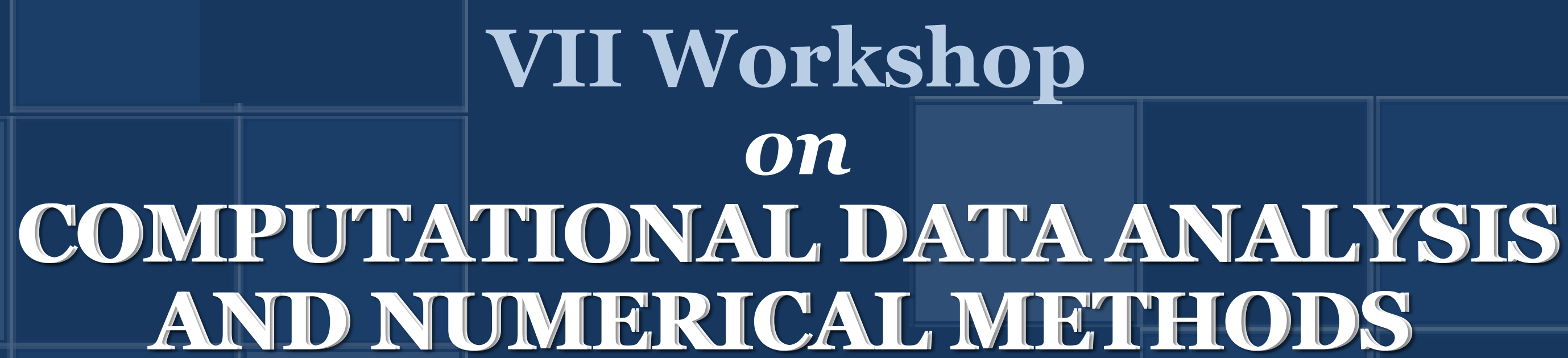


[illegible]

ipt
Instituto
Politécnico
de Tomar

BOOK OF ABSTRACTS

Polytechnic Institute of Tomar
Portugal
September 10–12, 2020

WELCOME TO THE VII WCDANM | 2020

Dear participants, colleagues and friends,

it is a great honour and a privilege to give you all a warmest welcome to the VII Workshop on Computational Data Analysis and Numerical Methods (VII WCDANM), which is organized by the Polytechnic Institute of Tomar (located in the center of Portugal in the beautiful city of Tomar), with the support of some Portuguese research centers, hoping that the final result may exceed the expectations of the participants, sponsors and organizers.

Due to the worldwide pandemic caused by the COVID-19 virus, for the first time, this meeting will be transmitted through videoconference (webinar). Nevertheless, the important contributions of Adélia Sequeira (University of Lisbon, Portugal), Sílvia Barbeiro (University of Coimbra, Portugal), Malay Banerjee (Indian Instituto of Techonology Kampur, India) and Indranhil Ghosh (University of North Carolina at Wilmington, USA) as Plenary Speakers, the high scientific level of oral and poster presentations and an active audience will certainly contribute to the success of the meeting. Part of the accepted papers (theoretical and applied) by the VII WCDANM involve big data, data mining, data science and machine learning, in different areas of research, some giving emphasis to coronavirus. A very special thanks to this small, yet important, scientific community, since this event could not be possible without any of these essential and complementary parts.

This year, there is also the possibility to attend a course on §Modeling Partial Least Squares Structural Equations (PLS-SEM) using SmartPLST given by Christian M. Ringle ((TUHH) Hamburg University of Technology, Germany), who kindly and readily accepted our invitation and to whom we are very grateful.

A special acknowledgment is also due to the Members of the Executive, Scientific and Organizing Committees. In particular, Anuj Mubayi (Arizona State University, USA), Milan Stehlík (Johannes Kepler University, Austria), Ana Nata, Isabel Pitacas and Manuela Fernandes (hosts from the Polytechnic Institute of Tomar, Portugal), A. Manuela Gonçalves (University of Minho, Portugal), Teresa Oliveira (Aberta University) and Fernando Carapau (University of Évora, Portugal) have been relentless in search for a balanced, broad and interesting program, having achieved an excellent result.

For the third consecutive year, the Journal of Applied Statistics (Taylor & Francis) and Neural Computing and Applications (Springer) are also associated to the event, being extremely important in the dissemination of the scientific results achieved at the meeting.

Given the above, it is a pleasure to be "together" with all of you in this web conference, hoping it may provide an intellectual stimulus and an opportunity for the scientific community to jointly work and disseminate scientific research, namely presenting approaches that may contribute to the solution of the pandemic we are experiencing in the expectation that the present might be past in the near future!

Tomar, September 10-12, 2020.

