# IWSM 2011

**Proceedings of the** 

# 26th International Workshop on Statistical Modelling

Valencia (Spain), July 11-15, 2011



# **Editors:**

David Conesa Anabel Forte Antonio López-Quílez Facundo Muñoz

# Proceedings of the 26th International Workshop on Statistical Modelling

July 11-15, 2011 València

David Conesa, Anabel Forte, Antonio López-Quílez, Facundo Muñoz (editors) Proceedings of the 26th International Workshop on Statistical Modelling. València, July 11-15, 2011 David Conesa, Anabel Forte, Antonio López-Quílez, Facundo Muñoz, eds. València 2011. ISBN 978-84-694-5129-8

### **Editors:**

David Conesa<sup>1</sup>, David.V.Conesa@uv.es Anabel Forte<sup>2</sup>, forte@eco.uji.es Antonio López-Quílez<sup>1</sup>, Antonio.Lopez@uv.es Facundo Muñoz<sup>1</sup>, Facundo.Munoz@uv.es

 Departament d'Estadística i Investigació Operativa Universitat de València (Estudi General)
 Facultat de Matemàtiques
 Dr. Moliner 50, 46100 Burjassot, Spain.

Departamento de Economía
 Universitat Jaume I
 Facultad de Ciencias Jurídicas y Económicas
 Campus del Riu Sec, E-12071 Castelló de la Plana, Spain.

Cover photo: Victor Roda

Printed by Copiformes S.L.

# Scientific Programme Committee

- Susie Bayarri (Chair) Universitat de València, Spain
- Carmen Armero Universitat de València, Spain
- Adrian Bowman
  University of Glasgow, UK
- Charmaine B. Dean Simon Fraser University, Canada
- María Durbán Universidad Carlos III de Madrid, Spain
- Claire Ferguson University of Glasgow, UK
- Herwig Friedl

  Graz University of Technology, Austria
- Gillian Heller

  Macquarie University, Australia
- John Hinde University of Galway, Ireland
- Thomas Kneib

  Carl von Ossietzky Universität Oldenburg, Germany
- Arnost Komàrek Charles University in Prague, Czech Republic
- Antonio López-Quílez Universitat de València, Spain
- Brian Marx
  Louisiana State University, USA
- Pere Puig Universitat Autònoma de Barcelona, Spain

# **Preface**

This volume contains all the papers of the 26th International Workshop on Statistical Modelling. Many things have changed since in 1986 an enthusiastic group of statisticians interested in statistical modelling started these series of workshops within a friendly and supportive academic atmosphere. New technologies, more attendants, but always with the same initial spirit: to promote and develop the use of statistical modelling in research and applications.

We are glad to present you these Proceedings, which clearly reflect the aliveness of that spirit. One the one hand, the five invited papers show new advances in theoretical research but always keeping an eye in their applied interest. One the other hand, the great amount of contributions (a total of 140) and their quality demonstrate that the workshop is in good shape. Authors should receive most of the credit for the quality of these Proceedings. Nevertheless, all submissions were carefully reviewed by the members of the Scientific Committee. Their detailed work has been reflected in a big improvement of the preliminary versions jointly with the final selection of contributions.

This 26th edition of the IWSM will be held in Valencia (Spain) in an informal environment (ADEIT- FUNDACIÓ UNIVERSITAT-EMPRESA of the Universitat de València) to encourage discussion and exchange of ideas which could result in future research. Valencia has a great tradition in Statistics and in particular in Bayesian Statistics. This why we are so happy to see that this way of thinking and doing statistics is quite present in these Proceedings reflecting its important role in the Society. We will also like to comment, that many of the contributions in these Proceedings are due to students, which clearly have the future in their hands.

Finally, we wish to acknowledge Carmen Armero, the chair of the local Committee for putting together all the pieces needed in the process of organising this event. Without her interest and passion it would have been impossible.

So welcome to Valencia. Enjoy the city and surroundings and have a great conference.

