SATELLITE MEETING ISI-COMMITTEE ON RISK ANALYSIS AND

XI WORKSHOP ON STATISTICS, MATHEMATICS AND COMPUTATION

- BOOK OF ABSTRACTS -

LISBON, JULY 10th
PORTALEGRE, JULY 11th-12th
2017

BOOK OF ABSTRACTS

Universidade Aberta de Lisboa
July 10, 2017
Lisboa - PORTUGAL

Instituto Politécnico de Portalegre
July 11–12, 2017
Portalegre - PORTUGAL

Sponsors

Participating Institutions

Committees

Executive Committee
Teresa Oliveira (Portugal)
Lidia Filus (USA)
Christos Kitsos (Greece)
Ivette Gomes (Portugal)

Scientific Committee
Adérito Marcos (Universidade Aberta, Portugal)
Amílcar Oliveira (Universidade Aberta, Portugal)
André Gustavo Pereira (Universidade Federal do Rio Grande do Norte, Brazil)
Anuj Mubayi (Arizona State University, USA)
Arminda Manuela Gonçalves (Universidade do Minho, Portugal)
Barbara Kimmel (University of Texas School of Public Health, USA
Carla Santos (Instituto Politécnico de Beja, Portugal)
Carlos Agra Coelho (Universidade Nova de Lisboa, Portugal)
Célia Nunes (Universidade da Beira Interior, Portugal)
Christina Ciecierski, Northeastern Illinois University, USA
Christos Kitsos (Technological Educacional Institution of Athens, Greece)
Christos Skiadas (Technical University of Crete, Greece)
Cristina Dias (Instituto Politécnico de Portalegre, Portugal)
Dário Ferreira (Universidade da Beira Interior, Portugal)
David Banks (Duke University, Durham, NC, USA)
Dinis Pestana (Faculdade de Ciências da Universidade de Lisboa, Portugal)
Dora Prata Gomes (Universidade Nova de Lisboa, Portugal)
Fernanda Figueiredo (Universidade do Porto, Portugal)
Fernando Carapau (Universidade de Évora, Portugal)
Filomena Teodoro (Escola Naval, Portugal)
Hammou el Barmi (Columbia University, USA)
Izabela Pruchnicka-Grabias (College of Eco. and Social, Banking Inst., Warsaw Poland)
Iwona Mejza (Poznan University of Life Sciences, Poland)
James R. Bozeman (Lyndon State College, USA)
Jerzy Filus (Oakton Community College, USA)
João Araújo (Universidade Aberta, Portugal)
João Miranda (Instituto Politécnico de Portalegre)
José António Pereira (Universidade do Porto, Portugal)
João Romacho (Instituto Politécnico de Portalegre)
Karl-Moder (University of Natural Resources and Life Sciences, Austria)
Lidia Filus (Northeastern Illinois University, USA)
Lisete Sousa, (Faculdade de Ciências da Universidade de Lisboa, Portugal)

Luís Miguel Grilo (Instituto Politécnico de Tomar, Portugal)
M. Ivette Gomes (Faculdade de Ciências da Universidade de Lisboa, Portugal)
Manuela Neves (ISA, Universidade de Lisboa, Portugal)
Marek Kimmel (Rice University, Houston, Texas, USA)
Maria Varadinov (Instituto Politécnico de Portalegre, Portugal)
Maria do Rosário Almeida (Universidade Aberta, Portugal)
Maria do Rosário Ramos (Universidade Aberta, Portugal)
Milan Stehlík (Johannes Kepler University in Linz, Austria)
Pawel Bartoszczuk (Warsaw School of Economics, Poland)
Robert G. Aykroyd (University of Leeds, UK)
Roberval Lima (Universidade Paulista e Universidade Federal do Amazonas, Brazil)
Sandra Ferreira (Universidade de Beira Interior, Portugal)
Sandra Nunes (Instituto Politécnico de Setúbal, Portugal)
Sneh Gulati (Florida International University, USA)
Stanislaw Mejza (Poznan University of Life Sciences, Poland)
Teresa Oliveira (Universidade Aberta, Portugal)
Tsuyoshi Nakamura (University of Nagasaki, Japan)
Vicente A. Núñez-Antón (Universidad del País Vasco, Spain)

Honour Committee

Rector of Universidade Aberta
President of Instituto Politécnico de Portalegre
President of Instituto Politécnico de Tomar
Coordinator of CEAUL-Centro de Estatística e Apli. da Univ. de Lisboa
President of ISI-Committee on Risk Analysis
President of ChicagoChec

Organizing Committee

Amílcar Oliveira (Universidade Aberta, Portugal)
Carla Santos (Instituto Politécnico de Beja, Portugal)
Célio Gonçalo Marques (Instituto Politécnico de Tomar, Portugal)
Fernando Carapau (Universidade de Évora, Portugal)
Luís Miguel Grilo (Instituto Politécnico de Tomar, Portugal)
Teresa Oliveira (Universidade Aberta, Portugal)

Local Organizing Committee (Instituto Politécnico de Portalegre, Portugal)

Adelaide Proença
Cristina Dias - Local Chair
Hermelinda Carlos
João Miranda
João Romacho
Maria José Varadinov

Technical Specifications

Title

Satellite Meeting ISI-Committee on Risk Analysis and XI Workshop on Statistics, Mathematics and Computation - Book of Abstracts

Editor

Authors

Teresa A. Oliveira, Amílcar Oliveira, Luís M. Grilo, Fernando Carapau, Cristina Dias and Carla Santos.

Published and printed by:

Preface

Dear participants, colleagues and friends,

WELCOME to the Satellite Meeting of the ISI-CRA, Committee on Risk Analysis of the International Statistical Institute organized jointly with the 11th Workshop on Statistics, Mathematics and Computation. We are delighted to promote another meeting under the topics of Risk Analysis and to jointly celebrate more than one decade of strong interaction between researchers from Portugal and abroad, leading to a successful commitment and enthusiasm on promoting research in the broad areas of Statistics, Mathematics and Computation.

It is a great pleasure to receive all our guests and contributors at the Universidade Aberta (UAb) main building, the Ceia Palace in Lisbon, in order to attend the beginning of our meetings on the 10th July, the ISI-CRA Satellite meeting in Honor of Professor David Banks and the 11th Workshop on Statistics Mathematics and Computation this year devoted to Professor Maria Antónia Turkman.

Keynote Speaker, Professor David Banks was, among many other positions, the coordinating editor of the Journal of the American Statistical Association, the President of the Committee on Risk Analysis and the President of ISBIS - International Society for Business and Industrial Statistics. His contributions were crucial on the development and expansion of Statistics by promoting it and encouraging new generations. Keynote Speaker, Professor Maria Antónia Turkman besides being an outstanding researcher in the area of Statistics, during several years she played a key role leading the Center of Statistics and Applications of the University of Lisbon and contributing for the respective international recognition. We are really grateful for her work, friendship and mentoring!

The activities and lectures will continue on 11-12 July, taking place in Portalegre, at the Instituto Politécnico de Portalegre (IPP), that kindly offered to accommodate the joint meeting. We express our gratitude to the Institute Coordinators for kindly accepted to embrace this Challenge. This joint meeting is a platform to exchange knowledge through international contacts involving Statisticians, Mathematicians, Computation professionals and Students, from several Universities, Organizations and other Institutions around the World. Our main objective is to assemble researchers and practitioners involved in the main areas of Statistics, Mathematics and Com-

putation, with special focus on Risk Analysis issues, by encouraging mutual exchange and interactions among a broad range of Science areas.

The format of the meeting involves Keynote Lectures, Plenary Talks, Organized Sessions and Poster Sessions. Special attention is given to applications and to new theoretical results having potential of solving real life problems. Discussion will be open in several areas, mainly with regard to challenges in Life Sciences, Medicine, Health Sciences, Cancer Research, Education, Management, Insurance and Industry.

Participants will have several journal opportunities for papers submission. Selected papers will appear in editions of Biometrical Letters, Journal Risks and in a Springer volume of the Series "Contributions to Statistics".

We are really grateful to all the participants for their valuable contribution, to Invited Speakers, Session Organizers and Authors who submitted abstracts, for the enthusiastic way how they assume their participation. We also acknowledge all the sponsors and contributors who made this meeting a reality.

Furthermore, we acknowledge our Honour Committee: Rector of UAb, President of IPP, Coordinator of the Center of Statistics and Applications of University of Lisbon, President of International Statistical Institute and President of the Committee on Risk Analysis of International Statistical Institute and for their precious support.

Many people and organizations contributed to the planning and execution of this joint meeting, and we are deeply grateful to all for that. However it is not an easy task to enumerate all the people who should be thanked. Nevertheless, there are a few groups and organizations that we would like to single out for special thanks. We are most grateful to UAb and to the IPP for hosting the meeting and to all the members of the Organizing Committee and of the Scientific Committee for their crucial help and suggestions. We wish to deeply thank to Cristina Dias, Luís M. Grilo and the Local Organizing Committee who were responsible for the meetings logistics at IPP and for organizing the social programme. Last but not least, special thanks are addressed to Amílcar Oliveira, Luís M. Grilo and Fernando Carapau for the invaluable contribution on organizing the Webpage, the final programme and

the Book of Abstracts details in an incredible short time. We believe that this joint meeting will be rewarding to all of us.

We wish you a very productive, successful, and enjoyable joint meeting, as well as a very nice stay in Lisbon and in Portalegre. Please, try also to find time to enjoy the delights of Portugal!!!

The Executive Committee (EC) - Chairs

Teresa Oliveira (Portugal)   Lidia Filus (USA)
Christos Kitsos (Greece)   Ivette Gomes (Portugal)

From Presidente of IPP

Welcome to the Instituto Politécnico de Portalegre!

It is an honor for the academic community of the Instituto Politécnico de Portalegre (IPP) to collaborate and host the ISI-CRA Satellite Meeting (in honor of Professor David Banks) and the 11th Workshop on Statistics, Mathematics and Computation (WSMC11). An initiative that is the result of a wide and diverse partnership and that has been consolidated in scientific terms. At the moment when Portugal brings together the entire academy to consolidate the importance of science for the development of societies, the promotion of tolerance and combating inequalities, I am particularly pleased that the IPP hosts this initiative, in partnership with other Institutions. I take this opportunity to express a word of appreciation to all partners for their contribution and availability. Congratulations to the Organizing Committee and I wish a good stay to all the participants. Enjoy the city and the region and make an appointment to return. We are waiting for you!

Joaquim Mourato
Presidente of IPP

Programme

Satellite Meeting ISI-CRA and WSMC11

July 10th 2017 (Monday) [PALÁCIO CEIA, UNIVERSIDADE ABERTA, LISBOA]

09:00 – 09:30 Registration Desk

09:30 – 10:00 Open Ceremony Satellite Meeting ISI CRA (SALA DE ATOS)

10:00 – 10:30 Keynote Speaker, David Banks (SALA DE ATOS)

Title: ***The Past, Present, and Future of Statistics***

10:30 – 11:00 Coffee break

11:00 – 11:30 Plenary Session, Carlos Braumann (SALA DE ATOS)

Title: ***General sustainable fishing models in randomly varying environments
and profit optimization***

11:30 – 12:00 Plenary Session, Carlos A. Coelho (SALA DE ATOS)

Title: ***Trimmed likelihood ratio statistics – an unexpected improvement in power***

12:00 – 14:00 Lunch

14:00 – 14:30 Keynote Speaker, Maria Antónia Turkman (SALA DE ATOS)

Title: ***Multivariate APC model in the analysis of stomach cancer incidence
in Southern Portugal***

14:30 – 15:00 Coffee break

15:00 – 15:30 Plenary Session, Rahim Moudmavand (SALA DE ATOS)

Title: ***Modelling Financial Data Using Modified Laplace Distribution***

15:30 – 16:00 Plenary Session, M. Ivette Gomes (SALA DE ATOS)

Title: ***Generalized means and peaks over random thresholds value-at-risk estimation***

16:00 – Closure of activities and departure to Portalegre

July 11th 2017 (Tuesday) [INSTITUTO POLITÉCNICO DE PORTALEGRE, PORTALEGRE]

09:00 – Registration Desk

09:30 – 10:00 Open Ceremony WSMC11 (Room: AUDITÓRIO)

10:00 – 10:30 Plenary Session, Rahim Mahmoudvand (Room: AUDITÓRIO)

Title: ***Determining Retention Limits in Reinsurance Using Bayesian Approach***

10:30 – 11:00 Coffee break

11:00 – 12:30 Parallel Sessions (Room: 2.01 and 2.06)

Organized Session S1; Room 2.01 Chair: Dora Gomes Title: ***Statistical Challenges and its applications***	Organized Session S2; Room 2.06 Chair: Filomena Teodoro Title: ***Computational Mathematics and Statistics in Sciences and Engineering I***
Comparison of the finite sample performance of several estimating methods for the Pareto distribution, Ayana Mateus ***On a Stochastic Model for a Cooperative Banking Scheme for Microcredit,*** Pedro Mota ***The Link between Exchange Rates and Commodity Prices is Not Always the Same: Lessons from Mozambique***, Alberto Mulenga ***Extreme value analysis: challenges and Applications,*** Manuela Neves	***Multicriteria Collection of Solid Urban Waste,*** Ana Jorge ***How to test different block diagonal structures in several covariance matrices,*** Filipe Marques ***An asymptotic approach of the likelihood-based exact inference for singly imputed synthetic data under the multiple linear regression model,*** Ricardo Moura ***Longitudinal modelling approach for MIBEL energy prices forecasting,*** Ana Borges

12:30 – 14:00 Lunch (IPP)

14:00 – 14:30 Plenary Session, Filomena Teodoro (Room: 2.01)

Title: ***Modeling caregivers literacy about pediatric arterial hypertension***

14:30 – 16:00 Parallel Sessions (Room: 2.01 and 2.06)

Organized Session S3; Room 2.01 Chair: Filomena Teodoro Title: ***Computational Mathematics and Statistics in Sciences and Engineering II***	Organized Session S4; Room 2.06 Chair: Dário Ferreira Title: ***Inference in Linear Models***
Determinants for hedge foreign exchange risk in portuguese and spanish companies, Eliana Costa e Silva ***Influence on scrap percentage of aluminium extrusion parameters,*** M. Fátima Almeida ***The impact of innovation Objectives on Product Innovation: CIS 2010,*** Aldina Correia ***Linear time series models for electricity pricing: the portuguese and spanish market,*** Maria de Fátima Pilar	***Mixed models with random sample sizes,*** Célia Nunes ***Global Confidence Regions for Mixed Models assuming Orthogonal Block Structure and Normality***, Sandra Ferreira ***Parametric models with random sample sizes***, Anacleto Mário

16:00 – 16:30 Coffee break + Poster Session

16:45 – Departure to village of Marvão by Bus

17:45 – Expected arrival in village of Marvão

17:45 – 19:30 Visit to the medieval village of Marvão (Castle)

19:30 – 22:00 Conference Dinner (Marvão)

22:00 – Return to Portalegre with arrival scheduled at 23:00

July 12th, 2017 (Wednesday) [INSTITUTO POLITÉCNICO DE PORTALEGRE, PORTALEGRE]

09:30 – 11:00 Parallel Sessions (Room: 2.01 and 2.06)

Organized Session S5; Room 2.01 Chair: Luís M. Grilo Title: *Computational Data Analysis and Statistical Inference*	Organized Session S6; Room 2.06 Chair: Amílcar Oliveira and Teresa Oliveira Title: *Statistical Modelling and Data Analysis*
Analysis of the influence of the period of activity and the size of the portfolios in their risk and diversification levels, João Romacho *Symmetric Stochastic Matrices width a Dominant Eigenvalue* Cristina Dias *On the enrichment of random factors models,* Carla Santos *Reverse Logistics within Pharmaceutical Supply Chains,* Maria Varadinov *Pharmaceutical Supply Chains: Training Early-Career Investigators in "Medicines Shortages",* João Miranda	*Canadian boreal forest fire disturbance and climate change,* Carla Francisco *Comparison of the Count Regression Models in Evaluation of the Effects of Periodontal Risk Factors on the Number of Teeth,* José António Pereira *Skewness and Kurtosis of the Extended Skew-Normal Distribution,* José António Seijas-Macias *Using the Rasch probabilistic outcome for assessing students' performance in an Information Security course,* Anacleto Correia *Performance of fishing policies for populations with weak Allee effects in a random environment,* Nuno Brites

11:00 – 11:30 Coffee break

11:30 – 12:00 Plenary Session, Amitava Mukherjee (Room: 2.01)

Title: *A Copula and Bootstrapped Test based scheme for simultaneous monitoring of bivariate processes with applications to manufacturing and Service Industries*

12:00 – 12:15 Closing Ceremony

12:30 – 14:00 Lunch (IPP)

14:00 – 17:30 Visit to the tapestry museum in city of Portalegre and a walking city tour

Contents

Keynote Speakers

Invited Speakers

Organized Sessions

Organized Session 6
   Statistical Modelling and Data Analysis
   Organizer: *Amílcar Oliveira and Teresa Oliveira*   58

Contributed Poster

Keynote Speakers

The Past, Present and Future of Statistics

David Banks[1]

[1]Duke University, USA

Abstract

Recently, I helped to edit a book with the same title as this talk. The book drew articles from 50 eminent statisticians who had won COPSS Awards for breakthrough work in statistical science. Some contributors wrote about their research, some about their career paths, and some about the history and future of our field. This address will try to synthesize the main ideas and themes, and point out the probable future, and its challenges, for the profession of statistics.

Multivariate APC model in the analysis of stomach cancer incidence in Southern Portugal

Papoila A.L.[1], Riebler A.[2], <u>Maria Antónia Turkman</u>[3], São João R.[4],
Ribeiro C.[5], Geraldes C.[6] and Miranda A.[7]

[1]Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Portugal e CEAUL
[2]Norwegian University of Science and Technology, Trondheim, Norway
[3]Centro de Estatística e Aplicações, Universidade de Lisboa (CEAUL), Portugal
[4]Escola de Gestão e Tecnologia, Instituto Politécnico de Santarém, Portugal e CEAUL
[5]Instituto Superior de Engenharia, Universidade do Algarve, Faro, Portugal e CEAUL
[6]Faculdade de Ciências Médicas, Universidade Nova de Lisboa, Portugal e CEAUL
[7]Inst. Port. de Oncologia de Lisboa de Francisco Gentil EPE, ROR-Sul, Portugal

E-mail address: *apapoila@hotmail.com, maturkman@fc.ul.pt*

Abstract

Stomach cancer occupies the fourth place among the different types of cancer that affect the population at world level and the second place in which pertains to mortality. It is known that its incidence varies with the location geographic area, with age and gender. Several studies have indicated the decline of the incidence of this type of cancer, possibly due to identification of helicobacter pylori infection as the main risk factor and its subsequent treatment (São João, 2015). In the south of Portugal it is among the first three types of cancer responsible for cancer mortality. There are several useful statistical methodologies in the study of trends in disease incidence rates. Particularly useful are models that simultaneously take into account three related components, namely age at diagnosis time (age), date of diagnosis (period), and birth date (cohort). These models are called Age-Period-Cohort (APC) models. The objective of this talk is to apply the Bayesian methodology to investigate, according to gender, relative temporal trends in the incidence of stomach cancer in the South of Portugal between 1998 and 2006, through the use of spatially multivariate APC models. Geographical variation is assessed by introducing a term of interaction between space and time. The data for this study were kindly provided by the Lisbon IPO. This talk is based on the work by Papoila et al. (2014), published in the Biometrical Journal. R-INLA (Rue et al., 2009) was thoroughly used in the analysis of the data and the script authored by Andreia Riebler is available in R INLA webpage (http://www.r-inla.org/examples/case-studies/papoila-et-al-2013).