Chairman of the Executive committee of VII WCDANM,



Luís Miguel Grilo

Instituto Politécnico de Tomar, Portugal

Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal

Committees

Organizing Committee

Luís Miguel Grilo, Instituto Politécnico de Tomar, Portugal (Chairman of the Workshop)
Ana Nata, Instituto Politécnico de Tomar, Portugal
Arminda Manuela Gonçalves, Universidade do Minho, Portugal
Isabel Pitacas, Instituto Politécnico de Tomar, Portugal
Maria Manuela Fernandes, Instituto Politécnico de Tomar, Portugal
Fernando Carapau, Universidade de Évora, Portugal
Teresa Oliveira, Universidade Aberta, Portugal

Executive Committee

Anuj Mubayi, Arizona State University, USA
Fernando Carapau, Universidade de Évora, Portugal
Luís M. Grilo, Instituto Politécnico de Tomar, Portugal
Milan Stehlík, Johannes Kepler University, Austria

Scientific Committee

Adélia Sequeira, IST, Universidade de Lisboa, Portugal
Alberto Pinto, Universidade do Porto, Portugal
Ana Cristina Nata, Instituto Politécnico de Tomar, Portugal
Ana Isabel Borges, Instituto Politécnico do Porto, Portugal
Ashwin Vaidya, Montclair State University, USA
Anuj Mubayi, Arizona State University, USA
Arminda Manuela Gonçalves, Universidade do Minho, Portugal
Amílcar Oliveira, Universidade Aberta, Portugal
Alexandra Moura, Instituto Superior de Economia e Gestão, Portugal
Alberto Simões, Universidade da Beira Interior, Portugal
Adelaide Figueiredo, Universidade do Porto, Portugal
Aldina Correia, Instituto Politécnico do Porto, Portugal
Bong Jae Chung, George Mason University, USA
Catarina Nunes, Universidade Aberta, Portugal
Carlos Braumann, Universidade de Évora, Portugal
Carla Santos, Instituto Politécnico de Beja, Portugal
Carlos Ramos, Universidade de Évora, Portugal
Carlos Agra Coelho, Universidade Nova de Lisboa, Portugal
Célia Nunes, Universidade da Beira Interior, Portugal
Cristina Dias, Instituto Politécnico de Portalegre, Portugal
Christos Kitsos, University of West Attica, Greece
Dário Ferreira, Universidade da Beira Interior, Portugal
Delfim F. M. Torres, Universidade de Aveiro, Portugal
Dora Gomes, Universidade Nova de Lisboa, Portugal
Eliana Costa e Silva, Instituto Politécnico do Porto, Portugal
Fátima de Almeida Ferreira, Instituto Politécnico do Porto, Portugal
Feliz Minhós, Universidade de Évora, Portugal
Fernanda Figueiredo, Universidade do Porto, Portugal
Filipe Marques, Universidade Nova de Lisboa, Portugal

Fernando Carapau, Universidade de Évora, Portugal
Filomena Teodoro, Escola Naval, Portugal
Henrique Pinho, Instituto Politécnico de Tomar, Portugal
Hermenegildo Oliveira, Universidade do Algarve, Portugal
Izabela Pruchnicka-Grabias, Warsaw School of Economics, Poland
Ilda Inácio, Universidade da Beira Interior, Portugal
Indranil Ghosh, University of North Carolina at Wilmington, USA
Isabel Silva Magalhães, Universidade do Porto, Portugal
Irene Rodrigues, Universidade de Évora, Portugal
João Tiago Mexia, Universidade Nova de Lisboa, Portugal
João Janela, Instituto Superior de Economia e Gestão, Portugal
José Carrilho, Universidade de Évora, Portugal
José Gonçalves Dias, Instituto Universitário de Lisboa, Portugal
José António Pereira, Universidade do Porto, Portugal
José Augusto Ferreira, Universidade de Coimbra, Portugal
Leonid Hanin, Idaho State University, USA
Luís Castro, Universidade de Aveiro, Portugal
Luís Bandeira, Universidade de Évora, Portugal
Luís M. Grilo, Instituto Politécnico de Tomar, Portugal
Malay Banerjee, Indian Institute of Technology Kampur, India
Maria Luísa Morgado, Universidade de Trás-os-Montes e Alto Douro, Portugal
Maria Clara Grácio, Universidade de Évora, Portugal
Maria do Rosário Ramos, Universidade Aberta, Portugal
Manuel L. Esquivel, Universidade Nova de Lisboa, Portugal
Manuel Branco, Universidade de Évora, Portugal
Manuela Oliveira, Universidade de Évora, Portugal
Milan Stehlík, Johannes Kepler University, Austria
Mouhaydine Tlemcani, Universidade de Évora, Portugal
Padmanabhan Seshaiyer, George Mason University, USA
Paulo Correia, Universidade de Évora, Portugal
Pedro Areias, Instituto Superior Técnico, Portugal
Pedro Lima, Instituto Superior Técnico, Portugal
Pedro Oliveira, Universidade do Porto, Portugal
Philippe Fraunié, Université de Toulon, France
Russell Alpizar-Jara, Universidade de Évora, Portugal
Sandra Ferreira, Universidade de Beira Interior, Portugal
Susana Faria, Universidade do Minho, Portugal
Tanuja Srivastava, Indian Institute of Technology, India
Teresa Oliveira, Universidade Aberta, Portugal
Tomas Bodnar, Czech Technical University in Prague, Czech Republic
Valter Vairinhos, Centro de Investigação Naval, Portugal
Victor Leiva, Pontificia Universidad Católica de Valparaíso, Chile

Sponsored by



Universidade do Minho
Escola de Ciências



Technical Specifications

Title

VII Workshop on Computational Data Analysis and Numerical Methods

Web-page

<http://www.wcdanm.ipt.pt/>

Editor

Instituto Politécnico de Tomar
Quinta do Contador – Estrada da Serra
2300-313 Tomar

Editors

Luís M. Grilo (Instituto Politécnico de Tomar)
Ana Nata (Instituto Politécnico de Tomar)
Manuela Fernandes (Instituto Politécnico de Tomar)
Isabel Pitacas (Instituto Politécnico de Tomar)
Fernando Carapau (Universidade de Évora)
A. Manuela Gonçalves (Universidade do Minho)
Teresa A. Oliveira (Universidade Aberta)

Authors

Many authors.

