Z., Tuerlinckx, F. and J. Vandekerckhove, A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*. RP08116.

[643] Schepers, J. and J. Hofmans, TwoMP: A MATLAB graphical user interface for two-mode partitioning. *Behavior Research Methods*. RP08117.

[644] Schoutens, W., The world of VG. *Proceedings of the Barcelona Financial Engineering Summer School 2008. Progress in Mathematics. Birkenhauser*, 2008. RP08048.

[645] Timmerman, M.E., Ceulemans, E., Lichtwarck-Aschoff, A. and K. Vansteelandt, Multilevel component analysis for studying intra-individual variability and inter-individual differences. *Dynamic Process Methodology in the Social and Developmental Sciences. J. Valsiner, P.C.M. Molenaar, M.C. D.P. Lyra and N. Chaudhary (Eds.), New York: Springer.* BP08009.

[646] 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 and Statistical Psychology*. RP08118.

[647] Valdivieso, L., Schoutens, W. and F. Tuerlinckx, Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type. *Statistical Inference for Stochastic Processes*, 2008. RP08050.

[648] Van Aerschot, W., Jansen, M. and A. Bulteel, Normal mesh based geometrical image compression. *Image and Vision Computing*, 2008. RP08049.

[649] Vandekerckhove, J. and F. Tuerlinckx, MATLAB for behavioral scientists: a novice's guide to MATLAB. *Experimental Psychology.* RP08119.

[650] Van Duffel, S., Chernih, A., Maj M. and W. Schoutens, A note on the suboptimality of path-dependent pay-offs in Lévy markets. *Applied Mathematical Finance*, 2008.RP08051.

[651] Wetzels, R., Vandekerckhove, J., Tuerlinckx, F. and E.J. Wagenmakers, Bayesian parameter estimation in the expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology*. RP08120.

[652] Wilderjans, T., Ceulemans, E. and I. Van Mechelen, Simultaneous analysis of coupled data blocks differing in size: a comparison of two weighting schemes. *Computational Statistics and Data Analysis*. RP08121.

C.2.4 Non-refereed publications (published)

[653] Hubert, M., Book review of "Data Depth: robust multivariate analysis, computational geometry and applications" by R. Liu, R. Serfling and D.L. Souvaine. *Biometrics*, **64**, 655-656. 2008. NR08001.

C.2.5 Non-refereed publications (in press)

[654] Hubert, M., Book review of "Applied multiway data analysis", by P. Kroonenberg, Wiley, New Jersey. *Journal of the American Statistical Association*, June issue. 2009. NRP08008.

C.3 Katholieke Universiteit Leuven, KUL-2

C.3.1 Technical reports

[655] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G. and M.G. Kenward, A sensitivity analysis for shared-parameter models, 2008. TR08058.

[656] Jara, A., Lesaffre, E., De Lorio, M. and F. A. Quintana, Bayesian semiparametric inference for multivariate doubly-interval-censored data, 2008. TR08009.

[657] Jara, A., Quintana, F.A. and E. San Martin, Linear effects mixed models with skew-elliptical distributions : a Bayesian approach, 2008. TR08010.

[658] Rizopoulos, D., Verbeke, G. and E. Lesaffre, Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data, 2008. TR08005.

[659] San Martin, E., Jara, A., Rolin, J.M. and M. Mouchart, On the analysis of Bayesian semiparametric IRT-type models, 2008. TR08029.

C.3.2 Refereed publications (published)

[660] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, **64**, 96-105, 2008. R08102.

[661] Bogaerts, K. and E. Lesaffre, Estimating local and global measures of association for bivariate interval censored data with a smooth estimate of the density. *Statistics in Medicine*, **27**, 5941-5955, 2008. R08089.

[662] Bogaerts, K. and E. Lesaffre, Modeling the association of bivariate interval-censored data using the copula approach. *Statistics in Medicine*, **27**, 6379-6392,2008. R08087.

[663] Borgermans, L., Goderis, G., Van Den Broeke, C., Mathieu, C., Aertgeerts, B., Verbeke, G., Carbonez, A., Ivanova, A., Grol, L. and J. Heyrman, A cluster randomized trial to improve adherence to evidence-based guidelines on diabetes and reduce clinical inertia in primary care physicians in Belgium: study protocol (NTR 1369). *Implementation Science*, **3**, 42, 2008. R08105.

[664] De Backer, M., Morren, M., Boonen, H., Boyden, B., Vertruyen, A., Lecomte, P., Paquay, C., Lesaffre, E. and J. Lambert, Belgian observational drug utilization study of pimecrolimus cream 1% in routine daily practice in atopic dermatitis. *Dermatology*, **217** , 156-163, 2008. R08109.

[665] Declerck, D., Leroy, R., Martens, L., Lesaffre, E., Garcia Zattera, M., Vanden Broucke, S., Debyser, M., and K. Hoppenbrouwers, Factors associated with prevalence and severity of caries experience in preschool children. *Community Dentistry and Oral Epidemiology*, **36**, 168-178, 2008. R08112.

[666] Defranc, A., Van den Broucke, S., Leroy, R., Hoppenbrouwers, K., Lesaffre, E., Martens, L., Debyser, M. and D. Declerck, Measuring oral health behaviour in Flemish health care workers: an application of the theory of planned behaviour. *Community Dental Health*, **25**, 107-114, 2008). R08091.

[667] De Wit, L., Putman, K., Baert, I., Lincoln, N., Angst, F., Beyens, H., Bogaerts, K., Brinkmann, N., Connell, L., Dejaeger, E., De Weerdt, W., Jenni, W., Kaske, C., Komárek, A., Lesaffre, E., Leys, M., Louckx, F., Schuback, B., Schupp, W., Smith, B. and H. Feys, Anxiety and depression in the first six months after stroke. A longitudinal multicentre study. *Disability and Rehabilitation*, **30**, 1858-1866, 2008. R08036.

[668] Dubois, C., Dens, J., Sinnaeve, P., Belmans, A., Van Cleemput, J., Mendez, M., Piessens, J., and W. Desmet, Results of percutaneous coronary interventions of the unprotected left main coronary artery in 143 patients and comparison of 30-day mortality to results of coronary bypass grafting. *American Journal of Cardiology*, **101**, 75-81, 2008. R08101.

[669] Fieuws, S., Verbeke, G., Maes, B. and Y. Vanrenterghem, Predicting renal graft failure using multivariate longitudinal profiles. *Biostatistics*, **9**, 419-431, 2008. R08111.

[670] Ghidey, W., Lesaffre, E.and G. Verbeke, A comparison of methods for estimating the random effects distribution of a linear mixedmodel. *Statistics in Medicine*, 2008. R08048.

[671] Hughes, S., Keene, O., Roes, K., Ashby, D., Beltangadu, M., Bird, S.M., Burton, C., Buyse, M., Day, S., van Ewijk, P., Fletcher, C., Grieve, A.P., Guinot, C., Hemmings, R., Hold, D., Hothorn, L., Jagers, P., Keiding, N., Korhonen, P., Lesaffre, E., Lewis, J.A., Molenberghs, G., Neumann, N., Olsen, K.J., van Osta, G., Pocock, S., Rockhold, F., Rossi, A., Senn, S., Stijnen, T., Vaeth, M., Wiklund, S.J. and J. Wolfram, Open Letter to the Heads of the European Regulatory Agencies. *British Medical Journal*, **336**, 250-250, 2008. R08145.

[672] Jara Vallejos, A., Quintana, F. and E. San Martin, Linear mixed models with skew-elliptical distributions: a Bayesian approach. *Computational Statistics and Data Analysis*, **52**, 5033-5045, 2008. R08237.

[673] 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.

[674] Lemiengre J., Dierckx de Casterlé B., Verbeke G., Van Craen K., Schotsmans P., and C. Gastmans, Ethics policies on euthanasia in nursing homes: a survey in Flanders, Belgium, *Social Science and Medicine*, **66**, 376-386, 2008. R08114.

[675] Leroy, R., Cecere, S., Lesaffre, E. and D. Declerck, Variability in permanent tooth emergence sequences in Flemish children. *European Journal of Oral Sciences*, **116**, 11-17, 2008. R08098.

[676] Leroy, R., Hoppenbrouwers, K., Jara Vallejos, A. and D. Declerck, Parental smoking behavior and caries experience in pre-school children. *Epidemiology*, **36**, 249-257, 2008. R08103.

[677] Lesaffre, E., Superiority, equivalence, and non-inferiority trials. *Bulletin of the Hospital for Joint Diseases Orthopaedic Institute*, **66**, 150-154, 2008. R08095.

[678] Lesaffre, E., Use and misuse of the p-value. *Bulletin of the Hospital for Joint Diseases Orthopaedic Institute*, **66**, 146-149, 2008. R08108.

[679] Lesaffre, E., Diya, L., Van den Heede, K., Sermeus, W. and A. Vleugels, Quality of care for cardiac surgery patients in Belgian acute hospitals: An illustration of some multilevel modelling issues. *Revue d'Epidemiologie et de Santé Publique*, **56**, S5-S6, 2008. R08107.

[680] Menten, J., Boelaert, M. and E. Lesaffre, Bayesian latent class models with conditionally dependent diagnostic tests: a case study. *Statistics in Medicine*, **27**, 4469-4488, 2008. R08041.

[681] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensitivity analysis for incomplete data. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods. Chapter 22. Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, eds. Chapman & Hall/CRC*, 501-552, 2008. R08242.

[682] 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*, **17**, 123-139, 2008. R08033.

[683] Rizopoulos, D., Verbeke, G., Lesaffre, E. and Y. Vanrenterghem, A two-part joint model for the analysis of survival and longitudinal binary data with excess zeros. *Biometrics*, **64**, 611-619, 2008. R08035.

[684] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63-74. R08019.

[685] Schacht, A., Bogaerts, K., Bluhmki, E. and E. Lesaffre, A new nonparametric approach for baseline covariate adjustment for two-group comparative studies. *Biometrics*, 2008. R08038.

[686] Sermeus, W., Delesie, L., Van den Heede, K., Diya, L., and E. Lesaffre, Measuring the intensity of nursing care: making use of the Belgian nursing minimum data set. *International Journal of Nursing Studies*, **45**, 1011-1021, 2008. R08090.

[687] Topal, B., Fieuws, S., Aerts, R., Vandeweyer, H. and F. Penninckx, Laparoscopic versus open liver resection of hepatic neoplasms: comparative analysis of short-term results. *Surgical Endoscopy: Ultrasound and Interventional Techniques*, **22**, 2208-2213, 2008. R08093.

[688] Tsonaka, S., Verbeke, G. and E. Lesaffre, A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, 2008. R08040.

[689] Van den Heede, K., Diya, L., Lesaffre, E., Vleugels, A. and W. Sermeus, Benchmarking nurse staffing levels: the development of a nationwide feedback tool. *Journal of Advanced Nursing*, **63**, 607-18, 2008. R08088.

[690] van Heumen, C., Kreulen, C., Bronkhorst, E., Lesaffre, E. and N. Creugers, Fiber-reinforced dental composites in beam testing. *Dental Materials*, **24**, 1435-1443, 2008. R08094.

[691] Van Landuyt, K., Peumans, M., Fieuws, S., De Munck, J., Vivan Cardoso, M., Ermis, R., Lambrechts, P., and B. Van Meerbeek, A randomized controlled clinical trial of a HEMA-free all-in-one adhesive in non-carious cervical lesions at 1 year. *Journal of Dentistry*, **36**, 847-55, 2008. R08086.

[692] Van de Voorde, P., Sabbe, M., Calle, P., Lesaffre, E., Rizopoulos, D., Tsonaka, S., Christiaens, D., Vantomme, A., De Jaeger, A.and D. Matthys, Paediatric trauma and trauma care in Flanders (Belgium). Methodology and first descriptive results of the PENTA registry. *European Journal of Pediatrics*, **167**, 1239-1249,2008. R08100.

[693] Van de Voorde, P., Sabbe, M., Rizopoulos, D., Tsonaka, S., De Jaeger, A., Lesaffre, E. and M. Peters, Assessing the level of consciousness in children: A plea for the Glasgow Coma Motor subscore. *Resuscitation*, **76**, 175-179, 2008. R08110.

[694] Verbeke, G., Molenberghs, G. and C. Beunckens, Formal and informal model selection with incomplete data. *Statistical Science*, **23**, 201-218, 2008. R08099.

C.3.3 Refereed publications (in press)

[695] Beunckens C., Sotto C., Molenberghs G. and G. Verbeke, An integrated sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08030.

[696] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multi-faceted sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08081.

[697] Ghidey W., Lesaffre E. and G. Verbeke, A comparison of methods for estimating the random effects distribution of a linear mixed model. *Statistical Methods in Medical Research*, 2008. RP08029.

[698] Lehnert S., Van Loo P., Thilakarathne P.J., Marynen P., Verbeke G. and F.C. Schuit, Evidence for co-evolution between human miRNAs and Alu-repeats. *Public Library of Science One (PLoS One)*, 2009. RP08036.

[699] Molas, M. and E. Lesaffre, A comparison of three random effects approaches to analyze repeated bounded outcome scores with an application in a stroke revalidation study. *Statistics in Medicine*, 2008. RP08023.

[700] Rizopoulos D., Verbeke G. and E. Lesaffre, Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society, Series B*, 2009. RP08032.

[701] Sotto C., Beunckens C., Molenberghs G., Jansen I., and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, 2008, RP08031.

[702] Van den Heede, K., Sermeus, W., Diya, L., Clarke, S., Lesaffre, E., Vleugels, A. and L. Aiken, Nurse staffing and patient outcomes in Belgian acute hospitals: Cross-sectional analysis of administrative data. *International Journal of Nursing Studies*, 2008. RP08024.

[703] Van Landuyt, K., Fieuws, S., Peumans, M., De Munck, J., Vivan Cardoso, M., Ermis, B., Lambrechts, P. and B. Van Meerbeek, RCT of a HEMA-free all-in-one adhesive in non-carious cervical lesions. *Journal of Dental Research*, **87**, 2008. RP08025.

[704] Van Nieuwenhuyse A., Crombez G., Burdorf A., Verbeke G., Masschelein R., Moens G. and P. Mairiaux. *Physical characteristics of the back are not predictive of low back pain in healthy workers: A prospective study, BioMed Central: Musculoskeletal Disorders*, 2009. RP08034.

[705] Vanneste E., Voet T., Le Caignec C., Ampe M., Konings P., Melotte C., Debrock S., Viikkula M., Schuit F., Fryns J.P., Verbeke G., Dhooghe T., Moreau Y. and J. Vermeesch, Human cleavage stage embryos are chromosomally unstable. *Nature Medicine*, 2009. RP08035.

[706] Verbeke G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling*, 2009. RP08033.

C.3.4 Non-refereed publications (published)

[707] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Shared-parameter models and missingness at random. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18*, 2008. NR08023.

[708] Molenberghs, G., Aerts, M., Beunckens, C., Creemers, A., Hens, N., Kenward, M.G., Sotto, C. and G. Verbeke, Missing not at random models and their missing at random counterparts, in various modelling frameworks. *ACTAS do XVI Congresso da SPE, UTAD, Vila REAL*, 2008. NR08026.

C.3.5 Books (in press)

[709] Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, *Longitudinal Data Analysis, A Handbook of Modern Statistical Methods*. Chapman and Hall/CRC, New York, 2009. BP08011.

C.4 Universiteit Gent, UG

C.4.1 Technical reports

[710] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015.

C.4.2 Refereed publications (published)

[711] Agulló, J., Croux, C. and S. Van Aelst, The multivariate least trimmed squares estimator. *Journal of Multivariate Analysis*, **99**, 311-338, 2008. R08058.

[712] Baele, G., van de Peer, Y. and S. Vansteelandt, A model-based approach to study nearest neighbour influences reveals complex substitution patterns in non-coding sequences. *Systematic Biology*, **57**, 675-692, 2008. R08204.

[713] Best, D.J., Rayner, J.C.W. and O. Thas, Comparison of some tests of fit for the Laplace distribution. *Computational Statistics and Data Analysis*, **52**, 5338-5343, 2008. R08218.

[714] Best, D.J., Rayner, J.C.W. and O. Thas, Tests of fit for the logarithmic distribution. *Journal of Applied Mathematics and Decision Sciences*, 8, 2008. R08215.

[715] Best, D.J., Rayner, J.C.W. and O. Thas, X2 and its components as tests of normality for grouped data. *Journal of Applied Statistics*, **35**, 481-492, 2008. R08216.

[716] Bogaert, V., Taes, Y., Konings, P., Van Steen, K., De Bacquer, D., Goemaere, S., Zmierczak, H., Crabbe, P. and J.M. Kaufman, *Clinical Endocrinology*, **69**, 129-135, 2008. R08209.

[717] 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*, **35**, 497-514, 2008. R08221.

[718] Carvalho, B., Postma, C., Mongera, S., Hopmans, E., Diskin, S., van de Wiel, M., Van Criekinge, W., Thas, O., Mathhai, A., Cuesta, M., Terhaar, J., Craanen, M., Schrock, E., Ylstra, B. and G. Meijer, Integration of DNA and expression microarray data unravels seven putative oncogenes on 20Q amplicon involved in colorectal adenoma to carcinoma progression. *Cellular Oncology*, **30**, 145-146, 2008. R08217.

[719] Catry, B., Duchateau, L., Van De Ven, J., Laevens, H., Opsomer, G., Haesebrouck, F. and A. De Kruif, Efficacy of metaphylactic florfenicol therapy during natural outbreaks of bovine respiratory disease. *Journal of Veterinary Pharmacology Therapy*, **31**, 479-487, 2008. R08228.

[720] Clement, L. and O. Thas, Testing for trends in the violation frequency of an environmental threshold in rivers. *Environmetrics*, **20**, 53-67, 2009. R08220.

[721] Coopman, F., Verhoeven, G., Saunders, J.H., van Bree, H. and L. Duchateau, Prevalence of hip dysplasia, elbow dysplasia and humeral head osteochondrosis in dog breeds in Belgium. *Veterinary Research*, **22**, 654-658, 2008. R08232.

[722] Crop, F., Van Rompaye, B., Paelinck, L., Vakaet, L., Thierens, H. and C. De Wagter, On the calibration process of film dosimetry: OLS inverse regression versus WLS inverse prediction. *Physics in Medicine and Biology*, **53**, 3971-3984, 2008. R08206.

[723] Coussement, K. and D. Van den Poel, Churn prediction in subscription services: an application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications*, **34**, 313-327, 2008. R08225.

[724] Coussement, K. and D. Van den Poel, Improving customer complaint management by automatic email classification using linguistic style features as predictors. *Decision Support Systems*, **44**, 870-882, 2008. R08223.

[725] Coussement, K. and D. Van den Poel, Integrating the voice of customers through call center emails into a decision support system for chum prediction. *Information and Management*, **45**, 164-174, 2008. R08222.

[726] Deschepper, E., Thas, O. and J.P. Ottoy, Tests and diagnostic plots for detecting lack-of-fit for circular-linear regression models. *Biometrics*, **64**, 921-902, 2008. R08034.

[727] Devos, Y., Thas, O., Cougnon, M., De Clercq, E., Cordemans, K. and D. Reheul, Feasibility of isolation perimeters for genetically modified maize. *Agronomical and Sustainable Development*, **28**, 195-206, 2008. R08213.

[728] Devos, Y., Cougnon, M., Thas, O. and D. Reheul, A method to search for optimal field allocations of transgenic maize in the context of co-existence. *Environmental Biosafety Research*, **7**, 97-104, 2008. R08219.

[729] De Wolf, J., Duchateau, L. and E. Schrevens, Dealing with sources of variability in the data-analysis of phenotyping experiments with transgenic rice. *Euphytica*, **160**, 325-337, 2008. R08231.

[730] Goetgeluk, S. and S. Vansteelandt, Conditional generalized estimating equations for the analysis of clustered and longitudinal data. *Biometrics*, **64**, 772-780, 2008. R08021.

[731] Goetgeluk, S., Vansteelandt, S. and E. Goetghebeur, Estimation of controlled direct effects. *Journal of the Royal Statistical Society - Series B*, **70**, 1049-1066, 2008. R08205.

[732] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, **35**, 1071-1079, 2008. R08230.

[733] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, **23**, 92-119, 2008. R08203.

[734] Moerkerke B. and E. Goetghebeur, Optimal screening for promising genes in 2-stage designs. *Biostatistics*, **9**, 700-714, 2008. R08022.

[735] Prinzie, A. and D. Van den Poel, Random forests for multiclass classification: random multinomial logic. *Expert Systems with Applications*, **34**, 1721-1732, 2008. R08224.

[736] Rayner, J.C.W., Thas, O. and B. De Boeck, A generalized Emerson recurrence relation. *Australian and New Zealand Journal of Statistics*, **50**, 235-240, 2008. R08037.

[737] Salibian-Barrera, M. and S. Van Aelst, Robust model selection using fast and robust bootstrap. *Computational Statistics and Data Analysis*, **52**, 5121-5135, 2008. R08057.

[738] Salibian-Barrera, M., Van Aelst, S. and G. Willems, Fast and robust bootstrap. *Statistical Methods and Applications*, **17**, 41-71, 2008. R08202.

[739] Saveyn, H., Curvers, D., Thas, O. and P. Van der Meeren, Optimization of sewage sludge conditioning and pressure dewatering by statistical modelling. *Water Research*, **42**, 1061-1074, 2008. R08211.

[740] Struelens, E., Tuyttens, F.A.M., Duchateau, L., Leroy, T., Cox, M., Vranken, E., Buyse, J., Zoons, J., Berckmans, D., Berg, F. and B. Sonck, Perching behaviour and perch height preference of laying hens in furnished cages varying in height. *British Poultry Science*, **49**, 381-389, 2008. R08229.

[741] Tas, O., De Rooster, H., Baert, E., Doom, M.H. and L. Duchateau, The accuracy of the Lactate Pro hand-held analyser to determine blood lactate in healthy dogs. *Journal of Small Animals Practice*, **49**, 504-508, 2008. R08226.

[742] Tuyttens, F.A.M., Wouters, F., Struelens, E., Sonck, B. and L. Duchateau, Synthetic lying mats may improve lying comfort of gestating sows. *Application in Animals Behaviour Science*, **114**, 76-85, 2008. R08227.

[743] Uyttendaele, M., Rajkovic, A., Van Houteghem, N., Boon, N., Thas, O., Debevere, J. and F. Devlieghere, Multi-method approach indicates no presence of sub-lethally injured Lysteria monocytogenes cells after mild heat treatment. *International Journal of Food Microbiology*, **123**, 262-268, 2008. R08212.

[744] Van Aelst, S., Khan, J.A. and R.H. Zamar, Fast robust variable selection. *COMPSTAT 2008: Proceedings in Computational Statistics, ed. P. Brito, Heidelberg: Physika-Verlag*, 359-370, 2008. R08239.

[745] Van den Eynde, F., Senturk, V., Naudts, K., Vogels, C., Bernagie, K., Thas, O., van Heeringen, C. and K. Audenaert, Efficacy of Quetiapine for impulsivity and affective symptoms in borderline personality disorder. *Journal of Clinical Psychopharmacology*, **28**, 147-155, 2008. R08214.

[746] Van Lent, K. Vanlerberghe, B., Van Oostveldt, P., Thas, O. and P. Van der Meeren, Determination of water droplet size distribution in butter: pulsed filed gradient NMR in comparison with confocal scanning laser microscopy. *International Dairy Journal*, **18**, 12-22, 2008. R08210.

[747] Van Nieuwerburgh, F., Goetghebeur, E., Vandewoestyne, M., and D. Deforce, Impact of allergic drop-out on evidential value of forensic DNA profiles using RMNE. *Bioinformatics Advance Access. Published online on November*, **23**, 2008. R08207.

[748] Vansteelandt, S., DeMeo, D., Su, J., Smoller, J., Murphy, A.J., Mc-Queen, M., Celedon, J., Weiss, S.T., Silverman, E.K. and C. Lange, Testing and estimating gene-environment interactions in family-based association studies. *Biometrics*, **64**, 458-467, 2008. R08020.

[749] Vansteelandt, S., Mertens, K., Suetens, C. and E. Goetghebeur, Marginal structural models for partial exposure regimes. *Biostatistics*, **10**, 46-59, 2009. R08208.

[750] Vansteelandt, S., VanderWeele, T., Tchetgen, E.J. and J.M. Robins, Multiply robust inference for statistical interactions. *Journal of the American Statistical Association*, **103**, 1693-1704, 2008. R08236.

C.4.3 Refereed publications (in press)

[751] Best, D.J., Rayner, J.C.W. and O. Thas, Anscombe's tests of fit for the negative binomial distribution. *Journal of Statistical Theory and Practice*, RP08123.

[752] Clement, L. and O. Thas, Nonparametric trend detection in river monitoring data: a spatio-temporal approach. *Environmetrics*, 2008. RP08122.

[753] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural and Biological Environmental Statistics*, 2008. RP08128.

[754] Legrand, C., Duchateau, L. Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis*, 2008/2009. RP08100.

[755] Paepe, D., Smets, P., van Hoek, I., Saunders, J., Duchateau, L. and S. Daminet, Within- and between-examiner agreement for two thyroid palpation techniques in healthy and hyperthyroid cats. *Journal of Feline Medical Surgery*, 2008. RP08129.

[756] Rayner, J.C.W. and O. Thas, Generalised smooth tests of goodness-of-fit. *Journal of Statistical Theory and Practice*, RP08124.

[757] Thas, O., Rayner, J.C.W. and D.J. Best, Informative statistical analyses using smooth goodness-of- fit tests. *Journal of Statistical Theory and Practice*, RP08130.

[758] Vansteelandt, S., Babanezhad, M. and E. Goetghebeur, Correcting instrumental variables estimators for systematic measurement error. *Statistica Sinica*, 2009. RP08126.

[759] Vansteelandt S., Carpenter J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 2008. RP08027.

[760] Wouters, H., Thas, O. and J.P. Ottoy, Data-driven smooth test and a diagnostic tool for lack-of-fit for circular data. *Australian and New Zealand Journal of Statistics*, 2009. RP08125.

C.4.4 Non-refereed publications (published)

[761] Massonnet, G., Janssen, P. and L. Duchateau, Using copulas to model four-dimensional udder enfection data. *Programme and abstract book 29th Annual conference of the International Society for Clinical Biostatistcs, Copenhagen, Denmark*, 2008. NR08028.

C.4.5 Books (published)

[762] Duchateau, L. and P. Janssen, *The Frailty Model*. Springer, New York, 2008. B08001.

[763] Lagakos, S. and E. Goetghebeur (co-author as member of the corresponding writing committee), *Methodological Challenges in Biomedical HIV Prevention Trials*. The National Academies Press, Washington DC, 2008. B08002.

C.4.6 Books (in press)

[764] Rayner, J.C.W., Thas, O. and D.J. Best, *Smooth tests of goodness-of-fit: using R*. Wiley, Second edition, 2009. BP08010.

C.5 Universiteit Hasselt, UH

C.5.1 Technical reports

[765] Anderluh, J. and T. Kadankova, Double-sided knock-in calls in an exponential compound Poisson framework, 2008. TR08007.

[766] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G. and M.G. Kenward, A sensitivity analysis for shared-parameter models, 2008. TR08058.

[767] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015.

[768] Kenward M.G. and G. Molenberghs, Last observation carried forward: a crystal ball?, 2008. TR08059.

[769] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008. TR08023.

[770] Sotto C., Beunckens C., Molenberghs G. and M.G. Kenward, MCMC-based estimation methods for continuous longitudinal data with non-random (non)-monotone missingness, 2008. TR08057.

[771] Veraverbeke, N., Omelka, M. and I. Gijbels, Weak convergence of the empirical copula process with covariates, 2008. TR08056.

C.5.2 Refereed publications (published)

[772] Abatih, E., Van Oyen, H., Bossuyt, N. and L. Bruckers, Variance estimation methods for health expectancy by relative socio-economic status. *European Journal of Epidemiology*, **23**, 243-249, 2008. R08175.

[773] 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*, **17**, 497-504, 2008. R08138.

[774] Alonso, A. and G. Molenberghs, Surrogate markers: hopes and perils. *Pharmacoeconomics and Outcomes Research*, **8**, 255-259, 2008. R08152.

[775] Alonso, A., Litière, S. and G. Molenberghs, A family of tests to detect misspecifications in the random effects distribution of generalized linear mixed models. *Computational Statistics and Data Analysis*, **52**, 4474-4486, 2008. R08148.

[776] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, **64**, 96-105, 2008. R08102.

[777] 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*, **52**, 1533-1548, 2008. R08157.