Keywords: Bayesian Inference, Age Period Cohort Models, Spatial Temporal Models, Stomach Cancer.

Acknowledgements

References

[1] Papoila A.L., Riebler A., Amaral Turkman, M. A., São João R., Ribeiro C., Geraldes C. and Miranda A. (2014). Stomach cancer incidence in Southern Portugal 1998-2006: a spatio temporal analysis. Biometrical Journal 56, 403-15.

[2] H. Rue, S. Martino, and N. Chopin (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). Journal of the Royal Statistical Society, Series B, 71(2), 319-392.

[3] São João, R. (2015). Métodos estatísticos aplicados à modelação de dados oncológicos. Tese de Doutoramento em Ciências da Vida (especialidade de Bioestatística e Informática). Faculdade de Ciências Médicas, Universidade Nova de lisboa.

Invited Speakers

Trimmed likelihood ratio statistics - an unexpected improvement in power

Carlos A. Coelho[1,2]

[1]Universidade Nova de Lisboa – Faculdade de Ciências e Tecnologia
[2]Mathematics Department and CMA–Centro de Matemática e Aplicações, FCT/UNL

E-mail address: *cmac@fct.unl.pt*

Abstract

The author shows how by using "trimmed" versions of some likelihood ratio test statistics we may obtain what may be at first sight quite unexpected gains in power, when compared with the power obtained when using the 'complete' likelihood ratio statistics. We call here "trimmed" likelihood ratio statistics, the statistics obtained from the common likelihood ratio statistics by deleting terms that appear more than once. Such "repeated" terms appear most commonly in likelihood ratio statistics used to test the equality of mean vectors when the covariance matrices are assumed to have some common structure like a compound symmetric or symmetric circulant structure, either scalar or by blocks [1–4].

Keywords: Exact distributions, likelihood inference, likelihood ratio test, Multivariate Analysis, near-exact distributions.

Acknowledgements

References

[1]Coelho, C. A. (2017) Likelihood Ratio Tests for Equality of Mean Vectors with Circular Covariance Matrices, in Oliveira, T. A., Kitsos, C., Oliveira, A., Grilo, L. M. (eds.) "Recent Studies on Risk Analysis and Statistical Modeling", Lecture Notes in Statistics, Springer (in print).
[2]Coelho, C. A. (2017) The Likelihood Ratio Test for Equality of Mean Vectors with Compound Symmetric Covariance Matrices, in Lecture Notes in Computer Science, Springer (accepted for publication).
[3]Coelho, C. A. (2017) Likelihood Ratio Test for the Equality of Mean Vectors when the Joint Covariance Matrix is Block-circulant or Block Compound Symmetric, in ICOSDA 2016 Abstract book, pp. 30.
[4]Coelho, C. A. (2017) The Likelihood Ratio Test for the Equality of Mean Vectors when the Covariance Matrices are Block Compound Symmetric, in Skiadas, C. H. (ed.), ASMDA 2017 Abstract book, pp. 5-6.

Determining Retention Limits in Reinsurance Using Bayesian Approach

Rahim Mahmoudvand[1]

[1]Department of Statistics, Bu-Ali Sina University, Hamedan, Iran

E-mail address: *r.mahmodvand@gmail.com*

Abstract

A popular method for finding retention limit in reinsurance treaties, is to find retention M that optimizes a risk measure. But the risk measure is a function of the parameters of the loss distribution. These parameters can be considered in Bayesian framework imply variable M. This paper consider the possibility of using Bayesian approach for finding retention limits in reinsurance treaties.

Keywords: Reinsurance, Retention Limits, Bayesian Approach, Value at Risk.

—–⊙—–
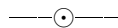
Setting out the optimal reinsurance contract is one of the significant challenges of both reinsured and reinsurance company. Let X denote loss amount and X_I and X_R show reinsured and reinsurer contributions, respectively; means that we have decomposed X as below:

$$X = X_I + X_R. \tag{1}$$

Main problem in reinsurance is the way of determining X_I. One solution is to consider

$$X_I = \min\{X, M\} = \begin{cases} X & , \ X \leq M, \\ M & , \ X > M. \end{cases} \tag{2}$$

Where M is called retention. Then, the problem will become to the way that we can use to determine retention. In order to do it, we should define a risk measure like value at risk (VaR) and determine the retention so that the VaR of the loss or the primary insurer's exposure becomes minimized. Up to this day, many studies have been published about this problem, see Payandeh Najafabadi and Panahi Bazaz (2016) and references therein. However, we have just seen in Payandeh Najafabadi and Panahi Bazaz (2016) that M can be obtained by Bayesian approach.

Main results

Denote by $T = X_I + \delta(M)$, the total cost of reinsured, where $\delta(M) = (1 + \rho)E(X - X_I)$ is the reinsurance premium, Cai and Tan (2007) showed that:

$$M = S_X^{-1}\left(\frac{1}{1+\rho}\right) \tag{3}$$

minimise VaR of T, where $S_X(x)$ is the survival function of X. The above solution is based on classical approach. Let us provide two main questions:

1. Could we consider M as a statistical parameter? Recall that parameter has specific meanings within various disciplines, including mathematics, computing and computer programming, engineering, statistics, etc. According to Everett and Skrondal (2010) a statistical parameter or population parameter is a quantity that indexes a family of probability distributions. It can be regarded as a numerical characteristic of a population or a statistical model.
2. Assume M can be considered as a statistical parameter. Then, the next question is that whether we are able to treat M as a random variable with a distribution to use Bayesian approach.

These concerns show that the study by Payandeh Najafabadi and Panahi Bazaz (2016) needs to be considered more carefully when we should use VaR. In my opinion, M can be considered in a Bayesian framework just because it is a function of the loss distribution parameters. In this case, using quadratic loss function, optimal M must minimises:

$$E\left\{\left[M - S_X^{-1}(1/(1+\rho))\right]^2 | x_1, \ldots, x_n\right\}, \tag{4}$$

imply that:

$$\hat{M}_{Bayes} = E\left\{S_X^{-1}\left(\frac{1}{1+\rho}\right) | x_1, \ldots, x_n\right\} \tag{5}$$

As an example, let loss amounts $X_1, \ldots, X_n$ given risk parameter θ, are exponentially distributed with mean θ. Then,

$$\hat{M}_{Bayes} = \log(1 + \rho)E(\theta | x_1, \ldots, x_n). \tag{6}$$

References

[1]Cai, J., Tan, K.S., (2007). Optimal retention for a stop-loss reinsurance under the VaR and CTE risk measures. Astin Bull. 37 (01), 93–112.
[2]Everitt, B. S.; Skrondal, A. (2010), The Cambridge Dictionary of Statistics, Cambridge University Press.
[3]Payandeh Najafabadi, A.T., and Panahi Bazaz, A. (2016). An optimal co-reinsurance strategy. Insurance: Mathematics and Economics, 69: 149–155.

General sustainable fishing models in randomly varying environments and profit optimization

Carlos A. Braumann[1,2] and Nuno M. Brites[1]

[1]Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e
Formação Avançada, Universidade de Évora
[2]Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora

E-mail address: *braumann@uevora.pt, brites@uevora.pt*

Abstract

In a randomly varying environment, we model the growth of a fish population by a very general stochastic differential equation (SDE) model, where an also very general sustainable harvesting policy is applied. We show existence and uniqueness of the solution and determine the conditions for non-extinction of the population and existence of a stationary density, i.e. for the existence of a sustainable stochastic equilibrium. Results are then applied to optimization issues for particular harvesting policies.

Keywords: General fishing models, profit optimization, random environments, stationary density, stochastic differential equations.

—–⊙—–
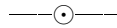
In a randomly varying environment, we model the growth of the harvested population by a very general stochastic differential equation (SDE) model, where an also very general sustainable fishing policy is applied:

$$dX(t) = g(X(t))X(t)dt - qE(X(t))X(t)dt + \sigma(X(t))X(t)dW(t)$$

with $X(0) = x_0 > 0$, where $W(t)$ is a standard Wiener process.

Use of very general models has the advantage of obtaining results and desirable properties of sustainable harvesting policies that are model robust. Results on this general model have already been presented in [1], [2]; they are here reviewed and extended, as well as applied to optimization issues.

Stratonovich calculus is used for a more convenient interpretation, but attention is called to the fact that results are equivalent if one uses Itô calculus (see [2]). We consider the natural growth of the harvested population to be of a very general density-dependent form, with a geometric average *per capita* natural growth rate $g(x)$ being a C^1 strictly decreasing function and with

$\sigma(x)\varepsilon(t)$ (where $\sigma(x)$ is a positive C^2 noise intensity and $\varepsilon(t) = dW(t)/dt$ is a standard white noise) describing the effect of environmental random fluctuations on that rate. The catchability coefficient is $q > 0$ and $E(x)$ is a non-negative C^1 function representing a very general sustainable harvesting effort when population size is x.

We show existence and uniqueness of the solution and determine the conditions for non-extinction of the population and existence of a stationary density, i.e. for the existence of a sustainable stochastic equilibrium. Results are then applied to optimization issues for constant effort fishing policies.

Acknowledgements

References

[1]Braumann, C. A. (2002) Variable effort harvesting models in random environments: generalization to density-dependent noise intensities. *Mathematical Biosciences* **177 & 178**: 229–245.
[2]Braumann, C. A. (2007) Harvesting in a random environment: Itô or Stratonovich calculus? *Journal of Theoretical Biology* **244**: 424–432.

Modeling caregivers literacy about pediatric arterial hypertension

M. Filomena Teodoro[1,2] and Carla Simão[3]

[1]CINAV, Center of Naval Research, Naval Academy, Portuguese Navy, Base Naval de Lisboa, Alfeite, Portugal
[2]CEMAT, Center for Computational and Stochastic Mathematics, Instituto Superior Técnico, Lisbon University, Portugal
[3]Pediatric Department, Santa Maria Hospital, Centro Hospitalar Lisboa Norte, Portugal

E-mail address: *maria.alves.teodoro@marinha.pt, carla.mail@netcabo.pt*

Abstract

Pediatric high blood pressure (PHBP) is an important public health issue with multiple problems on the health of children and adults [1]. It is an important issue that health professionals and family members have awareness of PHBH existence, the negative consequences associated with it, the risk factors and it prevention. Its expected the knowledge about this pathology increases with the level of education of the family members [4]. To investigate the caregivers knowledge degree about PHBP and to check if the assessment of pediatric blood pressure in health regular consultations is an usual practice, firstly [3], an experimental questionnaire introduced in [2] was built with the aim of easy and quick answers under the aim of a preliminary study about PHBP caregivers acquaintance. The statistical analysis and modeling of such questionnaire by Generalized Linear Model (GLM), took into account the presence of dichotomous data, discrete models which are usually estimated by logistic or probit regression. This approach can be found in [6,7] where each statistical significant question was modeled separately. Continuing such approach and exploring the idea of simplicity and reduced dimensionality, a factor analysis was applied and a multivariate analysis of variance was performed considering the obtained factors by exploratory factor analysis (EFA) as dependent variables [8]. An extension of such questionnaire was built and applied to a distinct population and it was filled online [5]. Some statistical models were estimated using several approaches, namely multivariate analysis (factorial analysis), also adequate methods to analyze the kind of data in study [9,10]. An approach by GLM is still ongoing. The results are promising but their analysis still need to be completed.

Keywords: Pediatric hypertension, caregiver, knowledge, factor analysis, analysis of variance, GLM.

Acknowledgements

References

[1]Bassareo, P.P. and Mercuro, G. (2014) Pediatric hypertension: An update on a burning problem. World Journal of Cardiology **6**(5): 253–259.

[2]Costa, J.R.B.(2015) *Hipertensão arterial em idade pediátrica: que conhecimento têm os prestadores de cuidados sobre esta patologia?*, Master Thesis, Medical Faculty, Lisbon University. (In portuguese)

[3]Lurbe, E. and Cifkovac, R.F. (2009) Management of high blood pressure in children and adolescents: recommendations of the European Society oh Hypertension. *Journal of Hypertension* **27**: 1719—1742.

[4]National High Blood Pressure Education Program Working Group on High Blood Pressure in Children And Adolescents (2004) The fourth report on the diagnosis, evaluation and treatment oh high blood pressure in children and adolescents. *Pediatrics* **114**: 555–576.

[5]Romana, A. (2017) *Hipertensão Arterial em Pediatria – Um estudo observacional sobre a literacia dos cuidadores*, Master Thesis, Medical Faculty, Lisbon University . (In portuguese)

[6]Teodoro, M. F. and Simão, C. (2017) Perception about Pediatric Hypertension. *Journal of Computational and Applied Mathematics* **312**: 209–215.

[7]Teodoro, M. F. and Simão, C. (2017) Completing the Analysis of a Questionnaire About Pediatric Blood Pressure. *Transactions on Biology and Biomedicine*, World Sci. Eng. Acad. Soc. **14**: 56–64.

[8]Teodoro, M. F. and Simão, C. (2017) Notes about Pediatrics Hypertension Literacy. *Transactions on Biology and Biomedicine*, World Sci. Eng. Acad. Soc. **14**: 89–97 .

[9]Teodoro, M. F., Romana, A. and Simão, C. An Issue of Literacy on Pediatric Hypertension. In T. Simos et al., editors, *Computational Methods in Science and Engineering*. New York, AIP Conference Proceedings. (to appear)

[10]Teodoro, M. F., Simão, C. and Romana, A. Questioning Caregivers About Pediatric High Blood Pressure. In Gervasi, O. et al. (Eds.), Computational Science and Its Applications – ICCSA 2017 Part V, *Lecture Notes in Computer Science* **10408**: chapter 4. (in press). DOI: 10.1007/978-3-319-62404-4 4435 – 462.

Generalized means and peaks over random thresholds value-at-risk estimation

M. Ivette Gomes[1,2], Fernanda Figueiredo[2,3] and Lígia
Henriques-Rodrigues[2,4]

[1]DEIO, Faculdade de Ciências, Universidade de Lisboa, Portugal
[2]Centro de Estatística e Aplicações da Universidade de Lisboa (CEAUL)
[3]Faculdade de Economia, Universidade do Porto
[4]Universidade de São Paulo, IME, São Paulo, Brasil

E-mail address: *ivette.gomes@fc.ul.pt, otilia@fep.pt, ligiahr@ime.usp.br*

Abstract

In many areas of application, it is a common practice to estimate the *value at risk* at a level q (VaR$_q$), a value, high enough, so that the chance of an exceedance of that value is equal to q, small, often smaller than $1/n$, with n the size of available sample, $\underline{\mathbf{X}}_n = (X_1, \ldots, X_n)$. Let us denote by $X_{1:n} \leq \cdots \leq X_{n:n}$ the associated ascending order statistics. Further assume that there exist sequences $\{a_n > 0\}_{n \geq 1}$ and $\{b_n \in R\}_{n \geq 1}$ such that the maximum, linearly normalized, i.e. $(X_{n:n} - b_n)/a_n$, converges in distribution to a non-degenerate random variable. Then (Gnedenko, 1943), the limit *cumulative distribution function* (CDF) is necessarily of the type of the general *extreme value* (EV) CDF, given by

$$G_\xi(x) = \begin{cases} \exp(-(1 + \xi x)^{-1/\xi}), & 1 + \xi x > 0, \text{ if } \xi \neq 0, \\ \exp(-\exp(-x)), & x \in R, \qquad\quad \text{ if } \xi = 0. \end{cases}$$

The CDF F is said to belong to the max-domain of attraction of G_ξ, and we write $F \in \mathcal{D}_\mathcal{M}(G_\xi)$. The parameter ξ is the *extreme value index* (EVI), the primary parameter of extreme events. We consider *heavy-tailed* models, i.e. *Pareto-type* underlying CDFs, with a positive EVI, working in $\mathcal{D}_\mathcal{M}^+ := \mathcal{D}_\mathcal{M}(\mathrm{EV}_{\xi>0})$. These heavy-tailed models are quite common in many areas of application, like biostatistics, finance, insurance, statistical quality control and telecommunications, among others. For heavy-tailed models, the classical EVI-estimators are the Hill (H) estimators (Hill, 1975),

$$\mathrm{H}(k) \equiv \mathrm{H}(k; \underline{\mathbf{X}}_n) := \frac{1}{k} \sum_{i=1}^{k} \ln X_{n-i+1:n} - \ln X_{n-k:n},, \qquad 1 \leq k < n.$$

The classical semi-parametric VaR-estimators were introduced in Weissman (1978), being given by

$$\widehat{\mathrm{VaR}}_q(k) := X_{n-k:n} \; (k/(nq))^{\mathrm{H}(k)},$$

but $\mathrm{H}(k)$ can be replaced in the previous formula by any consistent EVI-estimator. We now suggest ways to improve the performance of the existent VaR-estimators. The estimation of high quantiles depends strongly on the EVI-estimation, and recently, new classes of reliable EVI–estimators based on adequate generalized means (GM), and dependent on a tuning real parameter q have appeared in the literature (see Caeiro *et al.*, 2016, Penalva *et al.*, 2016, Paulauskas and Vaičiulis, 2017, and references therein), and will be used for the estimation of VaR$_q$, $q < 1/n$. Due to the fact that the GM VaR-estimators do not enjoy the adequate behaviour, in the sense that they do not suffer the appropriate linear shift in the presence of linear transformations of the data, as does any theoretical quantile and

the PORT-Weissman-Hill VaR-estimators introduced in Araújo-Santos *et al.* (2006), with PORT stating for *peaks over random thresholds*, we further discuss the so-called PORT GM VaR-estimators, more general than the class in Figueiredo *et al.* (2017).

Keywords: Generalized means; heavy-tailed parents; semi-parametric estimation, value-at-risk.

Acknowledgements

References

[1] Araújo Santos, P., Fraga Alves, M.I. and Gomes, M.I. (2006). Peaks over random threshold methodology for tail index and high quantile estimation. *Revstat* **4**:3, 227–247.

[2] Caeiro, F., Gomes, M.I., Beirlant, J. and de Wet, T. (2016). Mean-of-order p reduced-bias extreme value index estimation under a third-order framework. Extremes **19**:4, 561–589.

[3] Figueiredo, F., Gomes, M.I. and Henriques-Rodrigues, L. (2017). Value-at-risk estimation and the PORT mean-of-order-p methodology. *REVSTAT* **15**:2, 187–204.

[4] Gnedenko, B.V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.* **44**, 423-453.