Published in a PDF format by:

Instituto Politécnico de Tomar
Copyright © 2020 left to the authors of individual papers
All rights reserved.

ISBN: 978-989-8840-47-9

Contents

Welcome to the VII WCDANM 2020	i
Committees	iii
Sponsors	v
Technical Specifications	vii
<hr/>	
Invited Speakers	
<hr/>	
Malay Banerjee <i>Emergence of stationary pattern in Rosenzweig-MacArthur model</i>	1
Sílvia Barbeiro <i>Mathematical models for elastography</i>	2
Indranil Ghosh <i>A class of skewed distributions with applications in environmental data</i>	4
Adélia Sequeira <i>Modeling and simulation of the cardiovascular system: a mathematical challenge</i>	5
<hr/>	
Course	
<hr/>	
Christian M. Ringle <i>Course on partial least squares structural equation modeling (PLS-SEM) using SmartPLS</i>	9
<hr/>	
Organized Sessions	
<hr/>	
Organized Session 1	
Additive models, discriminant analysis and asymptotic results	
Organizer: Dário Ferreira	13
<hr/>	
Isaac Akoto <i>Singular multinomial distribution and discriminant analysis</i>	15
Patrícia Antunes <i>Bi-additive models, adjustment, confidence ellipsoids and prediction intervals</i> ...	16
Dário Ferreira <i>Sums of squares in prime factorial designs</i>	17
Sandra Ferreira <i>Asymptotic results for the discrete case</i>	18

João T. Mexia	
<i>Asymptotic results for the discrete case - Part II</i>	19
<hr/>	
Organized Session 2	
Statistical application: price modeling, COVID-19 and artificial intelligence	
Organizer: Manuela Oliveira	20
<hr/>	
Manuela Oliveira	
<i>COVID-19: How to estimate the percentages of asymptomatic and immune individuals</i>	21
Andressa Contarato Rodrigues	
<i>Selecting the most efficient model in predicting financial resources</i>	22
Andressa Contarato Rodrigues	
<i>Image analysis using neural convolutional network in health</i>	24
Tuany Esthefany Barcellos de Carvalho Silva	
<i>Impacts caused by COVID-19 in the Brazilian educational sector: An application of exploratory factor analysis</i>	25
<hr/>	
Organized Session 3	
Statistical Theory and Applications	
Organizer: Carla Santos	27
<hr/>	
Isabel Borges	
<i>Statistical modelling of the performance of Portuguese granites exposed to salt mist: the cases of RA and SPI granites</i>	29
Ana Cantarinha	
<i>Multivariate collective risk models. Inference and special case</i>	31
Cristina Dias	
<i>Analysis of adaptability and production stability of common wheat genotypes</i>	32
Nadab Jorge	
<i>The likelihood ratio test of independence for random sample sizes - power studies and computational issues</i>	34
Carla Santos	
<i>On the optimization of unbiased linear estimators in models with orthogonal block structure</i>	36
Maria Varadinov	
<i>Multivariate APC model in the analyses of the logistics activities within agri-food supply chains</i>	38
<hr/>	
Posters included in the Organized Session 3	
<hr/>	
Carla Santos	
<i>Coefficient of variation: properties, bounds and monotony</i>	43

Cristina Dias	
<i>Basic models with orthogonal structure</i>	45
Carla Santos	
<i>Comparing COVID-19 case fatality ratios through Marascuilo Procedure</i>	46
Maria Varadinov	
<i>The need to manage a RL flow to the traditional forward flow</i>	48
<hr/>	
Organized Session 4	
Computational Models and Machine Learning Approaches for solving Real-world Applications	
Organizer: Padmanabhan Seshaiyer	50
<hr/>	
Viraj Boreda	
<i>Efficient image segmentation and machine learning algorithms for improved malaria detection from blood smear images</i>	51
Udbhav Muthakana	
<i>Deep learning with neural networks for airborne spread of COVID-19 in enclosed spaces</i>	52
Padmanabhan Seshaiyer	
<i>Computational mathematics for solving real-world problems arising from COVID-19</i>	53
Diego Valencia	
<i>Computational techniques using machine learning for rehabilitation and relapse-detecting technology to assist recovering Opioid addiction patients</i>	54
<hr/>	
Organized Session 5	
Statistical Modeling	
Organizer: Milan Stehlík	55
<hr/>	
M. Ivette Gomes	
<i>Generalized means: progresses and challenges in statistics of extremes</i>	56
Luís M. Grilo	
<i>Comparison of different estimators for a reflective SEM: burnout syndrome in Portuguese workers</i>	58
Pavlina Jordanova	
<i>Ratios of order statistics and regularly varying tails</i>	60
M. Stehlík	
<i>Modelling and prediction of COVID-19 outbreaks</i>	62
<hr/>	
Contributed Talks	
<hr/>	
Ana Borges	
<i>The impact of entrepreneurial framework conditions on entrepreneurship early activities: an European longitudinal study</i>	65

