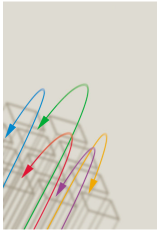
BELGIAN SCIENCE POLICY



Progress Report 2008

IAP-Network in Statistics

Contract P6/03

March 31, 2009

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

Contents

1 Accomplished Research Projects

1.1 Introduction and overview

1.1.1 Introduction

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

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

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

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

1.1.2 Overview

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

- In WP1, important contributions were made to the following topics: (1) quantile estimators of frontiers have been introduced and studied and the use of extreme value theory has been (partially) exploited and this opens new perspectives; (2) new semi-and nonparametric methods for estimating (non)smooth curves have been developed; there was a special focus on error variance and dispersion function estimation; (3) studies of the impact of penalty functions and data-driven choices of these in penalized regression; (4) development of a bootstrap procedure for functional data; (5) robust multivariate methods, including fact algorithms, model selection criteria and fast and robust bootstrap procedure; (6) modelling dependencies via copulas with specific attention to extreme-value copulas, modelling excesses, semiparametric modelling and dependencies between variables that depend on the given value of a covariate.
- The work of WP2 has been focused on modeling and doing inference for complex correlated data. Fundamental research results were provided on the nonparametric analysis of time series and spatial processes. Several types of complexity have been considered: unsmooth spectral density, nonstationarity, multivariate nonstationarity, censoring, multilevel

modelling, continuous-time processes and spatial data. In particular, the network provided new workable results on the adaptive, data-driven nonparametric estimation of temporally and spatially related data.

- The research endeavor of WP3 has placed emphasis on the dissemination and further development of models for incomplete and censored data and on sensitivity analysis for such, with focal points on clinical trials, epidemiological studies, surveys, and sociological and psychometric applications, often in conjunction with the propagation of proper methodology for longitudinal data, including latent structures, mixed-models, and meta-analytic methods, sometimes of high dimension, thus strongly interacting with WP1, WP4, and WP5. The output realized takes the form of journal publications, as well as 3 books and an edited volume. Dissemination has further taken place through short courses in Belgium, Europe, and the economically developed and developing parts of the world, and through dedicated IAP courses and seminars.

- Many statistical models in which latent structures are used to explain observed heterogeneity have been developed and investigated in WP4. The classical linear, generalized linear and non-linear mixed effects models have been generalized by allowing the random effects distribution to belong to a general class of continuous and/or discrete distributions. To this end, frequentist and Bayesian approaches have been employed and allowed to verify the impact of the correct specification of the random effects distribution on the estimation of various parts of the model. Latent heterogeneity has also been studied in survival models (including interval censoring and involving copula models), change point models, multivariate random effects, hierarchical and shared parameter models. New parameter estimation procedures have been developed as well as score tests for the random effects structure and techniques for prediction. Data from various disciplines (medicine, psychology, etc.) were taken to illustrate the performance of the various proposals.

- The work of WP5 has placed emphasis on bioinformatics, data mining and psychometrics. In bioinformatics, both the analysis and experimental design of genetic data is studied. More specifically, new statistical analysis techniques are developed for micro-array data. In data mining, a substantial part of the research consists of the development of robust methods for highdimensional data. Furthermore, principal and independent component analysis techniques are used to analyze chemometrical data. In psychometrics, hierarchical classes models and multi-mode models are studied. Different statistical tools, such as a fully Bayesian method, random effects techniques, and bootstrap techniques are used in this context.

1.2 Work package 1: Multivariate data with qualitative constraints

1.2.1 Boundaries, frontiers, and efficiency and productivity analysis

Results obtained on statistical inference in nonparametric frontier models have been collected in [59]. Most of the techniques use the bootstrap and [51] give the full theory for the asymptotics and for the consistency of the bootstrap when using DEA estimators. In [38], new properties of robust estimators of frontiers (based on nonstandard conditional quantiles) are derived.

Nonparametric truncated regression techniques, based on local likelihood theory, are in [56]. This is particularly useful when trying to explain in two-stage approaches the efficiency of firms in terms of some environmental variables: here we regress efficiency scores on these explanatory variables. A special issue of *Journal of Productivity Analysis* analyzes how the economies in the former Eastern European countries have been affected by the transition, in particular, in terms of productivity and of efficiency. As co-editors of this special issue, [53] introduce this volume.

New properties of FDH estimators and of quantile estimators of frontiers have been developed in [7] by using properties from Extreme Value Theory. Estimators of the frontier having a normal asymptotic distribution have been defined. Previous results on stochastic nonparametric frontier models have been adapted in [31] to full multivariate setups (multi-input and multi-output).

[1] provide a data-driven technique to derive optimal bandwidths for analyzing conditional efficiency measures. When a panel of data is available, a dynamic procedure has been proposed in [8] to predict the changes in productivity and its components (efficiency changes, technological changes and scale efficiency changes).

1.2.2 Nonparametric and semiparametric estimation of curves and surfaces, and estimation under qualitative constraints

Unknown functions and surfaces are not always smooth, and may show different types of irregularities in different regions. In [137] an overview is given on the use of local linear fitting for smoothing while preserving irregularities. In [104] such techniques are used to obtain estimates that have a better behavior near peaks.

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 [513] 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.

[108] looks into the problem of estimation of under- or overdispersion. Especially overdispersion of data is a well-known phenomenon when analysing data. A unified approach to nonparametrically estimating mean and dispersion function is provided in [108].

The papers [109] and [110] look into penalized regression with a difference type of penalty. These papers contribute in making a data-driven choice of the differencing order.

Consider a vector of criterion functions g depending on a parameter of interest μ and a nuisance parameter. In [553] a novel empirical likelihood (EL) method to do inference for μ, not requiring any smoothness assumptions on g is developed. Hence, g may involve indicators, which are encountered when considering e.g. differences of quantiles, copulas or ROC curves, which were previously studied by means of a smoothed EL approach. ROC curves are a useful tool to analyze the discrimination capability of a diagnostic variable in medical studies. The presence of a covariate related to the diagnostic variable can increase the discriminating power of the ROC curve. [519] models the effect of the covariate over the diagnostic variable by means of nonparametric location-scale regression models, and studies a new nonparametric estimator of the conditional ROC curve.

In [25] the error distribution is estimated in a nonparametric location-scale regression model with multivariate covariates. 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. [34] proposes a general method to estimate the dispersion function in a semiparametric way, and obtains generic conditions under which the proposed estimator satisfies certain asymptotic properties. [54] propose consistent estimators in a semiparametric transformation model, where the transformation is modelled parametrically, and all other components of the model are non- or semiparametric.

In [11] 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, and study an application to the construction of confidence intervals.

Statistical analysis of images is another research area within the project. In [539] 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 [435] cloud detection from geostationary satellite multispectral images through statistical methodologies is investigated. Discriminant analysis methods are considered to this purpose, endowed with a nonparametric density estimation and a linear transform into principal and independent components.

1.2.3 Nonparametric and semiparametric testing procedures

[40] and [41] consider a nonparametric location-scale regression model, in which the error is independent of the covariate, and the regression and variance function are smooth unknown functions. They construct tests for the validity of this model. The test proposed in [41] is based on the nonparametric estimation of the errors, whereas in [40] the test is based on differences of neighboring responses.

An important problem in regression is the development of goodness-of-fit tests for (semi)-parametric regression models. In [536], 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. On the other hand, [537] propose an empirical likelihood test that is able to test the goodness-of-fit of a class of parametric and semiparametric regression models. A third contribution is in [64] where the procedure 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.

1.2.4 Multivariate data, robust analysis and nonparametric inference

[249] introduces the least trimmed squares estimator for the multivariate regression model. The estimator reduces to the well-known LTS estimator in the case of a univariate response (multiple regression) and to the MCD estimator in the case of an intercept only model. The breakdown value, consistency at elliptical error distributions, the influence function and efficiency are derived. A fast algorithm is given and reweighted estimators are studied as well. [531] gives an overview of recent high-breakdown robust methods for multivariate settings such as covariance estimation, multiple and multivariate regression, discriminant analysis, principal components and multivariate calibration. [275] considers robust model selection criteria for linear regression that focus on robust measures of the expected prediction error estimated using the fast and robust bootstrap. The model selection criteria are consistent and robust and are feasible for real data applications. A fast and robust bootstrap method in general is provided in [276] which also considers its application to perform inference based on robust estimators for the 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. [282] discusses computationally efficient procedures for robust variable selection in linear regression. These robust variable selection procedures can easily handle missing (completely at random) data.