[5] Hill, B.M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Statist.* **3**, 1163-1174.

[6] Paulauskas, V. and Vaičiulis, M. (2017). A class of new tail index estimators. *Annals Institute of Statistical Mathematics* **69**:2, 461–487.

[7] Penalva, H., Caeiro, C., Gomes, M.I. and Neves, M.M. (2016). *An Efficient Naive Generalisation of the Hill Estimator: Discrepancy between Asymptotic and Finite Sample Behaviour*. Notas e Comunicações CEAUL 02/2016.

[8] Weissman, I. (1978). Estimation of parameters and large quantiles based on the k largest observations. *J. American Statistical Association* **73**, 812–815.

GModelling Financial Data Using Modified Laplace Distribution

Rahim Mahmoudvand[1]

[1]Department of Statistics, Bu-Ali Sina University, Hamedan, Iran

E-mail address: *r.mahmodvand@gmail.com*

Abstract

This paper, present a modified version of the classical Laplace distribution. This distribution applied for modelling two real datasets: exchange rate and insurance loss ratio. The results suggest that further improvement to classical Laplace distribution fitting is possible and the new model provides an attractive alternative to the classical Laplace distribution

Keywords: General fishing models, profit optimization, random environments, stationary density, stochastic differential equations.

———⊙———
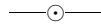
1 Introduction

The Laplace distribution belongs to the oldest distributions in probability theory. Its instances continue to enjoy applications in a variety of disciplines which range from image and speech recognition through ocean engineering to finance. Up to this day, many studies have been published with extensions and applications of the Laplace distribution. Extensions to a skewed model as well as to a multivariate setting can be found in the literature, see Kotz et al. (2001) and Mahmoudvand et al. (2015) and references therein. Nevertheless, the current forms of the Laplace distribution (both classical and generalised forms) have a sharp peak in the middle, which potentially restricts their usefulness. In the light of this issue, we present a new probability distribution in this paper that can be derived from the symmetric Laplace distribution and can be used for modelling and analysing real data, when a flat shape in the middle of the distribution can be observed. The modified classical Laplace distribution is a probability distribution on $(-\infty, +\infty)$ and its probability density function is given by

$$f(x; \theta, \sigma) = \frac{1}{3\sigma} \exp \left\{ -\frac{1}{2\sigma} \left[|x - \theta| + |x - \sigma - \theta| - \sigma \right] \right\} \; ; \; x \in \mathbb{R}. \qquad (7)$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$. Here, we use the notation $\mathcal{CL}(\theta, \sigma)$ for the classical Laplace distribution and $\mathcal{MCL}(\theta, \sigma)$ for the new modified version, respectively.

2 Empirical Results

Exchange rate data We consider the annual exchange rates between the United States Dollar (USD) and the Czech Republic Koruna (CZK, 1990-2014), United Kingdom Pound (GBP, 1975-2014), Swedish Krona (SEK, 1960-2014) and Danish Krone (DKK, 1960-2014). The data were obtained from the OECD (2014), available from the website `http://stats.oecd.org`. We examine the fits of $\mathcal{MCL}(\theta, \sigma)$ and $\mathcal{CL}(\theta, \sigma)$ for the logarithm of these data. K-S test statistics and associated p-values are reported in Table 1.

Loss ratio data We consider the logarithm of the loss ratio for private cars in motor insurance, based on about 200000 policies and 28721 associated claims in the period Mar 2010 to Mar 2013. The results of the K-S test and associated p-values are reported in Table 2.

References

[1] Kotz, S. Kozubowski, T. and Podgorski, K. (2001), The Laplace Distribution and Generalizations: A Revisit with New Applications. Springer.

[2] Mahmoudvand, R.; Faradmal. J.; Abbasi, N.; and Lurz, K. (2015). A New Modification of the Classical Laplace distribution. Journal of The Iranian Statistical Society, 14(2): 93-110.

Table 1. Kolmogorov-Smirnov results of fitting $\mathcal{MCL}(\theta, \sigma)$ and $\mathcal{CL}(\theta, \sigma)$ to the exchange rates data.

Quantity	CZK		DKK		SEK		GBP	
	MCL	CL	MCL	CL	MCL	CL	MCL	CL
K-S	0.1399	0.1937	0.1045	0.1437	0.1638	0.2220	0.1074	0.1649
p-value	**0.6612**	0.2687	**0.5852**	0.2059	**0.1044**	0.0089	**0.7057**	0.2029

Table 2. Kolmogorov-Smirnov results of fitting $\mathcal{MCL}(\theta, \sigma)$ and $\mathcal{CL}(\theta, \sigma)$ to the loss ratio data.

Sex	Year	length	$\mathcal{MCL}$		$\mathcal{CL}$	
			K-S	p-value	K-S	p-value
Male	2010-2011	6714	0.0258	0.0003	0.0424	7.1e-11
	2011-2012	8629	0.0861	2.1e-16	0.0379	3.2e-11
	2012-2013	6382	0.0316	5.8e-06	0.0436	4.8e-11
Female	2010-2011	2208	0.0250	0.1274	0.0429	0.0006
	2011-2012	3143	0.0303	0.0062	0.0505	2.3e-07
	2012-2013	1645	0.0336	0.0482	0.0456	0.0022

A copula and bootstrapped test based scheme for simultaneous monitoring of bivariate processes with applications to manufacturing and service industries

Amitava Mukherjee[1]

[1]XLRI – Xavier School of Management, Production, Operation and Decision Sciences Area, Jharkhand, India

Abstract

Simultaneous monitoring of the bivariate processes, such as, event frequency and magnitude, time between the two successive contract labour strikes and the duration of the strike; call centre response times and its service quality; machine downtime and the consequential financial loss or time between the two major earthquakes and the magnitude in Richter scale etc., is an important research topic in various areas ranging from production, operations and manufacturing to event environmental science. Most of the traditional approaches for simultaneous monitoring of event frequency and magnitude either assume that the event frequency and magnitude are two independent variables or consider a bivariate parametric model, such as bivariate gamma or bivariate normal. In reality, it is often difficult to justify such assumptions. Therefore, in the present work, we introduce a nonparametric bootstrapped p-value based Shewhart type procedure taking into account the dependence between two events or an event and its consequences. We assume that the underlying process distribution is unknown but random samples from retrospective phase are obtained as preliminary information. The concept of ranks and empirical copula are utilized in constructing the proposed nonparametric monitoring procedure. Our proposed procedure can also be applied in monitoring shifts in any general bi-variate processes. A special advantage of the proposed scheme is its follow-up procedure through which practitioners may identify the variable where the shift is more significant. Design and implementation procedure of the proposed scheme are described in details. We illustrate our proposed scheme with a real example and show the various construction steps. Some properties of the proposed procedure are studied though Monte-Carlo simulation.

Keywords: Bivariate Process, Call Centre, Cramér–von Mises Criterion, Empirical Copula, Monte-Carlo, Nonparametric, Phase-II, Ranks, Service Quality, Shewhart Chart.

Organized Sessions

Organized Session 1

Statistical Challenges and its applications

Organizer: Dora Gomes (Portugal)

Comparison of the finite sample performance of several estimating methods for the Pareto distribution

Ayana Mateus[1,2] and Frederico Caeiro[1,2]

[1] Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Portugal
[2] Centro de Matemática e Aplicações (CMA)

E-mail address: *amf@fct.unl.pt, fac@fct.unl.pt*

Abstract

The Pareto distribution was first introduced as a model for large incomes [4] and nowadays has been extensively used for modelling events in fields such as bibliometrics, demography, insurance, or finance, among others. Although there are several variants of this distribution, in this work we shall consider the classic Pareto distribution also known as Pareto type I. Estimation of the Pareto distribution has already been extensively addressed in the literature. The objective of this work is to compare the finite sample performance of several Pareto estimation methods, namely the Moment, Maximum Likelihood, probability weighted moments and log probability weighted moments method. The comparison will be made through a Monte-Carlo simulation study. We also present asymptotic results for the log probability weighted moment estimators.

Keywords: Monte Carlo simulation, Pareto distribution, Probability weighted moments.

Acknowledgements

References

[1] Arnold, B. C. (2008) Pareto and Generalized Pareto Distributions. *In D. Chotikapanich (ed.) Modeling Income Distributions and Lorenz Curves*, Springer, New York: 119–145.

[2] Arnold, B. C. , Balakrishnan, N. and Nagaraja, H. N. (1992) *A first course in order statistics*, New York, Wiley.

[3] Caeiro, F. and Gomes, M. I. (2010) Semi-parametric tail inference through probability-weighted moments. *Journal of Statistical Planning and Inference* **141**: 937–950.

[4] Pareto, V. (1897) Cours de économie politique. Rouge, Lausanne.

[5] Quandt, R. E. (1966) Old and new Methods of Estimation and the Pareto Distribution. *Metrika*, **10**: 55–82.

On a Stochastic Model for a Cooperative Banking Scheme for Microcredit

Manuel L. Esquível[1,2], Pedro Mota[1,2] and Joaquim P. Pina[3,4]

[1]Department of Mathematics, FCT NOVA
[2]Centre of Mathematics and Applications (CMA/FCT NOVA)
[3]Department of Applied Social Sciences, FCT NOVA
[4]Center for Advanced Studies in Management and Economics of the University of Évora, Branch FCT NOVA, CEFAGE-FCT NOVA

E-mail address: *mle@fct.unl.pt, pjpm@fct.unl.pt, jagl@fct.unl.pt*

Abstract

We propose and study a simple model for microcredit using two sums, of identically distributed random variables, the number of terms being Poisson distributed; the first sum accounts for the payments made to the *collective vault* by the participant and the second sum, subtracted to the first, accounts for the loans received by the participant. The process under study is the sum of the individual processes of a finite number of participants in the *collective vault*. Under a global independence hypothesis we show, by mean of moment generating functions, that if, for all the participants and at any time, on average, the sum of the loans is strictly smaller than the sum of the payments to the *collective vault* then the probability of the *collective* vault failing can be made arbitrarily small provided the loans only start to be accepted after a sufficiently large delay. We present numerical illustrations of *collective vaults* for exponential and chi-squared distributed random terms. For the practical management of such a *collective vault* it may be advisable to have loan granting rules that destroy independence of the random terms. We present a simulation study that shows the effect of such a breaking dependence loan granting rule on maintaining the stability of the *collective vault*.

Keywords: Collective vault; Microcredit and Ruin.

Acknowledgements

The Link between Exchange Rates and Commodity Prices is Not Always the Same: Lessons from Mozambique

Alberto Mulenga[1], Marta Faias[2,3], Pedro Mota[2,3] and Joaquim P. Pina[4,5]

[1]Department of Mathematics and Informatics, Faculty of Sciences Eduardo Mondlane University
[2]Department of Mathematics, FCT NOVA
[3]Centre of Mathematics and Applications (CMA/FCT NOVA)
[4]Department of Applied Social Sciences, FCT NOVA
[5]Center for Advanced Studies in Management and Economics of the University of Évora, Branch FCT NOVA, CEFAGE-FCT NOVA

E-mail address: *mulengamz@yahoo.com.br, a.mulenga@campus.fct.unl.pt mcm@fct.unl.pt, pjpm@fct.unl.pt, jagl@fct.unl.pt*

Abstract

We estimate multivariate autoregressive conditional heteroskedasticity models (M-GARCH) seeking to identify the patterns of comovement among exchange rates and commodity prices, particularly in different situations as to marked behavior in terms of currency appreciation/depreciation. We derive lessons from Mozambique, particularly tailored to the aimed setup and further adding up a previewed Ducth disease situation. Taking data for Mozambican New Metical (MZN) against South African Rand (ZAR) and also against Great Britain Pound (GBP), plus the quote of price of Coal for Africa (CZA), we employ diagonal and possibly varying correlation models of the M-GARCH family. Specifically, we run these three-variate models to a full sample from 2010-2014 and to four sub-samples −2010, time of depreciation; 2011, time of appreciation; 2012, stability period; 2013-2014, stability with some appreciation. Results suggest that the link across markets, namely exchange rates and commodity prices, is not always the same, it depends on the type of "systematic" behavior of the exchange rate.

Keywords: Exchange rates, commodities, M-GARCH, transmission across markets.

Acknowledgements

Extreme value analysis: challenges and applications

M.M. Neves[1,2], H. Penalva[3,4], S. Nunes[3,4] and D. Prata Gomes[5,6]

[1]Universidade de Lisboa, Instituto Superior de Agronomia
[2]Centro de Estatística e Aplicações, Lisboa, Portugal
[3]Escola de Ciências Empresariais do Instituto Politécnico de Setúbal
[4]Centro de Estatística e Aplicações, Lisboa, Portugal
[5]Universidade Nova de Lisboa, Faculdade de Ciências e Tecnologia
[6]Centro de Matemática e Aplicações, Lisboa, Portugal

E-mail address: *manela@isa.ulisboa.pt, helena.penalva@esce.ips.pt*
sandra.nunes@esce.ips.pt, dsrp@fct.unl.pt

Abstract

Some of the initial applications of extreme value theory (EVT) were dedicated to describe and predict the behaviour of extreme values in fields such as hydrology, meteorology and insurance. Implementation of EVT faces many challenges, including the scarcity of extreme data, the decision whether the distribution of the data is "fat-tailed", the choice of the threshold or beginning of the tail, and the choice of adequate methods for the estimation of the parameters. Other difficulties were related with the existence of "good" software for dealing with all those challenges. The main objective of this talk is to discuss the difficulties in using EVT through some applications in traditional fields and also in more recent ones. The main steps for performing an extreme value analysis, the use of some packages and functions in the R environment and different parameter estimation methods are presented.

Keywords: Extremal index, extreme value analysis, extreme value index, parameter estimation, R environment.

———⊙———

Extreme value theory (EVT) is concerned with the stochastic behaviour of extreme values. Originally developed to address problems in hydrology such as that one of deciding the height to build a dam in order to guard against a 100-year flood, the applications quickly spread out to many other fields such as meteorology (precipitation), insurance and risk management, telecommunications, environmental risk assessment, where catastrophic consequences for human lives can appear. One of the fields where it seems that EVT have been a limited role in the past, but nowadays is showing an increasing interest, is the ecological domain, see [4] and [1].

Let $X_1, X_2, , \cdots, X_n$ be independent and identically distributed (i.i.d.) random variables with common unknown distribution function F. First results in the asymptotic theory of sample extremes distribution date back to the beginning of the XX century, but were [2] and [3] who gave conditions for the existence of sequences $\{a_n\} \in R^+$ and $\{b_n\} \in R$ such that the partial maxima, $M_n \equiv \max(X_1, X_2, \ldots, X_n)$, suitably normalized, had a non-degenerate limiting distribution G, called Extreme Value distribution and is usually denoted by EV. This function is given by

$$EV_\xi(x) = \begin{cases} \exp[-(1+\xi x)^{-1/\xi}], \ 1+\xi x > 0, \ if \ \xi \neq 0 \ ; \\ \exp[-\exp(-x)], \qquad x \in R, \ if \ \xi = 0, \end{cases}$$

where the shape parameter ξ is called *the extreme value index*, a primary parameter in EVT, measuring the heaviness of the right-tail, $\overline{F} := 1 - F$. Besides this parameter, another one deserves a special attention; it is related to the clustering of high values, a situation very common in applications and that occurs with stationary sequences. This parameter, called *the extremal index*, θ, needs to be adequately estimated, not only by itself but because it influences other important parameters in EVT.

The application of EVT involves a number of challenges. The early stage of data analysis is very important in determining whether the series has the fat tail needed to apply the EVT results. The parameter estimates of the limit distributions EV depend also on the number of extreme observations used. The choice of a threshold should be large enough to satisfy the conditions to permit its application, while at the same time leaving sufficient observations for the estimation. Those steps will be illustrated in applications, in the sequel of [5] work.

Acknowledgements

References

[1] Garcia, C. and Borda-de-Água, L. (2017). Extended dispersal kernels in a changing world: insights from statistics of extremes. *Journal of Ecology*, **105**, 63–74.

[2] Gnedenko, B. V.(1943). Sur la distribution limite d'une série aléatoire *Annals of Mathematics*, **44**, 423-453.

[3] de Haan, L. (1970). *On Regular Variation and its Applications to the Weak Convergence of Sample Extremes*, Mathematical Centre Tract 32, Amesterdam, Dordrecht: D.Reidel.

[4] Katz,R.W., Brush, G.S. and Parlange, M.B. (2005) Statistics of Extremes: modelling ecological disturbances. *Ecology* **86** (5): 1124–1134.

[5] Penalva, H., Nunes, S. and Neves, M.M. (2016). Extreme Value Analysis–a brief overview with an application to flow discharge rate data in a hydrometric station in the Norh of Portugal. *Revstat* **14**, 2, 193-215.

Organized Session 2

Computational Mathematics and Statistics in Sciences and Engineering I

Organizer: M. Filomena Teodoro

Multicriteria collection of solid urban waste

Ana Maria Jorge[1,2] and Vladimir Bushenkov[2]

[1]ISCAL, Instituto Superior de Contabilidade e Administração de Lisboa, IPL, Portugal
[2]CIMA, Centro de Investigação em Matemática e Aplicações, Universidade de Évora, Portugal

E-mail address: *amjorge@iscal.ipl.pt, bushen@uevora.pt*

Abstract

The Traveling Salesman Problem(TSP) can be defined, in a simple way, as the determination of a set of paths (arcs) for a salesman go through beginning and ending at a starting point. The collection of municipal solid waste is a rather expensive task undertaken by municipalities, which involves choosing a particular circuit so that a vehicle, leaves a station, collects a set of containers and returns to it. This problem corresponds to the solution of the classical traveling salesman problem in which we consider a graph $G = (V, A)$, where V is the set of n nodes (containers) and A is the set of arcs (path) and matrices of Distances $D = d_{ij}$, time $T = t_{ij}$ and carbon monoxide $E = e_{ij}$ associated with each arc $(i, j) \in A$. There are several proposed formulations of integer linear programming for solving this problem. In this presentation, the authors present a multicriteria formulation chosing circuits for the collection vehicles, based on the distance traveled, the time spent performing the task and on environmental criteria such as emissions of gas greenhouse. The various circuits of collection point obtained by optimizing of the several criteria will be presented and analyzed.

Keywords: Traveling salesman problem, multicriteria optimization, collection of municipal solid waste.

Acknowledgements

References

[1]Kara, I., Bektas, T. (2006) Integer linear programming formulations of multiple salesman problems and its variations. European Journal of Operational Research 174: 1449-1458.

[2]Lotov, A., Bushenkov, V., Kamenev, G. (2004) Interactive Decision Maps – Approximation and Visualization of Pareto Frontier. New York, Springer US.

[3]Zsigraiova, Z., Semião, V., Beijico, F. (2013) Operation costs and pollutant emissions reduction by definition of new collection scheduling and optimization of MSW collection routes using GIS. The case study of Barreiro, Portugal. Waste Management 33: $793 - 806$.