[778] Bilau, M., De Henauw, S., Schroijen, C., Bruckers, L., Den Hond, E., Koppen, G., Matthys, C., Van de Mieroop, E., Keune, H., Baeyens, W., Nelen; V., Van Larebeke, N., Willems, J.L. and G. Schoeters, The relation between the estimated dietary intake of PCDD/Fs) and levels in blood in a Flemish population (50-65 years). *Environment International*, **35**, 9-13, 2009. R08179.

[779] 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.

[780] Bollaerts, K., Aerts, M., Faes, C., Grijspeerdt, K., Dewulf, J. and K. Mintiens, Human Salmonellosis: estimation of dose-illness from outbreak data. *Risk Analysis*, **28**, 427-440, 2008. R08046.

[781] 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 semiparametric quantiles regression. *Journal of the Royal Statistical Society, Series A*, **171**, 1-16, 2008. R08045.

[782] Braekers, R. and N. Veraverbeke, A conditional Koziol-Green model under dependent censoring. *Statistics and Probability Letters*, **78**, 927-937, 2008. R08135.

[783] Burzykowski, T., Surrogate endpoints: wishful thinking or reality? *Statistical Methods in Medical Research*, **17**, 463-466, 2008. R08160.

[784] Burzykowski, T., Buyse, M., Piccart-Gebhart, M.J., Sledge, G., Carmichael, J., Luck, H.J., Mackey J.R., Nabholtz, J.M., Paridaens, R., Biganzoli, L., Jasem, J., Bontenbal, M., Bonneterre, 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*, **26**, 1987-1992, 2008. R08028.

[785] Burzykowski, T., Buyse, M. Yothers, G., Skamoto, J. and D. Sargent, Exploring and validating surrogate endpoints in colorectal cancer. *Lifetime Data Analysis*, **14**, 54-64, 2008. R08026.

[786] Buyse, M., Burzykowski, T., Michiels, S. and K. Carroll, Individual and trial level surrogacy in colorector cancer. *Statistical Methods in Medical Research*, **17**, 467-475, 2008. R08161.

[787] Cadarso Suárez, C.M., Faes, C. and G. Molenberghs, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association*, **103**, 140-161, 2008. R08043.

[788] Claeskens, G., Nguti, R. and P. Janssen, Likelihood ratio and scoretest for a shared frailty model: a non-standard problem. *TEST*, **17**, 69-82, 2008. R08023.

[789] Cortiñas, J., Shkedy, Z. and G. Molenberghs, Alternative methods to evaluate trial level surrogacy. *Clinical Trials*, **5**, 194-208, 2008. R08149.

[790] Delhalle, L., De Sadeleer, L., Bollaerts, K., Farnir, F., Saegerman, C., Korsak, N., Dewulf, J., De Zutter, L. and G. Daube, Risk factors for Salmonella and hygiene indicators in the 10 largest Belgian pig slaughterhouses. *Journal of Food Protection*, **71**, 1320-1329, 2008. R08172.

[791] Faes, C., Aerts, M., Molenberghs, G., Geys, H., Teuns, G. and L. Bijnens, A high-dimensional joint model for longitudinal endpoints of different nature. *Statistics in Medicine*, **27**, 4408-4427, 2008. R08151.

[792] Faes, C., Molenberghs, G., Geys, H., Aerts, M., Cadarso-Súarez, C., Acuña, C. and M. Cano, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association, Case Studies*, **103**, 140-161, 2008. R08059.

[793] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, **35**, 1071-1079, 2008. R08230.

[794] Gyselaers, W., Verswijvel, G., Molenberghs, G. and W. Ombelet, Interlobar venous flow is different between left and right kidney in uncomplicated third trimester pregnancy. *Gynecologic and Obstetric Investigation*, **65**, 6-11, 2008. R08146.

[795] Hens, N., Aerts, M., Shkedy, Z., Kimani, P.K., Kojouharova, M., Van Damme, P. and P. Beutels, Estimating the impact of vaccination using age-time dependent incidence rates of Hepatitis B. *Epidemiology and Infection*, **136**, 341-351, 2008. R08165.

[796] Hens, N., Aerts, M., Shkedy, Z., Kimani, P.K., Kojouhariva, M., Van Damme, P. and P. Beutels, Modeling age-time dependent incidence rates of hepatitis B and estimating the prevalence and force of infaction using generalized additive models. *Epidemiology and Infection*, **136**, 341-351, 2008. R08169.

[797] Hens, N., Aerts, M., Shkedy, Z., Thetten, H., Van Damme, P. and P. Beutels, Modelling multi-sera data: the estimation of new joint and conditional epidemiological parameters. *Statistics in Medicine*, **27**, 2651-2664, 2008. R08163.

[798] Hughes, S., Keene, O., Roes, K., Ashby, D., Beltangadu, M., Bird, S.M., Burton, C., Buyse, M., Day, S., van Ewijk, P., Fletcher, C., Grieve, A.P., Guinot, C., Hemmings, R., Hold, D., Hothorn, L., Jagers, P., Keiding, N., Korhonen, P., Lesaffre, E., Lewis, J.A., Molenberghs, G., Neumann, N., Olsen, K.J., van Osta, G., Pocock, S., Rockhold, F., Rossi, A., Senn, S., Stijnen, T., Vaeth, M., Wiklund, S.J. and J. Wolfram, Open Letter to the Heads of the European Regulatory Agencies. *British Medical Journal*, **336**, 250-250, 2008. R08145.

[799] Jacobs, T., De Ridder, F., Rusch, S., Van Peer, A., Molenberghs, G. and L. Bijnens, Including information on the therapeutic window in bioequivalence specifications. *Pharmaceutical Research*, **25**, 2628-2638, 2008. R08154.

[800] Jacobs, T., Rossenu, S., Dunne, A., Molenberghs, G., Straetemans, R. and L. Bijnens, Combined models for data from in vivo-in vitro correlation experiments. *Journal of Biopharmaceutical Statistics*, **18**, 1197-1211, 2008. R08150.

[801] Jansen, I. and G. Molenberghs, A flexible marginal modeling strategy for non-monotone missing data. *Journal of the Royal Statistical Society, Series A*, **171**, 347-373, 2008. R08141.

[802] 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. *International Journal of Global Environmental Issues*, **8**, 90-111, 2008. R08178.

[803] Lin, D., Shkedy, Z., Burzykowski, T., Ion, R., Göhlmann, H.W., Bondt, A.D., Perer, T., Geerts, T., Van den Wyngaert, I. and L. Bijnens, An investigation on performance of Significance Analysis of Microarray (SAM) for the comparisions of several treatments with one control in the presence of small-variance genes. *Biometrical Journal*, **50**, 801-823, 2008. R08162.

[804] Litière, S., Alonso, A., Molenberghs, G. and H. Geys, The impact of a misspecified random-effects distribution on the estimation and performance of inferential procedures in generalized linear mixed models. *Statistics in Medicine*, **27**, 3125-3144, 2008. R08139.

[805] Maringwa, J.T., Geys, H., Shkedy, Z., Faes, C., Molenberghs, G., Aerts, M., Van Ammel, K., Teisman, A., and L. Bijnens, Application of semiparametric mixed midels in simultaneous confidence bands in a cardiovascular safety experiment with longitudinal data. *Journal of Biopharmaceutical Statistics*, **18**, 1043-1062, 2008. R08049.

[806] Maringwa, J.T., Geys, H., Shkedy, Z., Faes, C., Molenberghs, G., Aerts, M., Van Ammel, K., Teisman, A. and L. Bijnens, Tutorial in biostatistics: analysis of cross-over designs with serial correlation within periods using semiparametric mixed models. *Statistics in Medicine*, **27**, 6009-6033, 2008. R08153.

[807] Massonnet, G., Janssen, P. and T. Burzykowski, Fitting conditional survival models to meta-analytic data by using a transformation towards mixed-effects models. *Biometrics*, **64**, 8634-842, 2008. R08029.

[808] Méroc, E., Faes, C., Herr, C., Verheyden, B., Vanbinst, T., Vandenbussche, F., Hooyberghs, J. Aerts, M. De Clercq, K., Mintiens, K. and C. Staubach, Establishing the spread of bluetongue virus at the end of the 2006 epidemic in Belgium. *Veterinary Microbiology*, **131**, 133-144, 2008. R08166.

[809] Mintiens, K., Méroc, E., Faes, C., Cortiñas Abrahantes, J., Hendrickx, G., Staubach, C., Gerbier, G., Elbers, A.R.W., Aerts, M., De Claercq, K. and H. Deluyker, Impact of human interventions on the spread of Bluetongue virus serotype 8 during the 2006 epidemic in north-western Europe. *Preventive Veterinary Medicine*, **87**, 145-161, 2008. R08168.

[810] 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. *Journal of the Royal Statistical Society, Series B*, **70**, 371-388, 2008. R08073.

[811] 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.

[812] Mossong, J., Hens, N., Friederichs, V., Davidkin, I., Broman, M., Litwinska, A., Siennicka, J., Trzcinska, A., Van Damme, P., Beutels, P., Vyse, A., Shkedy, Z., Aerts, M., Massari, M. and G. Gabutti, Epidemiologue of parvovirus B19 infection in Europe: results from a sero-prevalence studies in Belgium, England and Wales, Finland, Italy and Poland. *Epidemiology and Infection*, **136**, 1059-1068, 2008. R08170.

[813] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensitivity analysis for incomplete data. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods. Chapter 22. Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, eds. Chapman & Hall/CRC*, 501-552, 2008. R08242.

[814] Namata, H., Aerts, M., Faes, C. and P. Theunis, Model averaging in microbioal risk assessment using fractional polynomials and generalized linear mixed models. *Risk Analysis*, **28**, 891-905, 2008. R08167.

[815] Namata, H., Méroc, E., Aerts, M., Faes, C., Cortiñas, A.J., Imberechts, H. and K. Mintiens, Salmonella in Belgian laying hens: an identification of risk factors. *Preventive Veterinary Medicine*, **83**, 323-336, 2008. R08164.

[816] Piccart-Gebhart, M.J., Burzyokowski, T., Buyse, M., Sledge, G., Carmichael, J., Luck, H.J., Mackey, J.R., Mabholtz, J.M., Paridaens, R., Biganzoli, L., Jassem, J., Bontenbal, M. Bonneterre, J., Chan, S., Atalay Basaran G. and P. Therasse, Effects of taxanes ablone or in combination with antracyclines as first-line therapy of patients with metastatic breast cancer. *Journal of Clinical Oncology*, **26**, 1980-1986, 2008. R08027.

[817] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63-74. R08019.

[818] Robaeys, G., Nevens, F., Van Eyken, P., Starkel, P., Colle, I., Bruckers, L., Van Ranst, M. and F. Buntinx, Intravenous substance use and the outcome if liver transplantation for chronic hepatitis C. *Journal of Hepatology*, **48** (Supp.2), 230, 2008. R08176.

[819] Schroijen, C., Baeyens, W., Schoeters, G., Den Hond, E., Koppen, G., Bruckers, L., Nelen, V., De Mieroop, E., Bilau, M., Covaci, A., Keune, H., Loots, I., Kleinjans, J., Dhooge, W. and Van Larebeke, Internal exposure to pollutants measured in blood and urine of Flemish adolescents in function of aera of residence. *Chemosphere*, **71**, 1317-1325, 2008 . R08174.

[820] 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. *International Journal for Parasitology*, **118**, 522-530, 2008. R08030.

[821] Sterna, J. and T. Burzykowski, Assessment of the usefulness of the fenestration method in case of disc extrusion in the cervical and thoraco-lumbar spine in chondrodystrophic dogs. *Polish Journal of Veterinary Sciences*, **11**, 55-62, 2008. R08183.

[822] 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*, **18**, 326-341, 2008. R08142.

[823] Tilahun, A., Maringwa, J.T., Geys, H., Alonso, A., Raeymaekers, L., Molenberghs, G., Van Den Kieboom, G., Drinkenburg, P. and L. Bijnens, Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. *Journal of Biopharmaceutical Statistics*, **19**, 1-17, 2009. R08155.

[824] Van Effelterre, T., Shkedy, Z., Aerts, M., Molenberghs, G., Van Damme, P. and P. Beutels, Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus. *Epidemiology and Infection*, **127**, 48-57, 2009. R08147.

[825] 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*, **18**, 1-22, 2008. R08140.

[826] Valkenborg, D., Van Sanden, S., Lin, D., Kasim, A., Zhu, Q., Haldermans, P., Jansen I., Shkedy, Z. and T. Burzykowski, A cross-validation study to select a classification procedure for clinical diagnosis base on proteomic mass spectrometry. *Statistical Applications in Genetics and Molecular Biology*, **7(2)**, article 12, 2008. R08158.

[827] Valkenborg, D., Janssen, I. and T. Burzykowski, A model-based method for the prediction of the isotopic distribution of peptides. *Journal of the American Society of Mass Spectrometry*, **19**, 703-712, 2008. R08159.

[828] Van Sanden, S., Lin, D. and T. Burzykowski, Performance of gene selection and classification methods in a microarray setting: A simulation study. *Communications in Statistics Simulation and Computation*, **37**, 409-424, 2008. R08031.

[829] Veraverbeke, N., Conditional residual life under random censorship. *Advances in Statistics. Felicitation volume in honour of B.K. Kale, B.C. Armold, U. Gather, S.M. Bendre, eds. MacMillan India, New Delhi*, 174-185, 2008. R08180.

[830] Verbeke, G., Molenberghs, G. and C. Beunckens, Formal and informal model selection with incomplete data. *Statistical Science*, **23**, 201-218, 2008. R08099.

[831] Vloeberghs, E., Van Dam, D., Frank, F., Serroyen, J., Molenberghs, G., Staufenbiel, M. and P.P. De Deyn, Altered ingestive behavior, weight changes and intact olfactory sense in an APP overexpression model. *Behavioral Neuroscience*, **122**, 491-497, 2008. R08137.

[832] Vyt, P., Maes, D., Quinten, C., Rysselaere, T., Deley, W., Aerts, M., de Kruif A. and A. Van Soom, Detailed motility evaluation of boar semen and it predictive value for reproductive performance in sows. *Flemish Veterinary Journal*, **77**, 291-299, 2008. R08173.

[833] 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.

[834] Wouters, K., Cortiñas, J., Molenberghs, G., Ahnaou, A., Bijnens, L. and H.I.M. Drinkenburg, A comparison of doubly hierarchical supervised learning procedures for multiple class longitudinal data from EEG experiments. *Journal of Biopharmaceutical Statistics*, **18**, 1120-1135, 2008. R08144.

C.5.3 Refereed publications (in press)

[835] Alonso, A., Molenberghs, G., Renard, D. and C. Darstein, Evaluation of biomarkers for pharmacological activity. *Journal of Biopharmaceutical Statistics*, 2008. RP08075.

[836] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multi-faceted sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08081.

[837] Beunckens C., Sotto C., Molenberghs G. and G. Verbeke, An integrated sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08030.

[838] Bollaerts, K., Messens, W., Delhalle, L., Aerts, M., Van der Stede, Y., Dewulf, J., Quoillin, S., Maes, D., Mintiens, K. and K. Grijspreedt, METZOON: development of a quantitative microbial risk assessment for human salmonellosis through household consumption of fresh minced pork meat in Belgium. *Risk Analysis*, 2008. RP08097.

[839] Boone, I., Van der Stede, Y., Bollaerts, K., Vose, D., Daube, G., Aerts, M. and K. Mintiens, NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for Salmonella in the pork production chain. *Risk analysis*, 2008. RP08098.

[840] Braekers, R. and I. Van Keilegom, Flexible modeling base on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, 2008. RP08090.

[841] Cortiñas Abrahantes, J., Bollaerts, K. Aerts, M. Van der Stede, Y., Ogunsanya, V. and K. Mintiens, Salmonella serosurveillance: different statistical methods to categorize pig herds based in serological data. *Preventive Veterinary Medicine*, 2008. RP08092.

[842] Decramer, M. and G. Molenberghs, Letter to the editor: does pharmacotherapy reduce rate of decline of lung function in COPD?. *American Journal of Respiratory and Critical Care Medicine*, 2008. RP08078.

[843] Faes, C., Aerts, M., Geys, H. and L. De Schaepdrijver, Modeling spatial learning in rats based on morris water maze experiment. *Pharmaceutical Statistics*, 2008. RP08102.

[844] Faes, C., Geys, H., Aerts, M. and G. Molenberghs, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association*, 2008. RP08101.

[845] Gaddah, A. and R. Braekers, Weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring. *Journal of Statistical Planning and Inference*, 2008. RP08089.

[846] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology*. RP08112.

[847] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural and Biological Environmental Statistics*, 2008. RP08128.

[848] Gyselaers, W., Molenberghs, G., Van Mieghem, W. and W. Ombelet, Doppler measurement of renal interlobar vein impedance index in uncomplicated and pre-eclamptic pregnancies: a pilot study. *Hypertension in Pregnancy*, 2008. RP08055.

[849] Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Van Damme, P. and P. Beutels, Mining social mixing patterns for infectious disease models based on a two-day population survey in Belgium. 2008. RP08095.

[850] Hens, N., Kvitkovicova, A., Aerts, M., Hubinka, D. and P. Beutels, Modelling distortions in seroprevalence data using change-point fractional polynomials. *Statistical Modelling*, 2008. RP08094.

[851] Hill, R.C. Choate, C.J., Scott, K.C. and G. Molenberghs, Comparison of the guaranteed analysis with the measured nutrient composition of commercial pet foods. *Journal of the American Veterinary Medical Association*, 2008. RP08076.

[852] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, A family of measures to evaluate scale reliability in a longitudinal setting. *Journal of the Royal Statistical Society, Series A*, **171**,2008. RP08067.

[853] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Reliability of a longitudinal sequence of scale ratings. *Psychometrika*, 2008. RP08077.

[854] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scale. *Journal of Psychiatric Research*, 2008. RP08082.

[855] Legrand, C., Duchateau, L. Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis*, 2008/2009. RP08100.

[856] Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso-Suárez, C., Pardo-Vázquez, J.L., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08079.

[857] Molanes, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, 2008. RP08068.

[858] Molenberghs, G., Incomplete data in clinical studies: analysis, sensitivity and sensitivity analysis (with discussion and rejoinder). *Drug Information Journal*, 2008. RP08080.

[859] Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A. and M. Buyse, A unified framework for the evaluation of surrogate endpoints in clinical trials. *Statistical Methods in Medical Research*, 2008. RP08085.

[860] Mossong, J., Hens, N., Jit, M., Beutels, P., Auranen, K., Mikolajczyk, R. Massari, M., Scalia Tomba, G., Wallinga, J., Sadkowska-Todys, Rosinska, M. and J.W. Edmunds, Social contacts and mixing patterns relevant to the spread of infectious diseases. *PloS Med*, **5**(3), 2008. RP08093.

[861] Mwambi, H., Ramroop, S., Shkedy, Z. and G. Molenberghs, An approach to estimating the force of infection and the rate of recovery from Respiratory Syncytial Virus (RSV) data. *Statistical Methods in Medical Research*, 2008. RP08083.

[862] Mwambi, K., Ramroop, S., Shkedy, Z. and G. Molenberghs, A frequentist appraoch to estimating the force of infection and the recovery rate for respiratory disease among infants in coastal Kenya. *Statistical Methods in Medical Research*, 2008. RP08088.

[863] Ogunjimi, B., Hens, N., Goeyvaerts, N., Aerts, M. and P. Beutels, Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. *Mathematical Biosciences*, 2008. RP08096.

[864] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, 2008. RP08047.

[865] Pardo Vázquez, J.L., Cadarso Suárez, C.M., Molenberghs, G., Geys, H., Faes, C., Maringwa, J.T., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08021.

[866] Serroyen, J., Bruckers, L., Rogiers, G. and G. Molenberghs, Characterizing persistent disturbing behavior using longitudinal and multivariate techniques. *Journal of Applied Statistics*, 2008. RP08086.

[867] Sotto, C., Beunckens, C., Molenberghs G., Jansen I., and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, 2008, RP08031.

[868] Stracjzek, A., Wannijn, J., Van Hees, M., Thijs, H. and Y. Thiry, Tolerance of hairy roots of carrots to U-chronic exposure in standardized in vitro device. *Environmental and Experimental Botany*, 2008. RP08103.

[869] Theeten, H., Hens, N., Aerts, M., Vandermeulen, C., Roelants, M., Hoppenbrouwers, K., Van Damme, P. and P. Beutels, Caregivers' willingness to pay to reduce the number of vaccine injections in infants. *The Pediatric Infectious Disease Journal*, 2008. RP08091.

[870] Verbeke G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling*, 2009. RP08033.

C.5.4 Non-refereed publications (published)

[871] Aerts, M., Bollaerts, K., Ribbens, K., Van der Stede, Y., Boone, I. and K. Mintiens, Semiparametric regression to identify farms with high Salmonella infection burden. *Proceedings of the 23rd International Workshop on Statistical Modelling, Eilers, P. (Ed.). Utrecht*, 89-92, 2008. NR08006.

[872] Bollaerts, K., Aerts, M., Faes, C., Dewulf, J. and K. Mintiens, Dose-illness models for human salmonellosis based on outbreak data. *Proceedings of 23rd International Workshop on Statistical Modelling. Paul Eilers (ed.), Utrecht, the Netherlands*, 1245-1300, 2008. NR08003.

[873] Bollaerts, K., Messens, W., Delhalle, L., Aerts, M., Dewulf, J., Debusser, E., Boone, I. and K. Grijspeerdt, Development of a modular quantitative microbial risk assessment to evaluate zoonotic risks in Belgium: salmonellosis through consumption of pork as an example. *Proceedings of 10th European Symposium on Statistical Methods for the Food Industry, Louvain-la-Neuve, Belgium*, 2008. NR08018.

[874] Buntinx, F., Aertgeerts, B., Aerts, M., Bruyninckx, R., Knottnerus, J.A., Van den Bruel, A. and J. Van den Ende, Multivariable analysis in diagnostic accuracy studies. What are the possiblities? *The Evidence Base of Clinical Diagnosis. J.A. Knottnerus and F. Buntinx. John Wiley and Sons*, 146-167, 2008. NR08008.

[875] Burzykowski, T. and D. Valkenborg, Processing of high-resolution MALDI-TOF mass spectrometry data by using an average isotopic distribution. *Proceedings of the Fourteenth National Conference on Application of Mathematics in Biology and Medine. Warsaw, Poland. (Bodnar, M., Forys, U., eds)*, 33-38, 2008. NR08017.

[876] Callaert, H., Understanding confidence intervals. *Proceedings of the Fifth Conference of the European Society for Research in Mathematics Education. CD-ROM edited by D. Pantazi and G. Philippou*, 2008. NR08002.

[877] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Shared-parameter models and missingness at random. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18*, 2008. NR08023.

[878] Faes, C., Geys, H. and P. Catalano, Joint models of discrete and continuous longitudinal data. *Handbooks of Modern Statistical Methods: Longitudinal Data Anlysis. Fitzmaurice, G., Davidian, M., Molenberghs, G. and Verbeke, G. (Eds.) London: Chapman Hall/CRC*, 327-348, 2008. NR08007.

[879] Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Van Damme, P. and P. Beutels, Estimating transmission parameters using social contact data and serological data. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18, 2008*, 2008. NR08022.

[880] Hens, N., Exploiting data for risk evaluation. *Scientific exploitation of databases within the framework of food safety risk assessment. Federal Food Safety Authority. Brussels, Belgium*, 2008. NR08027.

[881] Hens, N., Aerts, M., Melagaro, A., Gay, N. and J. Edmunds, Elucidation of the factors producing any observed differences in the epidemiological patterns of airborne infections across Europe. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18 2008*, 2008. NR08024.

[882] Hens, N., Wienke, A., Aerts, M. and G. Molenberghs, On the correlated frailty model for bivariate current status data. *Proceedings of the workshop on composite likelihood methods. Warwick, UK, 15-17 April 2008*, 2008. NR08021.

[883] Jacobs, T., Straetemans, R., Rossenu, S., Dunne, A., Molenberghs, G. and L. Bijnens, A new convolution-based approach to develop a level a in-vitro-in-vivo correlation (IVIVC) for an extended-release oral dosage form. *2007 Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM], Alexandria, VA: American Statistical Association*, 141-144, 2008. NR08004.

[884] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Reliability of a longitudinal sequence of scale ratings. *Proceedings of the 24th International Biometric Conference, Dublin, Ireland*, 2008. NR08029.

[885] Lejeune, O., Hens, N., Bilcke, J. and P. Beutels, Proceedings of the international workshop facing the challenge of infectious diseases. *Integrating Mathematical Modeling, Computa-*

tional Thinking and ICT Applications, ISI Foundations, Turin (Italy), October 13-18, 2008. NR08025.

[886] Leuridan, E., Hens, N., Ieven, G., Aerts, M. and P. Van Damme, Maternal antibodies against measles after natural infection or vaccination in parturient women. *Proceedings of the 26th Annual meeting of the European Society for Paediatric Infectious Diseases ESPID, Graz, Austria, 14-16 May 2008*, 2008. NR08020.

[887] Lin, D., Shkedy, Z., Burykowski, T., Aerts, M., Göhlmann, H.W.H., De Bondt, A., Perera, T., Geerts, T., Van den Wyngaert, I. and L. Bijnens, Information theory based marker evaluation. *2007 Proceedings of the American Statistical Association, Biopharmaceutical Section [CD-ROM], Alexandria, VA: American Statistical Association*, 472-479, 2008. NR08005.

[888] Massonnet, G., Janssen, P. and L. Duchateau, Using copulas to model four-dimensional udder enfection data. *Programme and abstract book 29th Annual conference of the International Society for Clinical Biostatistcs, Copenhagen, Denmark*, 2008. NR08028.

[889] Messens, W., Bollaerts, K., Delhalle, L., Aerts, M., Van der Stede, Y., Quoilin, S., De Busser, E., Dewulf, J., Maes, D., Boone, I., Mintiens, K. and K. Grijspeerdt, Development of a quantitative microbial risk assessment to evaluate zoonotic risks in Belgium: human salmonellosis through household consumption of minced pork meat. *First Belgian Symposium on Salmonella Research and Control in Pigs, Ghent Belgium, Book of proceedings. Editors: F. Boyens, F. Pasmans*, 43-47, 2008. NR08019.

[890] Molenberghs, G., Aerts, M., Beunckens, C., Creemers, A., Hens, N., Kenward, M.G., Sotto, C. and G. Verbeke, Missing not at random models and their missing at random counterparts, in various modelling frameworks. *ACTAS do XVI Congresso da SPE, UTAD, Vila REAL*, 2008. NR08026.

[891] Moons, E., Wets, G. and M. Aerts, Investigating the performance if rule-based models with increasing complexity on the prediction of trip generation and distribution. *Rossi, C. (Ed.): Frontiers in Brain, Vision and AI, In-The*, 167-182, 2008. NR08009.

C.5.5 Books (published)

[892] Duchateau, L. and P. Janssen, *The Frailty Model*. Springer, New York, 2008. B08001.

C.5.6 Books (in press)

[893] Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, *Longitudinal Data Analysis, A Handbook of Modern Statistical Methods*. Chapman and Hall/CRC, New York, 2009. BP08011.

C.6 Université Joseph Fourier, UJF–LMC–IMAG

C.6.1 Technical reports

[894] Antoniadis, A., Fryzlewicz, P. and F. Letué, The Dantzig selector in Cox's proportional hazards model, 2008. TR08047.

[895] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties, 2008. TR08050.

[896] Boshnakov, G. and S. Lambert-Lacroix, Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients, 2008. TR08048.

C.6.2 Refereed publications (published)

[897] Amato, U., Antoniadis, A., Cuomo, V., Cutillo, L., Franceze, M., Murino, L. and C. Serio, Statistical cloud detection from SEVIRI multispectral images. *Remote Sensitive Environment*, **112**, 750-766, 2008. R08067.

[898] Letué, F., A semiparametric shock model for a pair of event-related dependent censored failures times. *Journal of Statistical Planning and Inference*, **138**, 3869-3884, 2008. R08065.

C.6.3 Refereed publications (in press)

[899] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of spatially and temporally heterogeneous brain response images. *Journal of Computational and Graphical Statistics*, 2008. RP08001.

[900] Antoniadis, A., Gijbels, I. and J.M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, 2008. RP08038.

[901] Antoniadis, A., Paparoditis, E. and T. Sapatinas, Bandwidth selection for functional times series prediction. *Statistics and Probability Letters*, 2008. RP08003.

[902] Chatelain, F., Lambert-Lacroix, S. and J.Y. Tourneret, Pairwise likelihood estimation for multivariate mixed Poisson models generated by Gamma intensities. *Statistics and Computing*, 2008. RP08004.

[903] Gannaz, I. and O. Winterberger, Adaptive density estimation with dependent observations. *ESAIM: Probability and Statistics*, 2008. RP08005.