In [540] 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.

A review on the use of nonparametric methods and methods for robust analysis of data is given in [169].

1.2.5 Modelling and measuring of dependencies and copula functions

In [5] a complete and user-friendly directory of tails of Archimedean copulas is presented which can be used in the selection and construction of appropriate models with desired properties. The results are synthesized in the form of a decision tree.

In the world of multivariate extremes, estimation of the dependence structure still presents a challenging problem. In [39] a procedure for the bivariate case is presented that opens the road to a similar way of handling the problem in a truly multivariate setting. A method of moments estimator is proposed. A major issue in multivariate extreme value theory is the estimation of the spectral measure Φ_p with respect to the L_p norm. This problem is studied in [10].

The set of Pickands dependence functions for bivariate extreme-value copulas can be seen as a closed and convex parameter set embedded in a real Hilbert space. [43] looks into the estimator resulting from using a pilot estimate possibly falling outside the parameter set and projected onto this parameter set. Asymptotic properties of the initial estimator sequence in the Hilbert space topology transfer easily to those of the projected sequence and certain finite-dimensional approximations.

Modelling excesses over a high threshold using the Pareto or generalized Pareto distribution (PD/GPD) is the most popular approach in extreme value statistics. An extension of the (G)PD is in [514] yielding a statistical model that can be fitted to a larger portion of the data.

In the model where the copula is known and the marginal distributions are completely unknown, the empirical distribution functions are semiparametrically efficient if and only if the copula is the independence copula. Incorporating the knowledge of the copula into a nonparametric likelihood yields an estimation procedure which by simulations is shown to outperform the empirical distribution functions, the amount of improvement depending on the copula. See [30].

[544] propose to model the relationship between the error and the covariate in a nonparametric regression model by means of a copula model. This model offers a useful alternative for the model that assumes the independence between these two variables, since the latter assumption is often violated in practice.

[554] study an improved nonparametric estimator of the copula function. The improvement is shown to be very considerable, and weak convergence of the proposed estimator is established. The estimation procedure is then used in a goodness-of-fit testing problem for copula's. It often occurs that the dependence structure of two (or more) random variables depends on the given value of another variable (a covariate). This led to the concept of conditional copula function of which nonparametric estimation is studied in [524]. The resulting empirical copula process depends on the covariate value, and its weak convergence is established.

Contingency tables are useful for analyzing the associations between several variables. Exploring the effect of explanatory variables characterizing the individuals on the structure of association may be important (to detect, for instance, spurious dependence). In [14] and [15] single index models (SIM) are used to model the conditional probabilities in a two-way table. This allow to develop testing procedures, e.g. to test the conditional independence in the table.

1.2.6 Interactions with other Work packages

In WP1 quite some work has overlap with other WP's. We just mention a few of such papers (with between brackets the WP with which the work has overlap).

[439] proposes a method to select the bandwidth for functional time series prediction. The method consists of calculating the empirical risk of prediction using past segments of the observed series and to selecting as value of the bandwidth for prediction the bandwidth which minimizes this risk. (WP2)

A key motivation of [432] is to study the variable selection problem for Cox's proportional hazards function regression models using a framework that extends the theory, the computational advantages and the optimal asymptotic rate properties of the Dantzig selector to the class of Cox's proportional hazards under appropriate sparsity scenarios. (WP3)

In [440] the authors consider estimation of the parameters of multivariate mixed Poisson models. They study a maximum pairwise likelihood approach when the mixing distribution is a multivariate Gamma distribution. Simulations show that the method is well adapted. An application to change detection in low-flux images is also investigated. (WP4)

1.3 Work package 2: Temporally and spatially related data

1.3.1 The nonparametric analysis of time series

The analysis of irregular time series is a big achievement of the network. Many types of irregularities are present in real data. One of them is the irregularity of the spectral density function of a stationary time series . In [45], 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. In [103] the considered problem is that of spectral density estimation in stationary time series with short range dependence. Improved estimation of the peaks in a spectral density is obtained by modifying the maximum likelihood estimator (based on the Whittle likelihood) using local linear approximations. In [69] a general approach of shrinkage of spectral density matrix estimators is presented that considerably improves on both the mean-squared error and on the conditioning number of the resulting estimators. On the other hand, in [36] 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 [24] 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 stimated by the eigenvectors of a nonparametrically estimated spectral density matrix. The delicate issue of bandwidth selection is also dealt with, comparing a few selectors, and showing a very good performance of the proposed data-dependent bandwidth selector.

Bandwidth selection in nonparametric kernel testing is a long-standing problem. In [168], this work is established in a framework of a nonparametric regression model, where the covariate is a sequence of strictly stationary time series variables. The paper proposes a sound approach to the problem by finding an Edgeworth expansion of the asymptotic distribution of the concerned test. The authors provide an easy implementable procedure for the practical realization of the established methodology, and they illustrate its performance by means of simulated and real-data examples.

In [442] a model for general patterns of digit preference is presented. It assumes a smooth latent distribution, which has been observed through a misclassification process, reflecting the tendency of people to round imprecise numbers. Seasonal mortality tables with monthly data have been analyzed in [443]. The data are modelled by a combination of a trend and modulated (co)sine waves. Both trend and modulation are modelled as a smooth surface, varying over time and age. An extension of this idea is presented in [449], where the (co)sines are replaced by an unknown periodic function, to be estimated from the data. The subject of [446] is to transform a variable in such a way that after the transformation its density fits well to a given density. Spatial variability of crop yield in agricultural trials can be modelled by tensor products of P-splines and can be treated like a mixed model. This is explained and applied in [447]. Conditional mean curves and conditional median curves are familiar. To them conditional quantile curves have been added in recent years. Quantile curves are based on asymmetric weighting of absolute values of residuals.

It is not well known that the same idea can be applied to sums of squares of residuals. Details and an application are found in [451].

After observing a real valued stationary time series admitting a density, [441] provides an adaptative estimator of the density. The wavelet based method is applied for different cases, and in particular to non stationary but geometrically ergodic cases like dynamical systems and Markovian fields on the line. In [434] they consider the maximum entropy extension of a partially specified autocovariance sequence of a periodically correlated process. 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.

1.3.2 Statistical analysis of complex time series data

One source of complexity comes from the potential nonstationarity of the data. As an output of a joint work between KUL and UCL, [528] developed new nonparametric testing procedures that simultaneously test for structural breaks in the conditional mean and the conditional variance. The asymptotic distribution of an adaptive test statistic is established, as well as its asymptotic consistency and efficiency. Simulations illustrate the performance of the adaptive testing procedure. An application to the analysis of financial time series also demonstrates the usefulness of the proposed adaptive test in practice. In [61], an explicit wavelet-based model of local stationarity is introduced. This model enlarges the class of locally stationary wavelet processes and contains processes whose spectral density function may change very suddenly in time. A notion of time-varying wavelet spectrum is uniquely defined as a wavelet-type transform of the autocovariance function with respect to so-called autocorrelation wavelets. This leads to a natural representation of the autocovariance which is localized on scales. A pointwise adaptive estimator of the time-varying spectrum has been proposed. The behavior of the estimator is studied in homogeneous and inhomogeneous regions of the wavelet spectrum.

Multivariate time series data were also considered. In [55], a local coherence analysis between two nonstationary time series is proposed. Coherence is a widely used measure for characterizing linear dependence between a pair of signals. For nonstationary signals, the autospectrum, cross spectrum, and coherence between signals may evolve over time. In this paper, a consistent estimation procedure is developed using time-localized linear filtering. The proposed method automatically selects, via repeated tests of homogeneity, the optimal window width for estimating local coherence. It is pointwise adaptive in the sense that the width of the optimal interval is allowed to change across time. Under the locally stationary process framework, a central limit theorem on the Fisher-z transform of the time-localized band coherence is proved.

Censoring of time series data is another problem sometimes encountered with real data. In [42], the authors consider a nonparametric regression model in which the data are dependent and the response is subject to random right censoring. They consider the problem of nonparametrically estimating the conditional quantile function in this model. The proposed method is based on a local linear fit using the check function approach. This paper is very much related to the research topics studied under WP1 and WP3.