Carlos A. Braumann	
<i>Harvesting policies with stepwise effort in random environments</i>	67
Mariana Cardante	
<i>A cluster analysis of the business behavior and attitudes indicators defined by global entrepreneurship monitor</i>	69
Aldina Correia	
<i>The role of framework conditions in income level of countries</i>	71
Sérgio Cavaleiro Costa	
<i>A novel predator-prey strategy for optimization problems: preliminary study</i>	73
Sónia Dias	
<i>Clusterwise regression for interval data</i>	75
Manuel L. Esquível	
<i>A wavelet based neural network scheme (NNS) for supervised and unsupervised learning</i>	77
Susana Faria	
<i>Bivariate multilevel modelling of Portuguese student achievement</i>	79
Jerzy K. Filus	
<i>General theory and construction of k-dimensional survival functions</i>	80
Lidia Z. Filus	
<i>Theoretical aspects and constructions of bivariate probability distributions, most general case</i>	81
Raquel Francisco	
<i>Factors influencing exports and season sales tend in a footwear industry</i>	82
Sergey Frenkel	
<i>Ontological and probabilistic aspects of assessing the quality of predictors.</i>	84
Carlos Gomes	
<i>The role of perceptual variables and country-level culture on international entrepreneurship</i>	87
A. Manuela Gonçalves	
<i>TBATS models and linear regression models approaches for forecasting daily weather time series</i>	89
Gonçalo Jacinto	
<i>Individual growth modelling with stochastic differential equations</i>	91
Emanuel Lopes	
<i>A preliminary study on the transportation of non-urgent patients by a fire station</i>	93
Erika Johanna Martínez Salinas	
<i>Stochastic model to analyze the changes in the fisheries catch and species composition due to varying oceanic temperature along the Colombian Pacific coast</i>	95

João Mendes	
<i>Deep learning method to identify fire ignitions</i>	96
Telma Mendes	
<i>How does organizational ambidexterity influence the firms' internationalization speed? The contribution of network clustering</i>	98
Oumaima Mesbahi	
<i>New approach of total least square algorithm for nonlinear models</i>	100
Fernanda Pereira	
<i>Short-term forecast models for the maximum temperature time series</i>	102
J. A. Lobo Pereira	
<i>Modelling complex relationships in dentistry - The GAMLSS approach</i>	104
Sara Perestelo	
<i>Modelling spreading process of a wildfire in heterogeneous orography, fuel distribution and environmental conditions - a complex networks approach</i>	106
Oussama Rida	
<i>A stochastic diffusion process based on brody curve with exogenous factors</i>	108
Andrés Ríos-Gutiérrez	
<i>Modeling analysis and numerical estimation of a stochastic epidemic SEIR model with environmental stochasticity</i>	109
Marta Santos	
<i>Detection analysis of breaks in water consumption patterns: a simulation study</i>	111
Inês Sena	
<i>Trends identification in medical care</i>	113
Eliana Costa e Silva	
<i>European economies clustering based on the entrepreneurial framework conditions</i>	115
Rosa Silveira	
<i>A new RFM approach for customer segmentation using a SAF-T based business intelligence system</i>	117
A. M. Simões	
<i>A new kind of stability for the Bessel equation: the σ-semi-Hyers-Ulam stability</i>	119
Marta Teixeira	
<i>Cooperation, innovation and environmental sustainability: Portuguese companies research</i>	121
M. Filomena Teodoro	
<i>Using a multidimensional statistical approach to select experts opinion in a context of naval operations</i>	123

Fátima De Almeida	
<i>Minimize the production of scrap in the extrusion process</i>	127
Ana Cantarinha	
<i>Structured families of multivariate collective models</i>	129
Fernando Carapau	
<i>One-dimensional theories to study three-dimensional problems: theoretical and numerical aspects</i>	130
M. Ivette Gomes	
<i>Extreme value index estimation under non-regularity and through generalized means</i>	131
A. Manuela Gonçalves	
<i>Comparison of time series models' performance applied to economic data forecasting</i>	133
Conceição Leal	
<i>Geographically weighted panel logistic regression to model COVID-19 data</i>	135
Teresa A. Oliveira	
<i>Control charts for attribute control based on life distributions with applications on e-learning classes monitoring</i>	137
M. Filomena Teodoro	
<i>Mechanical behavior of the skin: A Statistical Approach</i>	138
Ricardo Vitorino	
<i>The global satisfaction of Portuguese expatriate workers: application of PLS-SEM</i>	139
Index of authors	141

Invited Speakers

Emergence of stationary pattern in Rosenzweig-MacArthur model

Malay Banerjee

Department of Mathematics & Statistics
IIT Kanpur, Kanpur, India

E-mail address: *malayb@iitk.ac.in*

The spatio-temporal Rosenzweig-MacArthur (RM) model with self-diffusion terms only is capable of producing a wide range of dynamic patterns, namely traveling wave, periodic traveling wave, spatio-temporal chaos. The RM type reaction kinetics is suitable for a wide range of resource-consumer type population interactions where the constituent species exhibit stationary distribution over their habitat. However, the spatio-temporal RM model fails to produce stationary Turing patterns as the requisite conditions for Turing instability are not satisfied. A slightly modified version of the spatio-temporal RM model is capable of producing a stationary Turing pattern. The main objective of this talk is to explain how Turing pattern can emerge for spatio-temporal RM model by incorporating (i) nonlocal consumption of resources by prey and (ii) density dependent cross-diffusion of both the population. These mechanisms are responsible for generation of Turing patterns for a wide range of prey-predator models.