How to test different block diagonal structures in several covariance matrices

Filipe J. Marques[1,2] and Carlos A. Coelho[1,2]

[1]Universidade NOVA de Lisboa
[2]Centro de Matemática e Aplicações (CMA), FCT, UNL

E-mail address: *fjm@fct.unl.pt, cmac@fct.unl.pt*

Abstract

The adequate calibration of different statistical models, such as growth curve or mixed models, requires an adequate choice of the covariance matrix structure. The complexity of some new models makes it important to be able to choose and to test elaborate patterns for covariance matrices. However, the testing procedures commonly used in this decision process is not easy due to the complicated structure of the exact distributions of the test statistics involved. Therefore, the required tests are often not performed or rather are performed using approximations for the distributions of the test statistics which, in most of the cases, are unable to guarantee the necessary accuracy of the results. In this work we will show how it is possible to create a procedure to test different block diagonal structures in several covariance matrices by splitting the null hypothesis into a set of conditionally independent hypotheses and how does this procedure makes it easy the development of very precise near-exact approximations. The numerical studies carried out demonstrate the adequacy and accuracy of these approximations.

Keywords: Covariance structure, hypothesis testing, likelihood ratio tests, near-exact distributions.

Acknowledgements

References

[1]Coelho, C. A. (2004) The Generalized Near-Integer Gamma Distribution: A Basis for 'Near-Exact' Approximations to the Distribution of Statistics which are the Product of an Odd Number of Independent Beta Random Variables. *Journal of Multivariate Analysis*, **89**: $191 - 218$.

[2]Coelho, C. A. and Marques, F. J. (2012) Near-exact distributions for the likelihood ratio test statistic to test equality of several variance-covariance matrices in elliptically contoured distributions. *Computational Statistics*, **27**: $627 - 659$.

[3]Marques, F. J. and Coelho, C. A. (2013) Obtaining the exact and near-exact distributions of the likelihood ratio statistic to test circular symmetry through the use of characteristic functions. *Computational Statistics*, **28**: $2091 - 2115$.

[4]Olkin, I. and Press, S. J. (1969) Testing and estimation for a circular stationary model. *The Annals of Mathematical Statistics*, **40**: $1358 - 1373$.

An asymptotic approach of the likelihood-based exact inference for singly imputed synthetic data under the multiple linear

Ricardo Moura[1,2]

[1]CINAV, Center of Naval Research, Naval Academy, Portuguese Navy, Base Naval de Lisboa, Alfeite, Portugal
[2]CMA, Faculty of Sciences and Technology, Nova University of Lisbon, Portugal

E-mail address: *pinto.moura@marinha.pt, rp.moura@fct.unl.pt*

Abstract

In this paper the author develops an asymptotic approach for the existing likelihood-based exact inference methods for singly imputed synthetic data generated via Posterior Predictive Sampling (PPS)and Plug-in Sampling, when the original data follows a multiple linear regression model. The use of existing inference methods from the literature implies that one has to obtain the distribution resorting to the use of empirical distributions which can be time-consuming. Thus, by proving that the statistic used in those methods is in fact asymptotically proportional to a chi-square distributed statistic, it is then possible, for large sample datasets, to only resort to a chi-square distribution where the degrees freedom is only dependent on the number of covariates which will not consume any time at all. Simulation studies compare the results obtained from the asymptotic methods with those for the exact test procedures under the PPS and Plug-in Sampling methods.

Keywords: Statistical disclosure control, ssymptotic distribution, multiple linear regression, synthetic data, multiple imputation.

Acknowledgements

References

[1]Klein, M. and Sinha, B. (2015) Inference for singly imputed synthetic data based on posterior predictive sampling under multivariate normal and multiple linear regression models. Sankhya B 77(2): 293-311.

[2]Moura, R., Klein, M., Coelho, C. and Sinha, B. (2016) Inference for multivariate regression model based on synthetic data generated using plug-in sampling. Technical Report - Centro de Matemática e Aplicações (CMA) 2/2016, (submitted for publication in Journal of the American Statistical Association).

[3]Moura, R., Klein, M., Coelho, C. and Sinha, B. (2017) Inference for Multivariate Regression Model based on synthetic data generated under Fixed-Posterior Predictive Sampling: comparison with Plug-in Sampling. Revstat 15 (2): 155-186.

Longitudinal modelling approach for MIBEL energy prices forecasting

Ana Borges[1], Eliana Costa e Silva[1], M. Filomena Teodoro[2,3], Marina A. P. Andrade[4] and Ricardo Covas[5,6]

[1]CIICESI, ESTG, Polytechnic of Porto, Portugal
[2]CINAV, Escola Naval - Marinha Portuguesa
[3]CEMAT, IST, Universidade de Lisboa, Portugal
[4]ISCTE-IUL/ISTAR-IUL, Instituto Universitário de Lisboa, Portugal
[5]CMA, Universidade Nova de Lisboa, Portugal
[6]EDP - Energias de Portugal

E-mail address: aib@estgf.ipp.pt, eos@estgf.ipp.pt

Abstract

The present work proposes to contribute to Electricity Price Forecasting (EPF) with a longitudinal statistical approach. By adjusting a mixed-effects longitudinal model (LM) to the hourly prices (HP) of Iberian Electricity Market (MIBEL) from March 10th 2014 to June 26th 2016, in a total of 20205 observations.

Keywords: Electricity Price Forecasting, Longitudinal Model, MIBEL.
———⊙———

EPF literature has mainly concerned on models that use information at daily level (aggregated data). However, we focus our interest on forecasting intra-day prices using hourly data (disaggregated data) in a multivariate approach rather than in the usually used univariate approach. Hence, we consider a LM that is able to incorporate the complex dependence structure of a multivariate price series. Longitudinal mixed-effects models are extremely popular in social, biological sciences and econometrics (panel data) and their popularity is explained by the flexibility they offer in modeling the within-group correlation often present in grouped data [2]. A day-ahead market consists in a system where agents submit their bids and offers for the delivery of electricity for each hour of the next day before a certain market closing time [3]. The present work studies the electricity prices of the MIBEL. The data is given as a strip of prices, all simultaneously observed once at a given time of each day. Longitudinal data models are regression models in which repeated observations of subjects are available. The daily market electricity prices can be given as a strip of prices (one for each hour of the day), all simultaneously observed once at a given time of each day. Therefore, the daily market prices can be interpreted as longitudinal, or panel data. Considering, in this particular analysis, the hours of the day as subjects, and the electricity prices for those hours for each day throughout the year, we are dealing with a balanced longitudinal data, i.e, repeated measurements for each subject (hour), taking at the same moment (day). This work is an extension of [1] analysis, where a vector autoregressive model with exogenous variables (VARX) was proposed. In order to compare the VARX approach with the longitudinal here proposed, we also analysed same the HP from March 10th 2014 to June 26th 2016, in a total of 20205 observations. Since LMs rely on the assumption, among others, of independent subjects (in this particular case, hours), we initially tested for correlation among the time series of electricity prices for the 24 hours of the day, by graphical interpretation of the correlogram (a correlation matrix) and by performing a factorial analysis. Similarly to [2], where different time period

slots were identified as important explanatory variables, three independent groups of hours were selected: i) the 1st until the 7th hour; ii) the 8th until the 18th hour; iii) the 19th until the 24 hour. We then adjusted a mixed-effects LM to the data, considering the three groups as subjects and, as reference time, the time since the first day of the year. Working on the logarithmic scale of price, as in [1]. We forecast electricity daily prices, for each group, estimating the best linear unbiased predictors, adding together the population predictions (based only on the fixed-effects estimates) and the estimated contributions of the random effects to the predictions at each group level. The results show that the LM approach considering a mixed-effects model, with the season of the year and the type of day (weekday versus weekend day) as fixed effects and the hour group as random effect, yield a model that explains the intra-day and intra-hour dynamics of the HP.

Acknowledgements

References

[1]Costa e Silva, E., Borges, A., Teodoro, M.F., Andrade, M.A.P. and Covas, R. (2017) Time Series Data Mining for Energy Prices Forecasting: An Application to Real Data. In Madureira, A.M. et al. (Eds.), Intelligent Systems Design and Applications, *Advances in Intelligent Systems and Computing* **557**.

[2]Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S. (2002) *Analysis of Longitudinal Data*, 2nd edition, Oxford, England, Oxford University Press.

[3]Pinheiro, J.C., and Bates, D.M. (2000) *Mixed-effects models in S and S-PLUS*, New York, Springer.

[4]Teodoro, M.F., Andrade, M.A.P., Costa e Silva, E., Borges, A. and Covas, R. Energy Prices Forcasting Using GLM, In Oliveira, T. et al. (Eds.), Recent studies on risk analysis and statistical modeling, *Cont. to Statistics* (in press).

[5]Weron, R. (2014) Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. Journal of Forecasting* **30** (4), 1030-1081.

Organized Session 3

Computational Mathematics and Statistics in Sciences and Engineering II

Organizer: M. Filomena Teodoro

Determinants for hedge foreign exchange risk in portuguese and spanish companies

Preciosa Carvalho[1], António Martins[2] and <u>Eliana Costa e Silva</u>[1]

[1]CIICESI, ESTG, Polytechnic of Porto, Portugal
[2]Universidade da Madeira, Portugal

E-mail address: Preciosa_Carvalho@hotmail.com, antonio.martins@staff.uma.pt
eos@estg.ipp.pt

Abstract

The exportation of services and products to the foreign market has become crucial for guarantying long-term survival of companies. This reality enhances the impact and relevance of exchange rate fluctuations on the value of the company and consequently on the way companies control this impact. The study of the exchange risk hedging practices of Portuguese and Spanish listed companies in 2013 and the determination of the factors that lead to companies making the decision to hedge Foreign Exchange (FX) risk, are the aims of this work. Data concerning Portuguese and Spain companies for the year 2013 were collect, for accessing the determinants for Portuguese and Spanish companies to decide hedge FX risk. The data consisted in 134 observations (i.e. companies), 37 Portuguese and 98 Spanish, for the year 2013. The variables size, leverage, liquidity and access external capital for each of the 134 companies are used as possible factors. Approximately 58% of the companies have taken FX risk in 2013, in opposing to the remaining 42% that have not. In order to infer the factors make the companies decide to hedge its transactions, univariate tests and multivariate logit regression are used. The variable binary hedge FX is consider as response variable while the independent variables are the quantitative variables size, leverage, liquidity and access external capital of the analysed listed companies. The tests suggest that there are statistical evidence that companies' size and leverage positively contribute to the companies' decision to hedge FX risk. These results are in accordance with e.g. [2], [3], [6] and [4]. In terms of companies' liquidity, the tests were inconclusive, while for access external capital has a significantly negative impact on hedge FX risk [2], [1]. Furthermore, size of the company presents a positive significant impact on the probability of the company hedge FX risk, while leverage presents also a positive impact although it is not statistically significant. Liquidity and access external capital yield a negative impact on FX risk, although not statistically significant.

Keywords: Logit Regression, multivariate statistics, foreign exchange risk.

Acknowledgements

References

[1]Cunha, F. (2009) *Hedging foreign currency and interest rate risks with derivatives: How much does it increase the firm's value?*, Lisboa, Instituto Superior de Ciências e da Empresa, ISCTE Business School: 1–41.

[2]Davies, D., Eckberg, C., & Marshall, A. (2006) The determinants of Norwegian exporters' foreign exchange risk management *European Journal of Finance* **12**(3): 217–240.

[3]Hutson, E., & Laing, E. (2014) Foreign exchange exposure and multinationality *Journal of Banking & Finance* **43**: 97–113.

[4]Li, H., Visaltanachoti, N., & Luo, R. H. (2014) Foreign Currency Derivatives and Firm Value: Evidence from New Zealand *Journal of Financial Risk Management* **3**: 92–112.

[5]Marshall, A., Kemmitt, M., & Pinto, H. (2013) The determinants of foreign exchange hedging in Alternative Investment Market firms, *The European Journal of Finance* **19**(2): 89–111.

[6]Pramborg, B. (2005) Foreign exchange risk management by Swedish and Korean non-financial firms: A comparative survey *Pacific-Basin Finance Journal* **13**(3): 343-366.

Influence on scrap percentage of aluminium extrusion parameters

M. Fátima Almeida[1,2,3], Aldina Coreia[1,2,3] and Nuno Carvalho[1,2,3]

[1]ESTG-School of Management and Technology, Polytechnic of Porto
[2]CIICESI-Center for Research and Innovation in Business Sciences and Information Systems
[3]CETRAD/UTAD-Centre for Transdisciplinary Development Studies, University of Trás-os-Montes and Alto Douro, Portugal

E-mail address: *mff@estg.ipp.pt, aic@estg.ipp.pt, 8...@estg.ipp.pt*

Abstract

In the industrial sector, optimization is one of the main factors of success. To use production and business data to improve profits, minimize coats, is an usual and vital procedure in industries. For that is was necessary to know the objective function variations and parameters which influences their behaviour. Aluminium extrusion industries are not exception, with big percentages of scrap, and the question of minimize it. In this study the influence of aluminium extrusion parameters in the variations on the scrap percentage produced, a crucial question in this industry, is approached. Due to all these advantages, this metal has great potential, capturing the interest of industry and researchers, [5,7]. In a previous work [2], the minimization of the de percentage of scrap in the extrusion process was addressed. In that work two steps are archived: the first was to monitoring the variables which are considered fundamental in literature to control the scrap produced, using the computer system; then we analysed the data using statistical techniques to identify critical variables in the process. The length and temperature of the billet, the extrusion speed, the extrusion ratio and the profiles temperature are the variables process to have into account, in this study, because they are the actual available parameters in the industry considered. After that, an having into account the bibliography research, other parameters are expected to affect the process, and the firm acquired a parameter monitoring software, wherewith other parameters were measured and can be considered in the study. In the present study we consider all the parameters monitoring in an industrial extrusion press, and we study their effect in the scrap percentage.

Keywords: Extrusion, Aluminium, Scrap Optimization.

———⊙———
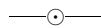
In literature we can identified some extrusion parameters influencing scrap percentage, some of them sumarize in [2]. For example, the authors in [4] identified that the level of scrap produced is related to the length of the billet, [3] refers the temperature of the billet, [8] indicates the extrusion speed, [1] indicate that Extrusion Ratio directly influences the output temperature of the profile and with regard to [6] the product quality is affected by the exit profile temperature.

In this work we intend to monitoring the aluminium extrusion parameters in order to minimize de metal waste. Having the indication, in our first work [2], about some variables influencing these process, and some additional information, now collected by the company with the monitoring software, wherewith other parameters were measured, we can consider

it in the study. The sample/production to be studied is about one profile (identified by the reference 9005), seven lots, between 27th March and 1st April of the current year, with two similar matrices including two exit orifices and the bars length is 20(mm). Among the measured values, we chose only those that are extrusion parameters, starting by analysing the correlations between them. There was thus a multicollinearity between them, so it was necessary to remove some of them from the study. We analyse the data, including 32 extrusion variables/parameters, using statistical techniques to identify critical variables in the process.

References

[1] Abdul-Jawwad, K. Abdul, and A. Bashir. A comprehensive model for predicting profile exit temperature of industrially extruded 6063 aluminum alloy. Materials and Manufacturing Processes, 26(2):193201, 2011.

[2] M.F. de Almeida, A. Correia, and N. Carvalho. Scrap optimization in an alu- minium extrusion industry. In CMMSE 2017: Proceedings of the 17th Interna- tional Conference on Mathematical Methods in Science and Engineering, vol- ume VI, pages 20822090. Elsevier, 2017.

[3] I. Flitta and T. Sheppard. Effect of pressure and temperature variations on fem prediction of deformation during extrusion. Materials science and technology, 21(3):339346, 2005.

[4] M. A. Hajeeh. Optimizing an aluminum extrusion process. Journal of mathe- matics and statistics, 9(2):7783, 2013.

[5] K. Masri and A. Warburton. Using optimization to improve the yield of an aluminium extrusion plant. Journal of the Operational Research Society, 49(11):11111116, 1998.

[6] P. K. Saha. Aluminum extrusion technology. Asm International, 2000.

[7] C. S. Zhang, G. Q. Zhao, H. Chen, and Y. J. Guan. Optimisation design of aluminium radiator extrusion die using response surface method. Materials Research Innovations, 15(sup1):s288s290, 2011.

[8] H. Zhu, M. J. Couper, and A. K. Dahle. Effect of process variables on the formation of streak defects on anodized aluminum extrusions: An overview. High Temperature Materials and Processes, 31(2):105111, 2012.

The impact of innovation Objectives on Product Innovation: CIS 2010

Aldina Correia[1,2,3], Anabela Marinho[1,2,3], Vítor Braga[1,2,3,4] and Alexandra Braga[1,2,4]

[1]ESTG - School of Management and Technology
[2]P.PORTO - Polytechnic of Porto
[3]CIICESI – Center for Research and Innovation in Business Sciences and Information Systems
[4]CETRAD/UTAD - Centre for Transdisciplinary Development Studies - University of Trás-os-Montes and Alto Douro, Portugal

E-mail addresses: *aic@estg.ipp.pt, 8150261@estg.ipp.pt*
vbraga@estg.ipp.pt, abraga@estg.ipp.pt

Abstract

In the business sector, innovation is one of the main factors of success. Innovation is considered to be an important decision for the company, since it involves the allocation of vital and costly resources, so it is essential that the objectives of innovation are clearly defined. In this paper, the subject of Product Innovation (PI) was approached in order to understand the impact of innovation objectives on companies that intend to focus on PI (goods and services). In order to achieve this objective, this analysis was based on the data set of the Community Innovation Survey (CIS 2010), covering the period from 2008 to 2010, which is made available through GPEARI / MCTES [1] – Office of Planning, Strategy, Evaluation and International Relations / Ministry of Science, Technology and Higher Education. According with the information in the CIS 2010 questionnaire *"An innovation is the introduction of a new or significantly improved product, process, organizational method, or marketing method by your enterprise. The innovation must be new to your enterprise, although it could have been originally developed by other enterprises"*. The minimum requirement for innovation is then that the product, process, marketing method or organizational method, is new (or substantially improved) for the company (OECD & Eurostat, 2005). This study focuses on PI (goods / services), which is defined as the process of bringing new products and / or new technologies to market (Lukas & Ferrell, 2000).

Keywords: CIS 2010, innovation, product innovation, innovation objectives.