Analysing time series data by multilevel models is useful for the practical use of inference, especially in the field of psychometrics. In [99], the authors propose a one/two-dimensional mixture model for DIF (1 dimension in the reference group, 2 dimensions in the focal group). A description of that and a means of estimating model parameters using easily available software are provided, with an application on test speededness. In [138], an item response model for dealing with test speededness is proposed. The model consists of two random processes, a problem solving process and a random guessing process.

Observed time series often appear to be non standard in their distribution. In [27], a sliding blocks estimator for the extremal index is proposed. In extreme value statistics for stationary sequences, blocks estimators are usually constructed by using disjoint blocks because exceedances over high thresholds of different blocks can be assumed asymptotically independent. In this paper the researchers focus on the estimation of the extremal index which measures the degree of clustering of extremes. They consider disjoint and sliding blocks estimators and compare their asymptotic properties. In particular they show that the sliding blocks estimator is more efficient than the disjoint version and has a smaller asymptotic bias. Other results on the tails of random sums of a heavy-tailed number of light-tailed terms was provided in [57].

The network also focussed on continuous-time processes. In [105], the Ratcliff diffusion model is used to decompose the effect of practice effect to the response time for repeatedly performing the same cognitive task. It is concluded that the practice effect consists of multiple subcomponents, and that abstracting these subcomponents in terms of a single output measure may be hazardous. This model was also considered in [156], where a Bayesian approach is used for diffusion process models of decision-making. A related MATLAB toolbox for Ratcliff diffusion models has been developed, see [155].

Continuous-time processes were also considered in multilevel, hierarchical modelling. In [118], 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. A crossed random effects diffusion model for speeded semantic categorization decisions is developed in [119]. In [123] a diffusion model for the analysis of continuous-time change in multivariate longitudinal data is proposed. The central idea is to model the data from a single person with an Ornstein-Uhlenbeck diffusion process. In [113], the authors develop a hierarchical continuous-time state-space model for repeated measurement data to model perpetually altering affective states. The model is described didactically and illustrated by an analysis of a diary study measuring affective quality.

1.3.3 Spatial data

Another goal of this work package is to analyse spatial data. In [516], 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. Weak consistency of the regression parameters estimator is derived, along with its asymptotic Gaussian limit. In [546] a new L_2 test for comparing spatial spectral densities is proposed. The goal of this paper is to test whether the dependence structure is the same for two spatial processes, that is, $H_0 : f_1 = f_2$, where each f_l denotes the spectral density of each observed process, for $l = 1, 2$. The authors study the asymptotic distribution for a Cramer-von-Mises type test statistic, under H_0 and under local alternatives. A simulation study showing the performance of the test is also provided.

[547] suggest to 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. As a particular case, they provide tests for independence. Asymptotic normal distribution of both statistics is obtained, under the null hypothesis. More general results on the second order properties of the multidimensional periodogram for regularly spaced data were also found in [556].

1.4 Work package 3: Incomplete data

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

1.4.1 Complex modeling approaches for missing data and sensitivity analysis tools

With **incomplete data**, specific modeling needs arise. First, it is important to ensure that proper, rather than simplistic, tools are used. A still commonly used method is last observation carried forward. Kenward and Molenberghs [520] 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 Mallinckrodt and Kenward [508].

Second, appropriate **modeling tools** need to be put forward. When fitting pattern-mixture models, for example, it may not always be straightforward how to derive marginal covariate effects, an issue studied by Sotto, Beunckens, Molenberghs, Jansen, and Verbeke [557]. In the same vein, Jansen and Molenberghs [339] proposed a flexible marginal modeling strategy for non-monotone missing data. Two key paradigms for dealing with incomplete data are using the predictive distribution of the unobserved outcomes on the one hand, such as in multiple imputation, and inverse probability weighting on the other, such as in conventional survey methodology. A comparison between both was undertaken by Beunckens, Molenberghs, and Sotto [315]. The LSHTM team has devoted additional research efforts to both multiple imputation, with a paper by Sterne *et al* [509] on the one hand, studying the use of multiple imputation in clinical and epidemiological studies, and a paper by Vansteelandt, Carpenter, and Kenward [558], investigating doubly-robust inverse probability weighting methods. The latter paper resulted from a collaboration with the UG team. As an alternative to existing but often computationally complex estimation methods, Sotto, Beunckens, Molenberghs, and Kenward [523] 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 [285], a new RMNE measure is proposed that likewise accounts for allelic drop-out in an observed forensic DNA profile. Also in [507], 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.

In [269], Goetgeluk, Vansteelandt, and Goetghebeur develop inverse probability weighting estimators, doubly robust estimators, and enriched doubly-robust estimators for controlled direct effects. In [287], Vansteelandt, Mertens, Suetens, and Goetghebeur 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. Vansteelandt, Babanezhad, and Goetghebeur show, in [296], 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 [550].

Third, the **dependence on assumptions** of models and their conclusions need to be carefully investigated. Such investigations are not a routine exercise, as demonstrated by Verbeke, Molenberghs, and Beunckens [538], who point out the dangers of using standard methodology and offer formal and informal sensitivity analysis tools. The shared-parameter framework has received considerable attention. In [226], Tsonaka, Verbeke and Lesaffre 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 semi-parametric method is used. Beunckens *et al* [525] formulate a so-called latent-class mixture model that brings together aspects of all three commonly used model families for incomplete data, i.e., selection models, pattern-mixture models, and shared-parameter models. This work offers an alternative basis for sensitivity analysis tools. Creemers, Hens, Aerts, Molenberghs, Verbeke, and Kenward [560] extended the class of shared-parameter models, so that it has become possible to incorporate the missing at random mechanism within it, a feat previously considered impossible.

Fourth, methods commonly used for the analysis of incomplete data rest on strong and unverifiable assumptions, calling for **sensitivity analysis**. The intrinsic nature of sensitivity, owing to non-identifiable model parts, is studied by Molenberghs, Beunckens, Sotto, and Kenward [533], who established that every missing not at random model has got a missing at random counterpart with equal fit. Beunckens, Sotto, Molenberghs, and Verbeke [542, 543] 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. Specifically for the context of shared-parameter models, and based upon the framework laid out in [560] Creemers, Hens, Aerts, Molenberghs, Verbeke, and Kenward [515], developed sensitivity analysis tools.

1.4.2 Censored survival data

Let us now turn to, possibly **censored time-to-event data**. In [320], Braekers and Veraverbeke extended the Koziol-Green model to dependent censoring. Weak convergence has been shown of a non-parametric estimator of the conditional distribution functional and its efficiency over the general copula-graphic estimator has been shown. Gaddah and Braekers [383] gave further consideration and proved the weak convergence of the conditional copula-graphic estimator. Furthermore, they develop a asymptotic confidence band for this estimator. In [367], Veraverbeke explored non-parametric estimators for several functionals of the distribution function when the observations are subject to right censoring and when covariate information is taken into account. In this paper, new results are obtained for the mean and median of the residual lifetime in a regression context. In non-parametric regression, it is common to assume independence between the error and the covariate. Braekers and Van Keilegom [224] generalized this by assuming a family of copula functions instead of independence. They established the asymptotic normality of the parameters of the copula family and the weak convergence of the conditional distribution of the response.

Estimation of copula functions is also a point of interest. In [553], Molanes, Van Keilegom and Veraverbeke obtained 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 has been done in Omelka, Gijbels and Veraverbeke [554]. Weak convergence is the main result and goodness-of-fit testing is one of the applications.

For censored data, non-parametric estimation of a density function, using ideas of pre-smoothing have been studied in [143]. Pre-smoothing of the Kaplan-Meier estimator has been shown to be more efficient. In [143] 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.

In [436], Antoniadis, Fryzlewicz, and Letué proposed a semi-parametric 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 proposed maximal partial likelihood estimators, establish the large sample properties of the estimators, and illustrated the results by a short simulation study and an application to a real data set in demography.

In [345], Massonnet, Janssen, and Burzykowski 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. In [561], Massonnet, Janssen, and Duchateau study a semi-parametric and a nonparametric two-stage estimation approach for copulas for four-dimensional survival data. They briefly discuss the asymptotic behaviour of the estimators obtained in the first and the second step of the estimation. A pseudo likelihood ratio test is used to select an appropriate copula from the power variance copula family. We propose a bootstrap algorithm to obtain a p-value for this test.