Keywords

Pattern Formation, Turing Instability, Nonlocal Interaction.

References

- [1] M. Banerjee and V. Volpert. Spatio-temporal pattern formation in Rosenzweig-MacArthur model: Effect of nonlocal interactions. *Eco. Comp.* **30**: 2–10, 2017.
- [2] M. Banerjee and V. Volpert. Prey-predator model with a nonlocal consumption of prey. *Chaos*, **26**: 083120, 2016.
- [3] M. Banerjee and L. Zhang. Stabilizing Role of Nonlocal Interaction on Spatio-temporal Pattern Formation. *Math. Model. Nat. Phenom.* **11(5)**: 135–150, 2016.
- [4] N. Mukherjee, S. Ghorai and M. Banerjee. Cross-diffusion induced Turing and non-Turing patterns in Rosenzweig-MacArthur model. *Lett. Biomath.* **6(1)**: 1–22, 2019.

Mathematical models for elastography

Sílvia Barbeiro

University of Coimbra, CMUC, Department of Mathematics, Coimbra, Portugal

E-mail address: *silvia@mat.uc.pt*

In this talk we discuss a mathematical model to reconstruct the mechanical properties of an elastic medium, for the elastography imaging technique. We start by addressing the numerical simulation of the mechanical wave propagation and induced displacements. This direct problem is the computational basis to solve the inverse problem which consists of determining the set of parameters that characterize the mechanical properties of the medium, knowing the displacement fields for a given excitation.

Keywords

Elastography, Coupled Acoustic Wave Propagation, Elastic Displacement Model, Inverse Problem.

The investigation reported in this talk arises in the framework of ElastoOCT project [1] where the development an optical coherence elastography (OCE) technique for imaging in vivo the mechanical properties of the retina is proposed. OCE is an emerging medical imaging modality created to visualize tissue elasticity non-invasively. This procedure combines mechanical excitation of the medium with optical coherence tomography (OCT) for measuring the corresponding elastic displacement. When using acoustic loading, an ultrasound source is coupled with an OCT device to this end. Usually, the mechanical properties of pathological tissue are distinct from those of healthy tissue, therefore elastography opens the prospect of the discovery of biomarkers for the early detection, before clinical manifestations, of neurodegenerative processes.

We are interested in applying OCE to the human retina. The mathematical simulation of this process includes the propagation of the acoustic wave from the source through the eye, the interaction of the acoustic pressure to generate an elastic wave in the retina and the propagation of the elastic wave in the retina, namely the induced displacements in the retinal tissue [2]. Having in mind the application, we investigate a process for obtaining the mechanical properties of the retina given the displacement field, that is, to solve the inverse problem of elastography. In our approach [3], we formulate the inverse problem as an optimization program, using mathematical model for solving the direct problem. The inverse problem consists of determining the set of parameters that characterize the mechanical properties of the medium knowing the displacement field for a given excitation. In practice, the objective is to infer the parameters that characterize the mechanical properties so that the difference between simulated displacements obtained with the mathematical model for the direct problem with those parameters and the data are minimized.

We present several computational results and discuss the performance of the methods.

Acknowledgements: This work was partially supported by the Centre for Mathematics of the University of Coimbra - UIDB/00324/2020, funded by the Portuguese Government through FCT/MCTES, and by FCT (Portugal) research project PTDC/EMD-EMD/32162/2017, COMPETE and Portugal2020.

References

- [1] A. M. Morgado, A. Batista, S. Barbeiro, R. Bernardes, J. M. Cardoso, C. Correia, J. Domingues, R. Henriques, C. Loureiro, M. Santos and P. Serranho. Optical coherence elastography for imaging retina mechanical properties, FCT project. <http://miguelmorgado.net/research/projects/on-going/elastoocct.html>
- [2] S. Barbeiro and P. Serranho. *The Method of Fundamental Solutions for the Direct Elastography Problem in the Human Retina*. Proceedings of the 9th Conference on Tretz Methods and 5th Conference on Method of Fundamental Solutions, SEMA SIMAI Springer Series 23, Chapter 5, 2020.
- [3] R. Henriques. *Mathematical model to reconstruct the mechanical properties of an elastic medium* (Master's thesis), Department of Mathematics, University of Coimbra, July 2020.

A class of skewed distributions with applications in environmental data

Indranil Ghosh¹ and Hon Keung Tony Ng²

¹University of North Carolina, Wilmington, North Carolina, USA

²Southern Methodist University, Dallas, Texas, USA

E-mail addresses: *ghoshi@uncw.edu; ngh@mail.smu.edu*

Keywords

Maximum Likelihood, Moments, Monte Carlo Simulation, Skewed Distribution, Truncation.

In environmental studies, many data are typically skewed and it is desired to have a flexible statistical model for this kind of data. In this article, we study a class of skewed distributions by invoking skewness into some known distributions. In particular, we consider using the logistic kernel to derive a class of univariate distribution called the truncated-logistic skew symmetric (TLSS) distribution. We provide some structural properties of the proposed distribution and develop the statistical inference for the TLSS distribution. A simulation study is conducted to investigate the efficacy of the maximum likelihood method. For illustrative purposes, two real data sets from environmental studies are used to exhibit the applicability of such a model.