David Conesa, Anabel Forte, Antonio López-Quílez, Facundo Muñoz Valencia, June 2011

# Contents

# Part 1. Invited papers

Berger et al. Risk Assessment for Pyroclastic Flows: Combining Deterministic and Statistical Modeling
Firth Quasi-variances and extensions
<b>Gómez</b> Some theoretical thoughts when using a composite endpoint to prove the efficacy of a treatment
Green et al. Identifying influential model choices in Bayesian hierarchical models
Jørgensen et al. The Ecological Footprint of Taylor's Universal Power Law
Part 2. Contributed papers
Aerts et al. Incomplete Clustered Data and Non-Ignorable Cluster Size
Alvaro-Meca et al. Bayesian Lee-Carter Model: A Spatio-Temporal Approach.
Andrés-Ferrer and Ney From Empirical Bayes to Leaving-One- Out
Aregay et al. Model Based Estimates of Long-Term Persistence of Induced HPV Antibodies: A Flexible Subject-Specific Approach
<b>Armero et al.</b> Bayesian model selection for assessing the progression of chronic kidney disease in transplanted children.
Badiella et al. Area under the ROC curve using logistic regression with random effects: Estimation and Inference
Barber et al. Optical properties of fresh date palm in different stages of maturity
Bárcena et al. Measuring the real estate bubble: a house price index for Bilbao.

Baxter et al. Missing data, multiple imputation and the UK National Vascular Database	71
Belgrave et al. A Comparison of Frequentist and Bayesian Approaches to Latent Class Modelling of Susceptibility to Asthma and Patterns of Antibiotic Prescriptions in Early Life	75
Boixadera et al. Who uses Complementary and Alternative Medicine? An analysis for cancer patients	79
Bowman and Crujeiras Assessing isotropy with the variogram	83
Brechmann et al. Simplified regular vines for modeling high- dimensional financial risk data	87
Brewer et al. Climate Envelopes for Species Distribution Models	93
Burke and MacKenzie XD survival regression models with frailty	99
Caballero-Águila et al. Least-squares signal estimation using correlated delayed observations transmitted by different sensors .	105
Caballero-Águila et al. Filtering algorithm for fractional order discrete systems with uncertain observations	109
Carrasco et al. The Log-Generalized Modified Weibull Regression Model	113
Castillo and Serra An exponential dispersion family to modelling critical phenomenon	117
Catelan and Biggeri Hierarchical Bayesian modelling to assess divergence in disease mapping	121
Conde and MacKenzie LASSO Penalised Likelihood in High- Dimensional Contingency Tables	127
Conesa et al. Describing the geography of Spanish bank branching.	133
Corberán-Vallet and Lawson Spatio-temporal disease modeling and surveillance with Bayesian hierarchical Poisson models	137
Corberán-Vallet et al. Time series modeling and Bayesian fore- casting with exponential smoothing models	141
Costa and Dias Assessment of e-government maturity in Portuguese municipalities using regression and clustering approaches	146

Creemers et al. Joint Modeling Longitudinal Health Care Costs and Time-to-Event Data in Matched Pairs
Cysneiros Bartlett-type Correction in Heteroscedastic Symmetric Nonlinear Models
Cysneiros et al. A Symbolic Robust Regression Model
Czado et al. Bayesian inference for copula based GARCH models
<b>Dejardin et al.</b> Bayesian Dose Escalation in phase I studies of Combinations of Drugs with Control
De Rooi and Eilers Using text mining tools to compose structure priors for inferring gene networks.
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$
Djeundje and Currie Smooth mixed models for nested curves .
<b>Dondelinger et al.</b> A Bayesian regression and multiple changepoint model for systems biology
Dooley et al. Analysis of an Observational Study
Eilers et al. Sea Level Trend Estimation by Seemingly Unrelated Penalized Regressions
Fabio et al. Generalized random intercept log-gamma exponential family models
Faria and Gonçalves Modelling Financial Data using Poisson Mixture Approach
Finazzi et al. A multivariate space-time model for heterogeneous air quality networks
Fonseca et al. Predictive distributions for non-regular parametric models
Forte et al. Objective Bayes Criteria for Variable Selection
Franco-Villoria et al. Conditional Probability of Flood Risk in Scotland
Fried et al. Outliers and interventions in INGARCH time series
Furche et al. Bivariate Ordinal Regression Models for the Analysis of Neural Data