UCL has been very active in the area of regression with incomplete, mostly censored, data. Hence, there is a close connection with WP1. [21] addressed estimating non-parametrically 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 presented a completely non-parametric 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 [22], the authors first proposed 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 assumed that the relation between X and Y is given by a single-index model, and proposed a new estimator of the parameters in this model. Further, consider a non-parametric location-scale regression model, where the error is independent of the covariate. The response and the covariate are subject to selection bias. [26] constructed tests for the hypothesis that the regression function belongs to some parametric family. The proposed tests compare the non-parametric MLE (NMLE) based on the residuals obtained under the assumed parametric model, with the NPMLE based on non-parametric residuals. In [32], a semi-parametric time-varying coefficients regression model was 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 constructed an omnibus goodness-of-fit test for this model, and they developed 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 non-parametric location-scale regression model. Under this model, [33] proposed a non-parametric estimator of the distribution of the error variable. They then use the proposed estimator to introduce non-parametric 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.

In [458], under a nonparametric regression model, Boente, González-Manteiga and Pérez González 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. In [460], Cao and González-Manteiga 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 are usually referred as recurrent event, whereas if they represent different states (i.e., multi-state models) they are usually modeled thought their intensity functions. In [472], Meira-Machado, Cadarso Suárez, and de Uña Álvarez present non-parametric 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 Pérez González, Vilar Fernández, and González-Manteiga in [495] is the non-parametric estimation of the regression function with correlated errors when observations are missing in the response variable. Two non-parametric 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.

1.4.3 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 Verbeke and Molenberghs [559], is **data enrichment**. These authors, in line with [533], demonstrated 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 by Alonso, Litière, and Molenberghs [313, 342] for the random-effects distribution in the generalized linear mixed model.

Goegebeur, De Boeck, Wollack, and Cohen [138] proposed 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. Goegebeur, De Boeck, and Molenberghs [548] applied 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) and and empirical Bayes estimate of test speededness random effects.

Carpenter *et al* [506] provided empirical evidence that the so-called Copas selection model can usefully be used in a large fraction of meta-analyses in clinical studies.

1.4.4 Books and interactions

The **interactions** and collaborations between network partners in relation to WP3 and between WP3 and other packages, are manifold. 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. A lot of the work on survival analysis incorporates latent structures (WP4), termed frailties in this context, and, at the same time, deals with censoring (WP3). The book by Duchateau (UG) and Janssen (UH) [563], dedicated to the frailty model, combines incompleteness and latent structures. Much of the work on longitudinal data (WP4) faces issues of incompleteness (WP3), as well. Geert Verbeke (KUL-2) and Geert Molenberghs (UH) are co-editors, jointly with Garrett Fitzmaurice (Harvard School of Public Health, Boston, MA) and Marie Davidian (North Carolina State University, Raleigh, NC), on the state-of-the art volume on recent advances in longitudinal data analysis (WP4), with an important focus on incomplete data and sensitivity analysis (WP3) [564].

The sensitivity analysis work is often to be situated within the context of mixed models or other modeling context for hierarchical data, thus placing itself at the crossroads of WP3 and WP4. It is here that the work of Kenward and Carpenter [502] and by Carpenter and Plewis [505] are aimed at. Rizopoulos, Verbeke, and Molenberghs [535] 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).

In a collaboration between UG and Harvard-based colleagues, [301] discusses various ways to improve the design, monitoring, and analysis of late-stage randomized clinical trials of biomedical interventions to prevent HIV infection. The main goal of these endeavors is to increase the chances that such trials will detect a beneficial effect on HIV risk and to more fully assess how adherence to the intervention influences the said risk, as well as to assess the intervention's effect on HIV risk behavior. Causal inference methodology and dynamic treatment regimes are at the heart of the methodology used.

The work on the evaluation of surrogate markers from clinical studies, described in the next section, is located at the interaction between WP3, WP4, and WP5.

1.5 Work package 4: Data with latent heterogeneity

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 by Verbeke and Molenberghs [559]. Nevertheless, those models have proven to be very useful in various contexts, and some illustrations of this will be summarized in Section 1.5.1. In a number of applications, latent structures are needed at various levels, implying models with so-called crossed random effects. Some examples will be discussed in Section 1.5.2. 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. An number of examples will be discussed in Section 1.5.3. 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 1.5.4. Although the models studied in this work package 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 will be discussed in Section 1.5.5.

1.5.1 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 ditribution, often assumed (multivariate) normal. Those random effects can have various interpretations such as levels of general pracitioners in a multicenter study (Borgermans et al. [201]), levels of hospitals in a study comparing various hospitals with respect to patient safety (Lesaffre et al. [217]), or patient effects in longitudinal studies (De Wit et al. [205]). Research areas where random-effects models are frequently applied include the evaluation of surrogate endpoints in clinical trial settings ([349, 323, 322, 324]) in medical research, and item response, diffusion, and Rasch models in psychometric research ([138, 105, 133, 173]). 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 Oravecz, Tuerlinckx, and Vandekerckhove [113] and Molas and Lesaffre [237]. 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. Vangeneugden *et al* [363] used generalized linear mixed model to extend psychometric generalizability theory. Related, work, in the context of reliability, was done by Laenen *et al* [390, 391, 422]. Other examples of mixed models can be found in Van Nijlen and Janssen [157], González, De Boeck, Tuerlinckx [479], and Menten, Boelaert and Lesaffre [218].

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 Beunckens et al. [525] 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, Vanneste et al. [243] have applied 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. Frederickx et al. [106] study Differential Item Functioning (DIF) by introducing an item mixture model based on a Rasch model with random item difficulties, and simulations show that this new model performs better than other traditional methods. Finally, Van Nijlen and Janssen [121] have applied mixture item response theory models to distinguish between quantitative and qualitative differences in mastery and to use latent groups, rather then manifest groupings like gender or grade.

Another area where models with latent structures have proven very useful is in the evaluation of surrogate markers in clinical studies, where complex hierarchical structures are often neeeded. Molenberghs *et al* [349] studied the highly hierarchical meta-analytic framework for the evaluation of surrogate endpoints. Alonso and Molenberghs [311] employed information theory to provide both an elegant, unifying, mathematically sophisticated, and computationally simple framework, based on classical information theory. The framework was studied in further detail by Tilahun *et al* [360] and, in an invited review paper, by Molenberghs *et al* [397]. Alonso and Molenberghs [312] and Burzykowski *et al* [321] discussed the hopes and perils that rest with surrogate marker evaluation. Burzykowski *et al* [322, 323, 324] carefully studied surrogate marker evaluation in the context of oncology. A variety of methods to evaluation so-callled trial-level surrogacy was studied by Cortiñas, Shkedy, and Molenberghs [327]. Application of surrogate-marker methodology to different contexts were studied by Tilahun *et al* [361] and by Alonso *et al* [373].

1.5.2 Crossed random effects

Crossed random effects occur when random variation is to be introduced at different levels in the model. For example, González, Tuerlinckx, and De Boeck [111] used crossed random effects models for the analysis of complex designs with ANOVA-type models containing multivariate random effects for the analysis of interactions between fixed and (nested) random factors. Vandekerckhove, Verheyen, Tuerlinckx [119] use a crossed random effects diffusion model in the analysis of of choice reaction times, with item-specific and person-specific random effects. Finally, Janssen [174] proposes a random-effects extension of the linear logistic test model to model the effect of item design matrices within the Rasch model. The estimation of the resulting crossed random-effects model is discussed, as well as related models within item response theory.

1.5.3 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 Fieuws et al. [207] use 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 a joint model allows for a much better prognosis than what was available before based on univariate models. Rizopoulos et al. [221] and Tsonaka, Verbeke and Lesaffre have used latent variables to jointly model a survival outcome with a longitudinally measured binary outcome. In [97], Braeken, Tuerlinckx, and De Boeck use a latent variable framework 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.