References

- [1] B. C. Arnold and R. J. Beaver. The skew-Cauchy distribution. *Statistics and Probability Letters* **49**: 285–290, 2000.
- [2] B. C. Arnold and R. J. Beaver. Skewed multivariate models related to hidden truncation and/or selective reporting (with discussion). *Test* **11**: 7–54, 2002.
- [3] A. Azzalini and A. Dalla Valle. The multivariate skew-normal distribution. *Biometrika* **83**: 715–726, 1996.
- [4] A. Azzalini and A. Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society, Series B*, **61**: 579–602, 1999.
- [5] J. T. A. S. Ferreira and M. F. J. Steel. A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association* **101**: 823–829, 2006.
- [6] M. Hallin and C. Ley. Skew-symmetric distributions and Fisher information—a tale of two densities. *Bernoulli* **18**: 747–763, 2012.

Modeling and simulation of the cardiovascular system: a mathematical challenge

Adélia Sequeira

Departamento de Matemática. Instituto Superior Técnico, Universidade de Lisboa,
Portugal
Centro de Matemática Computacional e Estocástica (CEMAT- IST/ULisboa), Portugal

E-mail address: *adelia.sequeira@tecnico.ulisboa.pt*

Cardiovascular diseases, such as heart attack and strokes, are the major causes of death in developed countries, with a significant impact in the cost and overall status of healthcare. Understanding the fundamental mechanisms of the pathophysiology and treatment of these diseases are matters of the greatest importance around the world. This gives a key impulse to the progress in mathematical and numerical modeling of the associated phenomena governed by complex physical laws, using adequate and fully reliable *in silico* settings.

In this talk we describe some mathematical models of the cardiovascular system and comment on their significance to yield realistic and accurate numerical results, using stable, reliable and efficient computational methods [1]. Results on the simulation of some image-based patient-specific clinical cases will also be presented [2].

Acknowledgements: Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects UIDB/04621/2020, UIDP/04621/2020 (CEMAT-IST/ULisboa) .

References

- [1] A. Fasano and A. Sequeira. Hemomath - The Mathematics of Blood. MS & A - Modeling, Simulation and Applications Series, Springer Verlag, ISBN: 978-3-319-60512-8, 2017.
- [2] V. G. Ardakani, X. Tu, A. M. Gambaruto, I. Velho, J. Tiago, A. Sequeira and R. Pereira. Near – Wall Flow in Cerebral Aneurysms. *Fluids* **4(2)**: 89, 2019.

Course

Course on partial least squares structural equation modeling (PLS-SEM) using SmartPLS

Christian M. Ringle

(TUHH) Hamburg University of Technology, Germany

E-mail address: *c.ringle@tuhh.de*

This online course is based on the book "A Primer on Partial Least Squares Structural Equation Modeling" by Joe Hair, Tomas Hult, Christian M. Ringle, and Marko Sarstedt.

Course topics:

- Foundations of structural equation modeling (SEM) and partial least squares SEM (PLS-SEM)
- Introduction to PLS-SEM and the SmartPLS software
- Assessing measurement model results (reflective and formative) & SmartPLS exercises
- Assessing structural model results & SmartPLS exercises
- Advanced model evaluation: Prediction-oriented assessment of PLS-SEM results
- Advanced model evaluation: Importance performance map analysis (IPMA) & SmartPLS exercises

Organized Sessions

Organized Session 1

Additive models, discriminant analysis and asymptotic results

Organizer: Dário Ferreira

Singular multinomial distribution and discriminant analysis

Isaac Akoto^{1,2}, João T. Mexia^{1,3}, Dário Ferreira^{4,5} and
Gracinda R. Guerreiro^{1,3}

¹FCT NOVA, Universidade NOVA de Lisboa, Campus de Caparica, Portugal

²Department of Mathematics and Statistics, University of Energy and Natural
Resources, Sunyani, Ghana

³Center of Mathematics and its Applications, FCT NOVA, Universidade NOVA de
Lisboa, Campus de Caparica, Portugal

⁴Department of Mathematics, University of Beira Interior, Covilhã, Portugal

⁵Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal

E-mail addresses: *i.akoto@campus.fct.unl.pt; jtm@fct.unl.pt; grg@fct.unl.pt; dario@ubi.pt*

Applying Discriminant Analysis to discrete variables usually leads to an enormous number of possible states if compared to the number of objects under study. This makes the models have a low performance due to the high number of parameters to be estimated. Moreover, usually an incorrect allocation of an element to a certain class has associated costs.

In this work, we address the problem of classifying an individual into one of several populations by considering “statistical decision functions” and derive a two-phase allocation rule, which minimizes the average allocation cost. This approach enables us to avoid having to consider a great number of cases. We apply our results to HIVDR to classify patients into two *a priori* defined classes of treatment.

Keywords

Decision Theory, Discrimination, HIVDR

Acknowledgements: This work was partially supported by the Portuguese Foundation for Science and Technology through the projects UIDP/MAT/00212/2020 and UIDP/MAT/00297/2020.

References

- [1] P. J. Brown, T. Fearn and H. S. Haque. Discrimination with many variables. *Journal of the American Statistical Association* **94**: 448, 1320–1329, 1999.
- [2] P. L. Cooley Bayesian and Cost. Considerations for Optimal Classification with Discriminant Analysis. *The Journal of Risk and Insurance*, 277–287, 1975.
- [3] L. Flury, B. Bourkai and B. D. Flury. The discrimination subspace model. *Journal of the American Statistical Association* **92(438)**: 758–766, 1997.
- [4] M. Lavine and M. West. A Bayesian method for classification and discrimination. *Can J Stat* **20(4)**: 451–461, 1992.
- [5] S. S. Ferreira and D. Ferreira and C. Nunes and J. T. Mexia. Discriminant Analysis and Decision Theory. *Far East Journal of Mathematical Sciences* **51(1)**: 69–79, 2011.
- [6] S. Srivastava, M. R. Gupta and B. A. Frigyk. Bayesian Quadratic Discriminant Analysis, *Journal of Machine Learning Research* **8**: 1277–1305, 2007.