Gallego et al. Modelling endocytosis by means of non-homogeneous temporal Boolean models.
García-Donato et al. A Prior for multiplicity control and closed- form Bayes factors in variable selection
García-Mora et al. Approximated Survival function in the Sum of Two Independent Homogeneous Markov Processes: Application to Bladder Carcinoma.
Gargoum On using the Hellinger distance in checking the validity of approximations based on dynamic generalized linear models
<b>George and Ünlü</b> Parameter Estimation in Skills-based Knowledge Space Theory and Cognitive Diagnosis Models: A Comparison
Gilchrist et al. Forecasting film revenues using GAMLSS
Gilthorpe et al. Importance of correctly specifying the random structure in growth mixture models
Gomes et al. Modeling swimming marks through Blocks and POT methods
Gonçalves and Costa Improvement of surface water quality variables modelling that incorporates a hydro-meteorological factor: a state-space approach
Gottard et al. Modelling fertility and education in Italy in the presence of time-varying frailty component
Grisotto et al. Empirical Bayes models to estimate contextual effects
Habteab Ghebretinsae et al. Generalized Frailty Model for Comet Assays
Ha et al. Interval Estimation of Random Effects in Frailty Models
Haggarty et al. Functional Clustering of Water Quality Data in Scotland
Hasso and Matawie Using Probability Models to Classify Software Patterns
Hernandez et al. Linear Model comparison with structured mean and dispersion parameters
Huertas et al. Joint Modelling of Two Sequential Times to Events With Longitudinal Information

Ibacache Pulgar and Paula Elliptical semiparametric mixed models	322
Kelly The change-point problem in regression with correlated data and change in variance	326
Komárek Capabilities of R package mixAK for clustering based on multivariate continuous and discrete longitudinal data	330
Lambert Additive location-scale model when the response and some covariates are interval censored	334
<b>Letón and Molanes-López</b> Second order delta method for estimating the Youden index and optimal threshold	338
Little et al. Modeling growth patterns of the swift tern using non-linear mixed effect models	342
Loquiha et al. Zero-Inflated Poisson and Negative Binomial Models Applied to Maternal Mortality Rate in Mozambique	346
Lynch and MacKenzie On Bivariate Survival Regression Models	352
Marchetti et al. Regression graph models: an application to joint modelling of fertility intentions among childless couples	358
Martínez-Beneito et al. A spatio-temporal monitoring system for Influenza-Like Illness incidence	364
Martínez-Coscollà et al. Bayesian hierarchical modelling for analyzing the efficiency in the European banking system.	368
Marx et al. Multidimensional Single-Index Signal Regression	372
Mauff and Little Multivariate Nonlinear Multi-Level Mixed Effect Models: Techniques and Application to Pharmacokinetic Data	378
Mayr et al. Boosting Generalized Additive Models for Location, Scale and Shape	384
Menten et al. Estimation of Infection Rates from Repeated ELISA Optical Density Data using Hidden Markov Models	390
Mirkov and Friedl Nonlinear and Spline Regression Models for Forecasting Gas Flow on Exits of Gas Transmission Networks	394
Mohd Din et al. Prediction of the rheumatoid arthritis activity score: a joint modeling approach	400
Molanes-López et al. Covariate-adjusted inference for the Youden index and associated classification threshold	404