1.5.4 Model extensions and/or flexible models

Many standard mixed and mixture models rely on very strict assumptions. In specific applications those assumptions need to be relaxed in order for the models to be realistic for specific data sets at hand. For example, Komárek and Lesaffre [211] use penalized Gaussian mixture model to analyze longitudinal random effects data with an emphasis on modeling the random effects distribution in a flexible manner. This allows to examine the impact of the Gaussian distribution on the estimation of the fixed effects part of the model. The conclusion was that the impact of the correct choice of random effects distribution varies from situation to situation. A similar model was used by Bogaerts and Lesaffre [200] for the estimation of the association between two interval-censored outcomes. They showed that the approach works well in many realistic situations and that it is superior to other existing approaches. In order to keep the model very flexible, all these authors allowed for (very) large amount of mixture components, and model fitting was based on a penalized likelihood approach. In a similar model but different context, Tsonaka, Verbeke, and

Lesaffre, [226] replaced the penalized likelihood approach by an estimation method based on the vertex exchange method, an algorithm borrowed from the finite mixture literature. Gampe and Eilers [448] compared mortality data from different countries (at the same age and in the same year) using models with a smooth, non-parametric, mixing distribution of the log mortality rates.

In a dental context, Mwalili, Lesaffre and Declerck [220] 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.

Vandekerckhove, Tuerlinckx, and Lee [118] have combined the Wiener diffusion process for choice response times with techniques from psychometrics in order to construct a hierarchical diffusion model. This lead to a modeling framework that is highly flexible and easy to work with. The same authors have also considered 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 [156].

1.5.5 Development of estimation methods and software

Since model extensions are often required to address subject-matter research questions adequately, standard estimation methods and/or standard software tools often do not suffice to fit the models at hand. In this work package, 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. Ghidey, Lesaffre, and Verbeke [208] have compared various methods for the estimation of the latent variable distribution, in the context of linear mixed models. Rizopoulos, Verbeke, and Lesaffre [238] propose to use fully exponential Laplace approximations, which have been shown to be fast, stable, and which yield valid inferences for the parameters of interest. An alternative has been presented by Rijmen, and Vomlel [152], who propose variational estimation, not requiring numerical evaluation of integrals, based on a lower bound approximation of the logistic model. Valdivieso, Schoutens, and Tuerlinckx [185] present extensive simulations to assess the properties of maximum likelihood estimates in a one-dimensional stationary process of Ornstein-Uhlenbeck type that is constructed via a self-decomposable distribution. In the context of frailty models for survival data, Massonnet, Janssen, and Burzykowski [345] have proposed a novel estimation method which uses a cumulative hazard based transformation to transform the data in such a way that the data can be analyzed using a linear mixed effects model. The properties of the method are studied by using a simulation study and a real life case study. Wetzels et al. [189] explore the statistical properties of the Expectancy Valence model of the Iowa gambling task. A Bayesian hierarchical estimation procedure is implemented and applied to data from an experimental study. Finally, Magis and Raiche [112] 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, Goldstein et al. [507] propose a MCMC algorithm for fitting a multilevel mixed model and apply this for multilevel multiple imputation.

As far as software development is concerned, De Boeck, Cho, and Wilson [99] propose an estimation method, possible with easily available softward, for the fitting of a one/two-dimensional mixture model for DIF. Vandekerckhove and Tuerlinckx [155] present a software tool freely downloadable, intended to make the Ratcliff diffusion model for reaction time and accuracy data more accessible to experimental psychologists. Braeken, Tuerlinckx, and De Boeck [98] developed a MATLAB IRT modeling (IRTm) toolbox that can fit a large variety of IRT models, with inclusion of recent developed copula IRT models to handle local item dependencies.

1.6 Work package 5: Highdimensional and compound data

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

1.6.1 Bioinformatics

At UG, Vansteelandt et al. [250] develop phylogenetic models that acknowledge dependence between neighboring nucleotides on the DNA/RNA string with corresponding MCMC inference techniques. Vansteelandt et al. [286] further develop gene-environment interaction tests for family-based association studies, which are robust to population admixture and for specification of the genetic model, other than the main genetic effect. Moerkerke and Goetghebeur [272] evaluated two separate approaches to avoid a dramatic loss in power when controling for the experiment-wise type I error when the association between many genetic markers and a trait is evaluated. First to increase power through a two stage design, allowing genes to be dropped or confirmed in a first phase and only use op the full sample size if there remains uncertaint. The second balances evidence from both the null and the alternative hypothesis before ranking the genes in order of promise. Here both approaches are combined in a set up for cost efficient screening. Van Nieuwerburgh et al. [249] evaluate two methods 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 drop-out can be assessed probabilistically. In this paper, a new RMNE measure is proposed that likewise accounts for allelic drop-out in an observed forensic DNA profile.

In phenotyping experiments with transgenic rice, different sources of variation, such as somaclonal variability and insertion variability occur. Dewolf et al. [267] propose mixed model techniques to separate the different sources of variability from each other to assess the effect of the inserted gene and its variability in a correct way.

At UH, Van Sanden et al. [366] 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. Lin et al. [341] investigated, 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.

Highdimensional data are encountered frequently in modern genetics. One problem, tackled by Rippe et al. [450] of UU, is to estimate genotypes of SNPs (single nucleotide polymorphisms) 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 Uh and Eilers [452].

At USC, Crujeiras and González-Manteiga [545] 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 baugmented Bootstrap. The performance of this estimator is checked by a simulation study and some real data applications.

1.6.2 Data mining

The development of robust techniques in data mining is partially a joint research effort of researchers of UG and KUL-1. Hubert et al. [531] review recent robuts methods for multivariate setting such as covariance estimation, multiple and multivariate regression, discriminant analysis, principal components, and multivariate calibration. Salibian-Barrera et al. [276] describe fast and robust bootstrap methods in general and consider its application to perform inference based on robust estimators for the 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.

At UH, Valkenborg et al. [364] 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. Valkenborg et al. [364] 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, Burzykowski et al. [413] review approaches, which use the concept of the isotopic distribution to discovering peptide related peaks in mass spectra.

At UCL, data mining in chemometrics is considered. 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. Debrus et al. [9] propose an integrated methodology based on Independent Components Analysis (ICA) and clustering to solve this problem and applies it ICA successively to HPLC-UV-DAD chromatograms. A new method for modeling chromatographic responses is presented by Lebrun et al. [52] 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. Rousseau et al. [58] 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.

At UU, Blasius et al. [444] advice on how the construction of biplots can be improved. One way to summarize and display highdimensional data is projection on principal components. Both observations and variable can be represented in biplots. Unfortunately most biplots are much less informative than they could be.

1.6.3 Psychometrics

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

Ceulemans and Van Mechelen [130] introduce a new model (CLASSI) to chart how individual differences in the stimuli-response (S-R) profiles are caused by individual differences in the stimuli-mediating variables (S-M) link and/or by individual differences in M-R link. An algorithm to fit CLASSI is described and evaluated. Depril et al. [134] propose three new algorithms to fit an overlapping additive clustering model in two-way two-mode data analysis. Simulations show that all three outperform the existing principal cluster model (PCL). González et al. [479] present a model which simulatenously accounts for the three ways in three-way data. Random effects are used to model the between-actor variability, and structural relations between the linking variables are investigated. Kuppens et al. [147] present a theoretical framework to explain individual differences in situation-specific emotional experience in terms of three different sources of variance. The relative contribution and nature of these sources was examined empirically for the experience of anger. Leenen et al. [148] present a stochastic extension of the hierarchical classes model for two-way two-mode binary data. A fully Bayesian method for fitting the new models is presented and tools for model selection and model checking are proposed. Schepers et al. [154] investigate four possible model selection heuristics in the framework of multi-mode partitioning models for N-way N-mode data. Performance of these four heuristics is systematically compared in a simulation study. Wilderjans et al. [158] present a global model for an integrated analysis of a three-way three-mode binary data array and a two-way two-mode binary data matrix that have one mode in common. The model is evaluated by means of a simulation study. Ceulemans and Kiers [164] suggest to consider that variance in three-way three-mode data can be decomposed into three components: random noise, strong CP or T3 structure, and weak T3 structure. Simulations based on re-analysis of real data sets show the improvements of this working hypothesis. Leenen and Ceulemans [178] compare INDCLASS and Tucker3-HICLAS hierarchical models for multi-way multi-mode data analyses. The underlying theory is introduced in substantive terms. Schepers and Hofmans [181] present a free, easy-to-use MATLAB graphical user interface (TwoMP) for two-mode partitioning models, specifically developed for nonspecialist users. The basic use of TwoMP is demonstrated using an example. Timmerman et al. [184] propose different bootstrap strategies for estimating confidence intervals of the parameters in multilevel simultaneous component analysis framework. Wilderjans et al. [190] compare two weighting schemes in the framework of coupled data blocks (N-way N-mode data blocks that have one or more modes in common) and a weighted overall objective function. Finally, Timmerman et al. [183] contributed a book chapter that deals with multilevel simultaneous component analysis, which models multivariate data that have been repeatedly gathered from more than one individual in an exploratory way. The goal of their modeling is to identify meaningful sources of the intra-individual variability in the observed variables, and to investigate whether and how the sources of intra-individual variability differ across individuals.