C.7 Universiteit Utrecht, UU

C.7.1 Refereed publications (published)

[904] Camarda, C.G., Eilers, P.H.C. and J. Gamper, Modelling general patterns of digit preference. *Statistical Modelling*, **8**, 385-402, 2008. R08068.

[905] Eilers, P.H.C., Gampe, J., Marx B.D. and R. Rau, Modulation models for seasonal time series and incidence tables. *Statistics in Medicine*, **27**, 3430-3441, 2008. R08069.

C.7.2 Refereed publications (in press)

[906] Blasius, J., Eilers, P.H.C. and J. Gower, Better biplots. *Computational Statistics and Data Analysis*, 2008. RP08006.

[907] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, 2008. RP08007.

C.7.3 Non-refereed publications (published)

[908] Camarda, C.G., Eilers, P.H.C. and J. Gampe, A warped failure time model for human mortality. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08010.

[909] Eeuwijk, F.A., Malosetti, M. Boer, M.P. and P.H.C. Eilers, Modeling spatial effects in field trials. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08013.

[910] Gampe, J. and P.H.C. Eilers, Variation in mortality: estimation via a meta-analytic approach. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08011.

[911] Marx, B.D., Eilers, P.H.C., Gampe, J. and R. Rau, Bilinear varying-coefficient models for seasonal time series and tables. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08015.

[912] Rippe, R.C.A., Eilers, P.H.C. and J.J. Meulman, Models for fluorescence signals on SNP arrays. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08012.

[913] Schnabel, S.K. and P.H.C. Eilers, Optimal expectile smoothing. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08016.

[914] Uh, H.W. and P.H.C. Eilers, Haplotype frequency estimation with the penalized composite link model. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08014.

C.8 Universidad de Santiago de Compostela, USC

C.8.1 Technical reports

[915] Crujeiras, R. M. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends, 2008. TR08044.

[916] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Goodness-of-fit tests for censored regression based on artificial data points, 2008. TR08060.

[917] González-Manteiga, W., Pardo-Fernandez, J. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models, 2008. TR08045.

[918] Pateiro-López, B. and A. Rodríguez-Casal, A., On the estimation of alpha-convex sets, 2008. TR08049.

C.8.2 Refereed publications (published)

[919] Barús, J.I., Febrero-Bande, M. and C.M. Cadarso Suárez, Flexible regression models for estimating postmortem interval (PMI) in forensic medicine. *Statistics in Medicine*, **27**, 5026-5038, 2008. R08078.

[920] Boente, G., González-Manteiga, W. and A. Pérez González, Robust nonparametric estimation with missing data. *Journal of Statistical Planning and Inference*, **139**, 571-592, 2009. R08184.

[921] Cadarso Suárez, C.M., Faes, C. and G. Molenberghs, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association*, **103**, 140-161, 2008. R08043.

[922] 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.

[923] Crujeiras, R. and W. González-Manteiga, Comments on: Augmenting the bootstrap to analyze high dimensional genomic data. *TEST*, **17**, 40-42, 2008. R08075.

[924] da Silva Domínguez, José Luis , Rilo Pousa, B., Mora, M.J., Cadarso Suárez, C.M., Santana P. and A. Urbano, Midline shift and lateral guidance angle in adults with unilateral posterior crossbite. *American Journal of Orthodontics and Dentofacial Orthopedics*, 804-808, 2008. R08083.

[925] Febrero-Bande, M. , Galeano, P. and W. González-Manteiga, Outlier detection in functional data by depth measures with application to identify abnormal NOx levels. *Environmetrics*, **19**, 4, 331-345, 2008. R08070.

[926] Fernández de Castro, B.M. and W. González-Manteiga, Boosting for real and functional samples. An application to an environmental problem. *Stochastic Environmental Research and Risk Assesment*, **22**, 1, 27-37, 2008. R08071.

[927] González-Manteiga, W., Ferraty F. and P. Vieu, Nonparametric functional data analysis: theory and practice. *Computational Statistics*, **23**, 2, 341-342, 2008. R08080.

[928] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Parametric conditional mean and variance testing with censored data. *H. Skiadas, Christos, ed. Recent Advances in Applied Stochastic Models and Data Analysis. N.p.: World Scientific Publishing Co. Pte Ltd*, 2008. R08104.

[929] González-Manteiga, W. , Lombardía, M.J. , Molina Perala, I. , Morales González, D. and L. Santamaría, Analytic and bootstrap approximations of prediction errors under a multivariate Fay-Herriot model. *Computational Statistics and Data Analysis*, **52**, 5242-5252, 2008. R08081.

[930] González-Manteiga, W., Lombardía, M.J., Molina, I., Morales, D. and L. Santamaría, Bootstrap mean squared error of a small-area EBLUP. *Journal of Statistical Computation and Simulation*, **78**, 443-462, 2008. R08234.

[931] Lombardía, M.J. and S. Sperlich, Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society Series B*, **70**, 913-930, 2008. R08233.

[932] Martinez Miranda, M.D., Raya-Miranda, R. , González-Manteiga, W. and A. González-Carmona, A bootstrap local bandwidth selector for additive models. *Journal of Computational and Graphical Statistics*, **17**, 38-55, 2008. R08072.

[933] Matias, J.M. , Febrero-Bande, M. , González-Manteiga, W. and J.C. Reboredo, Gradient boosting GARCH and neural networks for time series prediction. *Supervised and Unsupervised Ensemble Methods and Their Applications*, Studies in Comp. Intelligence, 153-164, 2008. R08079.

[934] Meira-Machado, L., Cadarso Suárez, C.M. and J. de Uña Álvarez, Inference in the progressive tree-state model. *International Journal of Mathematical Models and methods in Applied Sciences*, **2**, 447-454, 2008. R08085.

[935] Menezes, R. , García-Soidán, P.H. and M. Febrero-Bande, A kernel variogram estimator for clustered data. *Scandinavian Journal of Statistics*, **35**, 18-37, 2008. R08004.

[936] Pateiro-López, B. and A. Rodríguez-Casal, Length and surface area estimation under smoothness restrictions. *Advances in Applied Probability*, **40**, 348-358, 2008. R08076.

[937] Phillips C., Fondevila M., García-Magariños, M., Rodríguez A., Salas A., Carracedo A. and M.V. Lareu, Resolving paternity cases that show ambiguous STR results using autosomal. *Forensic Science International: Genetics*, **2**, 198-204, 2008. R08200.

[938] Roca Pardiñas, J., Cadarso Suárez, C.M., Garcia-Tahoces, P. and M.J. Lado, Categorical variables, interactions and generalized additive models. Applications in computer-aided diagnosis systems. *Computers in Biology and Medicine*, **38**, 4, 475-83, 2008. R08077.

[939] Roca Pardiñas, J., Cadarso Suárez, C.M., Garcia-Tahoces, P. and M.J. Lado, Selecting variables in non-parametric regression models for binary response. An application to the computerized detection of breast cancer. *Statistics in Medicine*, **28**, 2, 240-259, 2008. R08084.

[940] Salas A., Vega A., García-Magariños, M., Ruibal A., Benítez J. and A. Carracedo, The 'Pokemon' (ZBTB7) gene: no evidence of association. *Clinical Medicine: Oncology*, **2**, 357-362, 2008. R08201.

[941] Sperlich, S. and M. J. Lombardía, Semiparametric inference in generalized mixed effects models. *Journal of the Royal Statistical Society Series B*, **70**, 5, 913-930, 2008. R08074.

[942] Van Keilegom, I., González-Manteiga, W. and C. Sánchez Sellero, Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401-415, 2008. R08015.

[943] Van Keilegom, I., Sánchez Sellero, C.A. and W. González-Manteiga, Empirical likelihood based testing for regression. *Electronic Journal of Statistics*, **2**, 581-604, 2008. R08117.

C.8.3 Refereed publications (in press)

[944] Castellano Méndez, M., Franco, A., Cartelle, D., Febrero-Bande, M. and E. Roca, Identification of NOx and ozone episodes and estimation of ozone by statistical analysis. *Water, Air & Soil Pollution*, 2008. RP08019.

[945] Crujeiras, R., Fernández-Casal, R. and González-Manteiga, An L2-test for comparing spatial spectral densities. *Statistics and Probability Letters*, RP08011.

[946] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, Goodness-of-fit tests for the spatial spectral density. *Stochastic Environmental Research and Risk Assesment*, RP08015.

[947] Febrero-Bande, M., Galeano, P. and W. González-Manteiga, Measures of influence for the functional linear model with scalar response. *Journal of Multivariate Analysis*, RP08016.

[948] Galeano, P. and D. Peña, An unified approach to model selection, discrimination, goodness-of-fit and outliers in time series. *Advances in Mathematical and Statistical Modeling*, RP08010.

[949] González-Manteiga, W., Martinez Miranda, M.D. and R. Raya-Miranda, Sizer map for inference with additive models. *Statistics and Computing*, RP08012.

[950] González Barcala, F.J. and C.M. Cadarso Suárez, Spirometric reference equations for 6-18 year-olds in Galicia (N.W. Spain). *Arch Bronconeumol.*, RP08017.

[951] González-Manteiga, W. , Lombardía, M.J., Molina Perala, I., Morales González, D. and L. Santamaría, Small area estimation under Fay-Herriot models with nonparametric estimation of heteroscedasticity. *Statistical Modelling*, RP08013.

[952] Jorge Barreiro, F.J., Cadarso Suárez, C.M., Relova Quinteiro, J.L., Bregains, J., López Martín, E. and F.J. Ares Pena, The action of pulse-modulated GSM radiation increases regional changes in brain activity and c-Fos expression in cortical and subcortical areas in a rat model of picrotoxin-induced seizure-proneness. *Journal of Neuroscience Research*, 2008. RP08020.

[953] Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso-Suárez, C., Pardo-Vázquez, J.L., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08079.

[954] Martinez Urtaza, J., Gavilán, R.G., Huapaya, B., Blanco Abad, V., Cadarso Suárez, C. M., Figueiras Guzmán, A. and J. Trinanes, Emergence of Asiatic Vibrio diseases in South America in phase with El Niño. *Epidemiology*, 2008. RP08018.

[955] Meira-Machado, L., Cadarso Suárez, C.M., de Uña Álvarez, J. and P.K. Andersen, Multistate models for the analysis of time to event data. *Statistical Methods in Medical Research*, 2008. RP08022.

[956] Pardo Vázquez, J.L., Cadarso Suárez, C.M., Molenberghs, G., Geys, H., Faes, C., Maringwa, J.T., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08021.

[957] Pérez González, A., Vilar Fernández, J.M. and W. González-Manteiga, Asymptotic properties of local polynomial regression with missing data and correlated errors. *Annals ot the Institute of Statistical Mathematics*, RP08009.

[958] Porcu, E., Crujeiras, R., Mateu, J. and W. González-Manteiga, On the second order properties of the multidimensional peridogram for regularly spaced data. *Theory of Probability and its Applications*, RP08008.

[959] Reboredo, J.C., González-Manteiga, W., Febrero-Bande, M. and J.M. Matias,Boosting GARCH and neural networks for the prediction of heteroskedastic time series. *Journal of Computational and Applied Mathematics*, 2008. RP08014.

[960] Roca Pardiñas, J., Cadarso Suárez, C.M. and W. González-Manteiga, Resampling and testing in regression models with environmetrical applications. *Comprehensive Chemometrics (Brown S., Tauler R. and B. Walczak, Eds.)*, 2008. BP08001.

C.9 London School of Hygiene and Tropical Medicine, LSHTM

C.9.1 Technical reports

[961] Creemers, A., Hens, N., Aerts, M., Molenberghs. G., Verbeke. G. and M.G. Kenward, A sensitivity analysis for shared-parameter models, 2008. TR08058.

[962] Kenward, M.G. and G. Molenberghs, Last observation carried forward: a crystal ball?, 2008. TR08059.

[963] Sotto, C., Beunckens, C., Molenberghs, G. and M.G. Kenward, MCMC-based estimation methods for continuous longitudinal data with non-random (non)-monotone missingness, 2008. TR08057.

C.9.2 Refereed publications (published)

[964] Kenward, M.G. and J.R. Carpenter, Multiple Imputation. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods. Chapter 21. Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, eds. Chapman & Hall/CRC*, 477-500, 2008. R08241.

[965] 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. *Journal of the Royal Statistical Society, Series B*, **70**, 371-388, 2008. R08073.

[966] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensitivity analysis for incomplete data. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods. Chapter 22. Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, eds. Chapman & Hall/CRC*, 501-552, 2008. R08242.

C.9.3 Refereed publications (in press)

[967] Carpenter, J.R. and I. Plewis, Analysing longitudinal studies with non-response: issues and statistical methods. *Handbook of Methodological Innovation, Sage*, 2009. BP08007.

[968] Carpenter, J.R., Schwarzer, G., Rucker, G. and R. Kunstler, Empirical evaluation shows the Copas selection model provides a useful summary in 80% of meta-analyses. *Journal of Clinical Epidemiology*, 2009. RP08056.

[969] Goldstein, H., Carpenter, J.R., Kenward, M.G., and K. Levin, Multilevel models with multivariate mixed response types. *Statistical Modelling*, 2008. RP08028.

[970] Mallinckrodt, C.H. and M.G. Kenward, Conceptual considerations regarding choice of endpoints, hypotheses and analyses in longitudinal clinical trials?. *Drug Information Journal*, 2008. RP08026.

[971] Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M.R.P., Kenward, M.G., Wood, A.M. and J.R. Carpenter, Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*, 2009. RP08114.

[972] Vansteelandt, S., Carpenter, J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 2008. RP08027.

C.9.4 Non-refereed publications (published)

[973] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Shared-parameter models and missingness at random. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18*, 2008. NR08023.

[974] Molenberghs, G., Aerts, M., Beunckens, C., Creemers, A., Hens, N., Kenward, M.G., Sotto, C. and G. Verbeke, Missing not at random models and their missing at random counterparts, in various modelling frameworks. *ACTAS do XVI Congresso da SPE, UTAD, Vila REAL*, 2008. NR08026.

C.10 List of joint publications

C.10.1 Technical reports

[975] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties, 2008. TR08050. (KUL-1, UJF)

[976] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, 2008. TR08035. (UCL, KUL-1)

[977] Creemers A., Hens N., Aerts M., Molenberghs G., Verbeke G. and M.G. Kenward, A sensitivity analysis for shared-parameter models, 2008. TR08058.(LSHTM, UH, KUL-2)

[978] Crujeiras, R. M. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends, 2008. TR08044. (UCL, USC)

[979] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015. (UG, UH)

[980] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Goodness-of-fit tests for censored regression based on artificial data points, 2008. TR08060. (UCL, USC)

[981] González-Manteiga, W., Pardo-Fernandez, J. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models, 2008. TR08045. (UCL, USC)

[982] Kenward M.G. and G. Molenberghs, Last observation carried forward: a crystal ball?, 2008. TR08059. (LSHTM, UH)

[983] Molanes Lopez, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions, 2008. TR08023. (UCL, UH)

[984] San Martin, E., Jara, A., Rolin, J.M. and M. Mouchart, On the analysis of Bayesian semiparametric IRT-type models, 2008. TR08029. (UCL, KUL-2)

[985] Sotto C., Beunckens C., Molenberghs G. and M.G. Kenward, MCMC-based estimation methods for continuous longitudinal data with non-random (non)-monotone missingness, 2008. TR08057. (LSHTM, UH)

[986] Veraverbeke, N., Omelka, M. and I. Gijbels, Weak convergence of the empirical copula process with covariates, 2008. TR08056. (KUL-1, UH)

C.10.2 Refereed publications (published)

[987] Beunckens, C., Molenberghs, G., Verbeke, G. and C. Mallinckrodt, A latent-class mixture model for incomplete longitudinal Gaussian data. *Biometrics*, **64**, 96-105, 2008. R08102. (UH, KUL-2)

[988] Cadarso Suárez, C.M., Faes, C. and G. Molenberghs, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association*, **103**, 140-161, 2008. R08043. (USC, UH)

[989] Faes, C., Molenberghs, G., Geys, H., Aerts, M., Cadarso-Súarez, C., Acuña, C. and M. Cano, A flexible method to measure synchrony in neuronal firing. *Journal of the American Statistical Association, Case Studies*, **103**, 140-161, 2008. R08059. (UH, USC)

[990] Gao, J., Gijbels, I. and S. Van Bellegem, Nonparametric simultaneous testing for structural breaks. *Journal of Econometrics*, **143**, 123-142, 2008. R08005. (UCL, KUL-1)

[991] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences. *Journal of Applied Statistics*, **35**, 1071-1079, 2008. R08230. (UG, UH)

[992] González-Manteiga, W., Heuchenne, C. and C. Sánchez Sellero, Parametric conditional mean and variance testing with censored data. *H. Skiadas, Christos, ed. Recent Advances in Applied Stochastic Models and Data Analysis. N.p.: World Scientific Publishing Co. Pte Ltd*, 2008. R08104. (UCL, USC)

[993] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, **23**, 92-119, 2008. R08203. (UG, KUL-1)

[994] Hughes, S., Keene, O., Roes, K., Ashby, D., Beltangadu, M., Bird, S.M., Burton, C., Buyse, M., Day, S., van Ewijk, P., Fletcher, C., Grieve, A.P., Guinot, C., Hemmings, R., Hold, D., Hothorn, L., Jagers, P., Keiding, N., Korhonen, P., Lesaffre, E., Lewis, J.A., Molenberghs, G., Neumann, N., Olsen, K.J., van Osta, G., Pocock, S., Rockhold, F., Rossi, A., Senn, S., Stijnen, T., Vaeth, M., Wiklund, S.J. and J. Wolfram, Open Letter to the Heads of the European Regulatory Agencies. *British Medical Journal*, **336**, 250-250, 2008. R08145. (UH, KUL-2)

[995] 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. *Journal of the Royal Statistical Society, Series B*, **70**, 371-388, 2008. R08073. (LSHTM, UH)

[996] Molenberghs, G., Verbeke, G. and M.G. Kenward, Sensitivity analysis for incomplete data. *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods. Chapter 22. Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, eds. Chapman & Hall/CRC*, 501-552, 2008. R08242. (LSHTM, UH, KUL-2)

[997] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63-74. R08019. (UH, KUL-2)

[998] Van Keilegom, I., González-Manteiga, W. and C. Sánchez Sellero, Goodness-of-fit tests in parametric regression based on the estimation of the error distribution. *TEST*, **17**, 401-415, 2008. R08015. (UCL, USC)

[999] Van Keilegom, I., Sánchez Sellero, C. and W. González-Manteiga, Empirical likelihood based testing for regression. *Electronic Journal of Statistics*, **2**, 581-604, 2008. R08117. (UCL, USC)

[1000] Verbeke, G., Molenberghs, G. and C. Beunckens, Formal and informal model selection with incomplete data. *Statistical Science*, **23**, 201-218, 2008. R08099. (UH, KUL-2)

C.10.3 Refereed publications (in press)

[1001] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of spatially and temporally heterogeneous brain response images. *Journal of Computational and Graphical Statistics*, 2008. RP08001. (UJF, UCL)

[1002] Antoniadis, A., Gijbels, I. and J.M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, 2008. RP08038.(UJF, KUL-1)

[1003] Beirlant, J., Joossens, E. and J. Segers, Second order refined Peaks-over-Threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, 2009. RP08039. (KUL-1, UCL)

[1004] Beunckens C., Sotto C., Molenberghs G. and G. Verbeke, An integrated sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08030. (UH, KUL-2)

[1005] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multi-faceted sensitivity analysis of the Slovenian public opinion survey data. *Applied Statistics*, 2008. RP08081. (UH, KUL-2)

[1006] Braekers, R. and I. Van Keilegom, Flexible modeling base on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, 2008. RP08090. (UH, UCL)

[1007] Crujeiras, R. and W. González-Manteiga, Comments on: Augmenting the bootstrap to analyze high dimensional genomic data. *TEST*, **17**, 40-42, 2008. R08075. (USC, UCL)

[1008] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, An L2-test for comparing spatial spectral densities. *Statistics and Probability Letters*, RP08011. (USC, UCL)

[1009] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, Goodness-of-fit tests for the spatial spectral density. *Stochastic Environmental Research and Risk Assesment*, RP08015. (USC, UCL)

[1010] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology.* RP08112. (KUL-1, UH)

[1011] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural and Biological Environmental Statistics*, 2008. RP08128. (UH, UG)

[1012] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, 2008. RP08007. (UU, UCL)

[1013] Legrand, C., Duchateau, L. Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis*, 2008/2009. RP08100.(UH, UCL, UG)

[1014] Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso-Suárez, C., Pardo-Vázquez, J.L., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08079. (UH, USC)

[1015] Molanes, E.M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, 2008. RP08068. (UH, UCL)

[1016] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, 2008. RP08047. (UH, KUL-1)

[1017] Pardo Vázquez, J.L., Cadarso Suárez, C.M., Molenberghs, G., Geys, H., Faes, C., Maringwa, J.T., Leborán, V. and C. Acuña, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2008. RP08021. (USC, UH)

[1018] Porcu, E., Crujeiras, R., Mateu, J. and W. González-Manteiga, On the second order properties of the multidimensional peridogram for regularly spaced data. *Theory of Probability and its Applications*, RP08008. (USC, UCL)

[1019] Sotto, C., Beunckens, C., Molenberghs G., Jansen I., and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, 2008, RP08031. (KUL-2, UH)

[1020] Vansteelandt S., Carpenter J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, 2008. RP08027. (LSHTM, UG)

[1021] Verbeke G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling*, 2009. RP08033. (KUL-2, UH)

C.10.4 Non-refereed publications (published)

[1022] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Shared-parameter models and missingness at random. *Proceedings of the 24th International Biometric Conference, Dublin, July 13-18*, 2008. NR08023. (UH, LSHTM, KUL-2)

[1023] Massonnet, G., Janssen, P. and L. Duchateau, Using copulas to model four-dimensional udder enfection data. *Programme and abstract book 29th Annual conference of the International Society for Clinical Biostatistcs, Copenhagen, Denmark*, 2008. NR08028. (UH, UG)

[1024] Molenberghs, G., Aerts, M., Beunckens, C., Creemers, A., Hens, N., Kenward, M.G., Sotto, C. and G. Verbeke, Missing not at random models and their missing at random counterparts, in various modelling frameworks. *ACTAS do XVI Congresso da SPE, UTAD, Vila REAL*, 2008. NR08026. (UH, KUL-2, LSHTM)

C.10.5 Books (published)

[1025] Duchateau, L. and P. Janssen, *The Frailty Model*. Springer, New York, 2008. B08001. (UH, UG)

C.10.6 Books (in press)

[1026] Davidian, M., Fitzmaurice, G., Verbeke, G. and G. Molenberghs, *Longitudinal Data Analysis, A Handbook of Modern Statistical Methods*. Chapman and Hall/CRC, New York, 2009. BP08011. (UH, KUL-2)

D Publications of the network in 2009-2010

D.1 Université catholique de Louvain, UCL

D.1.1 Technical reports

[1027] Basrak, B., Krizmanic, D. and J. Segers, A functional limit theorem for partial sums of dependent random variables with infinite variance, 2010. TR10001.

[1028] Breuning, C. and J. Johannes, Minimax-optimal local nonparametric instrumental regression, 2010. TR10007.

[1029] Çetinyürek, A. and P. Lambert, Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized B-splines, 2009. TR09039.

[1030] Christiansen, M. and M. Denuit, First-order mortality rates and safe-side actuarial calculations in life insurance, 2010. TR10012.

[1031] Daouia, A., Florens, J.-P. and L. Simar, Regularization of nonparametric frontier estimators, 2009. TR09025.

[1032] Davydov, Y. and S. Liu, Transformations of multivariate regularly varying tail distributions, 2010. TR10011.

[1033] Degena, M., Lambriggerb, D.D. and J. Segers, Risk concentration and diversification: second-order properties, 2009. TR09028.

[1034] Denuit, M., Positive dependence of signals, 2010. TR10028.

[1035] Denuit, M. and L. Eeckhoudt, A general index of absolute risk attitude, 2010. TR10016.

[1036] Denuit, M. and L. Eeckhoudt, Bivariate stochastic dominance and common preferences of decision-makers with risk independent utilities, 2010. TR10021.

[1037] Denuit, M., Eeckhoudt, L. and O. Jokung, Transformations preserving stochastic dominance: Theory and applications, 2010. TR10027.

[1038] Denuit, M., Eeckhoudt, L. and M. Menegatti, Correlated risks, bivariate utility and optimal choices, 2010. TR10018.

[1039] Denuit, M., Eeckhoudt, L. and M. Menegatti, Adding independent risks in an insurance portfolio: Which shape for the insurer's preferences?, 2010. TR10023.

[1040] Denuit, M., Eeckhoudt, L. and B. Rey, Somes consequences of correlation aversion in decision science, 2010. TR10020.

[1041] Denuit, M., Haberman, S. and A. Renshaw, Comonotonic approximations to quantiles of life annuity conditional expected present values: extensions to general ARIMA models and comparison with the bootstrap, 2010. TR10014.

[1042] Denuit, M., Haberman, S. and A. Renshaw, Longevity-indexed life annuities, 2010. TR10026.

[1043] Denuit, M. and M. Mesfioui, Generalized increasing convex and directionally convex orders, 2010. TR10015.

[1044] Denuit, M. and B. Rey, Transformations of multivariate regularly varying tail distributions, 2010. TR10024.

[1045] Dette, H. and C. Heuchenne, Scale checks in censored regression, 2009. TR09086.

[1046] Eeckhoudt, L. and M. Denuit, Stronger measures of higher-order risk attitudes, 2010. TR10013.

[1047] Genest, C. and J. Segers, On the asymptotic covariance of the empirical copula process, 2009. TR09010.

[1048] Gudendorf, G. and J. Segers, Nonparametric estimation of an extreme-value copula in arbitrary dimensions, 2009. TR09026.

[1049] Guillotte, S., Perron, F. and J. Segers, Nonparametric Bayesian inference on bivariate extremes, 2009. TR09033.

[1050] Hafner, C., Laurent, S. and F. Violante, The diffusion limit of dynamic conditional correlation models, 2010. TR10004.

[1051] Hafner, C. and O.B. Linton, Efficient estimation of a multivariate multiplicative volatility model, 2009. TR09004.

[1052] Hafner, C. and H. Manner, Dynamic stochastic copula models: Estimation, inference and applications, 2009. TR09072.

[1053] Hafner, C. and O. Reznikova, On the estimation of dynamic conditional correlation models, 2010. TR10005.

[1054] Heuchenne, C. and I. Van Keilegom, Estimation of a general parametric location in censored regression, 2010. TR10010.

[1055] Hollevoet, K., Kellen, E., Thimpont, J., Germonpre, P., De Vuyst, P., Bosquee, L., Legrand, C., Kishi, Y., Nackaerts, K., Delanghe, J.R. and J.P. Van Meerbeeck, Diagnostic performance of soluble mesothelin and megakaryocyte potentiating factor as biomarkers of mesothelioma, 2009. TR09031.

[1056] Hunt, J., Equivalent local martingale measures and market incompleteness in a continuous time semi-Markov regime switching model, 2009. TR09027.

[1057] Johannes, J. and R. Schenk, On rate optimal local estimation in functional linear model, 2010. TR10006.

[1058] Johannes, J. and M. Schwarz, Adaptive circular deconvolution by model selection under unknown error distribution, 2009. TR09034.

[1059] Johannes, J. and M. Schwarz, Adaptive estimation in nonparametric instrumental regression, 2010. TR10008.

[1060] Kneip, A., Simar, L. and P.W. Wilson, A computationally efficient, consistent bootstrap for inference with non-parametric DEA estimators, 2009. TR09001.

[1061] Lambert, P., Smooth semi- and nonparametric Bayesian estimation of bivariate densities from bivariate histogram data, 2009. TR09038.

[1062] Laurent, S., On standardness and I-cosiness, 2009. TR09017.

[1063] Laurent, S., On Vershikian and I-cosy random variables and filtrations, 2009. TR09018.

[1064] Laurent, S., Some Poisson mixtures distributions with a hyperscale parameter, 2009. TR09035.

[1065] Laurent, S. and C. Legrand, A Bayesian framework for the ratio of two Poisson rates in the context of vaccine efficacy trials, 2009. TR09015.

[1066] Manner, H. and J. Segers, Tails of correlation mixtures of elliptical copulas, 2009. TR09087.

[1067] Manner, H. and O. Reznikova, Time-varying copulas : a survey, 2009. TR09020.

[1068] Meinguet, T. and J. Segers, Regularly varying time series in Banach spaces, 2010. TR10002.

[1069] Meyer, N., Legrand, C. and G. Giaccone, Sample size in randomized clinical trials in oncology: usefull basic statistical knowledge for the clinician, 2009. TR09032.