Moreira and Machado An R Package for the Estimation of the Bivariate Distribution for Censored Gap Times
Muggeo and Lovison Testing for a breakpoint in segmented regression: a pseudo-score approach
Muñoz and López-Quílez Geostatistical modelling with non- Euclidean distances
Murawska et al. Multi-state models for non Markov process
Mutsvari et al. Some approaches to correct for misclassification in the absence of an internal validation data set
Nicholls and Ryder Phylogenetic models for Semitic vocabulary.
Nicholls and Watt Partial Order Models for Episcopal Social Status in 12th Century England
Nysen et al. Testing Goodness-of-Fit of Parametric Models for Censored Data
Oller and Gómez Testing against ordered alternatives with interval-censored data
Palarea-Albaladejo and Martín-Fernández Examining distance-based grouping on the simplex sample space: the fuzzy clustering case
Pardo and Pérez The use of GEE for analyzing housing prices .
Peng and MacKenzie Precision of estimators in interval censored parametric survival models
Pennino et al. A Bayesian spatial approach to modelling fish species occurrence.
Pereira et al. The truncated inflated beta regression
Perra et al. A Bayesian analysis of survival times for stage IV non- small cells lung cancer
Pfeifer On probabilities of avalanches triggered by alpine skiers.  Models with random effects taking the stratified data into account.
Pita-Fernández et al. Cancer incidence in kidney transplant recipients
Pomann et al. Evaluating Change Detection in Data Streams .

Porcu et al. Modelling the Timing of Marital Dissolution in Italy: censored quantile regression with additive terms	
Prieto et al. Estimation of the density of the Antarctic Blue whales population using their sequences of sounds	
Ramsey and Futschik Optimal DNA Pooling for the Detection of Single Nucleotide Polymorphisms	
Riebler et al. Modelling seasonal patterns in longitudinal profiles with correlated circular random walks	
Rippe and Eilers Segmented smoothing with an $L_0$ penalty	
Rodríguez-Álvarez et al. Testing for covariate effects in ROC-GAM regression models based on bootstrap methods	
Rodríguez-Díaz et al. D-Optimum designs in random effect logistic regression models	
Rosen et al. Adaptive Spectral Estimation for Nonstationary Time Series	
Rushworth et al. Distributed lag models for hydrological data .	
Russo et al. Exact and approximate inferences for nonlinear mixed-effects heavy-tailed models	
Sabanés Bové et al. Hyper-g Priors for Generalised Additive Model Selection	
Schnabel et al. Optimal time scaling for plant growth analysis .	
Sellers Introducing a Model to Determine True Counts via the Conway-Maxwell-Poisson Distribution	
Sikorska et al. Fast genome-wide association analysis in longitudinal studies	
Singh and Huzurbazar Analysis of Gene Duplication Data	
Slaets et al. Flexible Modelling of Functional Data using Continuous Wavelet Dictionaries	
Smith and Bowman Boundary identification in 3D images	
Sobotka et al. Confidence intervals for geoadditive expectile regression models	
Stefanova Measuring Efficiency of Trial Designs with Unreplicated or Partially Replicated Test Lines	

Stöber and Czado A Markov switching model for vine copulas .
Sweeney and Haslett Bayesian residual analysis in Poisson regression models.
Tamura and Giampaoli Prediction for an observation in a new cluster for Multilevel Logistic Regression considering k random coefficients
Taylor and Einbeck Multivariate regression smoothing through the "falling net"
Tharmaratnam and Claeskens Robust model selection in additive penalized regression splines models
Thompson Statistical modeling of geographic risks for very low birth weights near Texas superfund sites
Ugarte et al. Spatio-temporal risk smoothing and forecasting with P-splines
Urbano et al. Bioassays models with natural mortality and random effects
Usuga et al. A study to compare HGLM and GAMLSS in mixed linear models
Van den Hout et al. A latent-class semi-parametric change point model for cognitive ability in older age
Van Oirbeek and Lesaffre Measuring the Brier score for frailty models
Ventrucci et al. A Dipole Model for MEG Data
Ventura and Racugno A Bayesian adjustment of the modified pro- file likelihood
Waldmann and Kneib Bayesian Structured Additive Quantile Regression
West et al. Groups within networks
Worton and Mclellan Robust mixture modelling of telemetry data in wildlife studies of home range
Yee and Hadi Row-Column Association Models
Ziegler-Graham and Rohde Use of Marginal Likelihoods in Sta- tistical Inference

# Modelling the Timing of Marital Dissolution in Italy: censored quantile regression with additive terms

Mariano Porcu<sup>1</sup>, Vito M. R. Muggeo<sup>2</sup>, Vincenza Capursi<sup>2</sup>