2 Network Activities

2.1 Web site and newsletter

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

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

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

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

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

2.2 Scientific meetings

2.2.1 Annual workshop

The annual workshop of 2008 entitled 'Missing Information in Survival Data Beyond Right Censoring' was organized by the University of Ghent on 18-19 September 2008. It 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'). The local organizer was Luc Duchateau. The workshop focused on new methodological developments for survival data when specific information is absent. Examples of missing information are interval-censoring, competing and semicompeting risks problems and frailty models. 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 work packages 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 will be organized by the KUL-2 group. It will take place on 19-20 November 2009, and the focus will be on the cornerstone 'Models'. The local organizers are Geert Verbeke and Geert Molenberghs. The annual workshop of 2010 will be organized by the KUL-1 partner and will be focused on the cornerstone 'Methods'. The final workshop in 2011 will be organized by the UCL partner.

2.2.2 Meetings

The following meetings were organized by the network in 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 Bionformatics (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 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. The idea is 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 modelling: smoothness and robustness' was held at KUL-1 (main organizer: I. Gijbels). Two of the ten invited speakers were members of the IAP-network. The international workshop was attended by approximately hundred participants including several members of the IAP-network.

2.3 Organization of the network: administrative meeting

The annual administrative meeting took place on 19 September 2008 at UG, during the annual workshop. The meeting was attended by:

- For Belspo : V. Feys
- For the follow-up committee : A. Davison (EPFL, Lausanne)
- For the network : A. Antoniadis (UJF), R. Crujeiras-Casais (UCL), L. Duchateau (UG), P. Eilers (UU), I. Gijbels (KUL-1), E. Lesaffre (KUL-2), L. Simar (UCL), S. Van Bellegem (UCL), I. Van Keilegom (UCL), N. Veraverbeke (UH), G. Verbeke (KUL-2).

R. Crujeiras-Casais also represented W. González Manteiga (USC).

The participants to this administrative meeting discussed issues related to past and future scientific activities organized by the network, scientific collaborations in the network, work valorization (such as web page, reports, ...), network organization, management and visibility. A detailed report on this meeting was sent to all participants.

2.4 Collaborations, working groups and seminars

2.4.1 Collaborations

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

- Geert Molenberghs (UH), Geert Verbeke (KUL-2) and Michael G. Kenward (LSHTM) are involved in an on-going project on the use of pseudo-likelihood methods when data are incomplete. To this effect, inverse probability weighting is being investigated. The same team members are involved in the development of a broad complex modeling framework that can deal with two important aspects simultaneously: (1) overdispersion owing to the non-Gaussian nature of the outcome; (2) within-unit correlation owing to the hierarchical nature of the design. This framework is also of use to Geert Molenberghs (UH) and Paul De Boeck (KUL-1) to derive explicit correlation functions for broad classes of item response theory models. Currently, correlations are either numerically approximated or calculated at latent scales.
- Paul Janssen (UH), Luc Duchateau (UG), Catherine Legrand (UCL) and PhD students of their respective research groups are working together on projects related to frailty models and competing risks in survival analysis. P. Janssen and L. Duchateau have also written together a Springer book on that topic that was published in 2008.
- Iven Van Mechelen (KUL-1) and Geert Verbeke (KUL-2) are both partners in a KUL Center of Excellence for computational systems biology. Their teams have made several methodological contributions in the context of statistical bioinformatics, which have lead to new biological and medical insights.

- There exists a lot of collaboration between the KUL-1 team (Irène Gijbels, Anneleen Verhasselt) and the team of Grenoble (Anestis Antoniadis). These collaborations resulted already in several joint papers. A PhD student of the KUL-1 team (A. Verhasselt) spent in spring 2008 2-3 months at the University of Grenoble, working on a joint project between KUL-1 and UJF.
- Ingrid Van Keilegom (UCL) continues her extensive collaborations with members of the USC group (Wenceslao González-Manteiga, Carmen Cadarso-Suárez, Rosa Crujeiras Casais) on problems related to goodness-of-fit testing for random effects models, nonparametric estimation and inference for models for successive survival times, and inference for censored spatial data. R. Crujeiras-Casais worked as an IAP postdoc at UCL in 2008 to work on these and other topics related to the network. Moreover, W. González-Manteiga and R. Crujeiras-Casais visited the UCL in February 2008 for three weeks.
- There are also ongoing collaborations between the teams of KUL-1 and UJF on variable selection within the context of mixture models.
- Irène Gijbels (KUL-1), Noël Veraverbeke (UH) and Marek Omelka (IAP-postdoc working 6 months at KUL-1, followed by 6 months at UH) continue working on problems related to nonparametric copula estimation.

2.4.2 Working groups

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

- The working group on copulas is very active, and met approximately every two months in 2008. The close collaboration between members of the KUL-1 and the UH within the context of this working group led already to several joint publications of members of these two universities.
- A working group on 'sensitivity analysis for incomplete data' is active and populated with members of UH and KUL-2. Meetings take place roughly once a term. The work is targeted at research problems, dissemination through education, consulting, and conferences. The scope encompasses modeling strategies for incomplete data and to sensitivity analysis for incomplete and coarse data. The group entertains contacts with the research community at large, the biopharmaceutical industry, and regulatory authorities, such as the Food and Drug Administration. Geert Molenberghs (UH) is a member of an international team convened by the National Academy of Sciences, under contract to the Food and Drug Administration, to provide guidelines on the treatment of incomplete data in regulated clinical trials.
- A working group on 'evaluation of surrogate markers in clinical studies' is active at UH. The point of gravity is located within UH, but a large number of researchers from other Belgian and foreign (France, UK, US) organizations and institutions participate on either a regular

basis or in an *ad hoc* fashion. Often, visitors from abroad are received. The work is closely connected to the remit of WP4, with ramifications to WP3 and WP5.

- A working group on 'Competing risks in survival analysis' consists of a group of about 15 researchers from UH, UG, UCL, as well as members from the EORTC (Brussels) and the University of Liège. Both applied as well as more theoretical researchers take part in the working group. In 2008, the working group met on January 17 (EORTC), March 19 (UH), and June 9 (UG).

2.4.3 Seminars

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

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

- *February 8, 2008: Wenceslao González-Manteiga (USC), 'Using SiZer map for testing problems', at UCL.
- February 8, 2008: John Einmahl (University of Tilburg, The Netherlands), 'Asymptotics for the Hirsch index', at UCL.
- May 13, 2008: Pieter Kroonenberg (Leiden University, The Netherlands), 'Analysing 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 31, 2008: Noël Veraverbeke (UH), 'Half a century of Kaplan-Meier estimation in survival analysis (1958 - 2008)', at KUL-1. The seminar (which was also announced as a colloquim talk) attracted more than 50 researchers, going from very mathematically oriented researchers to more applied statisticians, and was attended by Master students from several disciplines (pure mathematics, statistics, actuarial sciences, ...).
- November 14, 2008: Dimitris Rizopoulos (Erasmus University, Rotterdam, The Netherlands), 'Laplace approximations for joint models', at UCL.
- *November 14, 2008: Goele Massonnet (UH), 'Contributions to frailty and copula modelling with applications to clinical trials and dairy cows data', at UCL.

2.5 Short courses

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

A list of the short courses organized during the working year 2008 is given below.

- 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), 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.
- 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 'Penalised spline smoothing and geoadditive regression', by Thomas Kneib, Ludwig-Maximilians-University Münich, Germany, at KUL-1.