———⊙———

The data collection was done mainly from an online electronic platform. The response rate was 97% of the companies surveyed. The data collection period of the questionnaire ran from July 12, 2011 to April 11, 2012, to Portuguese companies belonging to CAE-Rev.3. Following the guidelines and recommendations of Eurostat, INE built a sample of 9245 companies. At the end of the data collection period, 6,160 responses were considered valid (DGEEC & DSECTSI, 2015). For this analysis, on PI, two new variables were created. The variables "IMPD01" and "INOVA" are the result of the sum between the

---

[1] GPEARI/MCTES – Gabinete de Planeamento, Estratégia, Avaliação e Relações Internacionais/Ministério da ciência, Tecnologia e Ensino Superior

variables "INPDGV" and "INPDSV", and are distinguished by their scale. In addition to these variables, variables that intervened directly with the objectives of innovation (CIS, 2010) were chosen. In this study the tools of multivariate statistical analysis, in particular MANOVA and Discriminant Analysis are applied, with the help of SPSS Software. The database consists of 158 variables, of which 146 are qualitative and 12 are quantitative. With the discriminant analysis, it was possible to verify that the objectives of the firm influence the introduction of new goods and/or new services, that is, are determinant to the firm belong to one of the two groups: 0 = The firm did not introduce any good and service; 1 = The firm introduced goods and/or services. With regard to MANOVA, the three innovation objectives with the greatest contribution to discriminant analysis were considered. It has been found that the fact that the company introduces new or significantly improved goods and / or services has a significant effect on extending the range of products, entering new markets or increasing market share and improving product quality. Thus, the results indicate that the objectives of innovation influence PI. In addition, depending on the goals of the company is possible to distinguish if the company is more likely to be innovative or not. It has also been found that PI has a significant effect on Innovation Objectives.

Acknowledgements

References

[1]CIS (2010) *Inquérito Comunitário à Inovação (Community Innovation Survey)*, Lisboa, Gabinete de Planeamento, Estratégia, Avaliação e Relações internacionais, Ministério da Ciência, Tecnologia e Ensino Superior.

[2]DGEEC & DSECTSI (2015) *Sumários Estatísticos: CIS 2010 – Inquérito Comunitário à Inovação [versão corrigida]*, Lisboa, Direção-Geral de Estatísticas da Educação e Ciência.

[3]Lukas, B. A. & Ferrell, O. C. (2000) The effect of market orientation on product innovation. *Journal of the Academy of Marketing Science* **28** (2): 239-247.

[4]OECD & Eurostat (2005) *Oslo Manual: Guidelines for Collecting and Interpreting Innovation Data* (3 ed), Paris, OECD Publishing.

Linear time series models for electricity pricing: the portuguese and spanish market

Maria de Fátima Pilar[1], Mariana Marques[1], Ana Teixeira[1] and Eliana Costa e Silva[1]

[1]*CIICESI, ESTG, Polytechnic of Porto, Portugal*

E-mail address: *8120109@estg.ipp.pt, 8110119@estg.ipp.pt 8130097@estg.ipp.pt, eos@estg.ipp.pt*

Abstract

In the last decade, the electricity industry has experienced significant changes towards deregulation and competition with the aim of improving economic efficiency [5]. The complexity of electricity markets calls for rich and flexible modeling techniques that help to understand market dynamics and to derive advice for the design of appropriate regulatory frameworks [6]. The Spanish and Portuguese electricity sector began to be organized in terms of an integrated system and an independent system in 1995, and the Iberian electricity market started on 1 January 1998. Both producers and consumers use day-ahead price forecasts to derive their respective bidding strategies to the electricity market. Therefore, accurate price estimates are crucial for producers to maximize their profits and for consumers to maximize their utilities [1]. The present empirical study analyzes temporal data concerning the electricity market from the OMIP database - The Iberian Energy Derivatives Exchange, which contains a set of electricity prices observations made sequentially over time. OMIP is the stock exchange of Iberian and non-Iberian products (including MIBEL), which manages the market jointly with OMIClear, a company incorporated and totally owned by OMIP, which assures the functions of Clearing and Counterparty Clearing House Central to market operations [3]. From the data available at the OMIP database the present work focus on hourly electricity prices and on "SPEL Base" index (i.e. the arithmetic average of the marginal prices hours of the day), for the period from January 1, 2016 to May 31, 2017. The hourly prices were aggregated in day period and night period, defined as the average of the electricity prices from 8am to 8pm - day period - and the average of the electricity prices of the remaining hours (See Figure 1). The day of the week (Sunday, Monday, ...) and the month (January, February, ...) were also variables considered. The ARIMA methodology was followed for modeling of the "SPEL Base" index, while the intra-day and intra-period dynamics was modeled using a vector autogressive (VAR) model approach. Finally, the exogenous variables day of the week and month were included in an attempt to improve the model fitting. For the consider period (January 2016 to May 2017) the best model was the VAR(9), while in previous work [2] for the period from March 2014
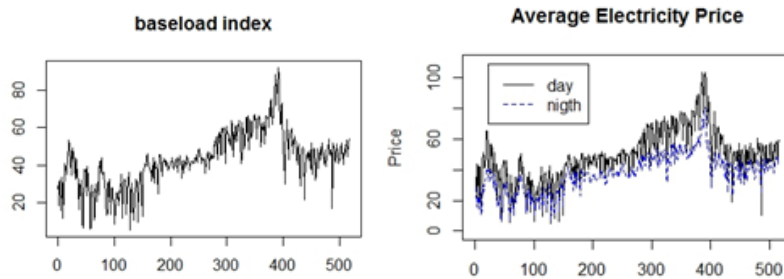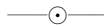
Figure 1: Eletricity prices from January 1, 2016 to May 31, 2017.

to May 2016, a VARX(7,0) was encountered. There is a clear need for further analysis.

Keywords: electricity pricing, multivariate time series, OMIP database, vector autoregressive models.

———⊙———

References

[1] Conejo, A. J., Contreras, J., Espínola, R., & Plazas, M. A. (2005). Forecasting electricity prices for a day-ahead pool-based electric energy market. *International Journal of Forecasting* **21**(3): 435–462.

[2] Costa e Silva, E., Borges, A., Teodoro, M.F., Andrade, M.A.P. & Covas, R. (2017) Time Series Data Mining for Energy Prices Forecasting: An Application to Real Data. In Madureira, A.M. et al. (Eds.), Intelligent Systems Design and Applications, *Advances in Intelligent Systems and Computing* **57**: 649-658.

[3] OMIP. (2017). Retrieved from OMIP - The Iberian Energy Derivatives Exchange: http://www.omip.pt (acessed at 30th June 2017)

[4] Tsay, R. S. (2014). *Multivariate Time Series Analysis: with R and financial applications*, John Wiley & Sons.

[5] Ventosa, M., Ballo, A., Ramos, A., & Rivier, M. (2005). Electricity market modeling trends. *Energy policy* bf33(7): 897–913.

[6] Weidlich, A., & Veit, D. (2008). A critical survey of agent-based wholesale electricity market models. *Energy Economics* **30**(4): 1728-1759.

[7] Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art. *International Journal of Forecasting* 30, 10301081.

Organized Session 4

Inference in Linear Models

Organizer: Dário Ferreira

Mixed models with random sample sizes

Célia Nunes[1], Sandra S. Ferreira[1], Dário Ferreira[1], Elsa Morreira[2] and
João Tiago Mexia[2]

[1]Department of Mathematics and Center of the Mathematics and Applications,
University of Beira Interior, Covilhã, Portugal
[2]Center of Mathematics and Applications, Faculty of science and Technology, New
University of Lisbon, Portugal

E-mail address: *celian@ubi.pt, sandraf@ubi.pt, dario@ubi.pt*
efnm@fct.unl.pt, jtm@fct.unl.pt

Abstract

When applying analysis of variance the sample sizes may not be previously known, so
it is more correct consider them as realizations of random variables. A motivation example
is the collecting of observations during a fixed time span in a study comparing, for example,
several pathologies of patients arriving at a hospital. The aim of this work is to extend
the theory of analysis of variance to those situations considering mixed effects models. We
will assume that the occurrences of observations correspond to counting processes so the
sample dimensions are considered as Poisson distributed. The applicability of the proposed
approach is illustrated through an application on real medical data from patients affected
by cancer.

Keywords: F-tests, mixed models, random sample sizes, cancer registries.

Acknowledgements

References

[1] Mexia, J.T., Nunes, C., Ferreira, D., Ferreira, S.S. and Moreira, E. (2011) Orthogonal
    fixed effects ANOVA with random sample sizes. Proceedings of the 5th International
    Conference on Applied Mathematics, Simulation, Modelling (ASM'11): 84–90.
[2] Moreira, E., Mexia, J.T., Fonseca, M. and Zmyślony, R. (2009) L models and multiple
    regressions designs. *Statistical Papers* **50**(4): 869–885.
[3] Moreira, E.E., Mexia, J.T. and Minder, C.E. (2013) F tests with random sample size.
    Theory and applications *Statistics & Probability Letters*. **83**(6): 1520–1526.
[4] Nunes, C., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2012) F-tests with a rare pathol-
    ogy. *Journal of Applied Statistics* **39**(3): 551–561.
[5] Nunes, C., Capistrano, G., Ferreira, D. and Ferreira, S.S. (2013) ANOVA with Random
    Sample Sizes: An Application to a Brazilian Database on Cancer Registries. 11th
    International Conference on Numerical Analysis and Applied Mathematics. AIP Conf.
    Proc. **1558**: 825–828.
[6] Nunes, C., Ferreira, D., Ferreira, S.S. and Mexia, J.T. (2014) Fixed effects ANOVA:
    an extension to samples with random size. *Journal of Statistical Computation and
    Simulation*. **84**(11): 2316–2328.

Global Confidence Regions for Mixed Models assuming Orthogonal Block Structure and Normality

<u>Sandra S. Ferreira</u> [1], Dário Ferreira[1], Célia Nunes[1] and João Tiago Mexia[2]

[1]Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal
[2]Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Portugal

E-mail address: *sandraf@ubi,pt; dario@ubi.pt; celian@ubi.pt; jtm@fct.unl.pt*

Abstract

The aim of this work is to show how to derive global regions for future observations, assuming the normality of them. To illustrate the theory we present an application which shows that the performances of the confidence regions obtained by the proposed approach are good.

Keywords: Orthogonal Block Structure, Uniformly Minimum Variance Unbiased Estimator, Variance components.

Acknowledgements

References

[1]Caliński, T., Kageyama, S. (2000). *Block Designs: A Randomization Approach, Volume I: Analysis, Lecture Notes in Statistics*, 150 Springer, New York.

[2]Caliński, T., Kageyama, S. (2003). *Block Designs: A Randomization Approach, Volume II: Design, Lecture Notes in Statistics*, 170 Springer, New York.

[3]Ferreira, S. S., Nunes, C., Ferreira, D. , Moreira, E. and Mexia, J.T. (2015). Estimation and Orthogonal Block Structure. *Hacettepe University Bulletin of Natural Sciences and Engineering*, Series B: Mathematics and Statistics, 45(58).

[4]Fonseca, M., Mexia, J. T. and Zmyślony, R. (2003). Estimating and testing of variance components: an application to a grapevine experiment. *Listy Biometryczne-Biometrical Letters*, 40 (1), 1-7.

[5]Mexia, J.T., Vaquinhas, R., Fonseca, M., Zmyślony, R. (2010). COBS: Segregation, Matching, Crossing and Nesting. Latest Trends on Applied Mathematics, Simulation, Modeling, 4th International Conference on Applied Mathematics, Simulation, Modelling (ASM'10), 249–255.

[6]Nelder, J.A. (1965a). The analysis of randomized experiments with orthogonal block structure. I. Block structure and the null analysis of variance. Proceedings of the Royal Society (London), Series A 273, 147–162.

[7]Nelder, J.A., (1965b). The analysis of randomized experiments with orthogonal block structure. II. Treatment structure and the general analysis of variance. Proceedings of the Royal Society (London), Series A 273, 163–178.

Parametric models with random sample sizes

Anacleto Mário[1], Célia Nunes[2], Dário Ferreira[2], Sandra S. Ferreira[2]
and João Tiago Mexia[3]

[1]PhD Student, University of Beira Interior, Portugal
[2]Department of Mathematics/Center of the Mathematics and Applications, University of
Beira Interior, Covilhã, Portugal
[3]Center of Mathematics and Applications, Faculty of science and Technology, New
University of Lisbon, Portugal

E-mail address: *anacleto.mario@ubi.pt, celian@ubi.pt, dario@ubi.pt
sandraf@ubi.pt, jtm@fct.unl.pt*

Abstract

In many relevant situations, such as in economic research, sample sizes may not be
previously known. The aim of this paper is to extend analysis of variance (ANOVA) to
those situations. Sample sizes are assumed as realizations of independent random variable
with Poisson, Binomial and Negative Binomial distributions. The applicability of the pro-
posed approach is illustrated considering a database on unemployment in four European
countries. The interest of this approach lies in avoiding false rejections obtained when
using the classical ANOVA.

Keywords: Random sample sizes, F- Tests, Poisson distribution, Binomial distribution,
Negative Binomial distribution, database on unemployment.

Acknowledgements

References

[1]Mexia, J. T., Nunes, C., Ferreira, D., Ferreira, S. S. and Moreira, E.(2011) Orthogonal
   fixed effects Anova with random sample sizes. *Proceedings of the 5th International
   Conference on Applied Mathematics, Simulation, Modelling (ASM'11)*, Corfu, Greece,
   84-90.
[2]Nunes, C., Ferreira, D., Ferreira, S. S., Mexia, J.T. (2010) *F* Tests with Random Sample
   Sizes. *8th International Conference on Numerical Analysis and Applied Mathematics.
   AIP Conference Proceedings*, 1281(II), 1241-1244.
[3]Nunes, C., Ferreira, D., Ferreira, S. S., Mexia, J.T. (2012) *F* Tests with a rare pathology.
   Journal of Applied Statistics, 39(3), 551-561.
[4]Nunes, C., Ferreira, D., Ferreira, S. S., Mexia, J.T. (2014) Fixed effects ANOVA: an
   extension to samples with random sizes. *Journal of Statistical computation and Sim-
   ulation*, 84(11), 2316-2328.

Organized Session 5

Computational Data Analysis and Statistical Inference

Organizer: Luís M. Grilo (Portugal)

Analysis of the influence of the period of activity and the size of the portfolios in their risk and diversification levels

João Romacho[1,2] and Cristina Dias[3,4]

[1,3]Instituto Politécnico de Portalegre, Portugal
[2]C3i-Coordenação Interdisciplinar para a Investigação e a Inovação
[4]CMA-Centro de Matemática e Aplicações da Universidade Nova de Lisboa, Portugal

E-mail address: *jromacho@estgp.pt, cpsilvadias@gmail.com*

Abstract

The risk and the portfolio diversification levels of equity mutual funds from different countries are analysed. It is studied the possible relationship between the age and the size of portfolios with their risk levels and concentration/diversification of portfolios. Belgium funds seem to be those that exhibit more concentrated portfolios, whereas Italian funds exhibit the most diversified portfolios. It is also identified a negative relation between the level of portfolio concentration and the total risk claimed by managers, that is to say that funds with higher portfolio concentration will be those that present a smaller risk. On the one hand, it is identified some evidence that funds with the lowest period of activity tend to take more risk and to exhibit less concentrated portfolios. On the other hand, there is no evidence of a relationship between the size of the portfolios with the levels of risk and the portfolio diversification.

Keywords: Mutual funds, Period of activity, Size, Risk, Portfolio diversification.

References

[1]Brands, S., Brown, S. and Gallagher, D. (2005): "Portfolio concentration and investment manager performance". International Review of Finance, 5 (3-4): 149-174.
[2]Chan, H., Faff, R., Gallagher, D. and Looi, A. (2009): "Fund size, transaction costs and performance: size matters?". Australian Journal of Management, 34 (1): 73-96.
[3]Chen, J., Hong, H., Huang, M. and Kubik, J. (2004): "Does fund size erode mutual fund performance? The role of liquidity and organization". American Economic Review, 94 (5): 1276-1302.
[4]Chou, P. and Lee, C. (2012): "Is concentration a good idea? Evidence from active fund management". Asia-Pacific Financial Market, 19 (1): 23-41.
[5]Cremers, K. and Petajisto, A. (2009): "How active is your fund manager? A new measure that predicts performance". Review of Financial Studies, 22 (9): 3329-3365.
[6]Cresson, J. (2002): "R2: a market-based measure of portfolio and mutual fund diversification". Quarterly Journal of Business and Economics, 41 (3-4): 115-143.
[7]Flavin, T. (2004): "The effect of the Euro on country versus industry portfolio diversification". Journal of International Money and Finance, 23 (7-8): 1137-1158.
[8]Huij, J. and Derwall, J. (2011): "Global equity fund performance, portfolio concentration, and the fundamental law of active management". Journal of Banking and Finance, 35 (1): 155-165.

[9]Kacperczyk, M., Sialm, C. and Zheng, L. (2005): "On the industry concentration of actively managed equity mutual funds". Journal of Finance, 60 (4): 1983-2011.

[10]Kacperczyk, M., Sialm, C. and Zheng, L. (2007): "Industry concentration and mutual fund performance". Journal of Investment Management, 5 (1): 50-64.

[11]Rudin, A. and Morgan, J. (2006): "A portfolio diversification index". Journal of Portfolio Management, 32 (2): 81-89.

[12]Sapp, T. and Yan, X. (2008): "Security concentration and active fund management: do focused funds offer superior performance?". Financial Review, 43 (1): 27-49.

[13]Shawky, H. and Smith, D. (2005): "Optimal number of stock holdings in mutual fund portfolios based on market performance". Financial Review, 40 (4): 481-495.

[14]Statman, M. (2004): "The diversification puzzle". Financial Analysts Journal, 60 (4): 44-53.

Symmetric Stochastic Matrices width a Dominant Eigenvalue

Cristina Dias[1,2], Carla Santos[3,4], João Romacho[5,6], Maria Varadinov[7,8], João Miranda[9,10] and João Tiago Mexia[11,12]

[1,5,7,9]Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre
[2,4,12]CMA - Centro de Matemática e Aplicações, FCT, UNL, Portugal
[3]Departamento de Matemática e Ciências Físicas do Instituto Politécnico de Beja
[6,8]C3i – Coordenação Interdisciplinar para a Investigação e a Inovação
[10]CERENA - Centro de Recursos Naturais e Ambiente Instituto Superior Técnico[11]Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

E-mail address: : *cpsilvadias@gmail.com, carla.santos@ipbeja.pt, jromacho@estgp.pt dinov@estgp.pt, jlmiranda@estgp.pt, jtm@fct.unl.pt*

Abstract

In this paper we will consider structured families of Symmetric stochastic matrices, which are associated to the treatments in a base design. The action of the factors in the base design on the structured vectors of the models in the family is studied. In Structured families with orthogonal base designs, the designs are associated to orthogonal partitions of R^m. Thus, as we shall see, we can apply the ANOVA and associated techniques in the analysis. Assuming that the family models have degree $r > 0$, it is possible to study the action of the factors of the base design on the first r structure vectors. Often the first eigenvalue is strongly dominant for most of the models, with $r = 1$ in such situations we have to analyse the action of the factors on the first structure vector.

Keywords: ANOVA, Structured families, Dominant eigenvalue.

Acknowledgements

References

[1]Areia, A. (2009) Séries emparelhadas de estudo. Ph.D. Thesis Évora University.

[2]Ito, P. K. (1980) Robustness of Anova and Macanova Test Procedures, P. Krishnaiah, R. (ed), Handbook of Statistics, Amsterdam: North Holland 1 : 199-236.