- Dipartimento Ricerche Economiche e Sociali, Università di Cagliari, ITALY email: mrporcu@unica.it
- <sup>2</sup> Dipartimento Scienze Statistiche e Matematiche 'S. Vianelli', Università di Palermo, ITALY

**Abstract:** The analysis of marital dissolution in Italy represents a quite interesting and challenging topic from a substantive standpoint; in fact, despite of the decreasing number of marriages and the increasing number of divorces, the traditional family based on the marriage of heterosexual partners is still considered as a fundamental institution of the society. Here we present a censored quantile regression model with additive terms to investigate the determinants of the timing of marital dissolution on a large and substantial sample from a survey carried on in Italy.

**Keywords:** censored quantile regression; Timing of Marital Dissolution; Survival data; Smoothing

# 1 Introduction

It is commonly asserted that the family is the fundamental institution of the Italian society. The main consequence of this assertion is a widespread political support addressed to the upholding of the classical family built on the marriage of heterosexual partners. In this context the analysis of the possible determinants of the marital dissolution is a largely debated issue: the study of the factors that could affect the end of the marriage is prominent in the social research. The topic is also of great interest for the policy-makers as the marital dissolution affects some of the key features of the modern societies, such as economy, gender equality and especially fertility. Although the study of the time-to-separation can provide quite useful information and insights to evaluate trends and changes in the formation and dissolution of the marriage, relatively few studies take explicitly into account the time dimension, see Cavanagh and Huston (2008), Gottman and Levenson (2000).

Using a large sample surveyed by the Italian national statistics institute (ISTAT), we aim to model the time-to-marital dissolution in a regression

quantile framework. While the Cox model represents the most used framework to model survival data, censored quantile regression (CQR) offers a more flexible alternative by focusing the attention on narrow slices, lower or upper tails, of the conditional survival distribution of interest (Koenker, 2008).

## 2 The Data

The data considered in this paper come by from the sample survey on Families and Social Subjects (FSS), carried on in Italy by the official statistics institute at the end of 2003 on a sample of over 19,000 Italian families (nearly 50,000 individuals). The survey was addressed to collect broad information on the Italian households, such as the shapes, the network of kinship, the relations among partners, the permanence of young adults in the family, and the working life.

TABLE 1. Some descriptive statistics on data analysed.

	Males (r	n = 4633	Females $(n = 5235)$		
Covariates	Separated	Non-Separ	Separated	Non-Separ	
Area (obs.)					
North	335	1640	445	1763	
Center	190	1164	223	1269	
South	124	1180	151	1384	
EDUCATION (obs.)					
1 <sup>st</sup> stage basic	75	591	63	684	
$2^{nd}$ stage basic	253	1529	268	1586	
Upper secondary	232	1473	384	1713	
Degree	89	391	104	433	
Age at Marriage (Year	·s)				
Mean (sd)	26.3(5.7)	27.7 (5.8)	23.3(5.0)	22.2(5.7)	
CHILDLESS					
%	33.4	12.8	32.1	12.9	
Work					
Yes at marr (%)	80.1	86.7	51.2	45.0	
Yes at separ (%)	88.0	_	64.7	·—	

The data from the 2003 FSS survey here analyzed represent the most recent information available on the topic: the data from the last FSS survey carried out in 2010 are not yet available. We have omitted from the sample persons married before the 1970 when the divorce, understood as the 'total dissolution of marital status', was not allowed. Moreover, to keep away from any potential confounding interaction between sex, covariates and timing we have considered females and males independently (e.g., Schoen and Canudas-Romo, 2006). Table 1 summarizes some descriptive statistics for the sample.