2.6 PhD and postdoctoral researchers

The postdocs financed by the IAP network have 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. Among those hired, we like to mention in particular the following two postdocs :

- Marek Omelka was an IAP postdoc coming from Prague University. He works for 6 months at KUL-1 (October 2007-April 2008) followed by 6 months at UH (April 2008 - October 2008). A working group has been set up around his work, and related projects.

- 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.

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

Below are a few examples of IAP members that took part in the PhD jury at other universities of the network. This participation is a very useful way to get familiar with the research carried out at other groups of the network, and will be extended even more in the future.

- At UCL, Irène Gijbels (KUL-1) was part of the jury of G. Geenens, and Paul Janssen (UH) was a member of the jury of B. Teodoreascu.
- At KUL-2, D. Rizopoulos, A. Jara Vallejos and S. Tsonaka defended their thesis. The committee encompassed members of UH (Geert Molenberghs, Marc Aerts) and of KUL-1 (Irène Gijbels).
- At UH, about 10 PhD students defended their thesis in 2008. Members from KUL-1 (Irène Gijbels, Francis Tuerlinckx), KUL-2 (Geert Verbeke), and UCL (Catherine Legrand) partook in the various juries.
- At UJF, Irène Gijbels (KUL-1) was a member of the Habilitation jury of S. Lambert.
- At USC, Ingrid Van Keilegom (UCL) partook in the PhD jury of B. Pateiro-López.

2.7 Prizes obtained by network members

- The network has been involved in two Belgian Francqui chairs for the academic year 2008–2009. Tom Snijders (member of the follow-up committee) is recipient of a chair awarded to KUL-1. Geert Molenberghs (UH) is recipient of a chair awarded to the University of Antwerp.
- Léopold Simar (UCL) is currently holder of the 'Chaire d'Excellence Pierre de Fermat', Université Toulouse I (2008-09), which is a research chair financed by the 'Conseil Régional de Midi-Pyrénées'.
- Ingrid Van Keilegom (UCL) obtained in 2008 a five-year grant from the ERC (European Research Council) for Starting Independent Researchers, and was elected Fellow of the IMS (Institute of Mathematical Statistics).
- Geert Molenberghs (UH) and Geert 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.

3 Technical Reports and Publications

Below we provide the scientific output related to the IAP-statistics network. We give both the Technical reports and the publications of network members in 2008 :

- 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 web site 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).

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 52 joint papers (all types confounded) in 2008.

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.

3.1 Université catholique de Louvain, UCL

3.1.1 Technical reports

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

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

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

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

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

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

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

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

[9] 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.

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

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

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

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

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

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

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

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

[18] 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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

3.1.2 Refereed publications (published)

[35] 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.

[36] 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.

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

[38] 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.

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

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

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

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

[43] 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.

[44] 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.

[45] 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.

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

[47] 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.

[48] 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.

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

[50] 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.

[51] 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

[52] 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.

[53] 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.

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

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

[56] 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.

[57] 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.

[58] 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.

[59] 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.

[60] 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.

[61] 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.

[62] 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.

[63] 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.

[64] 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.

3.1.3 Refereed publications (in press)

[65] 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.

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

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

[68] 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.

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

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

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

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

[73] 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.

[74] 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.

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

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

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

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

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

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

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

[82] 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.

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

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

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

[86] 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.

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

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

[89] 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.

[90] 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.

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

[92] 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.

3.1.4 Books (published)

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

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

3.2 Katholieke Universiteit Leuven, KUL-1

3.2.1 Technical reports

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[120] 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.

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

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

3.2.2 Refereed publications (published)

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

[124] 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.

[125] 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.

[126] 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.

[127] 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.

[128] 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.

[129] 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.

[130] 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.

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

[132] 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.

[133] 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.

[134] 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.

[135] 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.

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

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

[138] 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.

[139] 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.

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

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

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

[143] 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.

[144] 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.

[145] 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.

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

[147] 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.

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

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

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

[151] 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.

[152] 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.

[153] 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.

[154] 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.

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

[156] 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.

[157] 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.

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

[159] 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.

3.2.3 Refereed publications (in press)

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

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

[162] 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.

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

[164] 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.

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

[166] 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.

[167] 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.

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

[169] 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.

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

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

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

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

[174] 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.

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

[176] 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.

[177] 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.

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

[179] 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.

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

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

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

[183] 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.

[184] 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.

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

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

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

[188] 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.

[189] 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.

[190] 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.

3.2.4 Non-refereed publications (published)

[191] 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.

3.2.5 Non-refereed publications (in press)

[192] 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.

3.3 Katholieke Universiteit Leuven, KUL-2

3.3.1 Technical reports

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

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

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

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

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

3.3.2 Refereed publications (published)

[198] 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.

[199] 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.

[200] 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.

[201] 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.

[202] 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.

[203] 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.

[204] 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.

[205] 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.

[206] 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.

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

[208] 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.

[209] 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.

[210] 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.

[211] 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.

[212] 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.

[213] 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.

[214] 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.

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

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

[217] 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.

[218] 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.

[219] 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.

[220] 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.

[221] 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.

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

[223] 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.

[224] 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.

[225] 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.

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

[227] 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.

[228] 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.

[229] 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.

[230] 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.

[231] 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.

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

3.3.3 Refereed publications (in press)

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

[234] 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.

[235] 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.

[236] 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.

[237] 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.

[238] 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.

[239] 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.

[240] 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.

[241] 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.

[242] 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.

[243] 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.

[244] 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.

3.3.4 Non-refereed publications (published)

[245] 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.

[246] 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.

3.3.5 Books (in press)

[247] 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.

3.4 Universiteit Gent, UG

3.4.1 Technical reports

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

3.4.2 Refereed publications (published)

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

[250] 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.

[251] 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.

[252] 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.

[253] 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.

[254] 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.

[255] 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.

[256] 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.

[257] 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.

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

[259] 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.

[260] 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.

[261] 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.

[262] 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.

[263] 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.

[264] 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.

[265] 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.

[266] 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.

[267] 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.

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

[269] 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.

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

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

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

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

[274] 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.

[275] 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.

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

[277] 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.

[278] 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.

[279] 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.

[280] 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.

[281] 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.

[282] 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.

[283] 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.

[284] 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.

[285] 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.

[286] 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.

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

[288] 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.

3.4.3 Refereed publications (in press)

[289] 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.

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

[291] 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.

[292] 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.

[293] 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.

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

[295] 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.

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

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

[298] 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.

3.4.4 Non-refereed publications (published)

[299] 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.

3.4.5 Books (published)

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

[301] 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.

3.4.6 Books (in press)

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

3.5 Universiteit Hasselt, UH

3.5.1 Technical reports

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

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

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

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

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

[308] 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.

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

3.5.2 Refereed publications (published)

[310] 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.

[311] 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.

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

[313] 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.

[314] 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.

[315] 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.

[316] 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.

[317] 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.

[318] 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.

[319] 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.

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

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

[322] 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.

[323] 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.

[324] 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.

[325] 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.

[326] 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.

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

[328] 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.

[329] 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.

[330] 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.

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

[332] 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.

[333] 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.

[334] 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.

[335] 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.

[336] 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.

[337] 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.

[338] 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.

[339] 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.

[340] 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.

[341] 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.

[342] 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.

[343] 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.

[344] 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.

[345] 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.

[346] 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.

[347] 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.

[348] 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.

[349] 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.

[350] 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 seroprevalence studies in Belgium, England and Wales, Finland, Italy and Poland. *Epidemiology and Infection*, **136**, 1059-1068, 2008. R08170.

[351] 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.

[352] 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.

[353] 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.

[354] 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.

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

[356] 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.

[357] 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.

[358] 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.

[359] 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.

[360] 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.

[361] 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.

[362] 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.

[363] 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.

[364] 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.

[365] 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.

[366] 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.

[367] 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.

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

[369] 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.

[370] 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.

[371] 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.

[372] 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.

3.5.3 Refereed publications (in press)

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

[374] 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.

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

[376] 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.

[377] 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.

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

[379] 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.

[380] 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.

[381] 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.

[382] 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.

[383] 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.

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

[385] 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.

[386] 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.

[387] 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.

[388] 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.

[389] 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.

[390] 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.

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

[392] 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.

[393] 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.

[394] 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.

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

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

[397] 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.

[398] 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.

[399] 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.

[400] 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.

[401] 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.

[402] 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.

[403] 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.