[3]Mexia J. T. (1988) Standardized Orthogonal Matrices and Decomposition of the Sum of Squares for Treatments. Trabalhos de Investigação, nº2, Departamento de Matemática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

[4]Dias, C. (2013) Modelos e Famílias de Modelos para Matrizes Estocásticas Simétricas. Ph.D. Thesis, Évora University.

[5]Oliveira, M. M. and Mexia, J. T. (1999) F Tests for Hypothesis on theStructure Vectors of Series. Discussiones Mathematicae 19 (2) : 345-353.

[6]Scheffé, H. (1959) The Analysis of Variance, New York, John Wiley & Sons

On the enrichment of random factors models

<u>Carla Santos</u>[1,2], Célia Nunes[3], Cristina Dias[4,2], Maria José Varadinov[4] and João Tiago Mexia[2,5]

[1]Department of Mathematics and Physical Sciences, Polytechnical Institute of Beja, Portugal
[2]CMA-Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Portugal2
[3]Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Portugal
[4]College of Technology and Management, Polytechnical Institute of Portalegre, Portugal
[5]Faculty of Science and Technology, New University of Lisbon, Portugal

E-mail address: *carla.santos@ipbeja.pt, celian@ubi.pt*
cpsilvadias@gmail.com, dinov@estgp.pt

Abstract

To test the main effects and interactions in random effects models we consider an orthogonal partition in subspaces associated to the subset of factors. Availing ourselves of the much larger dimension of one of the subspaces in the orthogonal partition, we enrich the model by imbedding linear regressions.

Keywords: Factor crossing, linear regressions, orthogonal partition.

Acknowledgements

References

[1]Carvalho, F., Mexia, J.T., Santos, C., Nunes, C. (2015) Inference for types and structured families of commutative orthogonal block structures. Metrika 78: 337–372.

[2]M. Fonseca, M., Mexia, J.T., Zmyslony, R. (2006) Binary Operations on Jordan algebras and orthogonal normal models, Linear Algebra Appl. 117(1) ,75-86 .

[3]M. Fonseca, M., Mexia, J.T., Zmyslony, R. (2008) Inference in normal models with commutative orthogonal block structure, Acta et Commentationes Universitatis Tartuensis de Mathematica 12, 3–16.

[4]Khuri, A. I.,Mathew, T., Sinha, B. K. (1998) Statistical Tests for Mixed Linear Models, New York, Wiley.

[5]Mexia, J.T., Vaquinhas, R., Fonseca, M. and Zmyslony, R. (2010) COBS: segregation, matching, crossing and nesting, Latest Trends and Applied Mathematics, Simulation, Modelling, 4-th International Conference on Applied Mathematics, Simulation, Modelling (ASM'10) 249– 255 .

[6]Oliveira, M. M. (2002) Modelação de Séries Emparelhadas de Estudos com Estrutura Comum". PhD thesis. Universidade de Évora.

Reverse Logistics within Pharmaceutical Supply Chains

Maria Varadinov[1,2], João Miranda[3,4] and Cristina Dias[5,6]

[1,3,5]Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre, Portugal
[2]C3i-Coordenação Interdisciplinar para a Investigação e a Inovação
[4]CERENA-Centro de Recursos Naturais e Ambiente Instituto Superior Técnico
[6]CMA-Centro de Matemática e Aplicações da Universidade Nova de Lisboa, Portugal

E-mail address: *dinov@estgp.pt, jlmiranda@estgp.pt, cpsilvadias@gmail.com*

Abstract

The main goal of this proposed research is the reverse logistics (RL) process, which has been the subject of several studies over the years, but RL for the pharmaceutical supply chains (SC) in Portugal still is in a very early stage. RL is addressing the perspective of companies and other organizations that become active on end-of-use products, by accepting, discarding or recovering them (Dekker et al., 2004), due to economic, legal and social reasons. A survey using descriptive and empiric research methods is outlined, to consult on the implementation of RL practices in Portuguese pharmaceutical companies, and on the current mechanisms influencing RL activities, such as economic, social and legislative factors. Case studies and online questionnaires allow a comparative analysis of the Portuguese results with the existing resultsand studies conducted in other countries. The results prospects will indicate the relevance given by pharmaceutical companies to economic, social and legal reasons for the implementation of RL processes. Namely, the existence and the extension of environmental concerns, the economic motives, the customer needs, the satisfaction practices on the service level, or other specific items within the pharmaceutical SC.

Keywords: Reverse Logistics; Pharmaceutical Supply Chains; End-Of-Use Products; Methodologies.

Acknowledgements

References

[1]Varadinov, M.J. (2016) "A Gestão da Logística Inversa nas Empresas Portuguesas", Ph.D. Thesis, University of Extremadura, Badajoz, 2016.

[2]Miranda, J.L., Varadinov, M.J., and Rubio, S. (2014) Green Logistics, Reverse Logistics and Agro-Industries: Overview of scientific articles and international programs". Ariel, C. and Castro, W. (Ed.) Green Supply Chains: Applications in Agroindustries, Universidad Nacional de Colombia, Manizales: 97-106

[3]Dekker, R.; Fleischman, M.; Inderfurth, K. e Van Wassenhove, L. N. (2004). Quantitative Models for Closed-Loop Supply Chains (3-27). Reverse Logistics. Springer-Verlag. Germany.

Pharmaceutical Supply Chains: Training Early-Career Investigators in "Medicines Shortages"

João Miranda[1,2], Cristina Dias[3,4] and Maria Varadinov[5,6]

[1,3,5]Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre, Portugal
[6]C3i-Coordenação Interdisciplinar para a Investigação e a Inovação
[2]CERENA-Centro de Recursos Naturais e Ambiente Instituto Superior Técnico
[4]CMA-Centro de Matemática e Aplicações da Universidade Nova de Lisboa, Portugal

E-mail address: *dinov@estgp.pt, jlmiranda@estgp.pt, cpsilvadias@gmail.com*

Abstract

A Training School (TS) addressing Pharmaceutical Supply Chains (PharmSC) took place at Escola Superior de Tecnologia e Gestão (ESTG), Portalegre, Portugal, 03-07 of July-2017, on behalf of the COST Action European Medicines Shortages Research Network - addressing supply problems to patients, "Medicines Shortages" (CA15105), and organized by Instituto Politécnico de Portalegre (IPP), Centro de Recursos Naturais e Ambiente (CERENA/IST), and IBM Portugal. In a way to better introduce the COST Action "Medicines Shortages" and its main attributes (goals, methodology, work-plan, and tools), the TS program considered a Seminar open to the general public in the very first day. The Seminar participants also gained a complete overview of the TS, since the main topics for the technical sessions in the other days are also presented. The Seminar panels were addressing: i) "Medicines Shortages" and the impact on outcomes; ii) The needs and barriers within the Supply Chain (SC) actors; and iii) Pharmaceutical SC studies, methodologies and tools. The main topics for the technical sessions in the other days are also presented in the opening Seminar, and more details are available at www.tinyurl.com/PharmSC . These technical-sessions are "hands-on" training sessions with computational tools and case studies on Multi Criteria Decision Making (MCDM) for Pharmaceutical SC, on qualitative methodologies, so as on Suppliers Selection. Specific computational issues related with Decision Support Systems (DSS) are also treated, including IBM advanced tools for SC optimization (IBM ILOG/CPLEX) and IBM Watson/Bluemix for data science experiments. More advanced topics are presented in this second edition of the TS, namely, by discussing clinico-pharmacological needs and measuring the impact of disruptions and shortages on outcomes, so as outlining the prospects and further developments in the "Medicines Shortages" research lines.

Keywords: COST Action; Medicines Shortages; Pharmaceutical Supply Chains; Case Studies; Methodologies.

Acknowledgements

References

[1] Póvoa, A., Corominas, A., and Miranda, J.L. (2016) Optimization and DSS for SC, Springer.

[2] Miranda, J.L., Varadinov, M.J., and Rubio, S. (2014) Green Logistics, Reverse Logistics and Agro-Industries: Overview of scientific articles and international programs". Ariel, C. and Castro, W. (Ed.) Green Supply Chains: Applications in Agroindustries, Universidad Nacional de Colombia, Manizales: 97-106.

[3] Miranda, J.L. (2013) Odss.4SC: Optimization and DSS for Supply Chains. In: Mobility, Projects and Cooperation: The LLP best practices (2007-2012). ANPALV, Lisboa: 22-23.

Organized Session 6

Statistical Modelling and Data Analysis

Organizer: Amílcar Oliveira and Teresa Oliveira (Portugal)

Canadian boreal forest fire disturbance and climate change

Carla Francisco[1,2], Teresa A. Oliveira[3,4] and Steve Cumming[1,2]

[1]Laval University
[2]Faculty of Forestry, Geography and Geomatics, Québec, Canada
[3]Universidade Aberta, Palácio Ceia, Rua da Escola Politécnica, Lisboa, Portugal
[4]Center of Statistics and Applications, University of Lisbon, Portugal

E-mail address: *Carla.Francisco.1@ulaval.ca, Teresa.Oliveira@uab.pt
stevec.boreal@gmail.com*

Abstract

Wildfire is an important disturbance (e.g. Bergeron et al. 2004, Lefort et al. 2004) and it is likely that it will remain so, since climate change is predicted to strengthen fire activity throughout Canada (Flannigan et al. 2009, Wotton et al. 2010). We can sustain that the human being is today the great transforming agent of the earth's surface, however there has been a concern in making a purely physical analysis of the human influence impact on climate change. In Europe, glaciology is a area of greater concern, however in Russia and Canada, areas such geocriology gain more importance once it's dedicated to the study of permafrost. Since European settlement, Canada began to some extend promoting it's wildland fire management and with this statistical model aims to facilitate more reliable predictions of future fire regime characteristics under projected climate change, contributing to the development of a unique spatial − temporal database that would be of great value both to researchers and to fire management agencies. Previous studies have shown that expected daily fire arrivals can be modelled reasonably well using a Poisson distribution if we considered relatively small and homogeneous areas (Mandallaz and Ye 1997, Wotton et al. 2003, 2010). We can consider a point-process with conditional intensity function as a natural modeling structure, given the stochastic nature of the ignitions. The first stochastic model for predicting the occurrence of fires appears to have been developed by Bruce (1960), who utilized a negative binomial model that related counts to a fire danger rating index. Subsequently, Cunningham and Martell (1973) developed a Poisson model for counts of fires whose nonspatial conditional intensity function depended on fuel moisture, as measured by the Canadian Fine Fuel Moisture Code (FFMC) (Van Wagner, 1987). Given large differences in frequency and size, fires in Canada can be divided in two broad categories: lightning and human-caused (Stocks et al. 2003). In my fire model I run two simulations one for each category, in this way I have considered one for fire frequency and another one for fire size, to do so I have considered a total of four fire statistics simulations using data from the Canadian National Fire DataBase (CNFDB). Acording to Cumming (2001a) I expect that the distribution of logarithms of the fire sizes follow a truncated negative exponential. In my model I have considered the average number of fires between 1969 and 2000, by CAUSE type (Human or Lightning), the results seems suggest that CAUSE type is a good candidate for predicting the number of fires, because the mean value of the outcome appears to vary by CAUSE. The variances within each level of CAUSE are lower than the means within each level. These are the conditional means and variances. These differences suggest that over-dispersion is present and that a Negative Binomial model is more appropiate to estimate and update these inter-annual variation.

Keywords: Boreal forest, climate change, fire disturbance, negative Binomial, wildland fire management.

Acknowledgements

I want to thank my director Steve Cumming and my co-director Teresa Olivera for the time they spent to improve earlier versions of this paper.

References

[1]Bergeron, Y., Flannigan, M., Gauthier, S., Leduc, A., Lefort, P., (2004). Past, current and future fire frequency in the Canadian boreal forest: implications for sustainable forest management. Ambio 33, 356–360.

[2]Bruce, D. (1960). How many fires? Fire Control Notes 2445–50.

[3]Flannigan, M. D., M. A. Krawchuk, W. J. de Groot, B. M.Wotton, and L. M. Gowman, (2009). Implications of changing climate for global wildland fire. International Journal of Wildland Fire 18:483–507.

[4]Mandallaz D, Ye R, (1997). Prediction of forest fires with Poisson models. Canadian Journal of Forest Research 27: 1685-1694. - doi: 10.1139/x97-103.

[5]Stocks, B. J. et al., (2003). Large forest fires in Canada, 1959–1997. Journal of Geophysical Research 108, 8149. doi:10.1029/2001JD000484.

[6]Cumming, S. G., (2001a). A parametric model of the fire-size distribution. Canadian Journal of Forest Research 31, 1297–1303.

[7]Cunningham, A. A. and Martell, D. L., (1973). A Stochastic Mode for the Occurrence of Man-Caused Forest Fires, Can. J. Forest Res. 3, 282–287.

[8]Van Wagner, (1987). Forest fire research - hindsight and foresight, C.E. Pages 115-120 in Symposium on Wildland Fire, April 27-30, 1987, South Lake Tahoe, California, USA. USDA Forest Service, Washington, DC, USA.

Comparison of the Count Regression Models in Evaluation of the Effects of Periodontal Risk Factors on the Number of Teeth

J.L. Pereira[1], C. Morais[1], M. Mendes[1] and Teresa A. Oliveira[2,3]

[1]Department of Periodontology, School of Dental Medicine of Porto University, Porto, Portugal
[2]Universidade Aberta, Palácio Ceia, Rua da Escola Politécnica, Lisboa, Portugal
[3]Centro de Estatística e Aplicações (CEAUL)

E-mail address: *jpereira@fmd.up.pt*

Abstract

The number of remaining teeth in the mouth is of great importance in oral rehabilitation and depends on several factors. A number of conditions as gender, time of exposure to pathogenic factors (age), smoking habits and diabetes mellitus, are major factors that favor periodontal disease of which outcome is tooth loss [1]. The diagnosis of periodontal condition should be an explanatory factor of tooth loss, even considering that other oral and medical conditions are also associated with tooth loss. The aim of this study was to assess the performance of Poisson, quasi-Poisson and negative binomial models in modeling the number of teeth present (TP) or missing (MT) and find the best fit [2]. We conducted a retrospective study with data from the clinical records and orthopantomographies of patients of School of Dental Medicine of Porto, after the required authorisation of the ethic committee. The periodontal disesease was diagnosed and classified in mild, moderate, and severe according with the percentage of bone level around teeth roots observed in the orthopantomography. The data concernig gender, age, smoking, diabetic status and overall periodontal diagnosis were recorded in a .csv type spreadsheet and pocessed with software R [3]. Poisson, quasi-Poisson and negative binomial models were fitted to the data and compared by dispersion, deviance and Akaike Information Criteria (AIC) values. The results suggest that the best fit was the negative binomial modeling the number of MT (36 - TP) presenting dispersion, AIC and residual deviance of 1.02, 588.4 and 94.91 respectively. The results of this study indicate that besides to explore different models fitting the change of approach in quantifying the endpoint of oral diseases through a change of variable revealed usefull in obtain better fits.

Keywords: Count data models, Models comparison, Periodontitis, Tooth loss.

References

[1]Genco R.J., Borgnakke W. S. (2013) Risk factors for periodontal disease. Periodontology 2000 62 (1) : 59-94.
[2]Ver Hoefand J.M., Boveng P.L. (2007) Quasi-Poisson vs. Negative Binomial Regression: How Should We Model Overdispersed Count Data?. Ecology 88 (11): 2766-2772.
[3]R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Skewness and Kurtosis of the Extended Skew-Normal Distribution

J. António Seijas-Macias[1,2], Amílcar Oliveira[3,4] and Teresa A. Oliveira[3,4]

[1]Universidad da Coruña (Spain)
[2]Universidad Nacional de Educación a Distancia (UNED) - (Spain)
[3]Universidade Aberta (Portugal)
[4]Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal

E-mail address: *antonio.smacias@udc.gal*

Abstract

In the last years, a very interesting topic has arisen and became the research focus not only for many mathematicians and statisticians, as well as for all those interested in modelling issues: The Skew normal distributions' family that represents a generalization of normal distribution. The skew-normal probability distribution [1] was introduced by Azzalini in 1985 which including the normal distribution as a special case. The skew-normal is a generalization of the normal distribution where the more important characteristic is the presence of different level of skewness. Later on, the extended skew-normal distribution is defined as a generalization of skew-normal distribution [2]. These distributions are potentially useful for the data that presenting high values of skewness and kurtosis. Applications of this type of distributions are very common in model of economic data, especially when asymmetric models are underlying the data. Definition of this type of distribution is based in four parameters: location (ξ), scale (ω), shape (α) and truncation (τ). In [3] some properties of the density function are presented and the name of truncation for the new parameter τ is explained. The effect of the new parameter τ has consequences on the skewness and the kurtosis of the distribution of the random variable. The effect of the τ is not independent of the parameter α. For the smaller value of α variations on τ it produces less effect on the shape of the distribution. Basically, the two parameters determine values of skewness and kurtosis of the distribution. For R software [4] the package sn, developed by Azzalini, provides facilities to define and manipulate probability distributions of the skew-normal family and some related ones. The first version of the package was written in 1997, and in 2014 the version 1.0-0 was uploaded to CRAN. In this paper, we have used the implemented functions available in the version 1.4-0 of the package. In this paper, we analyse the evolution of skewness and kurtosis of extended skew-normal distribution as a function of two parameters (shape and truncation). We focus in the value of kurtosis and skewness and the existence of a range of values where tiny modification of the parameters produces large oscillations in the values. Calculations of the mean, variance, skewness and kurtosis are made from the cumulant generating function. The analysis shows while mean and variance present correct values and trend, skewness and kurtosis present an instability development for greater values of truncation. Moreover, some values of kurtosis could be erroneous. Packages implemented in software R confirm the existence of a range where value of kurtosis presents a random evolution.

Keywords: Computational Cumulants, Extended Skew-Normal Distribution, Kurtosis, Skewness.

Acknowledgements

References

[1]Azzalini, A. (1985) A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**: 171–178.

[2]Azzalini, A. and Capitanio, A. (2014) *The Skew-Normal and Related Families*, Cambridge, Cambridge University Press.

[3]Canale, A. (2011) Statistical aspects of the scalar extended skew-normal distributions. *Metron* **LXIX** (3): 279–295.

[4]Team, R.C. (2014) *A language and environment for statistical computing*, Vienna, R Foundation for Statistical Computing.

Using the Rasch probabilistic outcome for assessing students performance in an Information Security course

Anacleto Correia[1] and Victor Lobo[1]

[1]Portuguese Naval Academy/CINAV, Center of Naval Research, Portugal

E-mail address: *cortez.correia@marinha.pt, sousa.lobo@marinha.pt*

Abstract

Assessing students' knowledge is important in order to validate how effective a lecturer was conveying a course's subjects, as well as the ability of students to apprehend the lectured content. Improvements in learning have been sought, in recent years, through new pedagogical approaches using for instance Massive Open Online Courses (MOOC), with students approaching learning through interaction with multimedia contents and monitoring, in a regular basis, their own progress with quizzes at the end of each pedagogical unity. This study was applied to students attending an undergraduate course on Information Security. The students' knowledge of the lectured subjects in the course were measured through the Rasch probabilistic model. The study addressed firstly, the learning outcome patterns of students in the information security subjects based on an Entrance-Exit survey. This is followed by investigating students' perceived learning ability based on course's learning outcomes and students' actual learning ability based on their final examination scores.