[1070] Molanes Lopez, E. M., Cao, R. and I. Van Keilegom, Smoothed empirical likelihood confidence intervals for the relative distribution with left truncated and right censored data, 2009. TR09014.

[1071] Peng, L., Qi, Y. and I. Van Keilegom, Jackknife empirical likelihood method for copulas, 2009. TR09037.

[1072] Pigeon, M. and M. Denuit, Composite Lognormal-Pareto model with random threshold, 2010. TR10017.

[1073] Roueff, F. and R. Von Sachs, Locally stationary long memory estimation, 2009. TR09019.

[1074] Rousseau, R., Govaerts, B. and M. Verleysen, Combination of independent component analysis and statistical modeling for the identification of metabonomic biomarkers in 1H-NMR spectroscopy, 2009. TR09073.

[1075] Schubert, T. and L. Simar, Innovation and export activities in the German mechanical engineering sector: an application of testing restrictions in production analysis, 2009. TR09024.

[1076] Simar, L. and P.W. Wilson, Inference by subsampling in nonparametric frontier models, 2009. TR09036.

[1077] Simar, L. and P.W. Wilson, Inference by subsampling in nonparametric frontier models: Appendix, 2009. TR09036a.

[1078] Surmont, V., Aerts, J.G.J.V., Van Klaveren, R.J., Tournoy, K., Tan, K., Vernhout, R.M., Schmitz, P.I.M, Legrand, C., Hoogesteden, H.C. and J.P. Van Meerbeeck, A randomized phase II study of Gemcitabine and Carboplatin in advanced NSCLC comparing two schedules of the 21-day regimen of Gemcitabine and Carboplatin in advanced NSCLC, 2009. TR09030.

[1079] Trufin, J., Albrecher, H. and M. Denuit, Properties of risk measures derived from ruin theory, 2010. TR10022.

[1080] Trufin, J., Albrecher, H. and M. Denuit, Ruin problems in presence of underwriting cycles, 2010. TR10019.

[1081] Trufin, J., Albrecher, H. and M. Denuit, Ruin problems under IBNR dynamics, 2010. TR10025.

D.1.2 Refereed publications (published)

[1082] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of functional images. *Journal of Computational and Graphical Statistics*, **18**, 1, 216-237, 2009. R09231.

[1083] Badin, M., Daraio, C. and L. Simar, Optimal bandwidth selection for conditional efficiency measures: a data-driven approach. *European Journal of Operational Research*, **201**, 633-640, 2010. R10002.

[1084] Badin, L. and L. Simar, A bias corrected nonparametric envelopment estimator of frontiers. *Econometric Theory*, **25**, 5, 1289-1318, 2009. R09200.

[1085] Basrak, B. and J. Segers, Regularly varying multivariate time series. *Stochastic Processes and their Applications*, **119**, 1055-1080, 2009. R09006.

[1086] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, **139**, 2800-2815, 2009. R09026.

[1087] Bodin, N., Govaerts, B., Abboudi, T., Detavernier, C., De Saege, S., Larondelle, Y. and X. Rollin, Diet protein level affects the lysine requirement and retention efficiency at marginal lysine intakes in rainbow trout (Oncorhynchus mykiss) fry. *British Journal of Nutrition*, **102**, 37-53, 2009. R09001.

[1088] Böhm, H. and R. von Sachs, Shrinkage estimation in the frequency domain of multivariate time series. *Journal of Multivariate Analysis*, **100**, 913-935, 2009. R09003.

[1089] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, **100**, 1270-1281, 2009. R09160.

[1090] Charpentier, A. and J. Segers, Tails of multivariate Archimedean copulas. *Journal of Multivariate Analysis*, **100**, 1521-1537, 2009. R09010.

[1091] Chen, S.X. and I. Van Keilegom, A goodness-of-fit test for parametric and semi-parametric models in multiresponse regressions. *Bernoulli*, **15**, 4, 955-976, 2009. R09027.

[1092] Chen, S.X. and I. Van Keilegom, A review on empirical likelihood methods for regression. *TEST*, **18**, 415-447, 2009. R09024.

[1093] Chen, S.X. and I. Van Keilegom, Rejoinder on: A review on empirical likelihood methods for regression. *TEST*, **18**, 468-474, 2009. R09025.

[1094] Crujeiras, R. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends. *Computational Statistics and Data Analysis*, **54**, 452-465, 2010. R10004.

[1095] Debrus, B., Lebrun, P., Ceccato, A., Govaerts, B., Olsen, B., Rozet, E., Boulanger, B. and P. Hubert, A new statistical method for the automated detection of peaks in UV-DAD chromatograms of a sample mixture. *Talanta*, **79**, 77-85, 2009. R09002.

[1096] Dette H., Pardo-Fernandez, J.C. and I. Van Keilegom, Goodness-of-fit tests for multiplicative models with dependent data. *Scandinavian Journal of Statistics*, **36**, 782-799, 2009. R09022.

[1097] Einmahl, H.J. and J. Segers, Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, **37**, 2953-2989, 2009. R09015.

[1098] El Ghouch, A. and I. Van Keilegom, Local linear quantile regression with dependent censored data. *Statistica Sinica*, **19**, 1621-1640, 2009. R09023.

[1099] Genest, C. and J. Segers, Rank-based inference for bivariate extreme-value copulas. *The Annals of Statistics*, **37**, 5B, 2990-3022, 2009. R09014.

[1100] Hafner, C., Causality and forecasting of temporally aggregated multivariate GARCH model. *The Econometrics Journal*, **12**, 127-146, 2009. R09226.

[1101] Hafner, C. and P. Franses, A generalized dynamic conditional correlation model: Simulation and application to many assets. *Econometric Reviews*, **28**, 612-631, 2009. R09227.

[1102] Hafner, C. and H. Herwartz, Testing for linear vector autoregressive dynamics under multivariate generalized autoregressive heteroskedasticity. *Statistica Neerlandica*, **63**, 294-323, 2009. R09228.

[1103] Hafner, C. and A. Preminger, Asymptotic theory for a factor GARCH model. *Econometric Theory*, **25**, 336-363, 2009. R09229.

[1104] Hafner, C. and A. Preminger, Deciding between GARCH and stochastic volatility via strong decision rules. *Journal of Statistical Planning and Inference*, **140**, 790-805, 2010. R10008.

[1105] Hafner, C. and A. Preminger, On asymptotic theory of multivariate GARCH models. *Journal of Multivariate Analysis*, **100**, 2044-2054, 2009. R09230.

[1106] Hall, P. and I. Van Keilegom, Nonparametric "regression" when errors are positioned at end-points. *Bernoulli*, **15**, 3, 614-633, 2009. R09020.

[1107] Hjort, N.L., McKeague, I.W. and I. Van Keilegom, Extending the scope of empirical likelihood. *The Annals of Statistics*, **37**, 3, 1079-1111, 2009. R09017.

[1108] Jeong, S.O., Park, B.U. and L. Simar, Nonparametric conditional efficiency measures: asymptotic properties. *Annals of Operations Research*, **173**, 105-122, 2010. R10011.

[1109] Jullion, A., Lambert, P., Beck, B. and F. Vandenhende, Pharmacokinetic parameters estimation using adaptive Bayesian P-splines models. *Pharmaceutical Statistics*, **8**, 98-112, 2009. R09008.

[1110] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, **53**, 1388-1399, 2009. R09007.

[1111] Laurent, S., Further comments on the representation problem for stationary processes. *Statistics and Probability Letters*, **80**, 592-596, 2010. R10013.

[1112] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model. *Lifetime Data Analysis*, **15**, 59-78, 2009. R09004.

[1113] Molanes Lopez, E. M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, **36**, 413-432, 2009. R09019.

[1114] Mouchart, M. and M. Vandresse, A double-frontier approach for measuring market imperfection. *Special Issue of Annals of Operations Research on Efficiency and Productivity*, **173**, 137-144, 2010. R10010.

[1115] Mouchart, M., Russo, F. and G. Wunsch, Structural modelling, exogeneity, and causality. In: *Causal Analysis in Population Studies. Concepts, Methods, Applications*, Engelhardt, H., Kohler, H-P., Fürnkranz-Prskawetz, A. (eds.), 59-82, Springer, 2009. R09011.

[1116] Neumeyer, N. and I. Van Keilegom, Change-point tests for the error distribution in non-parametric regression. *Scandinavian Journal of Statistics*, **36**, 518-541, 2009. R09016.

[1117] Ojeda Cabrera, J. L. and I. Van Keilegom, Goodness-of-fit tests for parametric regression with selection biased data. *Journal of Statistical Planning and Inference*, **139**, 2836-2850, 2009. R09018.

[1118] Omrane, B. and C. Hafner, Information spillover, volatility, and the currency markets, International. *Econometric Review*, **1**, 47-59, 2009. R09225.

[1119] Robert, C.Y., Segers, J. and C.A.T., Ferro, A sliding blocks estimator for the extremal index. *Electronic Journal of Statistics*, **3**, 993-1020, 2009. R09021.

[1120] San Martin, E. and M. Mouchart, On joint completeness: sampling and Bayesian versions, and their connections. *Sankhya*, **69**, 780-807, 2009. R09012.

[1121] Schwarz, M. and S. van Bellegem, Consistent density deconvolution under partially known error distribution. *Statistics and Probability Letters*, **80**, 236-241, 2010. R10009.

[1122] Simar, L. and P.W. Wilson, Inference from cross-sectional stochastic frontier models. *Econometric Review*, **29**, 1, 62-98, 2010. R10012.

[1123] Timmermans C., von Sachs, R. and V. Delouille, Comparaison et classification de séries temporelles via leur développement en ondelettes de Haar asymétriques, *Actes des XVIemes Rencontres de la Société Francophone de Classification* , 57-60, 2009. R09009.

[1124] Van Keilegom, I. and L. Wang, Semiparametric modeling and estimation of heteroscedasticity in regression analysis of cross-sectional data. *Electronic Journal of Statistics*, **4**, 133-160, 2010. R10001.

[1125] Vriamont, N., Govaerts, B., Grenouillet, P., de Bellefon, C. and O. Riant, Design of a genetic algorithm for the simulated evolution of a library of asymmetric transfer hydrogenation catalysts. *Chemistry - A European Journal*, **15**, 6267-6278, 2009. R09013.

D.1.3 Refereed publications (in press)

[1126] Böhm, H., Ombao, H., von Sachs, R. and J. Sanes, Classification of multivariate nonstationary signals : The SLEX-shrinkage approach. *Journal of Statistical Planning and Inference* (to appear), 2010. RP10078.

[1127] Comte, F. and J. Johannes, Adaptive estimation in circular functional linear models. *Mathematical Methods of Statistics* (to appear), 2010. RP10066.

[1128] Daouia, A., Florens, J.P. and L. Simar, Frontier estimation and extreme values theory. *Bernoulli* (to appear), 2010. RP10076.

[1129] Daskovska, A., Simar, L. and S. Van Bellegem, Forecasting the Malmquist productivity index. *Journal of Productivity Analysis* (to appear), 2010. RP10073.

[1130] Degen, M., Lambrigger, D.D. and J. Segers, Risk concentration and diversification: Second-order properties. *Insurance: Mathematics and Economics* (to appear), 2009. RP09094.

[1131] El Ghouch, A., Van Keilegom, I. and I. McKeague, Empirical likelihood confidence intervals for dependent duration data. *Econometric Theory* (to appear), 2010. RP10002.

[1132] Ferraty, F., Van Keilegom, I. and P. Vieu, On the validity of the bootstrap in nonparametric functional regression. *Scandinavian Journal of Statistics* (to appear), 2010. RP09005.

[1133] Florens, J.P., Johannes, J. and S. Van Bellegem, Identification and estimation by penalization in nonparametric instrumental regression. *Econometric Theory* (to appear), 2010. RP10063.

[1134] Freyermuth, J.-M., Ombao, H. and R. von Sachs, Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *Journal of the American Statistical Association* (to appear), 2010. RP10079.

[1135] Geenens, G. and L. Simar, Nonparametric test for conditional independence in two-way contingency tables. *Journal of Multivariate Analysis* (to appear), 2010. RP10075.

[1136] Geenens, G. and L. Simar, Single-index modelling of conditional probabilities in two-way contingency tables. *Statistics* (to appear), 2010. RP10077.

[1137] González Manteiga, W., Pardo Fernández, J.C. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models. *Scandinavian Journal of Statistics* (to appear), 2010. RP10009.

[1138] Gudendorf, G. and J. Segers, Extreme-value copulas. *Workshop on Copula Theory and its Applications.* F. Durante, W. Haerdle, P. Jaworski and T. Rychlik, (eds.), Lecture Notes in Statistics - Proceedings. Springer (to appear), 2010. RP10081.

[1139] Hafner, C. and O. Reznikova, Efficient estimation of a semiparametric dynamic copula model. *Computational Statistics and Data Analysis* (to appear), 2010. RP10062.

[1140] Heuchenne, C. and I. Van Keilegom, Estimation in nonparametric location-scale regression models with censored data. *Annals of the Institute of Statistical Mathematics* (to appear), 2010. RP10003.

[1141] Heuchenne, C. and I. Van Keilegom, Goodness-of-fit tests for the error distribution in nonparametric regression. *Computational Statistics and Data Analysis* (to appear), 2010. RP10010.

[1142] Johannes, J. and S. Subba Rao, Nonparametric estimation for dependent data. *Journal of Nonparametric Statistics* (to appear), 2010. RP10065.

[1143] Johannes, J., Van Bellegem, S. and A. Vanhems, Convergence rates for ill-posed inverse problems with an unknown operator. *Econometric Theory* (to appear), 2010. RP10064.

[1144] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis (to appear)*, 2009. RP09001.

[1145] Linton, O., Mammen, E., Nielsen, J.P. and I. Van Keilegom, Nonparametric regression with filtered data. *Bernoulli* (to appear), 2010. RP10012.

[1146] Motta, G., Hafner, C. and R. von Sachs, Locally stationary factor models : identification and nonparametric estimation. *Econometric Theory* (to appear), 2010. RP10080.

[1147] Mouchart, M. and F. Russo, Causal explanation: recursive decompositions and mechanisms. *Causality in the sciences*, McKay Illari, P., Russo, F. and J. Williamson (eds.), Oxford University Press (to appear), 2010. RP10071.

[1148] Neumeyer, N. and I. Van Keilegom, Estimating the error distribution in nonparametric multiple regression with applications to model testing. *Journal of Multivariate Analysis* (to appear), 2010. RP10011.

[1149] Omey, E. and J. Segers, Generalised regular variation of arbitrary order. *Banach Center Publications* (to appear), 2009. RP09093.

[1150] Park, B.U., Jeong, S.-O. and L. Simar, Asymptotic distribution of conical-hull estimators of directional edges. *Annals of Statistics* (to appear), 2010. RP10074.

[1151] Simar, L. and V. Zelenyuk, Stochastic FDH/DEA estimators for frontier analysis. *Journal of Productivity Analysis* (to appear), 2010. RP10081.

[1152] Teodorescu, B. and I. Van Keilegom, Goodness-of-fit test in generalized conditional linear models under left truncation and right censoring. *Journal of Nonparametric Statistics* (to appear), 2010. RP10006.

[1153] Teodorescu, B., Van Keilegom, I. and R. Cao, Generalized conditional linear models under left truncation and right censoring. *Annals of the Institute of Statistical Mathematics* (to appear), 2010. RP10004.

[1154] Varron, D. and I. Van Keilegom, Uniform in bandwidth exact rates for a class of kernel estimators. *Annals of the Institute of Statistical Mathematics* (to appear), 2010. RP10008.

[1155] Wunsch, G., Russo, F. and M. Mouchart, Do we necessarily need longitudinal data to infer causal relations? *Bulletin de Méthodologie Statistique* **33**, (to appear), 2010. RP10072.

D.1.4 Non-refereed publications (in press)

[1156] Hafner, C. and H. Manner, Multivariate time series models for asset prices. *Handbook of Computational Finance*, Springer (to appear), 2010. NRP10002.

[1157] Hafner, C., GARCH modelling, In: *Encyclopedia of Complexity and System Science*. R.A. Meyers (ed.), Springer (to appear), 2010. NRP10001.

D.2 Katholieke Universiteit Leuven, KUL-1

D.2.1 Technical reports

[1158] Antoniadis, A., Gijbels, I. and A. Verhasselt, Variable selection in additive models using P-splines, 2009. TR09068.

[1159] Braeken, J., Tuerlinckx, F., Kuppens, P. and P. De Boeck, Coupled error terms in structural equation models for categorical data, 2009. TR09040.

[1160] Casas, I. and I. Gijbels, Unstable volatility: the break preserving local linear estimator, 2009. TR09067.

[1161] Ceulemans, E., Kuppens, P. and I. Van Mechelen, Capturing the structure of distinct types of individual differences in the situation-specific experience of anger, 2009. TR09041.

[1162] Ceulemans, E., Timmerman, E. and H.A.L. Kiers, A numerical convex hull based procedure for selecting among multilevel component solutions, 2009. TR09042.

[1163] Daouia, A. and I. Gijbels, Estimating frontier cost models using extremiles, 2009. TR09066.

[1164] Daouia, A. and I. Gijbels, Robustness and inference in nonparametric partial-frontier modeling, 2009. TR09065.

[1165] Depril, D. and I. Van Mechelen, Additive biclustering: A comparison of one new and two existing algorithms, 2009. TR09045.

[1166] Depril, D. and I. Van Mechelen, Identifiability of additive clustering models, 2009. TR09046.

[1167] Depril, D. and I. Van Mechelen, Lowdimensional overlapping clustering, 2009. TR09044.

[1168] Depril, D. and I. Van Mechelen, New ALS approaches for fitting the INDCLUS model, 2009. TR09043.

[1169] De Roover, K., Ceulemans E., Timmerman, M. E., Vansteelandt, K., Stouten, J. and P. Onghena, Clusterwise SCA-ECP for the analysis of structural differences in multivariate multilevel data, 2009. TR09047.

[1170] Desmet, L. and I. Gijbels, Curve fitting under jump and peak irregularities using local linear regression, 2009. TR09063.

[1171] Desmet, L. and I. Gijbels, Improved spectral density estimation at peaks, 2009. TR09064.

[1172] Gijbels, I., Omelka, M. and D. Sznajder, Positive quadrant dependence tests for copulas, 2009. TR09061.

[1173] Gijbels, I., Veraverbeke, N. and M. Omelka, Conditional copulas, association measures and their applications, 2009. TR09060.

[1174] Hubert, M., Rousseeuw, P.J. and T. Verdonck, A deterministic algorithm for the MCD, 2009. TR09059.

[1175] Hubert, M. and S. Van der Veeken, Robust classification for skewed data, 2009. TR09058.

[1176] Magis, D., Béland, S., Tuerlinckx, F. and P. De Boeck, A general framework and an R package for the detection of dichotomous differential item functioning, 2009. TR09050.

[1177] Magis, D. and P. De Boeck, A robust statistical approach to differential item functioning, 2009. TR09048.

[1178] Magis, D. and P. De Boeck, An outlier nonparametric based approach to identify uniform differential item functioning: comparison of several statistical procedures, 2009. TR09049.

[1179] Oravecz, Z. and F. Tuerlinckx, The linear mixed model and the hierarchical Ornstein-Uhlenbeck model: Some equivalences and differences, 2009. TR09051.

[1180] Schepers, J. and I. Van Mechelen, Capturing the nature of two-way interactions by means of a two-mode clustering method, 2009. TR09052.

[1181] Vandekerckhove, J., Tuerlinckx, F. and M.D. Lee, Hierarchical diffusion models for two-choice response times, 2009. TR09053.

[1182] van den Berg, R.A., Braaksma, M., van der Veen, D., van der Werf, M.J., Punt, P.J., van der Oost, J. and L.H. de Graaff, Identification of modules in Aspergillus niger by gene co-expression network analysis, 2009. TR09054.

[1183] Van Mechelen, I. and A.K. Smilde, A generic model for data fusion, 2009. TR09055.

[1184] Veraverbeke, N., Omelka, M. and I. Gijbels, Estimation of a conditional copula and association measures, 2009. TR09057.

[1185] Wilderjans, T.F., Ceulemans, E. and I. Van Mechelen, Simultaneous analysis of coupled data matrices subject to different amounts of noise, 2009. TR09056.

D.2.2 Refereed publications (published)

[1186] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, Special issue: Multiscale Methods and Statistics: A Productive Marriage, **19**, 1371-1387, 2009. R09048.

[1187] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, **139**, 2800-2815, 2009. R09026.

[1188] Braeken, J. and F. Tuerlinckx, A mixed model framework for teratology studies. *Biostatistics*, **10**, 744-755, 2009. R09029.

[1189] Braeken, J. and F. Tuerlinckx, Investigating latent constructs with item response models: A Matlab IRTm Toolbox. *Behavior Research Methods*, **41**, 1127-1137, 2009. R09030.

[1190] Ceulemans, E. and H.A.L. Kiers, Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical and Statistical Psychology*, **62**, 601-620, 2009. R09031.

[1191] Chiarandini, M. and Y. Goegebeur, Mixed models for the analysis of optimization algorithms. In: *Empirical Methods for the Analysis of Optimization Algorithms*. Bartz-Beielstein, T., Chiarandini, M., Paquete, L and M. Preuss (eds.), Berlin: Springer, 2009. R09028.

[1192] Corcuera, J.M., Guillaume, F., Leoni, P. and W. Schoutens, Implied Levy volatility. *Quantitative Finance*, **4**, 383-393, 2009. R09049.

[1193] De Brabanter, K., Pelckmans, K., De Brabanter, J., Debruyne, M., Suykens, J.A.K., Hubert, M. and B. De Moor, Robustness of kernel based regression: a comparison of iterative weighting schemes. *Proceedings of the 19th International Conference on Artificial Neural Networks (ICANN)*, 2009. R09050.

[1194] Debruyne, M. and M. Hubert, The influence function of the Stahel-Donoho covariance estimator of smallest outlyingness. *Statistics and Probability Letters*, **79**, 275-282, 2009. R09051.

[1195] Desmet, L. and I. Gijbels, Local linear fitting and improved estimation near peaks. *The Canadian Journal of Statistics*, **37**, 453-475, 2009. R09052.

[1196] Dierckx, G., Beirlant, J., De Waal, D. and A. Guillou, A new estimation method for Weibull-type tails based on the mean excess function. *Journal of Statistical Planning and Inference*, **139**, 6, 1905-1920, 2009. R09053.

[1197] Dutilh, G., Vandekerckhove, J., Tuerlinckx, F. and E.-J. Wagenmakers, A diffusion model decomposition of the practice effect. *Psychonomic Bulletin and Review*, **16**, 1026-1036, 2009. R09032.

[1198] Frederickx, S., Kuppens, P., Tuerlinckx, F. and I. Van Mechelen, Desequens: an R-package for the variance decomposition of sequential processes. *Behavior Research Methods*, **41**, 524-530, 2009. R09033.

[1199] Garcia, J., Goossens, S., Masol, V. and W. Schoutens, Lévy Base Correlation. *Wilmott Journal*, **1**, 2, 95-100, 2009. R09054.

[1200] Gijbels, I. and M. Hubert, Robust and nonparametric statistical methods. In: *Encyclopedia of Comprehensive Chemometrics*, Brown, S., Tauler, R., and R. Walczak (eds.), **1**, 189-211, Elsevier, Oxford, 2009. R09055.

[1201] González, J., Tuerlinckx, F. and P. De Boeck, Analyzing structural relations in multivariate dyadic binary data. *Applied Multivariate Research*, **13**, 77-92, 2009. R09034.

[1202] Guillaume, F., Jacobs, P. and W. Schoutens, Pricing and Hedging of CDO-squared tranches by using a one factor Levy model. *International Journal of Theoretical and Applied Finance*, **12**, 663-685, 2009. R09056.

[1203] Hubert, M., Robust Calibration. In: *Encyclopedia of Comprehensive Chemometrics*, Brown, S., Tauler, R. and B. Walczak (eds.) **3**, 315-343, Elsevier, Oxford, 2009. R09058.

[1204] Hubert, M. and M. Debruyne, Breakdown value. *WIREs: Computational Statistics*, **1**, 296-302, 2009. R09057.

[1205] Hubert, M., Rousseeuw, P.J. and T. Verdonck, Robust PCA for skewed data and its outlier map. *Computational Statistics and Data Analysis*, **53**, 2264-2274, 2009. R09059.

[1206] Ip, E. H., Smits, D. J. M. and P. De Boeck, Locally dependent linear logistic test model with person covariates. *Applied Psychological Measurement*, **33**, 555-567, 2009. R09035.

[1207] Jansen, M. and C. Damerval, Multiscale kernel smoothing using a lifting scheme. *Proceedings of the 17th European Signal Processing Conference* (Eusipco), Glasgow (Electronic proceedings). R09060.

[1208] Jansen, M., Nason, G. and B. Silverman, Multiscale methods for data on graphs and irregular multidimensional situations. *Journal of the Royal Statistical Society, Series B*, **71**, 1, 97-125, 2009. R09061.

[1209] Jönsson, H., Masol, V. and W. Schoutens, Normal inverse Gaussian model. In: *Encyclopedia of Quantitative Finance*, Cont R. (ed.). Wiley, 2009. R09062.

[1210] Jönsson, H. and W. Schoutens, Pricing of constant maturity credit default swaps under jump dynamics. *Journal of Credit Risk*, **5**, 1, 2009. R09063.

[1211] Jönsson, H., Schoutens, W. and G. Van Damme, Modeling default and prepayment using Lévy processes: an application to asset backed securities. *Radon Series on Computational and Applied Mathematics*, **8**, 1-23, 2009. R09064.

[1212] Leenen, I. and E. Ceulemans, Three-way hierarchical classes: A comparison of INDCLAS and Tucker3-HICLAS models. *Applied Multivariate Research*, **13**, 43-76, 2009. R09036.

[1213] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, **37**, 3023-3058, 2009. R09065.

[1214] Oravecz, Z., Tuerlinckx, F. and J. Vandekerckhove, A hierarchical Ornstein-Uhlenbeck model for continuous repeated measurement data. *Psychometrika*, **74**, 395-418, 2009. R09037.

[1215] San Martín, E., González, J. and F. Tuerlinckx, Identified parameters, parameters of interest and their relationships. *Measurement: Interdisciplinary Research and Perspective*, **7**, 97-105, 2009. R09038.

[1216] Schepers, J. and J. Hofmans, TwoMP: A MATLAB graphical user interface for two-mode partitioning. *Behavior Research Methods*, **41**, 507-514, 2009. R09039.

[1217] 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 and Statistical Psychology*, **62**, 299-318, 2009. R09040.

[1218] Valdivieso, L., Schoutens, W. and F. Tuerlinckx, Maximum likelihood estimation in processes of Ornstein-Uhlenbeck type. *Statistical Inference for Stochastic Processes*, **12**, 1-19, 2009. R09066.

[1219] Van Aerschot, W., Jansen, M. and A. Bulteel, Normal mesh based geometrical image compression. *Image and Vision Computing*, **27**, 4, 459-468, 2009. R09067.

[1220] van den Berg, R.A., Rubingh, C.M., Westerhuis, J.A., van der Werf, M.J. and A.K. Smilde, Metabolomics data exploration guided by prior knowledge. *Analytica Chimica Acta*, **651**, 173-18, 2009. R09042.

[1221] van den Berg, R.A., Van Mechelen, I., Wilderjans, T.F., Van Deun, K., Kiers, H.A.L. and A.K. Smilde, Integrating functional genomics data using maximum likelihood based simultaneous component analysis. *BMC Bioinformatics*, **10**, 340, 2009. R09043.

[1222] Van Deun, K., Hoijtink, H., Lieven, T., Van Lommel, L., Schuit, F. and I. Van Mechelen, Testing the hypothesis of tissue-selectivity: The intersection-union test and a Bayesian approach. *Bioinformatics*, **25**, 2588-2594, 2009. R09044.

[1223] Van Deun, K., Smilde, A.K., van der Werf, M.J., Kiers, H.A.L. and I. Van Mechelen, A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, **10**, 246, 2009. R09045.

[1224] Vanduffel, S., Chernih, A., Maj, M. and W. Schoutens, A note on the suboptimality of path-dependent pay-offs in Levy markets. *Applied Mathematical Finance*, **16**, 315-330, 2009. R09068.

[1225] Verlaak R., Hurlimann, W. and J. Beirlant, Quasi-likelihood estimation of benchmark rates for excess of loss reinsurance programs. *Astin Bulletin*, **39**, 2, 429-452, 2009. R09069.

[1226] Vrins, F. and W. Schoutens, Credit Default Swaps: The quest of the hedge. *Wilmott Magazine*, **1** (5-6), 245-254, 2009. R09070.