[404] Serroyen, J., Bruckers, L., Rogiers, G. and G. Molenberghs, Characterizing persistent disturbing behavior using longitudinal and multivariate techniques. *Journal of Applied Statistics*, 2008. RP08086.

[405] 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.

[406] 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.

[407] 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.

[408] 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.

3.5.4 Non-refereed publications (published)

[409] Aerts, M., Bollaerts, K., Ribbens, K., Van der Stede, Y., Boone, I. and K. Mintiens, Semi-parametric 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.

[410] 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.

[411] 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.

[412] 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.

[413] 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.

[414] 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.

[415] 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.

[416] 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.

[417] 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.

[418] 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.

[419] 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.

[420] 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.

[421] 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.

[422] 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.

[423] Lejeune, O., Hens, N., Bilcke, J. and P. Beutels, Proceedings of the international workshop facing the challenge of infectious diseases. *Integrating Mathematical Modeling, Computational Thinking and ICT Applications, ISI Foundations, Turin (Italy), October 13-18*, 2008. NR08025.

[424] 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.

[425] 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.

[426] 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.

[427] 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.

[428] 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.

[429] 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.

3.5.5 Books (published)

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

3.5.6 Books (in press)

[431] 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.

3.6 Université Joseph Fourier, UJF–LMC–IMAG

3.6.1 Technical reports

[432] Antoniadis, A., Fryzlewicz, P. and F. Letué, The Dantzig selector in Cox's proportional hazards model, 2008. TR08047.

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

[434] Boshnakov, G. and S. Lambert-Lacroix, Maximum entropy for periodically correlated processes from nonconsecutive autocovariance coefficients, 2008. TR08048.

3.6.2 Refereed publications (published)

[435] 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.

[436] 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.

3.6.3 Refereed publications (in press)

[437] 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.

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

[439] Antoniadis, A., Paparoditis, E. and T. Sapatinas, Bandwidth selection for functional times series prediction. *Statistics and Probability Letters*, 2008. RP08003.

[440] 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.

[441] Gannaz, I. and O. Winterberger, Adaptive density estimation with dependent observations. *ESAIM: Probability and Statistics*, 2008. RP08005.

3.7 Universiteit Utrecht, UU

3.7.1 Refereed publications (published)

[442] Camarda, C.G., Eilers, P.H.C. and J. Gamper, Modelling general patterns of digit preference. *Statistical Modelling*, **8**, 385-402, 2008. R08068.

[443] 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.

3.7.2 Refereed publications (in press)

[444] Blasius, J., Eilers, P.H.C. and J. Gower, Better biplots. *Computational Statistics and Data Analysis*, 2008. RP08006.

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

3.7.3 Non-refereed publications (published)

[446] 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.

[447] 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.

[448] 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.

[449] 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.

[450] 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.

[451] Schnabel, S.K. and P.H.C. Eilers, Optimal expectile smoothing. *Proceedings of the 23th International Workshop on Statistical Modelling, Utrecht*, 2008. NR08016.

[452] 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.

3.8 Universidad de Santiago de Compostela, USC

3.8.1 Technical reports

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

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

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

[456] Pateiro-López, B. and A. Rodríguez-Casal, A., On the estimation of alpha-convex sets, 2008. TR08049.

3.8.2 Refereed publications (published)

[457] 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.

[458] 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.

[459] 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.

[460] 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.

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

[462] 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.

[463] 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.

[464] 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.

[465] González-Manteiga, W., Ferraty F. and P. Vieu, Nonparametric functional data analysis: theory and practice. *Computational Statistics*, **23**, 2, 341-342, 2008. R08080.

[466] 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.

[467] 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.

[468] 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.

[469] 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.

[470] 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.

[471] 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.

[472] 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.

[473] 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.

[474] Pateiro-López, B. and A. Rodríguez-Casal, Length and surface area estimation under smoothness restrictions. *Advances in Applied Probability*, **40**, 348-358, 2008. R08076.

[475] 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.

[476] 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.

[477] 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.

[478] 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.

[479] 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.

[480] 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.

[481] 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.

3.8.3 Refereed publications (in press)

[482] 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.

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

[484] 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.

[485] 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.

[486] 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.

[487] González-Manteiga, W., Martinez Miranda, M.D. and R. Raya-Miranda, Sizer map for inference with additive models. *Statistics and Computing*, RP08012.

[488] 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.

[489] 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.

[490] 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.

[491] 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.

[492] 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.

[493] Meira-Machado, L., Cadarso Suárez, C.M., de Uña Álvarez, J. and P.K. Andersen, Multi-state models for the analysis of time to event data. *Statistical Methods in Medical Research*, 2008. RP08022.

[494] 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.

[495] 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.

[496] 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.

[497] 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.

[498] 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.

3.9 London School of Hygiene and Tropical Medicine, LSHTM

3.9.1 Technical reports

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

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

[501] 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.

3.9.2 Refereed publications (published)

[502] 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.

[503] 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.

[504] 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.

3.9.3 Refereed publications (in press)

[505] Carpenter, J.R. and I. Plewis, Analysing longitudinal studies with non-response: issues and statistical methods. *Handbook of Methodological Innovation, Sage*, 2009. BP08007.

[506] 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.

[507] Goldstein, H., Carpenter, J.R., Kenward, M.G., and K. Levin, Multilevel models with multivariate mixed response types. *Statistical Modelling*, 2008. RP08028.

[508] 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.

[509] 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.

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

3.9.4 Non-refereed publications (published)

[511] 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.

[512] 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.

3.10 List of joint publications

3.10.1 Technical reports

[513] Antoniadis, A., Gijbels, I. and M. Nikolova, Penalized likelihood regression for generalized linear models with nonquadratic penalties, 2008. TR08050. (KUL-1, UJF)

[514] Beirlant, J., Joossens, E. and J. Segers, Second-order refined peaks-over-threshold modelling for heavy-tailed distributions, 2008. TR08035. (UCL, KUL-1)

[515] 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)

[516] Crujeiras, R. M. and I. Van Keilegom, Least squares estimation of nonlinear spatial trends, 2008. TR08044. (UCL, USC)

[517] Goethals, K., Janssen, P. and L. Duchateau, Frailty models and copulas: similarities and differences, 2008. TR08015. (UG, UH)

[518] 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)

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

[520] Kenward M.G. and G. Molenberghs, Last observation carried forward: a crystal ball?, 2008. TR08059. (LSHTM, UH)

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

[522] 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)

[523] 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)

[524] Veraverbeke, N., Omelka, M. and I. Gijbels, Weak convergence of the empirical copula process with covariates, 2008. TR08056. (KUL-1, UH)

3.10.2 Refereed publications (published)

[525] 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)

[526] 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)

[527] 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)

[528] 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)

[529] 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)

[530] 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)

[531] Hubert, M., Rousseeuw, P.J. and S. Van Aelst, High-breakdown robust multivariate methods. *Statistical Science*, **23**, 92-119, 2008. R08203. (UG, KUL-1)

[532] 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)

[533] 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)

[534] 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)

[535] Rizopoulos, D., Verbeke, G. and G. Molenberghs, Shared parameter models under random-effects misspecification. *Biometrika*, **95**, 63-74. R08019. (UH, KUL-2)

[536] 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)

[537] 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)

[538] 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)

3.10.3 Refereed publications (in press)

[539] 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)

[540] Antoniadis, A., Gijbels, I. and J.M. Poggi, Smoothing non equispaced heavy noisy data with wavelets. *Statistica Sinica*, 2008. RP08038.(UJF, KUL-1)

[541] 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)

[542] 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)

[543] 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)

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

[545] 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)

[546] Crujeiras, R., Fernández-Casal, R. and W. González-Manteiga, An L2-test for comparing spatial spectral densities. *Statistics and Probability Letters*, RP08011. (USC, UCL)

[547] 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)

[548] 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)

[549] 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)

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

[551] 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)

[552] 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)

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

[554] 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)

[555] 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)

[556] 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)

[557] 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)

[558] 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)

[559] 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)

3.10.4 Non-refereed publications (published)

[560] 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)

[561] 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)

[562] 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)

3.10.5 Books (published)

[563] Duchateau, L. and P. Janssen, *The Frailty Model*. Springer, New York, 2008. B08001. (UH, UG)

3.10.6 Books (in press)

[564] 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)