### 3 Methods and Results

We aim to model the time-of-dissolution as a function of the following covariates in a CQR model: the categorical variables EDUCATION, AREA, and WORK AT MARRIAGE, and the numerical variables AGE AT MARRIAGE, NUMBER OF SONS and YEAR OF BIRTH. Until now, CQR has been discussed only with parametric linear terms, however for the aforementioned continuous covariates the linearity assumption is not tenable and more flexible alternatives are requested: we use B-spline bases with quadratic penalties on the coefficients to get smooth estimates of the nonlinear relationships. The additive CQR model with J nonparametric terms for the variables z and linear terms for the variables x, may be written as

$$Q_{\tau}(Y|x_i) = x_i^T \beta_{\tau} + \sum_{j=1}^{J} f_{\tau j}(z_{ij}),$$
 (1)

where the subscript  $\tau$  points the percentile of interest  $(0 < \tau < 1)$ . Notice that, unlike the usual model for the conditional mean, here the covariate effect (parametric or nonparametric) depends on the percentile  $\tau$ . The response Y measures the time span of their marriage up to the year of separation; we consider as uncensored the spouses (male or female) which stop living together regardless of the possibility of reconciliation; in fact in Italy only a slight proportion of separations ends with a reuniting of the couple (Castiglioni, 2008). We modify the iterative estimating algorithm described in Bottai and Zhang (2010) to include the additive (spline) terms in the linear predictor and to obtain parameter estimate of the additive CQR model (1). Although QR allows to model every quantile of the response conditional distribution, our analysis focuses on the lowest quantiles ( $\tau \leq 0.10$ ). Indeed, early dissolutions (i.e. the left tail of the survival distribution) are of major interest in the present study since the first years of marriage are known to be crucial for fertility, children social development, changes in lifestyle and also for their influence on the probability of remarriage.

TABLE 2. Point estimates for the parameters of the linear terms in the four CQR models.

	Males		Fem	ales
Linear Terms	$\tau = 0.05$	$\tau = 0.10$	$\tau = 0.05$	$\tau = 0.10$
EDUC (2 stage basic vs 1 stage basic)	-2.257	-1.666	-2.190	-3.971
EDUC (upper sec. vs. 1 stage basic)	-1.372	-1.138	-3.444	-5.668
EDUC (degree vs 1 stage basic)	-2.291	-2.474	-4.454	-6.503
Area (center vs. north)	0.197	1.081	0.770	1.116
Area (south vs. north)	2.319	2.730	1.317	1.486
Work at marr (yes vs. no)	1.500	2.157	0.469	0.130

For the four fitted additive CQR models (two quantiles 0.05 and 0.10 for males and females), Table 2 shows the point estimates for the parameters

of the linear terms and Figure 1 reports the fitted smooth effects of the three continuous covariates.

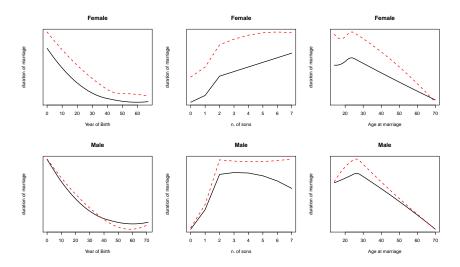


FIGURE 1. Fitted quantiles ( $\tau=0.05$ , continuous line;  $\tau=0.10$  dashed line) for males and females.

In short, for the *early* marital dissolutions (i.e., the low percentiles 5% and 10%) we observe strong and somewhat expected effects of the 'area' and of the 'educational level' for both male and female groups. On the other hand, the effect of the 'working status' is somewhat different. The plots in Figure 1 emphasize the nonlinear effects of the continuous covariates, by highlighting a different role of the number of sons variable.

# References

Bottai, M., Zhang, J. (2010) Laplace regression with censored data *Biometrical Journal*, **52**, 487-503.

Castiglioni, M., Dalla Zuanna, G. (2008) A Marriage-Cohort Analysis of Legal Separations in Italy. *Population*, **63**, 1, 173-193.

Cavanagh, S.E., Huston, A.C. (2008) The Timing of Family Instability and Children's Social Development. *Journal of Marriage and Family*, **70**, 1258-1269.

Gottman, J.M., Levenson, R.W. (2000) The Timing of Divorce: Predicting When a Couple Will Divorce Over a 14-Year Period. *Journal of Marriage and Family*, **62**, 737-745.

Koenker, R.W. (2008) Censored Quantile Regression Redux *Journal of Statistical Software*, **27**, 6, 1-25.