[1227] Wilderjans, T.F., Ceulemans, E. and I. Van Mechelen, Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics and Data Analysis*, **53**, 1086-1098, 2009. R09047.

[1228] Wilderjans, T.F., Ceulemans, E., Kiers, H.A.L. and K. Meers, LMPCA program: A graphical user interface for fitting the linked-mode PARAFAC-PCA model to coupled real-valued data. *Behavior Research Methods*, **41**, 1073-1082, 2009. R09046.

D.2.3 Refereed publications (in press)

[1229] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties. *The Annals of the Institute of Statistical Mathematics* (to appear), RP10013.

[1230] Ceulemans, E. and G. Storms, Detecting intra and inter categorical structure in semantic concepts using HICLAS. *Acta Psychologica* (to appear), 2009. RP09004.

[1231] Claeskens, G. and M. Jansen, Nonparametric estimation. In: *International Encyclopedia of Statistical Science*, Springer (to appear), 2010. RP10014.

[1232] Debruyne, M., Christmann, A., Hubert, M. and J.A.K. Suykens, Robustness and stability of reweighted kernel based regression. *Journal of Multivariate Analysis* (to appear), 2009. RP09018.

[1233] Debruyne, M., Hubert, M. and J. Van Horebeek. Detecting influential observations in kernel PCA. *Computational Statistics and Data Analysis* (to appear), 2009. RP09014.

[1234] Ding, H., Claeskens, G. and M. Jansen, Variable selection in partially linear wavelet models. *Statistical Modelling* (to appear), 2010. RP10015.

[1235] Gelman, A., Leenen, I., Van Mechelen, I., De Boeck, P. and J. Poblome, Bridges between deterministic and probabilistic models for binary data. *Statistical Methodology* (to appear), 2009. RP09005.

[1236] Gijbels, I., Censored data. *Interdisciplinary Reviews: Computational Statistics* (to appear), 2010. RP10017.

[1237] Gijbels, I., The Kaplan-Meier estimator. In: *International Encyclopedia of Statistical Science*, Springer (to appear), 2010. RP10020.

[1238] Gijbels, I. and I. Prosdocimi, Loess. *Interdisciplinary Reviews: Computational Statistics* (to appear), 2010. RP10019.

[1239] Gijbels, I., Prosdocimi, I. and G. Claeskens, Nonparametric estimation of mean and dispersion functions in extended generalized linear models. *TEST* (to appear), 2009. RP09049.

[1240] Gijbels, I. and A. Verhasselt, P-splines regression smoothing and difference type of penalty. *Statistics and Computing* (to appear), 2010. RP10016.

[1241] Gijbels, I. and A. Verhasselt, Regularisation and P-splines in generalised linear models. *Journal of Nonparametric Statistics* (to appear), RP10018.

[1242] Goegebeur, Y., De Boeck, P., Molenberghs, G. and M.G. Kenward, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology* (to appear), 2009. RP09006.

[1243] Hofmans, J., Editorial: Applications of functional measurement in psychology. *Psicologica* (to appear), 2009. RP09007.

[1244] Hubert, M. and M. Debruyne, Minimum covariance determinant. In: *Interdisciplinary Reviews: Computational Statistics* (to appear), 2009. RP09056.

[1245] Hubert, M. and P. Rousseeuw, Multivariate techniques, Robustness. *International Encyclopedia of Statistical Science* (to appear), 2009. RP09055.

[1246] Hubert, M. and S. Van der Veeken, Fast and robust classifiers adjusted for skewness. *Proceedings of Compstat* (to appear), 2010. RP10082.

[1247] Jansen, M. and G. Claeskens, The Cramér-Rao inequality. In: *International Encyclopedia of Statistical Science*, Springer (to appear), 2010. RP10021.

[1248] Janssen, R., Modeling the effect of item designs within the Rasch model. In: *New directions in psychological measurement with model-based approaches.* Embretson, S. and J. Roberts (eds.), Washington, DC: American Psychological Association (to appear), 2009. RP09003.

[1249] Kahraman, N., De Boeck, P. and R. Janssen, Modeling DIF in complex response data using test design strategies. *International Journal of Testing* (to appear), 2009. RP09008.

[1250] Magis, D. and G. Raîche, An iterative maximum a posteriori estimation of proficiency level to detect multiple local likelihood maxima. *Applied Psychological Measurement* (to appear), 2009. RP09009.

[1251] Theuns, P., Hofmans, J., Mazaheri, M., Van Acker, F. and J.L. Bernheim, Cross-national comparability of the WHOQOL-BREF: A measurement invariance approach. *Quality of Life Research* (to appear), 2009. RP09010.

[1252] Timmerman, M.E., Ceulemans, E., Lichtwarck-Aschoff, A. and K. Vansteelandt, Multilevel component analysis for studying intra-individual variability and inter-individual differences. In: *Dynamic process methodology in the social and developmental sciences*. Valsiner, J., Molenaar, P.C.M., Lyra, M.C.D.P. and N. Chaudhary (eds.), Springer (to appear), 2009. RP09002.

[1253] Vandekerckhove, J. and F. Tuerlinckx, MATLAB for behavioral scientists: A novice's guide to MATLAB. *Experimental Psychology* (to appear), 2009. RP09011.

[1254] Vandekerckhove, J., Verheyen, S. and F. Tuerlinckx, A crossed random effects diffusion model for speeded semantic categorization decisions. *Acta Psychologica* (to appear), 2009. RP09012.

[1255] Verboven, S. and M. Hubert, Matlab library LIBRA. *WIREs: Computational Statistics* (to appear), 2009. RP09015.

[1256] Verdonck, T., Rousseeuw, P. and M. Hubert, DetMCD in a regression framework. *Proceedings of Compstat* (to appear), 2010. RP10083.

[1257] Wetzels, R., Vandekerckhove, J., Tuerlinckx, F. and E.-J. Wagenmakers, Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task. *Journal of Mathematical Psychology* (to appear), 2009. RP09013.

D.2.4 Non-refereed publications (published)

[1258] Hubert, M., Book review of '*Applied Multiway Data Analysis*', P. Kroonenberg, Wiley, *Journal of the American Statistical Association*, **104**, 856, 2009. NR09001.

D.2.5 Books (published)

[1259] Cariboni, J. and W. Schoutens, *Lévy Jump Processes in Credit Risk*, Wiley, 2009. B09001.

D.3 Katholieke Universiteit Leuven, KUL-2

Note: G. Molenberghs works for 50% at KUL-2 and for 50% at UH. His papers are therefore mentioned in the lists of both groups.

D.3.1 Refereed publications (published)

[1260] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multifaceted sensitivity analysis of the Slovenian Public Opinion Survey data. *Applied Statistics*, **58**, 171-196, 2009. R09102.

[1261] Borgermans, L., Goderis, G., Van Den Broeke, C., Verbeke, G., Carbonez, A., Ivanova, A., Mathieu, C., Aertgeerts, B., Heyrman, J. and R. Grol, Interdisciplinary diabetes care teams operating on the interface between primary and specialty care are associated with improved outcomes of care: Findings from the Leuven Diabetes Project. *BMC Health Services Research*, **9**, 179, 2009. R09086.

[1262] Bossuyt, X., Coenen, D., Fieuws, S., Verschueren, P., Westhovens, R. and N. Blanckaert, Likelihood ratios as a function of antibody concentration for anti-cyclic citrullinated peptide antibodies and rheumatoid factor. *Annals of the Rheumatic Diseases*, **68**, 287-289, 2009. R09099.

[1263] Casteels, K., Fieuws, S., van Helvoirt, M., Verpoorten, C., Goemans, N., Coudyzer, W., Loockx, D. and F. de Zegher, Metformin therapy to reduce weight gain and visceral adiposity in children and adolescents with neurogenic or myogenic motor deficit. *Pediatric Diabetes*, 2009. R09095.

[1264] Debruyne, M., Christmann, A., Hubert, M. and J.A.K. Suykens, Robustness and stability of reweighted kernel based regression. *Journal of Multivariate Analysis*, **101**, 447-463, 2010. R10020.

[1265] Decramer, M. and G. Molenberghs, Letter to the Editor: Does pharmacotherapy reduce rate of decline of lung function in COPD? *American Journal of Respiratory and Critical Care Medicine*, **179**, 171-171, 2009. R09107.

[1266] De Wit, L., Molas, M., Lesaffre, E., Dejaeger, E., Feys, H., Jenni, W., Lincoln, N., Eckhof, H., Putman, K. and W. De Weerdt, The use of a biplot in studying outcomes after stroke. *Neurorehabilitation and Neural Repair*, **23**, 825-830, 2009. R09074.

[1267] Faes, C., Molenberghs, G., Aerts, M., Verbeke, G. and M.G. Kenward, The effective sample size and a novel small sample degrees of freedom method. *The American Statistician*, **63**, 389-399, 2009. R09113.

[1268] Fieuws, S. and G. Verbeke, Joint models for high-dimensional longitudinal data. In: *Longitudinal Data Analysis*, Fitzmaurice G., Davidian M., Verbeke G., and Molenberghs G. (eds.), Chapman and Hall, 367-394, 2009. R09089.

[1269] Fitzmaurice, G. and G. Verbeke, Parametric modeling of longitudinal data: Introduction and overview. In: *Longitudinal Data Analysis*, Fitzmaurice, G., Davidian, M., Verbeke, G., and Molenberghs, G. (eds.), Chapman and Hall, 31-42, 2009. R09087.

[1270] Floris, G., Debiec-Rychter, M., Sciot, R., Stefan, C., Fieuws, S., Machiels, K., Wozniak, A., Faa, G. and P. Schöffski, High efficacy of the (Histone)DeACetylase-inhibitor LBH589 towards gastro intestinal stromal tumors in a xenograft mouse model. *Clinical Cancer Research*, **15)**, 2009. R09094.

[1271] Gyselaers, W., Molenberghs, G., Van Mieghem, W. and W. Ombelet, Doppler measurement of renal interlobar vein impedance index in uncomplicated and pre-eclamptic pregnancies: a pilot study. *Hypertension in Pregnancy*, **28**, 23-33, 2009. R09105.

[1272] Hens, N., Wienke, A., Aerts, M. and G. Molenberghs, The correlated and shared frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, **28**, 2785-2800, 2009. R09111.

[1273] Ibrahim, J.G. and G. Molenberghs, Missing data methods in longitudinal studies: a review (with discussion and rejoinder). *Test*, **18**, 68, 2009. R09110.

[1274] Jacobs, B., De Roock,W., Piessevaux, H., Van Oirbeek,R., Biesmans, B., De Schutter,J., Fieuws, S., Vandesompele, J., Peeters, M., Van Laethem, J.-L., Humblet, Y., Pénault-Llorca, F., De Hertogh, G., Laurent-Puig, P., Van Cutsem, E. and S.Tejpar, Amphiregulin and Epiregulin mRNA expression in primary tumors predicts outcome in metastatic colorectal cancer treated with cetuximab. *Journal of Clinical Oncology*, **27**, 5068-5074, 2009. R09093.

[1275] Komárek A., Hansen B., Jansen, H.L.A. and E. Lesaffre, Prediction of binary response using multivariate longitudinal profiles: Study on chronic hepatitis B patients. *Proceedings of the 24th International Workshop on Statistical Modelling*, Ithaca, NY, Booth, J. G. (ed.), 187-192, 2009. R09076.

[1276] Komárek, A. and E. Lesaffre, The regression analysis of correlated interval-censored data: Review of accelerated failure time models with flexible distributional assumptions. *Statistical Modelling*, **9**, 299-319, 2009. R09075.

[1277] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, A family of measures to evaluate scale reliability in a longitudinal setting. *Journal of the Royal Statistical Society, Series A*, **172**, 237-253, 2009. R09108.

[1278] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Reliability of a longitudinal sequence of scale ratings. *Psychometrika*, **74**, 49-64, 2009. R09106.

[1279] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scale. *Journal of Psychiatric Research*, **43**, 730-738, 2009. R09103.

[1280] Lemmens, R., Cheng, F., Yicheng, N., Fieuws, S. and V. Thijs, Is there a decline in the vascular event rate after TIA or stroke in antiplatelet trials. *Cerebrovascular Disease*, **28**, 439-447, 2009. R09092.

[1281] Leroy, R., Cecere, S., Lesaffre, E. and D. Declerck, Caries experience in primary molars and its impact on the variability in permanent tooth emergence sequences. *Journal of Dentistry*, **37**, 865-871, 2009. R09072.

[1282] 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*, **9**, 99-118, 2009. R09073.

[1283] Luyten, L., Welkenhuysen, van Kuyck, K., Fieuws, S., Das, J., Sciot, R. and B. Nuttin, Likelihood ratios as a function of antibody concentration for anti-cyclic citrullinated peptide antibodies and rheumatoid factor. *Annals of the Rheumatic Diseases*, **68**, 287-289, 2009. R09098.

[1284] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 study Group, Evaluation of culture-based approaches for rapid detection of glycopeptide-resistant enterococci: a randomised, investigator-blinded study. *Clinical Microbiology and Infection*, **15**, 58-58, 2009. R09117.

[1285] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 Study Group, Rapid detection of extended-spectrum b-lactamaseproducing Enterobacteriaceae: a randomised, investigator-blinded evaluation of culture-based approaches. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09120.

[1286] Molenberghs, G., Incomplete data in clinical studies: Analysis, sensitivity, and sensitivity analysis (with discussion). *Drug Information Journal*, **43**, 409-446, 2009. R09100.

[1287] Molenberghs, G., Incomplete data in clinical studies: Analysis, sensitivity, and sensitivity analysis. Rejoinder. *Drug Information Journal*, **43**, 447-448, 2009. R09101.

[1288] Molenberghs, G., Beunckens, C., Jansen, I., Thijs, H., Verbeke, G. and M.G. Kenward, Missing data. In: *Handbook of Epidemiology*, I. Pigeot and W. Ahrends (eds.), Heidelberg: Springer, 2009. R09122.

[1289] Molenberghs, G. and M.G. Kenward, Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics and Data Analysis*, **54**, 585-597, 2009. R09114.

[1290] Molenberghs, G. and C. Ortman, Surrogate endpoints: application in pediatric clinical trials. In: *Concepts and Applications of Pediatric Drug Development*. Mulberg, A., Silber, S.A., and van den Anker, J.N. (eds.), New York: Wiley-Blackwell, 501-512, 2009. R09121.

[1291] Oskrochi, G., Lesaffre, E., Molas, M. and D. Shamley, An application of the multivariate linear mixed model to the analysis of shoulder complexity: EMG measurements in breast cancer patients. *Proceedings of the 24th International Workshop on Statistical Modelling*, Ithaca, NY, Booth, J. G. (ed.), 273-280, 2009. R09077.

[1292] Rizopoulos, D., Verbeke, G. and E. Lesaffre, Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society - Series B*, **71**, 637-654, 2009. R09070.

[1293] Rodriguez-Llanes, J.M., Verbeke, G. and C. Finlayson, Reproductive benefits of high social status in male macaques (Macaca). *Animal Behaviour*, **78**, 643-649, 2009. R09085.

[1294] Sabiiti, W., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Evaluation of commercially available molecular and culturebased assays for rapid detection of methicillin-resistant Staphylococcus aureus. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09119.

[1295] Sergeant, G., Ectors, N., Fieuws, S., Aerts, R. and B. Topal, Prognostic relevance of extracapsular lymph node involvement in pancreatic ductal adenocarcinoma. *Annals of Surgical Oncology*, **16**, 3070-3079, 2009. R09091.

[1296] Serroyen, J., Molenberghs, G., Verbeke, G. and M. Davidian, Nonlinear models for longitudinal data. *The American Statistician*, **63**, 378-388, 2009. R09112.

[1297] Sotto, C., Beunckens, C., Molenberghs, G., Jansen, I. and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, **69**, 305-336, 2009. R09104.

[1298] Thevissen, P., Fieuws, S. and G. Willems, Human dental age estimation using third molar developmental stages: does a Bayesian approach outperform regression models to discriminate between juveniles and adults?. *International Journal of Legal Medicine*, **124**, 35-42, 2009. R09090.

[1299] Thevissen, P., Pittayapat, P., Fieuws, S. and G. Willems, Estimating age of majority on third molars developmental stages in young adults from Thailand using a modified scoring technique. *Journal of Forensic Sciences*, **54**, 2009. R09096.

[1300] Thilakarathne, P.J., Verbeke, G., Engelen, K. and K. Marchal, A nonlinear mixed-effects model for estimating calibration intervals for unknown concentrations in two-color microarray data with spike-ins. *Statistical Applications in Genetics and Molecular Biology*, **8**, 2009. R09081.

[1301] Tilahun, A., Maringwa, J.T., Geys, H., Alonso, A., Raeymaekers, L., Molenberghs, G., Van Den Kieboom, G., Drinkenburg, P. and L. Bijnens, Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. *Journal of Biopharmaceutical Statistics, Series A*, **19**, 133-149, 2009. R09109.

[1302] Topal, B., Fieuws, S., Tomczyk, K., Aerts, R., Van Steenbergen, W., Verslype, C. and F. Penninckx, Clinical models are inaccurate in predicting bile duct stones in situ for patients with gallbladder. *Surgical Endoscopy*, **23**, 38-44, 2009. R09097.

[1303] 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*, **65**, 81-87, 2009. R09071.

[1304] Tsonaka, R., Verbeke, G. and E. Lesaffre, A semi-parametric shared parameter model to handle nonmonotone nonignorable missingness. *Biometrics*, **65**, 81-87, 2009. R09080.

[1305] Van den Heede, K., Lesaffre, E., Diya, L., Vleugels, A., Clarke, S.P., Aiken, L.H. and W. Sermeus, The relationship between inpatient cardiac surgery mortality and nurse numbers and educational level: analysis of administrative data. *International Journal of Nursing Studies*, **46**, 796-803, 2009. R09078.

[1306] Van Effelterre, T., Shkedy, Z., Aerts, M., Molenberghs, G., Van Damme, P. and P. Beutels, Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus. *Epidemiology and Infection*, **137**, 48-57, 2009. R09115.

[1307] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Comparison of the impact of direct plating versus a short or overnight pre-enrichment on detection of methicillin-resistant Staphylococcus aureus from clinical specimens. *Clinical Microbiology and Infection*, **15**, 208-208, 2009. R09118.

[1308] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 study group, Impact of a short period of pre-enrichment on detection and bacterial loads of methicillin-resistant Staphylococcus aureus from screening specimens. *Journal of Clinical Microbiology*, **47**, 3326-3328, 2009. R09116.

[1309] Vanneste, E., Voet, T., Le Caignec, C., Ampe, M., Konings, P., Melotte, C., Debrock, S., Amyere, M., Viikkula, M., Schuit, F., Fryns, J.-P., Verbeke, G., D'Hooghe, T., Moreau, Y. and J. Vermeesch, Chromosome instability is common in human cleavage-stage embryos. *Nature Medicine*, **15**, 577-583, 2009. R09084.

[1310] Van Nieuwenhuyse, A., Crombez, G., Burdorf, A., Verbeke, G., Masschelein, R., Moens, G. and P. Mairiaux, Physical characteristics of the back are not predictive of low back pain in healthy workers: A prospective study. *BioMed Central: Musculoskeletal Disorders*, **10**, 2009. R09083.

[1311] Verbeke, G. and M. Davidian, Joint models for longitudinal data: Introduction and overview. In: *Longitudinal Data Analysis*, Fitzmaurice, G., Davidian, M., Verbeke, G. and Molenberghs, G. (eds.), Chapman and Hall, 2009. R09088.

[1312] Verbeke, G. and E. Lesaffre, Repeated Measurements. In: *Methods and Applications of Statistics in the Life and Health Sciences*, N. Balakrishnan (ed.), Wiley, 695-708, 2009. R09079.

[1313] Verbeke, G., Molenberghs, G. and D. Rizopoulos, Random effects models for longitudinal data. In: *Longitudinal Research with Latent Variables.* K. van Montfort, H. Oud, and Al Satorra (eds.), New York: Springer, 2009. R09123.

D.3.2 Refereed publications (in press)

[1314] Alonso, A., Laenen, A., Molenberghs, G., Geys, H. and T. Vangeneugden, Reliability of single administered tests: a unified approach. *Biometrics* (to appear), 2009. RP09042.

[1315] Alqerban, A., Jacobs, R., Fieuws, S. and G. Willems, Comparison of the accuracy of two cone beam computed tomography systems versus panoramic imaging for localization of impacted maxillary canines and detection of root resorption on maxillary lateral incisors. *European Journal of Orthodontics* (to appear), 2010. RP10038.

[1316] Assam, P., Tilahun, A., Alonso, A. and G. Molenberghs, Using earlier measures in a longitudinal sequence as potential surrogate for a later one. *Computational Statistics and Data Analysis* (to appear), 2009. RP09047.

[1317] Breusegem, C., Spielberg, L., Van Ginderdeuren, R., Vandewalle, E., Renier, C., Van de Veire, S., Fieuws, S., Zeyen, T. and I. Stalmans, Preoperative topical non-steroidal anti-inflammatory drug or steroid and clinical outcomes after trabeculectomy. A randomized controlled trial. *Ophthalmology* (to appear), 2010. RP10037.

[1318] Bruckers, L., Serroyen, J., Molenberghs, G., Slaets, H. and W. Goeyvaerts, Latent-class analysis of persistent disturbing behavior patients using longitudinal profiles. *Applied Statistics* (to appear), 2009. RP09053.

[1319] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* (to appear), 2009. RP09030.

[1320] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Generalized shared-parameter models and missingness at random. *Statistical Modeling* (to appear), 2009. RP09043.

[1321] De Wolf, J., Duchateau, L., Verbeke, G., and E. Schrevens, Discovering transgenic elite events: Using information from early screening trials for improving experimental design. *Journal of Agricultural, Biological, and Environmental Statistics* (to appear), 2009. RP09017.

[1322] Diya, L., Lesaffre, E., Van den Heede, K., Sermeus, W. and A. Vleugels, Relationship between nurse staffing and hospital mortality: A multi-center nursing unit level study using clustered discrete-time logistic model. *Statistics in Medicine* (to appear), 2010. RP10028.

[1323] 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), 2010. RP10024.

[1324] Goderis, G., Borgermans, L., Heyrman, J., Van Den Broeke, C., Carbonez, A., Mathieu, C., Verbeke, G. and R. Grol, Monitoring modifiable cardiovascular risk in type 2 diabetes care in general practices: The use of an aggregated z-score. *Medical Care* (to appear), 2010. RP10032.

[1325] Guerra-Yi, M.E., Lorenzo-Luaces Alvarez, P., Faes, C., Galàn Alvarez, Y. and G. Molenberghs, Spatial analysis of breast and cervix cancer incidence in small geographical areas in Cuba, 19999-2003. *European Journal of Cancer Prevention* (to appear), 2009. RP09028.

[1326] Gyselaers, W., Molenberghs, G., Hanssens, M. and L. Peeters, Maternal renal interlobar vein impedance index is higher in early- than in late-onset preeclampsia. *Ultrasound in Obstetrics and Gynecology* (to appear), 2009. RP09032.

[1327] Gyselaers, W., Molenberghs, G., Mesens, T. and L. Peeters, Maternal hepatic vein Doppler velocimetry during uncomplicated and preeclamptic pregnancy. *Ultrasound in Medicine and Biology* (to appear), 2009. RP09025.

[1328] Hernández, M., Sorsa, T., Obregón, F., Tervahartiala, T., Valenzuela, M-A, Pozo, P., Dutzan, N., Lesaffre., E., Molas M. and J. Gamonal, MMP-13 enhances MMP-9 activation during progression of chronic periodontitis. *Journal of Clinical Periodontology* (to appear), 2010. RP10027.

[1329] Hill, R.C. Choate, C.J., Scott, K.C. and G. Molenberghs, Comparison of the guaranteed analysis with the measured nutrient composition of commercial pet foods. *Journal of the American Veterinary Medical Associationg* (to appear), 2009. RP09022.

[1330] Jacobs, T., Straetemans, R., Molenberghs, G., Bouwknecht, A. and L. Bijnens, A latent pharmacokinetic time profile to model dose-response survival data. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09027.

[1331] Jansen, I. and G. Molenberghs, Pattern-mixture models for categorical outcomes with non-monotone missingness. *Journal of Statistical Computation and Simulation* (to appear), 2009. RP09029.

[1332] Jara, A., Hanson, T.E. and E. Lesaffre, Robustifying generalized linear mixed models using mixtures of multivariate Polya trees. *Journal of Computational Statistics and Graphics* (to appear), 2010. RP10026.

[1333] Kasim, A., Lin, D., Van Sanden, S., Clevert, D., Bijnens, L., Gohlmann, H., Amaratunga, D., Hochreiter, S., Shkedy, Z. and W. Talloen, Informative or non-informative calls for gene expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* (to appear), 2009. RP09050.

[1334] Komárek, A., Hansen, B., Kuiper, E., van Buuren, H. and E. Lesaffre, Discriminant analysis using a multivariate linear mixed model with a normal mixture in the random effects distribution. *Statistics in Medicine* (to appear), 2010. RP10030.

[1335] Laenen, A. and A. Alonso, The functional living index - cancer: estimating its reliability based on clinical trial data. *Quality of Life Research* (to appear), 2010. RP10049.

[1336] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Coping wit memory effect and serial correlation when estimating reliability in a longitudinal framework. *Applied Psychological Measurement* (to appear), 2009. RP09033.

[1337] Lehnert, S., Van Loo, P., Thilakarathne, P.J., Marynen, P., Verbeke, G. and F.C. Schuit, Evidence for co evolution between human miRNAs and Alu-repeats. *Public Library of Science One (PLoS One)* (to appear), 2009. RP09016.

[1338] Leonard, D.,Penninckx, F., Fieuws, S., Jouret-Mourin, A., Sempoux, C., Jehaes, C. and E. Van Eycken, MDFactors predicting the quality of total mesorectal excision for rectal cancer. *Annals of Surgery* (to appear), 2010. RP10033.

[1339] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Considerations for comparing a test drug with standard of care in phase II clinical trials of central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09041.

[1340] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Considerations for using positive controls in phase II clinical trials of central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09040.

[1341] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Design archetypes for phase II clinical trials in central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09039.

[1342] Molas, M. and E. Lesaffre, Hurdle models for multilevel zero-inflated data via H-likelihood. *Statistics in Medicine* (to appear), 2010. RP10031.

[1343] Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A. and M. Buyse, A unified framework for the evaluation of surrogate endpoints in clinical trials. *Statistical Methods in Medical Research* (to appear), 2009. RP09020.

[1344] Molenberghs, G., Kenward, M.G. and G. Verbeke, Discussion of Lee, Y. and Nelder, J.A.: Likelihood inference for models with unobservables: another view. *Statistical Science* (to appear), 2009. RP09048.

[1345] Molenberghs, G., Kenward, M.G., Verbeke, G. and B. Teshome Ayele, Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* (to appear), 2009. RP09037.

[1346] Molenberghs, G. and G. Verbeke, Longitudinal and incomplete clinical studies. *Journal of the Japanese Society of Computational Statistics* (to appear), 2009. RP09035.

[1347] Mwambi, H., Ramroop, S., Shkedy, Z. and G. Molenberghs, An approach to estimating the force of infection and the rate of recovery fom Respiratory Syncytial Virus (RSV) data. *Statistical Methods in Medical Research* (to appear), 2010. RP10069.

[1348] Pittayapat, P., Thevissen, P., Fieuws, S., Jacobs, R. and G. Willems, Forensic oral imaging quality of hand-held dental X-ray devices: Comparison of two image receptors and two devices. *Forensic Science International* (to appear), 2010. RP10036.

[1349] Pryseley, A., Mintiens, K., Knapen, K., Van der Stede, Y. and G. Molenberghs, Estimating precision, repeatability, and reproducibility from Gaussian and non-Gaussian data: a mixed models approach. *Journal of Applied Statistics* (to appear), 2009. RP09031.

[1350] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival data. *Biometrics* (to appear), 2009. RP09024.

[1351] Sergeant, G., Topal, B. and S. Fieuws, On hazard ratios and magnitude of impact on prognosis. *Annals of Surgery* (to appear), 2010. RP10068.

[1352] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in linear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09026.

[1353] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in nonlinear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09051.

[1354] Shkedy, Z., Lin, D., Molenberghs, G., Goehlmann, H., Talloen, W. and L. Bijnens, Genomic biomarkers for depression: Feature-specific and joint biomarkers. *Statistics in Biopharmaceutical Research* (to appear), 2009. RP09036.

[1355] Siebens, K., Miljoen, H., Fieuws, S., Drew, B., De Geest, S. and C. Vrints , Implementation of the guidelines for the management of patients with chest pain through critical pathway approach improves length of stay and patient satisfaction but not anxiety. *Critical Pathways in Cardiology* (to appear), 2010. RP10034.