Keywords: Learning ability, Rasch measurement model, Information Security, MOOC, elearning.

———⊙———
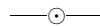
The Rasch analysis [1,2] applied to this study revealed that students perceived themselves as lacking the ability to understand about 75% of the information security concepts at the beginning of the course but eventually they revealed a good understanding of the topics at the end of the course [3,4]. Collected data of students' performance at the final examination [5], showed evidences that their ability in understanding the topics varies at different probability values given the initial knowledge of students and the level of difficulty of the questions. The majority of students found abstract topics and situations related with practical experience in organizations to be the most difficult to understand. Since the course was delivered using the Moodle platform, the study was supplemented by data collected from the log system of the eLearning platform. By correlating the students' level of performance in the course with the time they spent accessing contents made available by e-learning, a high degree of positive correlation was found

Acknowledgements

References

[1] M. Zamalia, M. I. Nurulasyikin, S. Shamsiah, A. Sanizah "Students' Learning Styles and Preferences in Learning Selected Statistical Topics Based on Interview and Multidimensional Scaling Techniques, 3rd International Conference on Education and Social Sciences (INTCESS 2016) ISI Web of Sciences, pp. 145—153, 2016.

[2] N. L. Abu Kassim, N.Z. Ismail, Z. Mahmud , M.S. Zainol, " Measuring students' understanding of statistical concepts using Rasch Measurement", International Journal of Innovation and Management Technology, vol.l,no.l,pp. 13-19, 2010.

[3] R. L. Linn, and N. E. Gronlund, Measurement and assessment in teaching (8th ed.) Englewood Cliffs, NJ: Merrill/Prentice Hall, 2000.

[4] J. D. Allen, Grades as valid measures of academic achievement of classroom learning, vol. 78(5), pp. 218-223. HeldrefPublication.

[5] Z. Mahmud, Diagnosis of perceived attitude, importance, and knowledge in statistics based on rasch probabilistic model. International Journal of Applied Mathematics and Informatics, Issue 5, vol. 4, pp. 291-298, 2011.

Performance of fishing policies for populations with weak Allee effects in a random environment

Nuno M. Brites[1] and Carlos A. Braumann[1,2]

[1]Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e
Formação Avançada, Universidade de Évora
[2]Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora

E-mail address: *brites@uevora.pt, braumann@uevora.pt*

Abstract

We use stochastic differential equations to model the growth of a population subjected to fishing and under week Allee efects. Two optimal harvesting policies are presented, one with variable effort, which is inapplicable (for practical reasons) in a random environment, and the other with constant effort, which is easily applicable. The performance of the two policies will be measured by the profit obtained over a finite time horizon. We will show that there is a slight reduction in profit when choosing the optimal policy with constant effort instead of the optimal policy with variable effort.

Keywords: Allee effects, constant effort, harvesting policies, profit optimization, stochastic differential equations.

———⊙———

In a random environment, we describe the growth of a population subjected to harvesting through stochastic differential equations (as in [1] and [3]).

We assume that the population is under the influence of weak Allee effects, that is, at very low values of population size, we observe lower *per capita* growth rates instead of the higher rates one would expected considering the higher availability of resources per individual (see, for example, [4]). The presence of weak Allee effects when population size is low may be due to the difficulty in finding mating partners or in constructing a strong enough group defense against predators.

We consider the population natural growth to follow a logistic-like model with Allee effects and that the rate of harvesting is proportional to the existing population and to the effort exerted in the capture.

The main goal of this work is to compare the performance of two fishing policies: one with variable effort, here named optimal policy, and the other with constant effort, denoted by sustainable optimal policy. The first allows the fishing effort to vary rapidly and abruptly depending on population size which, in a random environment, also varies constantly. This type of policy is inapplicable from the practical point of view. In addition, this policy requires the estimation of population size at each time instant (see [5]), which is usually an expensive, inaccurate, and time-consuming task. The second policy considers the application of a constant effort over time and predicts the sustainability of the population as well as the existence of a stationary density for its size (see [2]). This policy has the advantage of being applicable, easily implemented and does not require knowledge of population size at any given time. The performance of the two policies will be assessed by the profit obtained over a finite time horizon. Using realistic data based on a fish population, we will show that there is only a slight reduction in profit when choosing

the optimal sustainable policy with constant effort instead of the optimal and inapplicable policy with variable effort.

Acknowledgements

References

[1] Beddington, J. R. and May, R. M. (1977) Harvesting natural populations in a randomly fluctuating environment. *Science* **197**: 463–465.

[2] Braumann, C. A. (2002) Variable effort harvesting in random environments: generalization to density-dependent noise intensities. *Math. Biosciences* **177 & 178**: 229–245.

[3] Braumann, C. A. (1985) Stochastic differential equation models of fisheries in an uncertain world: extinction probabilities, optimal fishing effort, and parameter estimation. In *Mathematics in Biology and Medicine* (V. Capasso, E. Grosso, and S. L. Paveri-Fontana, editors), Springer, Berlin: 201–206.

[4] Carlos, C., Braumann, C. A. (2017) General population growth models with Allee effects in a random environment. *Ecological Complexity* **30**: 26–33.

[5] Suri, R. (2008) *Optimal harvesting strategies for fisheries: A differential equations approach*, PhD thesis, Massey University, Albany, New Zealand.

Contributed Poster

Student Performance in Mathematics - PISA 2015

Susana Faria[1,2] and Manuel João Castigo[1,3]

[1]University of Minho
[2]Centre of Mathematics
[3]Universidade Pedagógica de Moçambique

E-mail address: *sfaria@math.uminho.pt, cmanueljoao@gmail.com*

Abstract

Multilevel regression models are regression models which usually apply in situations where data are hierarchically structured. For that reason, these models are very important in the analysis of studies related with education because the population of these type of study are found structured in a hierarchical way (students nested into classes, classes nested into schools). In this work, a multilevel analysis is applied to data collected under the Programme for International Student Assessment (PISA) 2015 in Mathematics literacy in Portugal. With this application, we pretend to analyse the impact of student and school factors on mathematics achievement of Portuguese students in PISA 2015 and to understand which part of the student's achievement is due to regional disparities.

Keywords: Finite Mixtures of Linear Mixed Models; Model selection; Simulation study.

Acknowledgements

References

[1]Agasisti, T., Cordero-Ferrera, J.M. (2013) Educational disparities across regions: A multilevel analysis for Italy and Spain. *Journal of Policy Modeling*, **35**, 1079–1102.

[2]Agasisti, T., Ieva, F., Paganoni, A.M. (2017) Heterogeneity, school-effects and the North/South achievement gap in Italian secondary education: evidence from a three-level mixed model. *Statistical Methods and Applications*, **26 (1)**, 157–180.

[3]Goldstein, H. (2011) *Multilevel Statistical Models*, John Wiley and Sons.

[4]Pereira, M. C., Reis, H. (2012) What accounts for Portuguese regional differences in students performance? Evidence from OECD PISA. *Lisboa: Economic Bulletin*, Banco de Portugal.

A Simulation Study For Determining the Number of Components in Finite Mixtures of Linear Mixed Models

Susana Faria[1,2] and Luísa Novais[1]

[1]University of Minho
[2]CMAT-Centre of Mathematics of University of Minho

E-mail address: *sfaria@math.uminho.pt, luisa_novais92@hotmail.com*

Abstract

Finite mixture models are a well-known method for modelling data that arise from a heterogeneous population. In regression analysis, it has been a popular practice for modelling unobserved population heterogeneity through finite mixture regression models. Within the family of mixture regression models, finite mixtures of linear mixed models have also been applied in different areas of application. They conveniently allow to account for correlations between observations from the same individual and to model unobserved heterogeneity between individuals at the same time. Choosing the number of components for mixture models has long been considered as an important but difficult research problem. There is wide variety of literature available on the performance of model selection statistics for determining the number of components in mixture models. In this study the performance of various model selection methods was investigated in the context of Finite Mixtures of Linear Mixed Models.

Keywords: Finite Mixtures of Linear Mixed Models; Model selection; Simulation study.

Acknowledgements

References

[1]Bai X., Chen K., Yao W. (2016) Mixture of linear mixed models using multivariate t distribution, *Journal of Statistical Computation and Simulation*, **86:4**, 771–787.
[2]Celeux G., Martin O., Lavergne Ch. (2005) Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, **5**, 1–25.
[3]Fruhwirth-Schnatter, S. (2006) *Finite Mixture and Markov Switching Models*. Springer Series in Statistics.
[4]Young D., Hunter D. R. (2015) Random effects regression mixtures for analyzing infant habituation, *Journal of Applied Statistics*, **42:7**, 1421–1441.

Prediction of Structural Components in Environmental Time Series

A. Manuela Gonçalves[1], Marco Costa[2] and Olexandr Baturin[3]

[1]CMAT-Centre of Mathematics, DMA-Department of Mathematics and Applications,
University of Minho, Portugal
[2]CIDMA-Centre for Research and Development in Mathematics and Applications,
University of Aveiro, Portugal
[3]DMA-Department of Mathematics and Applications, University of Minho, Portugal

E-mail address: *mneves@math.uminho.pt, marco@ua.pt, olexandr.baturin@gmail.com*

Abstract

A structural time series model is one which is set up in terms of dynamic components which have a direct interpretation. In this study, in the context of a water quality monitoring problem, it is proposed an approach for the structural time series analysis based on the state space models associated to the Kalman Filter. The main goal is to analyze and evaluate the temporal evolution of the environmental time series, and to identify trends or possible changes in water quality within a dynamic monitoring procedure, by identifying unexpected changes that are important for the process of management and evaluation of water quality.

Keywords: Dynamic structural components, structural time series, state-space models, Kalman filter.

Acknowledgements

References

[1]Costa, M., Gonçalves, A.M., Teixeira, L. (2016) Change-point detection in environmental time series based on the informational approach. *Electronic Journal of Applied Statistical Analysis*, **9(2)**:267–296.
[2]Durbin, J., Koopman, S. (2001) *Time Series Analysis by State Space Methods.* Oxford University Press, Oxford.
[3]Harvey, A.C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press. Cambridge.
[4]Petris, G., Petrone, S., Campagnoli, P. (2009) *Dynamic Linear Models with R.* Springer, New York.

Statistical Modeling in the Management of Wastewater Treatment Plants

Ana Gonzaga[1], A. Manuela Gonçalves[2], Cristiana Barbosa[3] and Pedro Bastos[4]

[1]DMA-Department of Mathematics and Applications, University of Minho, Portugal
[2]CMAT-Centre of Mathematics, DMA-Department of Mathematics and Applications, University of Minho, Portugal
[3,4]Águas do Norte, Grupo Águas de Portugal, Portugal

E-mail address: *a76204@alunos.uminho.pt, mneves@math.uminho.pt*
cristiana.barbosa@ADP.PT, pedro.bastos@ADP.PT

Abstract

The progressive deterioration of water resources and the large amount of polluted water generated in modern societies gives Wastewater Treatment (WWT) processes a fundamental importance in water prevention and control. Inside a biological Wastewater Treatment Plant (WWTP), the activated sludge process is the most commonly used technology to remove organic pollutants from wastewater (by means of a bacterial biomass suspension). This is the most cost-effective technology, it is very flexible and it can be adapted to different kinds of wastewater. Therefore, it is very important to understand and to model the management processes involved that can lead to benefits for the overall WWTP, in particular in cost-effectiveness. In this work the discussion focuses on the dynamic monitoring procedure based on the statistical modeling approach, in order to quantify and to characterize significant statistical patterns of interaction between wastewater flows (that are tributary to the WWTPs), hydro-meteorological variables (such as rainfall), and physicochemical variables. A statistical exploratory analysis and linear models were performed in order to obtain an accurate prediction and forecast of the relevant predictors (wastewater effluent variables) in the flows' behaviour and which have the greatest impact on cost reduction. The statistical modeling procedure was applied to a set of nine Wastewater Treatment Plants located in the Northwest region of Portugal (five in rural regions and four in urban regions), and data set consists of monthly measurements during a period of two years from January 2015 to December 2016. By accommodating the well-know seasonal regimes of dry and wet seasons, the statistical results will provide a better representation of the plant's real situation in order to design a efficient management process.

Keywords: Wastewater flows, physicochemical variables, seasonality, costs, correlations, linear models.

Acknowledgements

References

[1] Costa, M, Gonçalves, A.M. (2012) Combining Statistical Methodologies in Water Quality Monitoring in a Hydrological Basin - Space and Time Approaches. In *Water Quality Monitoring and Assessment*, ed. Kostas Voudouris and Dimitra Voutsa, 121–142. ISBN: 978-953-51-0486-5. InTech Published.

[2] Gonçalves, A.M., Alpuim, T. (2011) Water Quality Monitoring using Cluster Analysis and Linear Models. *Environmetrics*, **22(8)**: 933–945.

Modeling health perception based on psychosocial risks. A data driven approach

Luís M. Grilo[1,2] and Valter Vairinhos[3,4]

[1]Instituto Politécnico de Tomar, Departamento de Matemática e Física, Portugal
[2]CMA - Centro de Matemática e Aplicações, FCT, UNL, Portugal
[3]CINAV, Center of Naval Research, Portugal
[4]ICLAB- ICAA- Intelectual Capital Accreditation Association, Santarém, Portugal

E-mail address: *lgrilo@ipt.pt, valter.vairinhos@sapo.pt*

Abstract

The percentage of workers exposed to psychosocial stressors has been increasing through the ages in every activity sector, with negative consequences for workers, organizations and national economies. An adapted version of the Copenhagen Psychosocial Questionnaire was used to assess the psychosocial risks and their impact on workers' health and wellbeing of a Portuguese company. The answers to the 41 questions of this questionnaire were expressed in a Likert-type scale of five categories and generated a data set of around 5000 observations by 41 variables. The data have a character entirely observational. It must be pointed out that authors did not have access to any kind of biographic information such as age, job position or sex of respondents. Having in mind the characterization and model building of workers perceptions about its occupational health, implicit in their answers, several multivariate methodologies were employed to describe and synthetize the data. Specifically, from cluster analysis, a set of 6 variables clusters emerged. Those clusters where studied for possible psychological and social meaning, statistical homogeneity and unidimensional behavior, being the conclusion that it makes sense to see those clusters as the manifestations of latent variables with clear psychosocial meaning. Based on the results of this empirical approach, a path model was formulated, expressing a priori perceptions and beliefs about causal relations among those latent variables, consistent with literature and authors experience. That model was estimated using the Partial Least Squares (PLS) technique, coupled with bootstrap, implemented by R packages. Although PLS Path Modeling (PLS-PM) is, basically, a descriptive technique, the use of bootstrap methodology allows limited but credible forms of inference leading to the conclusion that the main causal hypothesis expressed with the model were supported by data. The results obtained, despite the limitations alluded to, are considered very interpretable, useful and encouraging, given the available alternatives.

Keywords: Causality, Bootstrap, Partial Least Squares, Path Modeling, R environment.

Acknowledgements

References

[1]Bakker, A., Demerouti, E. and Sanz-Vergel, A.I. (2014). Burnout and Work Engagement: The JD–R Approach. Annual Review of Organizational Psychology and Organizational Behavior, 1(1), 140114155134003.

[2]Hair, J. F, Hult, G. T. M., Ringle, C. M., Sarstedt, M. (2014). A primer on partial least squares structural equation modeling (PLS-SEM). SAGE Publications, Inc.

[3]Kristensen, T.S., Borritz, M., Villadsen, E. and Christensen, K.B. (2005). The Copenhagen Burnout Inventory: A new tool for the assessment of burnout. Work&Stress 19: 192 207.

[4]Maslach, C. and Leiter, M.P. (2016). Understanding the burnout experience: recent research and its implications for psychiatry. World Psychiatry152:103-111.

[5]Maslach, C. and Leiter, M.P. (2008). Early predictors of job burnout and engagement. Journal of Applied Psychology 93: 498-512.

Cognitive Appraisal as a Mediator in the Relationship Between Work-Family Conflicts and Burnout

Jéssica Rodrigues[1], A. Manuela Gonçalves[2], Susana Faria[2], A. Rui Gomes[3] and Clara Simões[4],

[1,2,3,4]University of Minho, Portugal
[1]DMA-Department of Mathematics and Applications
[2]CMAT-Centre of Mathematics, DMA-Department of Mathematics and Applications
[3]School of Psychology
[4]School of Nursing

E-mail address: *jessicarrodrigues93@gmail.com, mneves@math.uminho.pt sfaria@math.uminho.pt, rgomes@psi.uminho.pt, csimaes@ese.uminho.pt*

Abstract

This study tested the mediating role of cognitive appraisal in the relationship between work-family conflicts and burnout. This mediation was tested using a cross-sectional study based on self-reported measures. The total sample consisted of 438 portuguese teachers who teach from kindergarten through high school and completed an evaluation protocol with measures of work-family conflicts, cognitive appraisal, and burnout. To test the mediating role of cognitive appraisal in the relationship between work-family conflicts and burnout, we used Structural Equation Modeling (SEM). The results confirmed cognitive appraisal partially mediated the relationship between work-family conflicts and burnout. The findings indicated that cognitive appraisal is an important underlying mechanism in explaining adaptation at work.

Keywords: Burnout; cognitive appraisal; teachers; work-family conflicts.

Acknowledgements

References

[1]Byrne, B.M. (2010) *Structural equation modeling with AMOS: Basic concepts, applications, and programming.* New York: Routledge.

[2]Glaser, W. and Tracy, D.H. (2013) Work-family conflicts, threat-appraisal, self-efficacy and emotional exhaustion. *Journal of Managerial Psychology,* **28(2)**, 164–182.

[3]Gomes, A.R., Faria, S. and Gonçalves, A.M. (2013) Cognitive appraisal as a mediator in the relationship between stress and burnout. *Work and Stress,* **27(4)**, 351–367.

[4]Simães, C., McIntyre, T. and McIntyre, S. (2009) Portuguese adaptation of the Work-Family Conflict and Family-Work Conflict scales in Nurses: a Preliminary Study. *Psychology and Health,* **24(SI1)**, 364–364.

Satisfaction Index in Engineering Courses in Portugal: Institutions' ratio versus Students point of view ratio

Raquel Oliveira[1], A. Manuela Gonçalves[2] and Rosa M. Vasconcelos[3]

[1,2,3]University of Minho, Portugal
[1]CMAT - Centre of Mathematics, Portugal, IPCA-EST
[2]CMAT - Centre of Mathematics, DMA - Department of Mathematics and Applications
[3]2C2T - Centre for Textile Science and Technology, DET - Department of Textile Engineering

E-mail address: *rmro_17@hotmail.com, mneves@math.uminho.pt, rosa@det.uminho.pt*

Abstract

In this paper, we continue previous studies on students' allocation satisfaction in the Portuguese public higher education system in the academic engineering programs by comparing two ratios, one provided by the Portuguese Education Ministry - the Institutions' point of view, and the other proposed by the authors - the Students' point of view. The data set used covers the results of the national contest from 2007 to 2016, provided by the Portuguese Ministry of Education. Non-parametric tests were performed to assess whether there are significant differences between the considered ratios. The results seem to confirm that there are some differences between the two ratios, which may mean that the proposed ratio should be considered for a better understanding of the satisfaction of the allocation of students in engineering courses in Portugal.