Bi-additive models, adjustment, confidence ellipsoids and prediction intervals

Patrícia Antunes¹, Sandra S. Ferreira^{1,2}, Dário Ferreira^{1,2} and João T. Mexia³

¹Center of Mathematics and Applications, University of Beira Interior, Covilhã, Portugal

²Department of Mathematics, University of Beira Interior, Covilhã, Portugal

³Center of Mathematics and its Applications, Faculty of Science and Technology, New University of Lisbon, Monte da Caparica, Portugal

E-mail addresses: *antunes.patricia03@gmail.com; sandraf@ubi.pt; dario@ubi.pt; jtm@fct.unl.pt*

Models given by

$$\underline{Y} = X\underline{\beta} + \sum_{i=1}^w X_i \underline{Z}_i,$$

where the \underline{Z}_i are independent, with independent and identically distributed components, are bi-additive models, since their covariance matrices are the sum of the corresponding matrices for the $X_i \underline{Z}_i, i = 1, \dots, w$. This structure enable the use of cumulants in the estimation of model parameters.

Besides this Edgeworth Expansions were obtained and used to derive Confidence Ellipsoids for the model distribution and Prediction Intervals for "future" observations.

Keywords

Cumulants, Inference, Mixed Models.

Acknowledgements: This work has received funding from the Portuguese Foundation for Science and Technology under the project UIDB/MAT/00212/2020 and UIDB/MAT/00297/2020.

References

- [1] C. C. Craig. On A Property of the Semi-Invariants of Thiele. *Ann. Math. Statist.* **2**: 154–164, 1931.
- [2] A. DasGupta. *Edgeworth Expansions and Cumulants. In: Asymptotic Theory of Statistics and Probability.* Springer Texts in Statistics, Springer, New York, 2008.
- [3] F. Ramos. Burnout: um terço dos inquiridos em risco, <https://www.deco.proteste.pt/saude/doencas/noticias/burnout-um-terco-dos-portugueses-em-risco>, 2018, accessed on 03.Feb.2020.

Sums of squares in prime factorial designs

Dário Ferreira¹, Sandra Ferreira¹, Célia Nunes¹ and João T. Mexia²

¹Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Portugal

²Centre of Mathematics and its Applications, FCT, New University of Lisbon, Portugal

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

Prime factorial designs are p^n designs, where p is a prime number or a power of a prime. Since [7], these designs have been studied over the years. An example of a recent study is [3], where results about tests for hypotheses, in models with orthogonal block structure, were applied to prime factorial designs. These kind of models have the advantage of allowing the simultaneous study of a larger number of interactions compared to classical approaches. In this work we show how to use recurrence relations in order to obtain the sums of squares, for cases with any number of factors and levels, even when it is neither a prime or a power of a prime. An example is included.

Keywords

Inference, Mixed Models, Recurrence Method.

Acknowledgements: This work was partially supported by the Portuguese Foundation for Science and Technology, FCT, through the projects UIDB/MAT/00212/2020 and UIDB/MAT/00297/2020.

References

- [1] R. A. Bailey, S. S. Ferreira, D. Ferreira and C. Nunes. Estimability of Variance Components when all Model Matrices Commute, *Linear Algebra and its Applications* **492**, 144–160, 2015.
- [2] F. Carvalho. Strictly associated models, prime basis factorials: an application. *Discussiones Mathematicae Probability and Statistics* **31**: 77–86, 2011.
- [3] D. Ferreira, S. S. Ferreira, C. Nunes and J. T. Mexia. Tests and relevancies for the hypotheses of an orthogonal family in a model with orthogonal block structure. *Journal of Statistical Computation and Simulation* **90(3)**: 412–419, 2020.
- [4] J. T. Mexia. Standardized Orthogonal Matrices and the Decomposition of the sum of Squares for Treatments. *Trabalhos de Investigação 2*, Departamento de Matemática FCT-UNL, 1988.
- [5] J. A. Nelder. 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, 1965.
- [6] J. A. Nelder. 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, 1965.
- [7] F. Yates. The Design and Analysis of Factorial Experiments. Technical Communication 35. Harpenden: Imperial Bureau of Soil Science, 1937.

Asymptotic results for the discrete case

Sandra Ferreira¹, Isaac Akoto², Célia Nunes¹, Dário Ferreira¹ and
João T. Mexia³

¹Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Portugal

²FCT NOVA, Universidade NOVA de Lisboa, Campus de Caparica, Portugal

³Center of Mathematics and its Applications FCT, New University of Lisbon, Portugal

E-mail addresses: *sandraf@ubi.pt; celian@ubi.pt; dario@ubi.pt; i.akoto@campus.fct.unl.pt; jtm@unl.fct.pt*

Given an experiment with m possible results, the corresponding samples follow the discrete multinomial distribution. We present for these distributions normal asymptotics distributions. We show how to write quadratic forms as inner products in order to obtain confidence ellipsoids and simultaneous confidence intervals for the vector of the results probabilities. We apply our approach to inference on single and pair of samples.