[1356] Star, H., Thevissen, P., Jacobs, R., Fieuws,S., Solheim, T. and G. Willems, Human dental age estimation by calculation of pulp/tooth volume ratios yielded on clinically acquired cone beam computed tomography (CBCT) images of mono radicular teeth. *Journal of Forensic Sciences* (to appear), 2010. RP10035.

[1357] Tilahun, A., Lin, D., Shkedy, Z., Geys, H., Alonso, A., Peeters, W.P., Talloen, W., Drinkenburg W., Goehlmann, H.W.H., Gorden, E., Bijnens L. and G. Molenberghs, Selection and evaluation of genomic biomarkers for depression in a microarray setting. *Journal of Biopharmaceutical Research* (to appear), 2009. RP09052.

[1358] Troxel, A., Lipsitz, S., Fitzmaurice, G., Ibrahim, J., Sinha, D. and G. Molenberghs, A weighted combination of pseudo-likelihood estimators for longitudinal binary data subject to nonignorable non-monotone missingness. *Statistics in Medicine* (to appear), 2009. RP09046.

[1359] Tsonaka, S., Rizopoulos, D., Verbeke, G. and E. Lesaffre, Non-ignorable models for intermittently missing categorical longitudinal responses. *Biometrics* (to appear), 2010. RP10029.

[1360] VandeVoorde P., Sabbe, M., Rizopoulos, D., Tsonaka, S., De Jaeger, E., Lesaffre, E., Peters, M. and The Penta study group, Assessing the level of consciousness in children: a plea for the Glasgow Coma Motor Subscore. *Resuscitation* (to appear), 2010. RP10022.

[1361] VandeVoorde P., Sabbe, M., Rizopoulos, D., Tsonaka, S., De Jaeger, E., Lesaffre, E., Peters, M. and The Penta study group, Paediatric trauma and trauma care in Flanders (Belgium). Methodology and first descriptive results of the Penta registry. *European Journal of Pediatrics* (to appear), 2010. RP10023.

[1362] Vangeneugden, T., Molenberghs, G., Laenen, A., Geys, H., Beunckens, C., and C. Sotto, Marginal correlation in longitudinal binary data based on generalized linear mixed models. *Communications in Statistics, Theory and Methods* (to appear), 2009. RP09034.

[1363] Vangeneugden, T., Molenberghs, G., Verbeke, G. and C. Demétrio, Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Journal of Applied Statistics* (to appear), 2009. RP09038.

[1364] Verbeke, G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling* (to appear), 2009. RP09021.

[1365] Wouters, K., Cortiñas Abrahantes, J., Molenberghs, G., Geys, H., Ahnaou, A., Drinkenburg, W.H.I.M. and L. Bijnens, Correction for model selection bias using a modified model averaging approach for supervised learning methods applied to EEG experiments. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09045.

[1366] Zhang, Y., Todem, D., Kim, K. and E. Lesaffre, Bayesian latent variable models for spatially correlated tooth-level binary data in caries research. *Statistical Modelling* (to appear), 2010. RP10025.

D.3.3 Non-refereed publications (published)

[1367] Creemers, A., Hens N., Aerts, M., Molenberghs, G., Verbeke, G. and M. Kenward, A sensitivity analysis in the shared parameter modeling framework. *Proceedings of the 30th Annual Conference of the International Society for Clinical Biostatistics.* University of Economics, Prague, 2009. NR09007.

[1368] Laenen, A., Alonso, A., Molenberghs, G. and Vangeneugden, Estimating the reliability of rating scales from clinical trial data. *Proceedings of the 16th International Meeting of the Psychometric Society*. Cambridge, 2009. NR09002.

[1369] Vankerckhoven, V., Lammens, C., Chapelle, S., Vanderstraeten, A., Hens, N., Molenberghs, G. and H. Goossens, Analysis of the gut microbiotica in irritable bowel syndrome patients and healthy volunteers using selective media and real-time PCR. In: *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 1249, 2009. NR09003.

[1370] Vankerckhoven, V., van Voorden, S., Hens, N., Molenberghs, G., Goossens, H. and E. Wiertz, Activation of Toll-like receptor 2 and Toll-like Receptor 2/6 by lacobacilli. *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 2009. NR09004.

D.3.4 Books (published)

[1367] Dmitrienko, A., Molenberghs, G., Christy Chuang-Stein, J.L. and W.W. Offen, *Analysis of Clinical Trial Data using SAS: A Practical Guide. (Japanese Translation)*, Kodansya Scientific and SAS Press, 2009. B09004.

[1368] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Longitudinal Data Analysis*, Handbooks of Modern Statistical Methods. New York: Chapman and Hall, 2009. B09002.

[1369] Verbeke, G. and G. Molenberghs, *Linear Mixed Models for Longitudinal Data* (Paperback Edition). New York: Springer, 2009. B09003.

D.4 Universiteit Gent, UG

D.4.1 Refereed publications (published)

[1370] Alqallaf, F., Van Aelst, S., Yohai, V.J. and R.H. Zamar, Propagation of outliers in multivariate data. *Annals of Statistics,* **37**, 311-331, 2009. R09146.

[1371] Babanezhad, M., Vansteelandt, S. and E. Goetghebeur, Comparison of causal effect estimators under exposure misclassification. *Journal of Statistical Planning and Inference*, **140**, 1306-1319, 2010. R10021.

[1372] Baele, G., van de Peer, Y. and S. Vansteelandt, Efficient context-dependent model building based on clustering posterior distributions for non-coding sequences. *BMC Evolutionary Biology*, **9**, 2009. R09129.

[1373] Bekaert, M., Benoit, D., Decruyenaere, J. and S. Vansteelandt, A note on statistical association and causality derived from epidemiological ICU data - Reply. *Intensive Care Medicine*, **36**, 550, 2010. R10019.

[1374] Bekaert, M., Vansteelandt, S. and K. Mertens, Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. *Lifetime Data Analysis*, **16**, 45-70, 2010. R10018.

[1375] Carvalho, B., Postma, C., Mongera, S., Hopmans, E., Diskin, S., van de Wiel, M., Van Criekinge, W., Thas, O., Mathhai, A., Cuesta, M., Terhaar, J., Craanen, M., Schrock, E., Ylstra, B. and G. Meijer, Multiple putative oncogenes at the chromosome 20q amplicon contribute to colorectal adenoma to carcinoma progression. *GUT*, **58**, 79-89, 2009. R09140.

[1376] Clement, L. and O. Thas, Nonparametric trend detection in river monitoring network data: a spatio-temporal approach. *Environmetrics*, **20**, 283-297, 2009. R09137.

[1377] Clement, L. and O. Thas, Testing for trends in the violation frequency of an environmental threshold in rivers. *Environmetrics*, **20**, 53-67, 2009. R09138.

[1378] Comté, L., Vansteelandt, S., Tousset, E., Baxter, G. and B. Vrijens, Linear and loglinear structural mean models to evaluate the benefits of an on-demand dosing regimen. *Clinical Trials*, **6**, 403-415, 2009. R09132.

[1379] Delva, W., Vercoutere, A., Louac, C., Lamahc, J., Vansteelandt, S., De Koker, P., Claeys, P., Temmerman, M. and L. Annemans, Psychological well-being and socio-economic hardship among AIDS orphans and other vulnerable children in Guinea. *AIDS Care*, **21**, 1490-1498, 2009. R09135.

[1380] Flahou, B., Hellemans, A., Meyns, T., Duchateau, L., Chiers, K., Baele, M., Pasmans, F., Haesebrouck, F. and R. Ducatelle, Protective immunization with homologous and heterologous antigens against Helicobacter suis challenge in a mouse model. *Vaccine*, **27**, 1416-1421, 2009. R09153.

[1381] García-Escudero, L.A., Gordaliza, A., San Martín, R., Van Aelst, S. and R.H. Zamar, Robust Linear Clustering. *Journal of the Royal Statistical Society Series B*, **71**, 301-319, 2009. R09145.

[1382] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological and Environmental Statistics*, **14**, 1-14, 2009. R09151.

[1383] Hoffmann, T.J., Lange, C., Vansteelandt, S. and N.M. Laird, Gene-environment interaction test for dichotomous traits in nuclear families. *Genetic Epidemiology*, **33**, 691-699, 2009. R09133.

[1384] Hoffmann, T.J., Lange, C., Vansteelandt, S., Raby, B.A., Demeo, D.L., Silverman, E.K. and N.M. Laird, Parsing the Effects of Individual SNPs in Candidate Genes with Family Data. *Human Heredity*, **69**, 91-103, 2010. R10017.

[1385] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model. *Lifetime Data Analysis*, **15**, 59-78, 2009. R09004.

[1386] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, **139**, 3865-3877, 2009. R09150.

[1387] Mehrzad, J., Duchateau, L. and C. Burvenich, Phagocytic and bactericidal activity of blood and milk-resident neutrophils against Staphylococcus aureus in primiparous and multiparous cows during early lactation. *Veterinary Microbiology*, **134**, 106-112, 2009. R09154.

[1388] Rayner, J.C.W. and O. Thas, Generalised smooth tests of goodness of fit. *Journal of Statistical Theory and Practice*, **3**, 565-679, 2009. R09139.

[1389] Roelant, E., Van Aelst, S. and C. Croux, Multivariate generalized S-estimators. *Journal of Multivariate Analysis,* **100**, 876-887, 2009. R09147.

[1390] Roelant, E., Van Aelst, S. and G. Willems, The minimum weighted covariance determinant estimator. *Metrika,* **70**, 177-204, 2009. R09148.

[1391] Sjolander, A., Humphreys, K., Vansteelandt, S., Belococco, R. and J. Palmgren, Sensitivity analysis for principal stratum direct effects with an application to a study of physical activity and coronary heart disease. *Biometrics*, **65**, 514-520, 2009. R09126.

[1392] Thas, O., Rayner, J.C.W., Best, D.J. and De Boeck, B., Informative statistical analyses using smooth goodness of fit tests. *Journal of Statistical Theory and Practice*, **3**, 705-733, 2009. R09142.

[1393] Thas, O., Comment on : Goodness-of-fit tests in mixed modes : smooth tests of goodness-of-fit for the random effects distribution in linear mixed models. *Test*, **18**, 260-264, 2009. R09141.

[1394] Van Aelst, S., and P.J. Rousseeuw, Minimum volume ellipsoid. *Wiley Interdisciplinary Reviews: Computational Statistics* **1**, 71-82, 2009. R09149.

[1395] VanderWeele, T.J. and S. Vansteelandt, Conceptual issues concerning mediation, interventions and composition. *Statistics and its Interface*, **2**, 457-468, 2009. R09134.

[1396] VanderWeele, T.J., Vansteelandt, S. and J.M. Robins, Marginal structural models for sufficient cause interactions. *American Journal of Epidemiology* **171**, 506-514, 2010. R10016.

[1397] Vansteelandt, S., Estimating direct effects in cohort and case-control studies. *Epidemiology*, **20**, 851-860, 2009. R09131.

[1398] Vansteelandt, S., Discussion on: ‘Identifiability and estimation of causal effects in randomized trials with noncompliance and completely non-ignorable missing data’. *Biometrics*, **65**, 686-689, 2009. R09130.

[1399] Vansteelandt, S., Babanezhad, M. and E. Goetghebeur, Correcting Instrumental Variables Estimators for Systematic Measurement Error. *Statistica Sinica*, **19**, 1223-1246, 2009. R09127.

[1400] Vansteelandt, S., Carpenter, J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, **6**, 37-48, 2009. R09221.

[1401] Vansteelandt, S., Goetgeluk, S., Lutz, S., Waldman, I., Lyon, H., Schadt, E.E., Weiss, S.T. and C. Lange, On the adjustment for covariates in genetic association analysis: A novel, simple principle to infer causality. *Genetic Epidemiology*, **33**, 394-405, 2009. R09128.

[1402] Vansteelandt, S., Mertens, K., Suetens, C. and E. Goetghebeur, Marginal structural models for partial exposure regimes. *Biostatistics*, **10**, 46-59, 2009. R09124.

[1403] Van Nieuwerburgh F., Goetghebeur E., Vandewoestyne M., and D. Deforce, Impact of allelic dropout on evidential value of forensic DNA profiles using RMNE. *Bioinformatics*, **25**, 225-229, 2009. R09136.

[1404] Verheyden, K., Noppe, H., Vanhaecke, L., Wille, K., Vanden Bussche, J., Bekaert, K., Thas, O., Janssen, C. and H. De Brabander, Excretion of endogenous boldione in human urine : influence of phytosterol consumption. *Journal of Steroid Biochemistry and Molecular Biology*, **117**, 8-14, 2009. R09143.

[1405] Wouters, H., Thas, O. and J-P. Ottoy, Data-driven smooth test and a diagnostic tool for lack-of-fit for circular data. *Australian & New Zealand Journal of Statistics*, **51**, 461-480, 2009. R09144.

[1406] Yewhalaw, D., Legesse, W., Van Bortel, W., Gebre-Selassie, S., Kloos, H., Duchateau, L. and N. Speybroeck, Malaria and water resource development: the case of Gilgel-Gibe hydroelectric dam in Ethiopia. *Malaria Journal*, **8**, 1-10, 2009. R09152.

[1407] Zhan, Y., Vansteelandt, S., Van der Elst, J. Coetsier, T., Dhont, M. and P. De Sutter, The efficacy of the embryo transfer catheter in IVF and ICSI is operator-dependent: a randomized clinical trial. *Human Reproduction*, **24**, 880-887, 2009. R09125.

D.4.2 Refereed publications (in press)

[1408] Bruckers, L., Serroyen, J., Molenberghs, G., Slaets, H. and W. Goeyvaerts, Latent-class analysis of persistent disturbing behavior patients using longitudinal profiles. *Applied Statistics* (to appear), 2009. RP09053.

[1409] Buyze, J., Van Rompaye, B. and E. Goetghebeur, Designing a sequentially randomized study with adherence enhancing interventions for diabetes patients. *Statistics in Medicine* (to appear), 2010. RP10041.

[1410] De Bock, K.W., Coussement, K. and D. Van Den Poel, Ensemble classification based on generalized additive models. *Computational Statistics and Data Analysis* (to appear), 2010. RP10045.

[1411] De Boeck, B., Thas, O. and J.P. Ottoy, Reweighted smooth tests of goodness of fit. *Communications in Statistics - Theory and Methods* (to appear), 2010. RP10043.

[1412] De Boeck, B., Thas, O., Rayner, J.C.W. and D.J. Best, Smooth tests for the gamma distribution. *Journal of Statistical Computation and Simulation* (to appear), 2010. RP10044.

[1413] De Wolf, J., Duchateau, L., Verbeke, G., and E. Schrevens, Discovering transgenic elite events: Using information from early screening trials for improving experimental design. *Journal of Agricultural, Biological, and Environmental Statistics* (to appear), 2009. RP09017.

[1414] Khan, J.A., Van Aelst, S. and R.H. Zamar, Fast robust estimation of prediction error based on resampling. *Computational Statistics and Data Analysis* (to appear), 2010. RP10047.

[1415] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model - Validation of prognostic indices using the frailty model. *Lifetime Data Analysis* (to appear), 2009. RP09001.

[1416] Manoochehr, B., Vansteelandt, S. and E. Goetghebeur, Comparison of causal effect estimators under exposure misclassification. *Journal of Statistical Planning and Inference* (to appear), 2010. RP10040.

[1417] Moerkerke, B., Vansteelandt, S. and C. Lange, A doubly-robust test for gene-environment interaction in family-based studies. *Biostatistics* (to appear), 2009. RP09057.

[1418] Tolley, E., Harrison, P., Goetghebeur E. and al., Adherence and its measurement in phase 2/3 microbicide trials. *AIDS and Behavior* (to appear), 2010. RP10039.

[1419] Van Aelst, S., Vandervieren, E. and G. Willems, Stahel-Donoho estimators with cellwise weights. *Journal of Statistical Computation and Simulation* (to appear), 2010. RP10046.

[1420] Van Rompaye, B., Jaffar, S. and E. Goetghebeur, Design and testing for clinical trials faced with misclassified causes-of-death. *Biostatistics* (to appear), 2010. RP10042.

D.4.3 Books (published)

[1421] Rayner, J.C.W. and O. Thas, *Smooth Tests of Goodness-of-Fit : using R.* Wiley, Singapore, 2009. B09005.

[1422] Thas, O., *Comparing Distributions.* Springer, New-York, 2009. B09006.

D.5 Universiteit Hasselt, UH

Note: G. Molenberghs works for 50% at KUL-2 and for 50% at UH. His papers are therefore mentioned in the lists of both groups.

D.5.1 Technical reports

[1423] Braekers, R. and A. Gaddah, Flexible modeling in the Koziol-Green model by a copula function, 2009. TR09008.

[1424] Gaddah, A. and R. Braekers, An extension of the Koziol-Green model under dependent censoring, 2009. TR09009.

[1425] Gijbels, I., Omelka, M. and D. Sznajder, Positive quadrant dependence tests for copulas, 2009. TR09061.

[1426] Gijbels, I., Veraverbeke, N. and M. Omelka, Conditional copulas, association measures and their applications, 2009. TR09060.

[1427] Veraverbeke, N., Omelka, M. and I. Gijbels, Estimation of a conditional copula and association measures, 2009. TR09057.

D.5.2 Refereed publications (published)

[1428] Acuña, C., Leborán, V., Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso Suárez, C.M. and J.L. Pardo Vázquez, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal*, 2009. RP09091.

[1429] Aelvoet, W., Terryn, N., Francis, W., Redivo, M. and C. Faes, Miscoding: a threat to the hospital care system. How to detect it ? *Epidemiology and Public Health*, **57**, 169-177, 2009. R09199.

[1430] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multifaceted sensitivity analysis of the Slovenian Public Opinion Survey data. *Applied Statistics*, **58**, 171-196, 2009. R09102.

[1431] Bilau, M., De Henauw, S., Schroijen, C., Bruckers, L., Den Hond, E., Koppen, G., Matthys, C., Van De Mieroop, E., Keune, H., Bayens, W., Nelen, V., Van Larebeke, N., Willems, J.L. and G. Schoeters, The relation between the estimation dietary intake of PCDD/Fs and levels in blood in a Flemish population (50-65 years). *Environment International*, **35**, 9-13, 2009. R09186.

[1432] Bilcke J., Van Damme P., Van Ranst M., Hens N., Aerts M. and P. Beutels, Estimating the incidence of symptomatic rotavirus infections: a systematic review and meta-analysis. *PLoS ONE*, **4**, 2009. R09204.

[1433] Bollaerts, L., Messens, W., Delhalle, L., Aerts, M., Van der Stede, Y., Dewulf, J., Quoilin, S., Maes, D., Mintiens, K. and K. Grijspeerdt, METZOON: Development of a qualitative microbial risk assessment for human salmonellosis through household consumption of fresh minced pork meat in Belgium. *Risk Analysis*, **29**, 820-840, 2009. R09191.

[1434] Boone, I., Van der Stede, Y., Bollaerts, K., Messens, W., Vose, D., Daube, G., Aerts, M. and K. Mintiens, Expert judgement in a risk assessment model for Salmonella spp. in pork: The performance of different weighting schemes. *Preventive Veterinary Medicine*, **92**, 224-234, 2009. R09180.

[1435] Boone, I., Van der Stede, Y., Bollaerts, K., Vose, D., Maes, D., Dewulf, J., Messens, W., Daube, G., Aerts, M. and Mintiens, K., NUSAP method for evaluating the data quality in a quantitative microbial risk assessment model for Salmonella in the pork production chain. *Risk Analysis*, **29**, 502-517, 2009. R09171.

[1436] Boulet, G.A.V., Benoy, I.H., Depuydt, C.E., Horvath, C.A.J., Aerts, M., Hens, N., Vereecken, A.J. and J.J Bogers, HPV16 load and integration as biomarkers for CIN2+ lesions in HPV16-positive women. *Cancer Epidemiology, Biomarkers and Prevention*, **18**, 2992-2999, 2009. R09205.

[1437] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, **100**, 1270-1281, 2009. R09160.

[1438] Cao, R., Villar, J.M., Devia, A., Veraverbeke, N. Boucher, J-Ph. and J. Beran, Modelling consumer credit risk via survival analysis. *Statistics and Operations Research Transactions*, **33**, 31-47, 2009. R09157.

[1439] Cortiñas Abrahantes, J., Bollaerts, K., Aerts, M., Ogunsanya, V. and Y. Van der Stede, Salmonella serosurveillance: Different statistical methods to categorize pig herds based on serological data. *Preventive Veterinary Medicine*, **89**, 59-66, 2009. R09172.

[1440] Croes, K., Baeyens, W., Bruckers, L., Den Hond, E., Koppen, G., Nelen, V., Van de Mieroop, E., Keune, H., Dhooge, W., Schoeters, G. and N. Van Larebeke, Hormone levels and sexual development in Flemish adolescents residing in areas differing in pollution pressure. *International Journal of Hygiene and Environmental Health*, **212**, 612-625, 2009. R09189.

[1441] Decramer, M. and G. Molenberghs, Letter to the Editor: Does pharmacotherapy reduce rate of decline of lung function in COPD? *American Journal of Respiratory and Critical Care Medicine*, **179**, 171-171, 2009. R09107.

[1442] Den Hond, E., Govarts, E., Bruckers, L. and G. Schoeters, Determinants of polychlorinated aromatic hydrocarbons in serum in three age classes - methodological implications for human biomonitoring. *Environmental Research*, **109**, 495-502, 2009. R09187.

[1443] De Wit, L., Molas, M., Lesaffre, E., Dejaeger, E., Feys, H., Jenni, W., Lincoln, N., Eckhof, H., Putman, K. and W. De Weerdt, The use of a biplot in studying outcomes after stroke. *Neurorehabilitation and Neural Repair*, **23**, 825-830, 2009. R09074.

[1444] Faes, C., Aerts, M., Geys, H. and G. Teuns, Modeling spatial learning in rats based on morris water maze experiments. *Pharmaceutical Statistics*, 2009. R09201.

[1445] Faes, C., Molenberghs, G., Aerts, M., Verbeke, G. and M.G. Kenward, The effective sample size and a novel small sample degrees of freedom method. *The American Statistician*, **63**, 389-399, 2009. R09113.

[1446] Gaddah, A. and R. Braekers, Weak convergence for the conditional distribution function in a Koziol-Green model under dependent censoring. *Journal of Statistical Planning and Inference*, **139**, 930-943, 2009. R09159.

[1447] Galecki, A.T., Chen, S., Faulkner, J.A., Ashton-Miller, J. and T. Burzykowski, Statistical power calculations for clustered continuous data. *International Journal of Knowledge Engineering and Soft Data Paradigms*, **1**, 40-48, 2009. R09193.

[1448] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological and Environmental Statistics*, **14**, 1-14, 2009. R09151.

[1449] Gyselaers, W., Molenberghs, G., Van Mieghem, W., and W. Ombelet, Doppler measurement of Renal Interlobar Vein Impedance Index in uncomplicated and pre-eclamptic pregnancies: a pilot study. *Hypertension in Pregnancy*, **28**, 23-33, 2009. R09105.

[1450] Hens, N., Goeyvaerts, N., Aerts, M., Shkedy, Z., Van Damme, P. and P. Beutels, Mining social mixing patterns for infectious discease models based on a two-day population survey in Belgium. *BMC Infectious Diseases*, **9**, 5, 2009. R09168.

[1451] Hens, N., Minalu, G., Goeyvaerts, N., Aerts, M., Mossong, J., Edmunds, W.J. and P. Beutels, Estimating the impact of school closure on social mixing behaviour and the transmission of close contact infections in eight European countries. *BMC Infectious Diseases*, **9**, 187, 2009. R09184.

[1452] Hens, N., Wienke, A., Aerts, M. and G. Molenberghs, The correlated and shared frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, **28**, 2785-2800, 2009. R09111.

[1453] Ibrahim, J.G. and G. Molenberghs, Missing data methods in longitudinal studies: a review (with discussion and rejoinder). *Test*, **18**, 68-., 2009. R09110.

[1454] Jacobs, N., Claes, N., Thijs, H., Dendale P. and I. De Bourdeaudhuij, Effect of a tailored behaviour change programme in highly educated adults: a randomised controlled trial. *European Journal of Cardiovascular Prevention and Rehabilitation*, **16**, 2009. R09198.

[1455] Janssen, P., Swanepoel, J. and N. Veraverbeke, New tests for exponentiality against new better than used in p-th quantile. *Journal of Nonparametric Statistics*, **21**, 85-97, 2009. R09158.

[1456] Kadankov, V., Kadankova, T. and N. Veraverbeke, Intersections of an interval by a difference of a compound Poisson process and a compound renewal process. *Stochastic Models*, **25**, 270-300, 2009. R09155.

[1457] Koppen, G., Den Hond, E., Nelen, V., Van De Mieroop, E., Bruckers, L., Bilau, M., Keune, H., Van Larebeke, N., Covaci, A., Van De Weghe, H., Schroijen, C., Desager, K., Stalpart, M., Baeyens, W. and G. Schoeters, Organochlorine and heavy metals in newborns: results from the Flemish Environmental and Health Survey (FLEHS 2002-2006). *Environment International*, **35**, 1015-1022, 2009. R09190.

[1458] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, A family of measures to evaluate scale reliability in a longitudinal setting. *Journal of the Royal Statistical Society, Series A*, **172**, 237-253, 2009. R09108.

[1459] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Reliability of a longitudinal sequence of scale ratings. *Psychometrika*, **74**, 49-64, 2009. R09106.

[1460] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scale. *Journal of Psychiatric Research*, **43**, 730-738, 2009. R09103.

[1461] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model. *Lifetime Data Analysis*, **15**, 59-78, 2009. R09004.

[1462] Lin, D., Shkedy, Z., Burzykowski, T., Talloen, W. and L. Bijnens, A comparison of procedures for controlling the false discovery rate in the presence of small variance genes: a simulation study. *Communications in Statistics-Simulation and Computation*, **38**, 2111-2122, 2009. R09207.

[1463] Lin, D., Shkedy, Z., Yekutieli, D., Dilba, G., Burzykowski, T., Goehlmann, H.W.H., De Bondt A., Perera, T., Geerts, T., and L. Bijnens, Multiple contract test for detecting monotonic dose-response relationship and FDR-adjusted confidence intervals for selected parameters in a microarray setting. *Online Journal of Bioinformatics*, **10**, 67-73, 2009. R09192.

[1464] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 study Group, Evaluation of culture-based approaches for rapid detection of glycopeptide-resistant enterococci: a randomised, investigator-blinded study. *Clinical Microbiology and Infection*, **15**, 58-58, 2009. R09117.

[1465] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 Study Group, Rapid detection of extended-spectrum b-lactamaseproducing Enterobacteriaceae: a randomised, investigator-blinded evaluation of culture-based approaches. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09120.

[1466] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, **139**, 3865-3877, 2009. R09150.

[1467] Michiels, S., Le Maître, A., Buyse, M., Burzykowski, T., Maillard, E., Bogaerts, J., Vermorken, J.B., Budach, W., Pajak, T.F., Ang, K.K., Bourhis, J. and J.P. Pignon, on behalf of the MARCH and MACH-NC collaborative groups, Surrogate endpoints for overall survival in locally advanced head and neck cancer: meta-analysis of individual patient data. *Lancet Oncology*, **10**, 341-350, 2009. R09195.

[1468] Molanes Lopez, E. M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, **36**, 413-432, 2009. R09019.

[1469] Molenberghs, G., Incomplete data in clinical studies: Analysis, sensitivity, and sensitivity analysis (with discussion). *Drug Information Journal*, **43**, 409-446, 2009. R09100.

[1470] Molenberghs, G., Incomplete data in clinical studies: Analysis, sensitivity, and sensitivity analysis. Rejoinder. *Drug Information Journal*, **43**, 447-448, 2009. R09101.

[1471] Molenberghs, G., Beunckens, C., Jansen, I., Thijs, H., Verbeke, G. and M.G. Kenward, Missing data. In: *Handbook of Epidemiology*, I. Pigeot and W. Ahrends (eds.), Heidelberg: Springer, 2009. R09122.

[1472] Molenberghs, G. and M.G. Kenward, Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics and Data Analysis*, **54**, 585-597, 2009. R09114.

[1473] Molenberghs, G. and C. Ortman, Surrogate endpoints: application in pediatric clinical trials. In: *Concepts and Applications of Pediatric Drug Development.* Mulberg, A., Silber, S.A., and van den Anker, J.N. (eds.), New York: Wiley-Blackwell, 501-512, 2009. R09121.

[1474] Namata, H., Welby, S., Aerts, M., Faes, C., Cortiñas Abrahantes, J., Imberechts, H., Vermeersch, K., Hooyberghs, J., Méroc, E. and K. Mintiens, Identification of risk factors for the prevalence and persistence of Salmonella in Belgian broiler chicken flocks. *Preventive Veterinary Medicine*, **90**, 211-222, 2009. R09173.