Keywords: Students satisfaction index, applicants satisfaction index, higher education, non-parametric tests.

Acknowledgements

References

[1]OECD, Organization for economic co-operation and Development (2006) Reviews of National policies for education: Tertiary education in Portugal. Examiner's Report. Available at http://www.dges.mctes.pt/NR/rdonlyres/8B016D34-DAAB-4B50-ADBB-25AE105AEE88/2564/Backgoundreport.pdf.

[2]http://www.europa.eu.

[3]European Ministers of Education. The Bologna Declaration, 1999.

[4]http://www.dges.mctes.pt/DGES/pt/Estudantes/Acesso/Estatisticas/.

[5]http://www.dges.mctes.pt.

[6]http://www. dges.mctes.pt/DGES/pt/Estudantes/Acesso.

[7]Oliveira, R., Gonçalves, A.M., Vasconcelos, R.M. (2013) An empirical study on the index of satisfaction of student allocation in the Portuguese undergraduate engineering courses. *Book of Abstracts ASMDA2013*, 4 pages.

[8]Oliveira, R., Gonçalves, A.M., Vasconcelos, R.M. (2015) A New Perspective of Students Allocation Satisfaction in Engineering Courses in Portugal. *AIP Conference Proceedings* **1648**: 840002-1–840002-4.

[9]Oliveira, R., Gonçalves, A.M., Vasconcelos, R.M. (2016) Index of satisfaction in engineering courses in Portugal based on the student's perspective. *AIP Conference Proceedings* **1738**: 470003-1–470003-4.

[10]Higgins, J.H. (2004) *Introduction to Modern Nonparametric Statistics*. Thomson, Toronto.

[11]Siegel, S., Castellan, N.J. (1988) *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill, 2Edition.

Data mining and statistical analysis of educational data

Pedro Melgueira[1,2,4], Nuno M. Brites[1,2], Irene Pimenta Rodrigues[1,2,3] and Ligia Silva Ferreira[1,2,3]

[1] Universidade de Évora, Portugal
[2]Centro de Tecnologias Educativas
[3]Laboratório de Informática, Sistemas e Paralelismo
[4]Departamento de Informática

E-mail address: *pedromelgueira@gmail.com, brites@uevora.pt*
ipr@uevora.pt, lsf@uevora.pt

Abstract

We present some data mining techniques applied to Educational Data. From the Moodle data repository of the University of Évora, we apply supervised learning techniques with the aim of predicting the students success from their interaction with Moodle. At the end, we show very interesting conclusions when unsupervised learning techniques are applied.

Keywords: Big data, classification, data mining, decision trees, learning management systems.

——⊙——

The advancement of the technologies related to the internet has enabled E-Learning to gain popularity as a way of transmitting knowledge. Institutions such as Universities and Companies have been using E-Learning to disseminate educational content to remote locations, reaching out to students and employees who are physically distant. The Moodle platform is an example of a Learning System Management (LMS). LMSs provide on-line platforms where teachers and trainers can publish content organized into activities, conduct assessments, and other activities so that the students involved can learn and be assessed. In addition, LMS generates and stores large amounts of data, named Educational Data, from not only user activities but also the LMS itself.

In this work we will present some data mining techniques applied to Educational Data. In fact, from the Moodle data repository of the University of Évora, we will apply supervised learning techniques with the aim of predicting the students success from their interaction with Moodle. We will also see interesting conclusions when unsupervised learning techniques are applied.

Preliminary statistical analysis of barotraumatism occurence and hyperbaric oxygen therapy

M. Filomena Teodoro[1,2], Sofia Teles[2], Marta C. Marques[3] and Francisco G. Guerreiro[2,4]

[1]CEMAT, Instituto Superior Técnico, Lisbon University, Lisboa, Portugal
[2]CINAV, Portuguese Naval Academy, Base Naval de Lisboa, Alfeite, Portugal
[3]Medicine Faculty, Lisbon University, Portugal
[4]CMSH, Centro de Medicina Subaquática e Hiperbárica, Marinha Portuguesa

E-mail address: *maria.alves.teodoro@marinha.pt*

Abstract

Hyperbaric oxygen therapy (HBOT) is a therapeutic modality consisting of the intermittent administration of 100% oxygen within a chamber under pressure conditions above sea level pressure (1 ATA), allowing an increase of perfusion of 0_2 in the tissues. Consequently, there exists a reduction in edema and tissue hypoxia, aiding the treatment of ischemia and infection[1,2]. Between several complications, middle ear barotrauma (BTOM) is the most frequent but its incidence, risk factors and severity are not yet well known. This work was started in [3] being studied the clinical characteristics of 1732 patients who underwent treatment at the Portuguese Navy's Center for Underwater and Hyperbaric Medicine (HMCS) between 2012 and 2016, in order to better characterize this problem with regard to incidence, severity and recurrence, as well as to identify possible risk factors such as age, sex, clinical indication for HBOT, personal history of allergic rhinitis, and symptomatology of nasal obstruction at the time of the occurrence. There was an incidence of 8.3% with BTOM between patients. Most of occurrences were unilateral 62%, the remaining 21% cases were bilateral. BTOM occurred in the first 3 sessions in 36% of cases and 44% in up to 5 sessions. The recurrence rate was 28%. There were constructed statistical models so one could get the the statistical significant relations between gender, personal clinical history and BTOM. Were identified some risk factors. An approach by general least squares (GLM) is still ongoing. The results are promising but their analysis still need to be completed.

Keywords: Hyperbaric oxygen therapy, barotraumatism, middle ear, risk factors, analysis of variance, GLM.

Acknowledgements

References

[1]Acott, C. (1999) A Brief History of Diving and Decompression Illness. *SPUMS J.* **29**(2): 98–109.

[2]Mathieu, D. (ed.) (2006) *Handbook on hyperbaric medicine*, Dordrecht, The Netherlands, Springer.

[3]Teles, S.(2017) *O Barotraumatismo do Ouvido Média em Doentes a Fazer Oxigenoterapia Hiperbárica*, Master Thesis, Medical Faculty, Lisbon University.

[4]Turkman, M.A. and Silva, G. (2000) *Modelos Lineares Generalizados da teoria a prática*, Lisboa, Sociedade Portuguesa de Estatística.

Structured families of symmetric stochastic matrices

Hermelinda Carlos[1,2], Cristina Dias[3,4], Carla Santos[5,6] and João Tiago Mexia[7,8]

[1,5,7,9]Escola Superior de Tecnologia e Gestão do Instituto Politécnico de Portalegre, Portugal
[3,5,8]CMA–Centro de Matemática e Aplicações da Universidade Nova de Lisboa, Portugal
[6]Departamento de Matemática e Ciências Físicas do Instituto Politécnico de Beja, Portugal
[2]C3i – Coordenação Interdisciplinar para a Investigação e a Inovação
[8]Departamento de Matemática da Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal

E-mail address: : *hecarlos@estgp.pt, cpsd@estgp.pt*
carla.santos@ipbeja.pt, jtm@fct.unl.pt

Abstract

In this work we study structured families of models, whose matrices correspond to the treatments of a base design. We also consider families of models divided into subfamilies that correspond to these treatments. We are mainly interested in basic models with orthogonal structure. We present this structure and show how to apply these models in the study of structured families.

Keywords: Structured families, Subfamilies, Orthogonal structure.

Acknowledgements

References

[1]Areia, A. (2009) Séries emparelhadas de estudo. Ph.D. Thesis Évora University.

[2]Ito, P. K. (1980) Robustness of Anova and Macanova Test Procedures, P. Krishnaiah, R. (ed), Handbook of Statistics, Amsterdam: North Holland 1 : 199-236.

[3]Mexia J. T. (1988) Standardized Orthogonal Matrices and Decomposition of the Sum of Squares for Treatments. Trabalhos de Investigação, nº2, Departamento de Matemática, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa.

[4]Dias, C. (2013) Modelos e Famílias de Modelos para Matrizes Estocásticas Simétricas.Ph.D. Thesis, Évora University.

[5]Oliveira, M. M. and Mexia, J. T. (1999) F Tests for Hypothesis on theStructure Vectors of Series. Discussiones Mathematicae 19 (2) : 345-353.

[6]Scheffé, H. (1959) The Analysis of Variance, New York, John Wiley & Sons.

1D third-grade fluid model

Fernando Carapau[1,2], Paulo Correia[1,2] and Luís M. Grilo[3,4]

[1]Universidade de Évora, Portugal
[2]CIMA-Centro de Investigação em Matemática e Aplicações, Universidade de Évora
[3]Instituto Politécnico de Tomar, Departamento de Matemática e Física, Portugal
[4]CMA - Centro de Matemática e Aplicações, FCT, UNL, Portugal

E-mail address: *flc@uevora.pt, pcorreia@uevora.pt, lgrilo@ipt.pt*

Abstract

A specific modified constitutive equation for a third-grade fluid is proposed so that the model be suitable for applications where shear-thinning or shear-thickening may occur. For that, we use the Cosserat theory approach reducing the exact three-dimensional equations to a system depending only on time and on a single spatial variable. This one-dimensional system is obtained by integrating the linear momentum equation over the cross-section of the tube, taking a velocity field approximation provided by the Cosserat theory. From this reduced system, we obtain the unsteady equations for the wall shear stress and mean pressure gradient depending on the volume flow rate, Womersley number, viscoelastic coefficient and flow index over a finite section of the tube geometry with constant circular cross-section.

Keywords: One-dimensional model, generalized third-grade model, shear-thickening fluid, shear-thinning fluid, Cosserat theory.

Acknowledgements

References

[1] C. Truesdell and W. Noll, *The non-linear field theories of mechanics*, 2nd edition, Springer, New York, 1992.

[2] R.S. Rivlin and J.L. Ericksen, Stress-deformation relations for isotropic materials, *J. Rational Mech. Anal.*, **4**, pp. 323–425 (1955).

[3] R.L. Fosdick and K.R. Rajagopal, Thermodynamics and stability of fluids of third grade, *Proc. R. Soc. Lond. A.*, **339**, pp. 351–377 (1980).

[4] J.E. Dunn and K.R. Rajagopal, Fluids of differential type: critical review and thermodynamic analysis, *Int. Journal of Engineering Science*, **33**, pp. 689–729 (1995).

[5] D. Mansutti and K.R. Rajagopal, Flow of a shear thinning fluid between intersecting planes, *Int. J. Non-Linear Mech.*, **26** (5), pp. 769–775 (1991).

[6] D. Mansutti, G. Pontrelli and K.R. Rajagopal, Steady flow of non-Newtonian fluids past a porous plate with suction or injection, *Int. J. Numer. Methods Fluids*, **17** (11), pp. 927–941 (1993).

[7] D.A. Caulk and P.M. Naghdi, Axisymmetric motion of a viscous fluid inside a slender surface of revolution, *Journal of Applied Mechanics*, **54** (1), pp. 190–196 (1987).

Sea Ice Index - looking at the trend

M. Rosário Ramos[1,2] and Clara Cordeiro[3,4]

[1]Universidade Aberta, Portugal
[2]Centro de Matemática, Aplicações Fundamentais e Investigação
Operacional(CMAF-CIO), Universidade de Lisboa, Portugal
[3]Faculdade de Ciências e Tecnologia da Universidade do Algarve
[4]Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal

E-mail address: *MariaR.Ramos@uab.pt, ccordei@ualg.pt*

Abstract

This study analyses the trends of the Sea Ice Extent Index in the Northern (NH) and Southern (SH) Hemispheres, through its time series of records available by NSIDC: National Snow and Ice Data Center (USA). Sea Ice Index is a source for consistently processed ice extent and concentration images and data values since 1979. The research on this theme is current and diverse, as it is part of the problem of Climate Change. Research on trend analysis produced several parametric and nonparametric approaches to accommodate the specificities of hydrological, environmental and other climate data. In recent years, versions of the tests that use resampling techniques have shown to be versatile and outperform the classical tests in some situations, namely in different structures of serial correlation or there is both deterministic and stochastic seasonality.The analysis of an existing trend in NH and SH time series is performed and compared on the base of two trend tests, the t-test (OLS linear regression) and the Mann-Kendall test (nonparametric), applied in its original form and combined with a resampling technique. Data are monthly Sea Ice Index from November 1979 to March 2017. In a first stage the seasonal component is estimated through a variant of the STL - Seasonal Decomposition of Time Series by Loess, which contemplates non-constant seasonality; then seasonality is removed from the original series before the trend test. The comparison between methods is performed. The order of autocorrelation structure in each time series is estimated by the best fitting model obtained through the AIC information criterion.

Keywords: Trend analysis, sieve bootstrapp, seasonal-trend decomposition, Sea Ice Index, Mann-Kendall, t-test.

Acknowledgements

References

[1]R.B. Cleveland, W.S. Cleveland, J.E. McRae, and I.Terpenning. (1990) Stl: A seasonal-trend decomposition procedure based on loess. *J. Off. Stat* **6** (1): 3–73.
[2]Harcum J. B., Loftis J. C., Ward R. C. (1992) Selecting trend tests for water quality series with serial correlation and missing values. *Water Resour. Bull.* **28** 469–478.

[3] Noguchi, K., Gel, Y.R., Duguay, C.R. (2011) Bootstrap-based tests for trends in hydrological time series, with application to ice phenoloy data. *J. Hydrology.* **410** 150–161.

[4] Ramos, M.R., Cordeiro C. (2014) Trend Tests: a Tendency to resampling. *Proceedings of ITISE 2014-International Work Conference on Time Series*, 322–328, Granada, Spain, ISBN 978-84-15814-97-9.

[5] Fetterer, F., K. Knowles, W. Meier, and M. Savoie. (2016), updated daily. Sea Ice Index, Version 2. Boulder, Colorado USA. NSIDC: National Snow and Ice Data Center. doi: http://dx.doi.org/10.7265/N5736NV7. [acessed 08/04/17].

[6] McLeod, A.I.(2011) kendall: Kendall rank correlation and Mann-Kendall trend test. R package version 2.2, http://cran.r-project.org/package=kendall

[7] Thorsten Pohlert.(2016) trend: Non-Parametric Trend Tests and Change-Point Detection. R package version 0.2.0. http://cran.r-project.org/package=trend

[8] Hyndman RJ. (2017) forecast: Forecasting functions for time series and linear models. R package version 8.0, http://github.com/robjhyndman/forecast.

Iberian energy prices forecasting using a time series approach

Marina A. P. Andrade[1], M. Filomena Teodoro[2,3], Ana Borges[4], Eliana Costa e Silva[4] and Ricardo Covas[5,6]

[1]*ISTAR-IUL, ISCTE, Instituto Universitário de Lisboa*
[2]*CEMAT, Instituto Superior Técnico, Lisbon University, Portugal*
[3]*CINAV, Portuguese Naval Academy, Portuguese Navy, Portugal*
[4]*CIICESI, ESTG, Polytechnic of Porto, Portugal*
[5]*CMA, Lisbon New University, Portugal*
[6]*EDP - Energias de Portugal*

E-mail address: *maria.alves.teodoro@marinha.pt, mariana.andrade@iscte.pt*

Abstract

This work is a consequence of the challenge proposed at the 119th European Study Group with Industry in June 2016, by the company EDP, Energia de Portugal, and the work presented in [2]. Here we introduce a preliminary approach using Time Series Methodology. In short term forecasting electricity price context, the techniques more commonly applied are autoregressive and moving average models ARMA, which can be combined with the stationary form of ARIMA models. ARX, ARMAX, ARIMAX and SARIMAX are the extension of these models when exogenous factors [3] are considered (e.g. generation capacity, load profiles and meteorological conditions). In previous works [2, 1] were obtained interesting results, where different time period slots were identified as important explanatory variables. Using this information, we included this variables in the new ARIMAX models here presented. These exogenous factors allowed to improve the forecasting measure quality, even though we are still in a preliminary stage of analysis. In order to compare the results with previously ones obtained, the hourly prices data were considered from March 10th 2014 to May 29th 2016.

Keywords: Electricity price forecasting, ARIMA, ARIMAX.
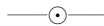
———⊙———
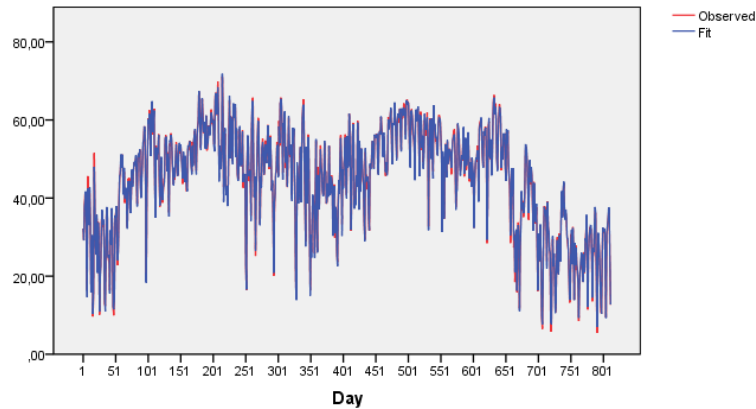
Figure 1: Preliminary results: electricity price estimates. Estimated period: March 10*th* 2014 to May 29*th* 2016.

References

[1] Costa e Silva, E., Borges, A., Teodoro, M.F., Andrade, M.A.P. & Covas, R. (2017) Time Series Data Mining for Energy Prices Forecasting: An Application to Real Data. In A.M. Madureira et al. (Eds.), Intelligent Systems Design and Applications, *Advances in Intelligent Systems and Computing* **57**: 435 462.

[2] Teodoro, M.F., Andrade, M.A.P., Costa e Silva, E., Borges, A. & Covas, R. Energy Prices Forcasting Using GLM, Recent studies on risk analysis and statistical modeling , *Contributions to Statistics* (in press).

[3] Weron, R. (2014) Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* **30**(4): 1030–1081.

Index of authors

HOSTED BY:

Instituto Politécnico Portalegre

AbERTA
www.uab.pt

CEAUL
Centro de Estatística e Aplicações
Universidade de Lisboa

IN COLLABORATION WITH:

Escola Superior Tecnologia Gestão
IPPortalegre

ipt
Instituto Politécnico de Tomar

isi

CIMA
Projecto UID/MAT/04674/2013

UNIVERSIDADE
de MADEIRA

Chicago
Cancer
Health Equity
Collaborative

ISI COMMITTEE
ON RISK ANALYSIS

SPONSORS:

CEAUL
Centro de Estatística e Aplicações
Universidade de Lisboa

DELTA
CAFÉS

Vitalis

São
MAMEDE
AZEITE VIRGEM EXTRA

ADEGA DE
PORTALEGRE
EST WINERY 1951

PSE
PRODUTOS E SERVIÇOS DE ESTATÍSTICA