[1475] Ogunjimi, B., Hens, N., Goeyvaerts, N., Aerts, M. Van Damme, P. and P. Beutels, Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. *Mathematical Biosciences*, **218**, 80-87, 2009. R09183.

[1476] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, **37**, 3023-3058, 2009. R09065.

[1477] Robaeys, G., Nevens, F., Starkel, P., Colle, I., Van Eyken, P., Bruckers, L., Van Ranst, M. and F. Buntinx, Previous intravenous substance use and outcome of liver transplantation in patients with chronic Hepatitis C infection. *Transplantation Proceedings*, 589-594, 2009. R09185.

[1478] Sabiiti, W., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Evaluation of commercially available molecular and culturebased assays for rapid detection of methicillin-resistant Staphylococcus aureus. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09119.

[1479] Sanchez, L., Cortiñas Abrahantes, J., Pelaez, O., Gutierrez, H., Concepción, D. and P. Van der Stuyft, Breteau index threshold levels indicating risk for dengue transmission in areas with low aedes infestation. *Tropical Medicine and International Health*, 2009. R09181.

[1480] Sanchez, L., Perez, D., Cruz, G., Castro, M., Kouri, G., Shkedy, Z., Vanlerberghe, V. and P. Van der Stuyft, Intersectoral coordination, community empowerment and dengue prevention: six years of controlled interventions in Playa Municipality, Havana, Cuba. *Tropical Medicine and International Health*, **14**, 1356-1364, 2009. R09206.

[1481] Serroyen, J., Molenberghs, G., Verbeke, G. and M. Davidian, Nonlinear models for longitudinal data. *The American Statistician*, **63**, 378-388, 2009. R09112.

[1482] Smeets, K., Opdenakker, K., Remans, T. Van Sanden, S., Van Belleghem, F., Semane, B., Horemans, N., Guisez, Y., Vangronsveld, J. and A. Cuypers, Oxidative stress-related responses at transcriptional and enzymatic levels after exposure to Cd or Cu in a multipollution context. *Journal of Plant Physiology*, **166**, 1982-92, 2009. R09170.

[1483] Sotto, C., Beunckens, C., Molenberghs, G., Jansen, I. and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, **69**, 305-336, 2009. R09104.

[1484] Straczek A., Wannijn, J., Van Hees, M., Thijs, H. and Y. Thiry, Tolerance of hairy roots of carrots to U chronic exposure in a standardized in vitro device. *Environmental and Experimental Botany*, **65**, 82-89, 2009. R09197.

[1485] Theeten, H., Hens, N., Aerts, M., Vandermeulen, C., Roelants, M., Hoppenbrouwers, K., Van Damme, P. and P. Beutels, Caregiver concerns about the number of vaccine injections and willingness to pay to reduce them. *The Pediatric Infectious Disease Journal*, **28**, 61-63, 2009. R09202.

[1486] Theeten, H., Hens, N., Aerts, M., Vandermeulen, C., Roelants, M., Hoppenbrouwers, K., Van Damme, P. and P. Beutels, Common attitudes about concomitant vaccine injections for infants and adolescents in Flanders, Belgium. *Vaccine*, **27**, 1964-1969, 2009. R09203.

[1487] Tilahun, A., Maringwa, J.T., Geys, H., Alonso, A., Raeymaekers, L., Molenberghs, G., Van Den Kieboom, G., Drinkenburg, P. and L. Bijnens, Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. *Journal of Biopharmaceutical Statistics*, **19**, 133-149, 2009. R09109.

[1488] Valkenborg, D., Thomas, G., Krols, L., Kas, L. and T. Burzykowski, A strategy for the prior processing of high-resolution mass spectral data obtained from high-dimensional combined fractional diagonal chromatography. *Journal of Mass Spectrometry*, **44**, 516-529, 2009. R09194.

[1489] Van Effelterre, T., Shkedy, Z., Aerts, M., Molenberghs, G., Van Damme, P. and P. Beutels, Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus. *Epidemiology and Infection*, **137**, 48-57, 2009. R09115.

[1490] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Comparison of the impact of direct plating versus a short or overnight pre-enrichment on detection of methicillin-resistant Staphylococcus aureus from clinical specimens. *Clinical Microbiology and Infection*, **15**, 208-208, 2009. R09118.

[1491] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 study group, Impact of a short period of pre-enrichment on detection and bacterial loads of methicillin-resistant Staphylococcus aureus from screening specimens. *Journal of Clinical Microbiology*, **47**, 3326-3328, 2009. R09116.

[1492] Vankrunkelsven, P., Kellen, E. Lousbergh, D., Cloes, E., Op de Beeck, L., Faes C., Bruckers, L., Mertens, R., Coeberg, J.W., Van Leeuwen, F.E. and F. Buntinx, Reduction in hormon replacement therapy use and declining breast cancer incidence in the Belgian province of Limburg. *Breast Cancer Research and Treatment*, **118**, 425-432, 2009. R09188.

[1493] Verbeke, G., Molenberghs, G. and D. Rizopoulos, Random effects models for longitudinal data. In: *Longitudinal Research with Latent Variables.* K. van Montfort, H. Oud, and Al Satorra (eds.), New York: Springer, 2009. R09123.

[1494] Wang, L. and N. Veraverbeke, Bayes prediction based on right censored data. *Statistical Papers*, **50**, 137-149, 2009. R09156.

D.5.3 Refereed publications (in press)

[1495] Aerts, M., Hens, N. and J.S. Simonoff, Model selection in regression based on presmoothing. *Journal of Applied Statistics* (to appear), 2009. RP09076.

[1496] Alonso, A., Laenen, A., Molenberghs, G., Geys, H. and T. Vangeneugden, Reliability of single administered tests: a unified approach. *Biometrics* (to appear), 2009. RP09042.

[1497] Assam, P., Tilahun, A., Alonso, A. and G. Molenberghs, Using earlier measures in a longitudinal sequence as potential surrogate for a later one. *Computational Statistics and Data Analysis* (to appear), 2009. RP09047.

[1498] Bollaerts, K., Messens, W., Aerts, M., Dewulf, J., Maes, D., Grijspeerdt, K. and Y. Van der Stede, Evaluation of scenarios for reducing human salmonellosis through household consumption of fresh minced pork meat. *Risk Analysis* (to appear), 2009. RP09080.

[1499] Bruckers, L., Serroyen, J., Molenberghs, G., Slaets, H. and W. Goeyvaerts, Latent-class analysis of persistent disturbing behavior patients using longitudinal profiles. *Applied Statistics* (to appear), 2009. RP09053.

[1500] Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., Lane, P. , Matcham, J., Morgan, D., Philips, A., Roger, J., Sullivan, B., White, I. and L.M. Yu, of the PSI missing data expert group, Missing data: discussion points from the PSI missing data expert group. *Pharmaceutical Statistics* (to appear), 2009. RP10061.

[1501] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* (to appear), 2009. RP09030.

[1502] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Generalized shared-parameter models and missingness at random. *Statistical Modeling* (to appear), 2009. RP09043.

[1503] Goegebeur, Y., De Boeck, P. and G. Molenberghs, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology* (to appear), 2009. RP09006.

[1504] Goeyvaerts, N., Hens, N., Ogunjimi, B., Aerts, M., Shkedy, Z., Van Damme, P. and P. Beutels, Estimating infectious disease parameters from data on social contacts and serological status. *Journal of the Royal Statistical Society - Series C* (to appear), 2010. RP10048.

[1505] Guerra-Yi, M.E., Lorenzo-Luaces Alvarez, P., Faes, C., Galàn Alvarez, Y. and G. Molenberghs, Spatial analysis of breast and cervix cancer incidence in small geographical areas in Cuba, 19999-2003. *European Journal of Cancer Prevention* (to appear), 2009. RP09028.

[1506] Gyselaers, W., Molenberghs, G., Hanssens, M. and L. Peeters, Maternal renal interlobar vein impedance index is higher in early- than in late-onset preeclampsia. *Ultrasound in Obstetrics and Gynecology* (to appear), 2009. RP09032.

[1507] Gyselaers, W., Molenberghs, G., Mesens, T. and L. Peeters, Maternal hepatic vein Doppler velocimetry during uncomplicated and preeclamptic pregnancy. *Ultrasound in Medicine and Biology* (to appear), 2009. RP09025.

[1508] Hens, N., Aerts, M., Faes, C., Shkedy, Z., Lejeune, O., Van Damme, P. and P. Beutels, 75 years of estimating the force of infection. *Epidemiology and Infection* (to appear), 2009. RP09078.

[1509] Hens, N., Kvitkovicova, A., Aerts, M., Hlubinka, D. and P. Beutels, Modelling distortions in seroprevalence data using change-point fractional polynomials. *Statistical Modelling* (to appear), 2009. RP09077.

[1510] Hill, R.C. Choate, C.J., Scott, K.C. and G. Molenberghs, Comparison of the guaranteed analysis with the measured nutrient composition of commercial pet foods. *Journal of the American Veterinary Medical Associationg* (to appear), 2009. RP09022.

[1511] Jacobs, T., Straetemans, R., Molenberghs, G., Bouwknecht, A. and L. Bijnens, A latent pharmacokinetic time profile to model dose-response survival data. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09027.

[1512] Jansen, I. and G. Molenberghs. Pattern-mixture models for categorical outcomes with non-monotone missingness. *Journal of Statistical Computation and Simulation* (to appear), 2009. RP09029.

[1513] Kasim, A., Lin, D., Van Sanden, S., Clevert, D., Bijnens, L., Gohlmann, H., Amaratunga, D., Hochreiter, S., Shkedy, Z. and W. Talloen, Informative or non-informative calls for gene expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* (to appear), 2009. RP09050.

[1514] Kenward, M.G. and G. Molenberghs, Last Observation carried forward: a crystal ball ? *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09054.

[1515] Laenen, A. and A. Alonso, The functional living index - cancer: estimating its reliability based on clinical trial data. *Quality of Life Research* (to appear), 2010. RP10049.

[1516] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Coping wit memory effect and serial correlation when estimating reliability in a longitudinal framework. *Applied Psychological Measurement* (to appear), 2009. RP09033.

[1517] Legrand C., Duchateau L., Janssen P., Ducrocq V. and R. Sylvester, Validation of prognostic indices using the frailty model. *Lifetime Data Analysis* (to appear), 2009. RP09001.

[1518] Liang, X., Jacobs, R., Hassan, B., Li L.M., Pauwels, R., Corpas, L., Couto Souza P., Martens, M. Shahbazian, M., Alonso, A. and I. Lambrichts, A comparative evaluation of cone bean computed tomography (CBCT) and multi-slice CT (MSCT)-Part I: on subjective image quality. *European Journal of Radiology* (to appear), 2009. RP09079.

[1519] Lin, D., Shkedy, Z., Burzykowski, T., Aerts, M., Goehlmann, H.W.H., De Bondt, A., Perera, T., Geerts, T. and L. Bijnens, Classification of trends in dose-response microarray experiments using information theory selection methods. *The Open Applied Informatics Journal* (to appear), 2009. RP09081.

[1520] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Considerations for comparing a test drug with standard of care in phase II clinical trials of central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09041.

[1521] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Considerations for using positive controls in phase II clinical trials of central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09040.

[1522] Mallinckrodt, C.H., Detke, M.J., Prucka, W.R., Ruberg, S.J. and G. Molenberghs, Design archetypes for phase II clinical trials in central nervous system disorders. *Drug Information Journal* (to appear), 2009. RP09039.

[1523] Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A., and M. Buyse, A unified framework for the evaluation of surrogate endpoints in clinical trials. *Statistical Methods in Medical Research* (to appear), 2009. RP09020.

[1524] Molenberghs, G., Kenward, M.G. and G. Verbeke, Discussion of Lee, Y. and J.A. Nelder: Likelihood inference for models with unobservables: another view. *Statistical Science* (to appear), 2009. RP09048.

[1525] Molenberghs, G., Kenward, M.G., Verbeke, G. and B. Teshome Ayele, Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* (to appear), 2009. RP09037.

[1526] Molenberghs, G. and G. Verbeke, Longitudinal and incomplete clinical studies. *Journal of the Japanese Society of Computational Statistics* (to appear), 2009. RP09035.

[1527] Mwambi, H., Ramroop, S., Shkedy, Z. and G. Molenberghs, An approach to estimating the force of infection and the rate of recovery from Respiratory Syncytial Virus (RSV) data. *Statistical Methods in Medical Research* (to appear), 2010. RP10069.

[1528] Pryseley, A., Mintiens, K., Knapen, K., Van der Stede, Y. and G. Molenberghs, Estimating precision, repeatability, and reproducibility from Gaussian and non-Gaussian data: a mixed models approach. *Journal of Applied Statistics* (to appear), 2009. RP09031.

[1529] Rizopoulos, D., Verbeke, G., and G. Molenberghs, Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival data. *Biometrics* (to appear), 2009. RP09024.

[1530] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in linear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09026.

[1531] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in nonlinear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09051.

[1532] Shkedy, Z., Lin, D., Molenberghs, G., Goehlmann, H., Talloen, W. and L. Bijnens, Genomic biomarkers for depression: Feature-specific and joint biomarkers. *Statistics in Biopharmaceutical Research* (to appear), 2009. RP09036.

[1533] Tilahun, A., Lin, D., Shkedy, Z., Geys, H., Alonso, A., Peeters, W.P., Talloen, W., Drinkenburg W., Goehlmann, H.W.H., Gorden, E., Bijnens L. and G. Molenberghs, Selection and evaluation of genomic biomarkers for depression in a microarray setting. *Journal of Biopharmaceutical Research* (to appear), 2009. RP09052.

[1534] Troxel, A., Lipsitz, S., Fitzmaurice, G., Ibrahim, J., Sinha, D. and G. Molenberghs, A weighted combination of pseudo-likelihood estimators for longitudinal binary data subject to nonignorable non-monotone missingness. *Statistics in Medicine* (to appear), 2009. RP09046.

[1535] Vangeneugden, T., Molenberghs, G., Laenen, A., Geys, H., Beunckens, C. and C. Sotto, Marginal correlation in longitudinal binary data based on generalized linear mixed models. *Communications in Statistics, Theory and Methods* (to appear), 2009. RP09034.

[1536] Vangeneugden, T., Molenberghs, G., Verbeke, G. and C. Demétrio, Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Journal of Applied Statistics* (to appear), 2009. RP09038.

[1537] Verbeke, G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling* (to appear), 2009. RP09021.

[1538] Wouters, K., Cortiñas Abrahantes, J., Molenberghs, G., Geys, H., Ahnaou, A., Drinkenburg, W.H.I.M. and L. Bijnens, Correction for model selection bias using a modified model averaging approach for supervised learning methods applied to EEG experiments. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09045.

D.5.4 Non-refereed publications (published)

[1539] Creemers, A., Hens N., Aerts, M., Molenberghs, G., Verbeke, G. and M. Kenward, A sensitivity analysis in the shared parameter modeling framework. *Proceedings of the 30th Annual Conference of the International Society for Clinical Biostatistics*. University of Economics, Prague, 2009. NR09007.

[1540] Del Fava, E., Shkedy, Z., Hens N., Faes, C., Bollaerts, K., Aerts, M., Manfredi, P., Teunis, P., Beutels, P. and P. Van Damme, Estimating the prevalence and the force of infection of parvovirus B19 in Belgium using Bayesian hierarchical mixture models. *Second IBS Channel Network Conference*, Ghent, 2009. NR09008.

[1541] Jacobs, N., Claes, N., Thijs, H., Dendale, P. and I. De Bourdeaudhuij, Effect of a tailored behaviour change programme in highly educated adults: a randomised trial, Kesteloot, H. and Ector, H. (eds.). *Acta Cardiol.*, **64**, 17-18, 2009. NR09013.

[1542] Jacobs, N., Claes, N., Van Mierlo, J., Thijs, H., Dendale, P. and I. De Bourdeaudhuij, Predictors of effectiveness of a tailored behaviour change programme with muliple delivery modes. In: *Psychology and Health*, Moss-Morris, R. and Yardley, L. (eds.). , **24**, 67-68, 2009. NR09012.

[1543] Jacobs, N., Thijs, H., Dendale, P., De Bourdeaudhuij, I. and N. Claes, One year results of a tailored behaviour change programme in highly educated adults: a randomised controlled trial. In: *International Society of Behavioral Nutrition and Physical Activity ISBNPA'09*, Biddle, S., Guldan, G., Gans, K., Almeida, M.D. and French, S. (eds.). 235-235, 2009. NR09014.

[1544] Laenen, A., Alonso, A., Molenberghs, G. and Vangeneugden, Estimating the reliability of rating scales from clinical trial data. *Proceedings of the 16th International Meeting of the Psychometric Society*. Cambridge, 2009. NR09002.

[1545] Loens K., Hens N., Ursi D., Goossens H., Aerts M. and M. Ieven, Application of an expanded gold standard and latent class analysis to decide on the strategy for the diagnosis of M.pneumoniae infections. *19th European Congress of Clinical Microbiology and Infectious Diseases*. Helsinki, 2009. NR09006.

[1546] Theeten, H., Roelants, M., Hens, N., Aerts, M., Depoorter, A.-M., Vandermeulen, C., Boonen, M., Hoppenbrouwers, K., and P. Van Damme, Coverage of recommended vaccines in infants in Flanders, Belgium, anno 2008. In: *Proceedings of the 27th Annual Meeting of the European Society for Paediatric Infectious Diseases*. Brussels, 2009. NR09011.

[1547] Theeten, H., Roelants, M., Hens, N., Vandermeulen, C., Boonen, M., Hoppenbrouwers, K., Depoorter, A.M. and P. Van Damme, Timeliness of infant vaccination in Flanders Belgium. In: *Eurovaccine*, Stockholm, 2009. NR09010.

[1548] Vankerckhoven, V., Lammens, C., Chapelle, S., Vanderstraeten, A., Hens, N., Molenberghs, G. and H. Goossens, Analysis of the gut microbiotica in irritable bowel syndrome patients and healthy volunteers using selective media and real-time PCR. In: *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 1249, 2009. NR09003.

[1549] Vankerckhoven, V., van Voorden, S., Hens, N., Molenberghs, G., Goossens, H. and E. Wiertz, Activation of Toll-like receptor 2 and Toll-like Receptor 2/6 by lacobacilli. *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 2009. NR09004.

[1550] Wallinga, J., Hens, N. and T. Tamme, Who infected whom? *Second IBS Channel Network Conference*, Ghent, 2009. NR09009.

D.5.5 Books (published)

[1551] Dmitrienko, A., Molenberghs, G., Christy Chuang-Stein, J.L. and W.W. Offen, *Analysis of Clinical Trial Data using SAS: A Practical Guide. (Japanese Translation)*, Kodansya Scientific and SAS Press, 2009. B09004.

[1552] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Longitudinal Data Analysis*, Handbooks of Modern Statistical Methods. New York: Chapman and Hall, 2009. B09002.

[1553] Verbeke, G. and G. Molenberghs, *Linear Mixed Models for Longitudinal*, Data (Paperback Edition). New York: Springer, 2009. B09003.

D.6 Université Joseph Fourier, UJF–LMC–IMAG

D.6.1 Technical reports

[1554] Antoniadis, A., Amato, U., Samarov, A. and A. Tsybakov, Kernel based discriminant analysis and noisy ifa models for dimensionality reduction, 2009. TR09074.

[1555] Antoniadis, A., Gijbels, I. and A. Verhasselt, Variable selection in additive models using P-splines, 2009. TR09068.

[1556] Boshnakov, G. and S. Lambert-Lacroix, Periodic levinson-durbin algorithm for entropy maximisation, 2009. TR09077.

[1557] Brouste, A., Istas, J. and S. Lambert-Lacroix, On simulation of manifold indexed fractional gaussian fields, 2009. TR09078.

[1558] Gamboa, F., Klein, T. and C. Prieur, Conditional large and moderate deviations for sums of discrete random variables. Combinatoric applications, 2009. TR09079.

[1559] Garcia, N., Galves, A. and C. Prieur, Perfect simulation of the $\bar{d}$-minimal coupling between ordered pairs of binary chains of infinite order, 2009. TR09080.

[1560] Mestre, O., Prieur, C., Gruber, C. and H. Caussinus, A method for homogenisation of daily temperature observations, 2009. TR09081.

D.6.2 Refereed publications (published)

[1561] Antoniadis, A., Some statistical approaches for the analysis of biomedical functional data. *ECMI newsletter*, **46**, 2152-2155, 2009. R09233.

[1562] Antoniadis, A., Amato, U., De Feis, I., Guido, M., Matricardi, M. and C. Serio, Functional sliced inverse regression to infer temperature, water vapour and ozone from iasi data. *Atmospheric Chemistry and Physics Discussions*, **9**, 7589-7613, 2009. R09161.

[1563] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of functional images. *Journal of Computational and Graphical Statistics* , **18**, 216-237, 2009. R09231.

[1564] Antoniadis, A., Brossat, X., Cugliari, J. and J.M. Poggi, Multivariate functional time series clustering. *Compstat 2010*, 2009. R09041.

[1565] Antoniadis, A., Garcia-Magarinos, M., Cao, R. and W. Gonzalez-Manteiga, Lasso logistic regression, gsoft and the cyclic coordinate descent algorithm. Application to gene expression data. *Statistical Applications in Genetics and Molecular Biology*, 2009. R09082.

[1566] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, Special issue: Multiscale Methods and Statistics: A Productive Marriage, **19**, 1371-1387, 2009. R09048.

[1567] Antoniadis, A., Paparoditis, E. and T. Sapatinas, Bandwidth selection for functional times series prediction. *Statistics and Probability Letters*, **79**, 733-740, 2009. R09166.

[1568] Boshnakov, G. and S. Lambert-Lacroix, Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients. *Journal of Time Series Analysis*, **30**, 467-486, 2009. R09163.

[1569] Chatelain, F., Lambert-Lacroix, S. and J.Y. Tourneret, Pairwise likelihood estimation for multivariate mixed poisson models generated by gamma intensities. *Statistics and Computing*, **19**, 283-301, 2009. R09164.

[1570] Dedecker, J. and C. Prieur, Some unbounded functions of intermittent maps for which the central limit theorem holds. *Special Issue: In the memory of José de Sam Lazaro*, ALEA **5**, 29-45, 2009. R09165.

D.6.3 Refereed publications (in press)

[1571] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties. *Annals of the Institute of Mathematical Statistics* (to appear), 2010. RP10013.

[1572] Antoniadis, A., Letué, F. and P. Fryzlewicz, The Dantzig selector for regression with right censored data. *Scandinavian Journal of Statistics* (to appear), 2010. RP10070.

[1573] Guillotin-Plantard, N. and C. Prieur, Central limit theorem for sampled sums of dependent random variables. *ESAIM-Probab* (to appear), 2009. RP09097.

[1574] Guillotin-Plantard, N. and C. Prieur, Limit theorem for random walk in weakly dependent random scenery. *Annales de l'Institut Henri Poincaré - Probabilités et Statistiques* (to appear), 2009. RP09096.

D.7 Universiteit Utrecht, UU

D.7.1 Refereed publications (published)

[1575] Blasius, J., Eilers, P.H.C. and J. Gower, Better biplots. *Computational Statistics and Data Analysis*, **53**, 3145-315, 2009. R09209.

[1576] Eilers, P.H.C., Li, B. and B.D. Marx, Multivariate calibration with single-index signal regression. *Chemometrics and Intelligent Laboratory Systems*, **96**, 196-202, 2009. R09210.

[1577] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, **53**, 1388-1399, 2009. R09007.

[1578] Schnabel, S.K. and P.H.C. Eilers, An analysis of life expectancy and economic production using expectile frontier zones. *Demographic Research*, **21**, 109-134, 2009. R09212.

[1579] Schnabel, S.K. and P.H.C. Eilers, Optimal expectile smoothing. *Computational Statistics and Data Analysis*, **53**, 4168-4177, 2009. R09213.

[1580] Van Deun, K., Hoijtink, H., Lieven, T., Van Lommel, L., Schuit, F. and I. Van Mechelen, Testing the hypothesis of tissue-selectivity: The Intersection-Union Test and a Bayesian approach. *Bioinformatics*, **25**, 2588-2594, 2009. R09044.

[1581] van de Wiel, M.A., Brosens, R., Eilers, P.H.C., Kumps, C., Meijer, G.A., Menten, B., Sistermans, E., Speleman, F., Timmerman, M.E. and B. Ylstra, Smoothing waves in array CGH tumor profiles. *Bioinformatics*, **25**, 1099-1104, 2009. R09214.

D.8 Universidad de Santiago de Compostela, USC

D.8.1 Refereed publications (published)

[1582] Antoniadis, A., Garcia-Magarinos, M., Cao, R. and W. González-Manteiga, Lasso logistic regression, gsoft and the cyclic coordinate descent algorithm. application to gene expression data. *Statistical Applications in Genetics and Molecular Biology*, 2009. R09082.

[1583] Boente, G., González-Manteiga, W. and A. Pérez González, Robust nonparametric estimation with missing data. *Journal of Statistical Planning and Inference*, **139**, 571-592, 2009. R09216.

[1584] Castellano Méndez, M., Franco, A., Cartelle, D., Febrero-Bande, M. and E. Roca, Identification of NOx and ozone episodes and estimation of ozone by statistical analysis. *Water, Air and Soil Pollution*, **198**, 95-110, 2009. R09217.

[1585] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, Goodness-of-fit tests for the spatial spectral density. *Stochastic Environmental Research and Risk Assesment*, **24**, 67-79, 2010. R10014.

[1586] Crujeiras, R. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends. *Computational Statistics and Data Analysis*, **54**, 452-465, 2010. R10004.

[1587] Febrero-Bande, M., Galeano, P. and W. González-Manteiga, Measures of influence for the functional linear model with scalar response. *Journal of Multivariate Analysis*, **101**, 327-339, 2010. R10015.

[1588] González-Manteiga, W., Mateu, J., Crujeiras, R. and E. Porcu, On the second order properties of the multidimensional peridogram for regularly spaced data. *Theory of Probability and its Applications. SIAM*, **53**, 349-356, 2009. R09215.

bibitemR10005González-Manteiga, W., Ordoñez, C., Taboada Castro, J. and J.M. Matías, Partially linear support vector machines applied to the prediction of mine slope movements. *Mathematical and Computer Modelling*, **51**, 206-215, 2010. R10005.

[1589] Pateiro-López, B. and A. Rodríguez-Casal, Surface area estimation under convexity type assumptions. *Journal of Nonparametric Statistics*, **21**, 729-741, 2009. R09219.

[1590] Reboredo, J.C., González-Manteiga, W., Febrero-Bande, M. and J.M. Matías, Boosting GARCH and neural networks for the prediction of heteroskedastic time series. *Mathematical and Computer Modelling*, **51**, 256-271, 2010. R10003.

[1591] Roca Pardiñas, J., Cadarso Suárez, C.M., Garcia-Tahoces, P. and M.J. Lado, Selecting variables in non-parametric regression models for binary response. An application to the computerized detection of breast cancer. *Statistics in Medicine* **28**, 240-259, 2009. R09218.

D.8.2 Refereed publications (in press)

[1592] Acuña, C., Leborán, V., Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso Suárez, C.M. and J.L. Pardo Vázquez, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal* (to appear), 2009. RP09091.

[1593] Bianco, A., Boente, G., González-Manteiga, W. and A. Pérez González, Estimation of the marginal location under a partially linear model with missing responses. *Computational Statistics and Data Analysis* (to appear), 2010. RP10054.

[1594] Cadarso Suárez, C.M., Botana López, M. and R. Pérez Fdez, On the estimation of the spectral density for continuous spatial processes. *Statistics* (to appear), 2009. RP09083.

[1595] Cadarso Suárez, C.M., Hervada Vidal, X., Rodríguez Muíños, M.A. and T. Seoane Pillado, Free software in the medical investigation. *Medicina Clínica* (to appear), 2009. RP09087.

[1596] Chernick, M.R., González-Manteiga, W., Crujeiras, R. and B.E. Barrios, Bootstrap methods. *International Encyclopedia of Statistical Science.* (to appear), 2010. RP10053.

[1597] Cid-Álvarez, B., Gude, F., Cadarso Suárez, C.M., Rodríguez-Álvarez, M.X., García-Acuña, J.M. and J.R. González-Juanatey, Admission and fasting plasma glucose for estimating risk of death of diabetic and non-diabetic patients with acute coronary syndrome: nonlinearity of hazard ratios, and time-dependent comparison. *American Heart Journal* (to appear), 2009. RP09086.

[1598] Crujeiras, R. and R. Fernández-Casal, On the estimation of the spectral density for continuous spatial processes. *Statistics* (to appear), 2010. RP10050.

[1599] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, Nonparametric test for separability of spatio-temporal processes. *Environmetrics. Wiley* (to appear), 2009. RP09092.

[1600] de Uña Álvarez, J., Rodríguez-Casal, A. and H.Y. Liang, Nonlinear wavelet estimator of the regression function under left truncated dependent data. *Journal of Nonparametric Statistics* (to appear), 2010. RP10055.

[1601] Ginarte, M., García Doval, I., Monteagudo, B., Cabanillas, M., Labandeira, J., Flórez, A., Cabo, F., Álvarez, J., Zulaica, A., Allegue, L.F., Abalde, M., Rosón, E., de la Torre, C. and M.X. Rodríguez-Álvarez, Observer agreement in toenail disorders: implications for diagnosis and clinical research. *British Journal of Dermatology* (to appear), 2009. RP09088.

[1602] González-Manteiga, W., Lombardía, M. J., Molina Perala, I., Morales González, D. and L. Santamaría, Small area estimation under Fay-Herriot models with nonparametric estimation of heteroscedasticity. *Statistical Modelling* (to appear), 2010. RP10051.

[1603] González-Manteiga, W. and C.A. Martínez, Comments on : Dynamic relations for sparsely sampled Gaussian processes. *TEST* (to appear), 2010. RP10056.

[1604] González-Manteiga, W., Pardo Fernández, J.C. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models. *Scandinavian Journal of Statistics* (to appear), 2010. RP10009.

[1605] Gude, F., Kneib, T., Meira-Machado, L. and C.M. Cadarso Suárez, Flexible hazard ratio curves for continuous predictors in multi-state models: An application to breast cancer. *Statistical Modelling* (to appear), 2009. RP09082.

[1606] Lombardía Cortiña, M.J., Domínguez Gómez, R., González Manteiga, W. and J.M. Prada Sánchez, Estimación de áreas pequenas: o ingreso medio mensual por comarca nos fogares galegos. *Xunta de Galicia. Consellería de Facenda* (to appear), 2009. RP09090.

[1607] Muñoz Barús, J.I., Rodríguez-Calvo, M.S., Suárez-Peñaranda, J., Vieira, D.N., Cadarso Suárez, C.M. and M. Febrero-Bande, PMICALC: An R code-based software for estimating postmortem interval (PMI) compatible with Windows, Mac and Linux operating systems. *Forensic Science International* (to appear), 2009. RP09085.

[1608] Pateiro-López, B. and A. Rodríguez-Casal, Generalizing the convex hull of a sample: the R package alphahull. *Journal of Statistical Software* (to appear), 2010. RP10052.

[1609] Pérez González, A., Vilar Fernández, J.M. and W. González-Manteiga, Asymptotic properties of local polynomial regression with missing data and correlated errors. *Annals ot the Institute of Statistical Mathematics* (to appear), 2009. RP09089.

[1610] Pérez González, A., Vilar Fernández, J.M. and W. González-Manteiga, Nonparametric variance function estimation with missing data. *Journal of Multivariate Analysis* (to appear), 2010. RP10057.

[1611] Setti, I., Rodríguez Castro, A., Pata, M.P., Cadarso Suárez, C.M., Yacoubi, B., Bensmael, L., Moukrim, A. and J. Martinez Urtaza, Characteristics and dynamics of salmonella contamination along the coast of Agadir, Morocco. *Applied and Environmental Microbiology* (to appear), 2009. RP09084.

[1612] Sperlich, S. and M.J. Lombardía, Local polynomial inference for small area statistics: estimation, validation and prediction. *Journal of Nonparametric Statistics* (to appear), 2010. RP10058.

D.8.3 Books (in press)

[1613] Chernick, M.R., González-Manteiga, W., Crujeiras, R., and E.B. Barrios, Bootstrap methods. *International Encyclopedia of Statistical Science* (to appear), 2010. BP10001.

[1614] Crujeiras, R., Faraldo Roca, P., Lombardía, M.J., Fernández de Castro, B.M., Carreira, V., Matías, J.M. and S. Naya Fernández, Estudo da inserción laboral dos titulados no Sistema Universitario de Galicia 2005-2006, 148, ACSUG (to appear), 2009. BP09001.

[1615] Faraldo Roca, P., Crujeiras, R., Carreira, V., Fernández de Castro, B.M., Iglesias Patiño, C.L., Lombardía, M.J., Matías, J.M. and S. Naya Fernández, Impacto das prácticas na inserción laboral dos titulados na Universidad de Santiago de Compostela. Consello Social (to appear), 2009. BP09002.

D.9 London School of Hygiene and Tropical Medicine, LSHTM

D.9.1 Technical reports

[1616] Carpenter, J.R., Roger, J.H. and M.G. Kenward, Relevant, accessible sensitivity analyses using multiple imputation, 2010. TR10003.

[1617] Daniel, R.M. and M.G. Kenward, Doubly robust multiple imputation, 2009. TR09069.

[1618] Daniel, R.M., Kenward, M.G., Cousens, S. and B. de Stavola, Using directed acyclic graphs to guide analysis in missing data problems, 2009. TR09071.

[1619] Kenward, M.G. and G. Rosenkranz. Joint modelling of outcome, observation time and missingness, 2009. TR09070.

D.9.2 Refereed publications (published)

[1620] Carpenter, J.R. and I. Plewis, Coming to terms with non-response in longitudinal studies. *SAGE handbook of Methodological Innovation*, M. Williams and P. Vogt (eds.), London, SAGE, 2010. R10006.

[1621] Faes, C., Molenberghs, G., Aerts, M., Verbeke, G. and M.G. Kenward, The effective sample size and a novel small sample degrees of freedom method. *The American Statistician*, **63**, 389-399, 2009. R09113.

[1622] Goldstein, H., Carpenter, J.R., Kenward, M.G. and K. Levin, Multilevel models with multivariate mixed response types. *Statistical Modelling*, **9**, 173-197, 2009. R09222.

[1623] Mallinckrodt, C.H. and M.G. Kenward, Conceptual considerations regarding choice of endpoints, hypotheses, and analyses in longitudinal clinical trials. *Drug Information Journal*, **43**, 449-458, 2009. R09224.

[1624] Molenberghs, G., Beunckens, C., Jansen, I., Thijs, H., Verbeke, G. and M.G. Kenward, Missing data. In: *Handbook of Epidemiology*, I. Pigeot and W. Ahrends (eds.), Heidelberg: Springer. 2009. R09122.

[1625] Molenberghs, G. and M.G. Kenward, Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics and Data Analysis*, **54**, 585-597, 2009. R09114.

[1626] Nevalainen, J., Kenward, M.G. and S.M. Virtanen, Missing values in longitudinal dietary data: a multiple imputation approach based on a fully conditional specification. *Statistics in Medicine*, **28**, 3657-3669, 2009. R09220.

[1627] Nur, U., Shack, L.G., Rachet, B., Carpenter, J.R. and M.P. Coleman, Modelling relative survival in the presence of incomplete data: a tutorial. *International Journal of Epidemiology*, **39**, 118-128, 2010. R10007.

[1628] Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M. and Carpenter, J.R., Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *British Medical Journal*, **339**, b2393, 2009. R09223.

[1629] Vansteelandt, S., Carpenter, J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, **6**, 37-48, 2009. R09221.

D.9.3 Refereed publications (in press)

[1630] Burzykowski, T., Carpenter, J., Coens, C., Evans, D., France, L., Kenward, M., Lane, P., Matcham, J., Morgan, D., Philips, A., Roger, J., Sullivan, B., White, I. and L.M. Yu, of the PSI missing data expert group, Missing data: discussion points from the PSI missing data expert group. *Pharmaceutical Statistics* (to appear), 2009. RP10061

[1631] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* (to appear), 2009. RP09030.

[1632] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Generalized shared-parameter models and missingness at random. *Statistical Modeling* (to appear), 2009. RP09043.

[1633] Goegebeur, Y., De Boeck, P., Molenberghs G. and M.G. Kenward, Person fit for test speededness: Normal curvature, likelihood ratio tests, and empirical Bayes estimates. *Methodology* (to appear), 2009. RP09006.

[1634] Kenward, M.G. and G. Molenberghs, Last observation carried forward: a crystal ball ? *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09054.

[1635] Molenberghs, G., Kenward, M.G. and G. Verbeke, Discussion of Lee, Y. and Nelder, J.A.: Likelihood inference for models with unobservables: another view. *Statistical Science* (to appear), 2009. RP09048.

[1636] Molenberghs, G., Kenward, M.G., Verbeke, G. and B. Teshome Ayele, Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* (to appear), 2009. RP09037.

[1637] Schwarzer, G. Carpenter, J.R. and G. Rucker, Empirical evaluation suggests Copas selection model preferable to trim-and-fill method for selection bias in meta-analysis. *Journal of Clinical Epidemiology* (to appear), 2010. RP10059.

[1638] Youngjo, L., Manegseok, N. and M.G. Kenward, Robust estimation of dropout models using hierarchical likelihood. *Journal of Statistical Computation and Simulation* (to appear), 2010. RP10060.

D.9.4 Non-refereed publications (published)

[1639] Creemers, A., Hens N., Aerts, M., Molenberghs, G., Verbeke, G. and M. Kenward, A sensitivity analysis in the shared parameter modeling framework. *Proceedings of the 30th Annual Conference of the International Society for Clinical Biostatistics*. University of Economics, Prague, 2009. NR09007.

D.10 List of joint publications

D.10.1 Technical reports

[1640] Antoniadis, A., Gijbels, I. and A. Verhasselt, Variable selection in additive models using P-splines, 2009. TR09068. (KUL-1, UJF)

[1641] Gijbels, I., Omelka, M. and D. Sznajder, Positive quadrant dependence tests for copulas, 2009. TR09061. (KUL-1, UH)

[1642] Gijbels, I., Veraverbeke, N. and M. Omelka, Conditional copulas, association measures and their applications, 2009. TR09060. (KUL-1, UH)

[1643] Veraverbeke, N., Omelka, M. and I. Gijbels, Estimation of a conditional copula and association measures, 2009. TR09057. (KUL-1, UH)

D.10.2 Refereed publications (published)

[1644] Antoniadis, A., Bigot, J. and R. von Sachs, A multiscale approach for statistical characterization of functional images. *Journal of Computational and Graphical Statistics*, **18**, 1, 216-237, 2009. R09231. (UCL, UJF)

[1645] Antoniadis, A., Garcia-Magarinos, M., Cao, R. and W. González-Manteiga, Lasso logistic regression, gsoft and the cyclic coordinate descent algorithm. application to gene expression data. *Statistical Applications in Genetics and Molecular Biology*, 2009. R09082. (UJF, USC)

[1646] Antoniadis, A., Gijbels, I. and J.-M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, Special issue: Multiscale Methods and Statistics:A Productive Marriage, **19**, 1371-1387, 2009. R09048. (KUL-1, UJF)

[1647] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions. *Journal of Statistical Planning and Inference*, **139**, 2800-2815, 2009. R09026. (UCL, KUL-1)

[1648] Beunckens, C., Sotto, C., Molenberghs, G. and G. Verbeke, A multifaceted sensitivity analysis of the Slovenian Public Opinion Survey data. *Applied Statistics*, **58**, 171-196, 2009. R09102. (KUL-2, UH)

[1649] Braekers, R. and I. Van Keilegom, Flexible modeling based on copulas in nonparametric median regression. *Journal of Multivariate Analysis*, **100**, 1270-1281, 2009. R09160. (UCL, UH)

[1650] Crujeiras, R. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends. *Computational Statistics and Data Analysis*, **54**, 452-465, 2010. R10004. (UCL, USC)

[1651] De Wit, L., Molas, M., Lesaffre, E., Dejaeger, E., Feys, H., Jenni, W., Lincoln, N., Eckhof, H., Putman, K. and W. De Weerdt, The use of a biplot in studying outcomes after stroke. *Neurorehabilitation and Neural Repair*, **23**, 825-830, 2009. R09074. (KUL-2, UH)

[1652] Faes, C., Molenberghs, G., Aerts, M., Verbeke, G. and M.G. Kenward, The effective sample size and a novel small sample degrees of freedom method. *The American Statistician*, **63**, 389-399, 2009. R09113. (KUL-2, UH, LSHTM)

[1653] Goethals, K., Ampe, B., Berkvens, D., Laevens, H., Janssen, P. and L. Duchateau, Modelling interval-censored, clustered cow udder quarter infection times through the shared gamma frailty model. *Journal of Agricultural, Biological and Environmental Statistics*, **14**, 1-14, 2009. R09151. (UG, UH)

[1654] Hens, N., Wienke, A., Aerts, M. and G. Molenberghs, The correlated and shared frailty model for bivariate current status data: an illustration for cross-sectional serological data. *Statistics in Medicine*, **28**, 2785-2800, 2009. R09111. (KUL-2, UH)

[1655] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, A family of measures to evaluate scale reliability in a longitudinal setting. *Journal of the Royal Statistical Society*, Series A, **172**, 237-253, 2009. R09108. (KUL-2, UH)

[1656] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Reliability of a longitudinal sequence of scale ratings. *Psychometrika*, **74**, 49-64, 2009. R09106. (KUL-2, UH)

[1657] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Using longitudinal data from a clinical trial in depression to assess the reliability of its outcome scale. *Journal of Psychiatric Research*, **43**, 730-738, 2009. R09103. (KUL-2, UH)

[1658] Lambert, P. and P.H.C. Eilers, Bayesian density estimation from grouped continuous data. *Computational Statistics and Data Analysis*, **53**, 1388-1399, 2009. R09007. (UCL, UU)

[1659] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model. *Lifetime Data Analysis*, **15**, 59-78, 2009. R09004. (UCL, UG, UH)

[1660] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 study Group, Evaluation of culture-based approaches for rapid detection of glycopeptide-resistant enterococci: a randomised, investigator-blinded study. *Clinical Microbiology and Infection*, **15**, 58-58, 2009. R09117. (KUL-2, UH)

[1661] Malhotra-Kumar, S., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M. and H. Goossens on behalf of the MOSAR WP2 Study Group, Rapid detection of extended-spectrum b-lactamaseproducing Enterobacteriaceae: a randomised, investigator-blinded evaluation of culture-based approaches. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09120. (KUL-2, UH)

[1662] Massonnet, G., Janssen, P. and L. Duchateau, Modelling udder infection data using copula models for quadruples. *Journal of Statistical Planning and Inference*, **139**, 3865-3877, 2009. R09150. (UG, UH)

[1663] Molanes Lopez, E. M., Van Keilegom, I. and N. Veraverbeke, Empirical likelihood for non-smooth criterion functions. *Scandinavian Journal of Statistics*, **36**, 413-432, 2009. R09019. (UCL, UH)

[1664] Molenberghs, G., Beunckens, C., Jansen, I., Thijs, H., Verbeke, G. and M.G. Kenward, Missing data. In: *Handbook of Epidemiology*, I. Pigeot and W. Ahrends (eds.), Heidelberg: Springer, 2009. R09122. (KUL-2, UH, LSHTM)

[1665] Molenberghs, G. and M.G. Kenward, Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics and Data Analysis*, **54**, 585-597, 2009. R09114. (UH, LSHTM)

[1666] Omelka, M., Gijbels, I. and N. Veraverbeke, Improved kernel estimation of copulas: weak convergence and goodness-of-fit testing. *The Annals of Statistics*, **37**, 3023-3058, 2009. R09065. (KUL-1, UH)

[1667] Sabiiti, W., Cortiñas Abrahantes, J., Lammens, C., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Evaluation of commercially available molecular and culturebased assays for rapid detection of methicillin-resistant Staphylococcus aureus. *Clinical Microbiology and Infection*, **15**, 440-440, 2009. R09119. (KUL-2, UH)

[1668] Serroyen, J., Molenberghs, G., Verbeke, G. and Davidian, M., Nonlinear models for longitudinal data. *The American Statistician*, **63**, 378-388, 2009. R09112. (KUL-2, UH)

[1669] Sotto, C., Beunckens, C., Molenberghs, G., Jansen, I. and G. Verbeke, Marginalizing pattern-mixture models for categorical data subject to monotone missingness. *Metrika*, **69**, 305-336, 2009. R09104. (KUL-2, UH)

[1670] Tilahun, A., Maringwa, J.T., Geys, H., Alonso, A., Raeymaekers, L., Molenberghs, G., Van Den Kieboom, G., Drinkenburg, P. and L. Bijnens, Investigating association between behavior, corticosterone, heart rate, and blood pressure in rats using surrogate marker evaluation methodology. *Journal of Biopharmaceutical Statistics*, **19**, 133-149, 2009. R09109. (KUL-2, UH)

[1671] Van Deun, K., Hoijtink, H., Lieven, T., Van Lommel, L., Schuit, F. and I. Van Mechelen, Testing the hypothesis of tissue-selectivity: The Intersection-Union Test and a Bayesian approach. *Bioinformatics*, **25**, 2588-2594, 2009. R09044. (KUL-1, UU)

[1672] Van Effelterre, T., Shkedy, Z., Aerts, M., Molenberghs, G., Van Damme, P. and P. Beutels, Contact patterns and their implied basic reproductive numbers: an illustration for varicella-zoster virus. *Epidemiology and Infection*, **137**, 48-57, 2009. R09115. (KUL-2, UH)

[1673] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 Study Group, Comparison of the impact of direct plating versus a short or overnight pre-enrichment on detection of methicillin-resistant Staphylococcus aureus from clinical specimens. *Clinical Microbiology and Infection*, **15**, 208-208, 2009. R09118. (KUL-2, UH)

[1674] Van Heirstraeten, L., Cortiñas Abrahantes, J., Lammens, C., Lee, A., Harbarth, S., Molenberghs, G., Aerts, M., Goossens, H. and S. Malhotra-Kumar on behalf of the MOSAR WP2 study group, Impact of a short period of pre-enrichment on detection and bacterial loads of methicillin-resistant Staphylococcus aureus from screening specimens. *Journal of Clinical Microbiology*, **47**, 3326-3328, 2009. R09116. (KUL-2, UH)

[1675] Vansteelandt, S., Carpenter, J.R. and M.G. Kenward, Analysis of incomplete data using inverse probability weighting and doubly robust estimators. *Methodology*, **6**, 37-48, 2009. R09221. (UG, LSHTM)

[1676] Verbeke, G., Molenberghs, G. and D. Rizopoulos, Random effects models for longitudinal data. In: *Longitudinal Research with Latent Variables*. K. van Montfort, H. Oud and Al Satorra (eds.), New York: Springer, 2009. R09123. (KUL-2, UH)

D.10.3 Refereed publications (in press)

[1677] Acuña, C., Leborán, V., Maringwa, J.T., Faes, C., Geys, H., Molenberghs, G., Cadarso Suárez, C.M. and J.L. Pardo Vázquez, Application of penalized smoothing splines in analyzing neuronal data. *Biometrical Journal* (to appear), 2009. RP09091. (UH, USC)

[1678] Alonso, A., Laenen, A., Molenberghs, G., Geys, H. and T. Vangeneugden, Reliability of single administered tests: a unified approach. *Biometrics* (to appear), 2009. RP09042. (KUL-2, UH)

[1679] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties. *The Annals of the Institute of Statistical Mathematics* (to appear), 2010. RP10013. (KUL-1, UJF)

[1680] Assam, P., Tilahun, A., Alonso, A. and G. Molenberghs, Using earlier measures in a longitudinal sequence as potential surrogate for a later one. *Computational Statistics and Data Analysis* (to appear), 2009. RP09047. (KUL-2, UH)

[1681] Bruckers, L., Serroyen, J., Molenberghs, G., Slaets, H. and W. Goeyvaerts, Latent-class analysis of persistent disturbing behavior patients using longitudinal profiles. *Applied Statistics* (to appear), 2009. RP09053. (KUL-2, UH)

[1682] Burzykowski, T., Carpenter, J. Coens, C., Evans, D., France, L., Kenward, M., Lane, P. , Matcham, J., Morgan, D., Philips, A., Roger, J., Sullivan, B., White, I. and L.M. Yu, of the PSI missing data expert group, Missing data: discussion points from the PSI missing data expert group. *Pharmaceutical Statistics* (to appear), 2009. RP10061. (UH, LSHTM)

[1683] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, A sensitivity analysis for shared-parameter models for incomplete longitudinal outcomes. *Biometrical Journal* (to appear), 2009. RP09030. (KUL-2, UH, LSHTM)

[1684] Creemers, A., Hens, N., Aerts, M., Molenberghs, G., Verbeke, G. and M.G. Kenward, Generalized shared-parameter models and missingness at random. *Statistical Modeling* (to appear), 2009. RP09043. (KUL-2, UH, LSHTM)

[1685] De Wolf, J., Duchateau, L., Verbeke, G., and E. Schrevens, Discovering transgenic elite events: Using information from early screening trials for improving experimental design. *Journal of Agricultural, Biological, and Environmental Statistics* (to appear), 2009. RP09017. (KUL-2, UG)

[1686] Goegebeur, Y., De Boeck, P., Molenberghs, G. and M.G. Kenward, Person fit for test speededness: normal curvatures, likelihood ratio tests and empirical Bayes estimates. *Methodology* (to appear), 2009. RP09006. (KUL-1, UH, LSHTM)

[1687] González-Manteiga, W., Pardo Fernández, J.C. and I. Van Keilegom, ROC curves in nonparametric location-scale regression models. *Scandinavian Journal of Statistics* (to appear), 2010. RP10009. (UCL, USC)

[1688] Guerra-Yi, M.E., Lorenzo-Luaces Alvarez, P., Faes, C., Galàn Alvarez, Y. and G. Molenberghs, Spatial analysis of breast and cervix cancer incidence in small geographical areas in Cuba, 1999-2003. *European Journal of Cancer Prevention* (to appear), 2009. RP09028. (KUL-2, UH)

[1689] Jacobs, T., Straetemans, R., Molenberghs, G., Bouwknecht, A. and L. Bijnens, A latent pharmacokinetic time profile to model dose-response survival data. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09027. (KUL-2, UH)

[1690] Jansen, I. and G. Molenberghs. Pattern-mixture models for categorical outcomes with non-monotone missingness. *Journal of Statistical Computation and Simulation* (to appear), 2009. RP09029. (KUL-2, UH)

[1691] Kasim, A., Lin, D., Van Sanden, S., Clevert, D., Bijnens, L., Gohlmann, H., Amaratunga, D., Hochreiter, S., Shkedy, Z. and W. Talloen, Informative or non-informative calls for gene expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* (to appear), 2009. RP09050. (KUL-2, UH)

[1692] Kenward, M.G. and G. Molenberghs, Last observation carried forward: a crystal ball ? *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09054. (UH, LSHTM)

[1693] Laenen, A., Alonso, A., Molenberghs, G., Vangeneugden, T. and C. Mallinckrodt, Coping wit memory effect and serial correlation when estimating reliability in a longitudinal framework. *Applied Psychological Measurement* (to appear), 2009. RP09033. (KUL-2, UH)

[1694] Legrand, C., Duchateau, L., Janssen, P., Ducrocq, V. and R. Sylvester, Validation of prognostic indices using the fraitly model - Validation of prognostic indices using the frailty model. *Lifetime Data Analysis* (to appear), 2009. RP09001. (UCL, UG, UH)

[1695] Molenberghs, G., Burzykowski, T., Alonso, A., Assam, P., Tilahun, A. and M. Buyse, A unified framework for the evaluation of surrogate endpoints in clinical trials. *Statistical Methods in Medical Research* (to appear), 2009. RP09020. (KUL-2, UH)

[1696] Molenberghs, G., Kenward, M.G. and G. Verbeke, Discussion of Lee, Y. and Nelder, J.A.: Likelihood inference for models with unobservables: another view. *Statistical Science* (to appear), 2009. RP09048. (KUL-2, UH, LSHTM)

[1697] Molenberghs, G., Kenward, M.G., Verbeke, G. and B. Teshome Ayele, Pseudo-likelihood estimation for incomplete data. *Statistica Sinica* (to appear), 2009. RP09037. (KUL-2, UH, LSHTM)

[1698] Molenberghs, G. and G. Verbeke, Longitudinal and incomplete clinical studies. *Journal of the Japanese Society of Computational Statistics* (to appear), 2009. RP09035. (KUL-2, UH)

[1699] Mwambi, H., Ramroop, S., Shkedy, Z. and G. Molenberghs, An approach to estimating the force of infection and the rate of recovery fom Respiratory Syncytial Virus (RSV) data. *Statistical Methods in Medical Research* (to appear), 2010. RP10069. (KUL-2, UH)

[1700] Rizopoulos, D., Verbeke, G., and G. Molenberghs, Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival data. *Biometrics* (to appear), 2009. RP09024. (KUL-2, UH)

[1701] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in linear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09026. (KUL-2, UH)

[1702] Serroyen, J., Molenberghs, G., Aerts, M., Vloeberghs, E., De Deyn, P.P. and G. Verbeke, Flexible estimation of serial correlation in nonlinear mixed models. *Journal of Applied Statistics* (to appear), 2009. RP09051. (KUL-2, UH)

[1703] Shkedy, Z., Lin, D., Molenberghs, G., Goehlmann, H., Talloen, W. and L. Bijnens, Genomic biomarkers for depression: Feature-specific and joint biomarkers. *Statistics in Biopharmaceutical Research* (to appear), 2009. RP09036. (KUL-2, UH)

[1704] Tilahun, A., Lin, D., Shkedy, Z., Geys, H., Alonso, A., Peeters, W.P., Talloen, W., Drinkenburg W., Goehlmann, H.W.H., Gorden, E., Bijnens L. and G. Molenberghs, Selection and evaluation of genomic biomarkers for depression in a microarray setting. *Journal of Biopharmaceutical Research* (to appear), 2009. RP09052. (KUL-2, UH)

[1705] Vangeneugden, T., Molenberghs, G., Laenen, A., Geys, H., Beunckens, C., and C. Sotto, Marginal correlation in longitudinal binary data based on generalized linear mixed models. *Communications in Statistics, Theory and Methods* (to appear), 2009. RP09034. (KUL-2, UH)

[1706] Vangeneugden, T., Molenberghs, G., Verbeke, G. and C. Demétrio, Marginal correlation from an extended random-effects model for repeated and overdispersed counts. *Journal of Applied Statistics* (to appear), 2009. RP09038. (KUL-2, UH)

[1707] Verbeke, G. and G. Molenberghs, Arbitrariness of models for augmented and coarse data, with emphasis on incomplete-data and random-effects models. *Statistical Modelling* (to appear), 2009. RP09021. (KUL-2, UH)

[1708] Wouters, K., Cortiñas Abrahantes, J., Molenberghs, G., Geys, H., Ahnaou, A., Drinkenburg, W.H.I.M. and L. Bijnens, Correction for model selection bias using a modified model averaging approach for supervised learning methods applied to EEG experiments. *Journal of Biopharmaceutical Statistics* (to appear), 2009. RP09045. (KUL-2, UH)

D.10.4 Non-refereed publications (published)

[1709] Creemers, A., Hens N., Aerts, M., Molenberghs, G., Verbeke, G. and M. Kenward, A sensitivity analysis in the shared parameter modeling framework. *Proceedings of the 30th Annual Conference of the International Society for Clinical Biostatistics*. University of Economics, Prague, 2009. NR09007. (KUL-2, UH, LSHTM)

[1710] Laenen, A., Alonso, A., Molenberghs, G. and T. Vangeneugden, Estimating the reliability of rating scales from clinical trial data. *Proceedings of the 16th International Meeting of the Psychometric Society*. Cambridge, 2009. NR09002. (KUL-2, UH)

[1711] Vankerckhoven, V., Lammens, C., Chapelle, S., Vanderstraeten, A., Hens, N., Molenberghs, G. and H. Goossens, Analysis of the gut microbiotica in irritable bowel syndrome patients and healthy volunteers using selective media and real-time PCR. In: *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 2009. NR09003. (KUL-2, UH)

[1712] Vankerckhoven, V., van Voorden, S., Hens, N., Molenberghs, G., Goossens, H. and E. Wiertz, Activation of Toll-like receptor 2 and Toll-like Receptor 2/6 by lacobacilli. *19th European Congress of Clinical Microbiology and Infectious Diseases*, Helsinki, 2009. NR09004. (KUL-2, UH)

D.10.5 Books (published)

[1713] Fitzmaurice, G., Davidian, M., Molenberghs, G. and G. Verbeke, *Longitudinal Data Analysis*, Handbooks of Modern Statistical Methods. New York: Chapman and Hall, 2009. B09002. (KUL-2, UH)

[1714] Verbeke, G. and G. Molenberghs, *Linear Mixed Models for Longitudinal Data* (Paperback Edition). New York: Springer, 2009. B09003. (KUL-2, UH)