Keywords

Asymptotics Distributions, Confidence Ellipsoids, Inference.

Acknowledgements: This work was partially supported by the Portuguese Foundation for Science and Technology, FCT, through the projects UIDB/MAT/00212/2020 and UIDB/MAT/00297/2020.

References

- [1] O. Kallenberg. *Foundations of modern probability*. 2nd Edn. Springer-Verlag, New York, 2002.
- [2] J. R. Schott. *Matrix analysis for statistics*, 3rd ed. Wiley series in probability and mathematical statistics, Wiley, 2017.
- [3] F. Ramos. Burnout: um terço dos inquiridos em risco, <https://www.deco.proteste.pt/saude/doencas/noticias/burnout-um-terco-dos-portugueses-em-risco>, 2018, accessed in 03.Feb.2020.

Asymptotic results for the discrete case - Part II

João T. Mexia¹, Isaac Akoto², Sandra Ferreira³, Célia Nunes³ and
Dário Ferreira³

¹Center of Mathematics and its Applications, FCT, New University of Lisbon, Portugal

²FCT NOVA, Universidade NOVA de Lisboa, Campus de Caparica, Portugal

³Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Portugal

E-mail addresses: *jtm@unl.fct.pt; i.akoto@campus.fct.unl.pt; sandraf@ubi.pt; celian@ubi.pt; dario@ubi.pt*

Continuing the study of inference we consider structured families of models. The samples are assumed to be independent and corresponding to the same set of results. In the pairs of samples we compare homologue probabilities and linear combinations of probabilities but we go further and we will consider samples of a structured family that corresponds to the treatments of a fixed effects model. We study the action of the factors in the base design, on linear combinations of the results probabilities.

Keywords

Asymptotics Distributions, Inference, Structured Family.

Acknowledgements: This work was partially supported by the Portuguese Foundation for Science and Technology, FCT, through the projects UIDB/MAT/00212/2020 and UIDB/MAT/00297/2020.

References

- [1] O. Kallenberg. *Foundations of modern probability*. 2nd Edn., Springer-Verlag, New York, 2002.
- [2] J. R. Schott. *Matrix analysis for statistics*, 3rd ed, Wiley series in probability and mathematical statistics, Wiley, 2017.
- [3] F. Ramos. Burnout: um terço dos inquiridos em risco, <https://www.deco.proteste.pt/saude/doencas/noticias/burnout-um-terco-dos-portugueses-em-risco>, 2018, accessed in 03.Feb.2020.

Organized Session 2

Statistical application: price modeling, COVID-19
and artificial intelligence

Organizer: Manuela Oliveira

COVID-19: How to estimate the percentages of asymptomatic and immune individuals

Manuela Oliveira^{1,2}, João T. Mexia^{4,5}, Eugénio Garção³ and
Luís M. Grilo^{5,6,7,8}

¹Universidade de Évora, Departamento de Matemática, Portugal

²Centro de Matemática e Aplicações (CIMA), Portugal

³Universidade de Évora, Departamento de Física, Portugal

⁴Universidade Nova de Lisboa, Portugal

⁵Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal

⁶Instituto Politécnico de Tomar (IPT) & Universidade Aberta, Portugal

⁷Centro de Investigação em Cidades Inteligentes (Ci2), IPT, Portugal

⁸CIICESI, ESTG, Politécnico do Porto, Portugal

E-mail address: *mmo@uevora.pt; jtm@unl.fct.pt; jesg@uevora.pt; lgrilo@ipt.pt*

Epidemiology focuses on the identification of patterns in disease occurrence in order to provide information that may be useful to help prevent it. The research of how a disease may be transmitted is influential for that identification. In the case of COVID-19 this research should focus on the relations between the disease and the populations of susceptible individuals that might be infected. In this context, we propose an approach based on a stratified sampling scheme to estimate the percentages of asymptomatic and immune persons per region in Portugal.

Keywords

COVID-19, Asymptomatic Percentage, Immune Percentage, Sampling Design.

Acknowledgements: Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/04674/2020 (CIMA).

References

- [1] Entidade Reguladora da Saúde (ARS). *Estudo sobre as Unidades de Saúde Familiar e as Unidades de Cuidados de Saúde Personalizados*, 2016.
- [2] J. L. Fernandes, V. L. Costa, P. Oliveira and A. Almeida. Impact of patient panel size on the quality of care provided for family physicians . *Acta Médica* Aceite para publicação, 2020.
- [3] W. G. Cochran, *Sampling Techniques 3rd Edition*. John Wiley and Sons, New York. 1977.

Selecting the most efficient model in predicting financial resources

Andressa Contarato Rodrigues¹, Pedro Ivo Rodrigues¹ and
Marco Aurélio dos Santos Sanfins¹

¹Universidade Federal Fluminense, Brazil

E-mail addresses: *andressacontarato@gmail.com; pir.pedro@gmail.com; marcosanfins@id.uff.br*

Finance encompasses a set of processes, markets, and institutions, among others, with the purpose of transferring funds between people, companies and even the Government. Given its importance, it is always necessary to use sophisticated analysis and complex algorithms to understand the market, in order to predict future conditions that may affect from the life of a person to an entire country. Models widely used in this field come from the area of Time Series [4]. However, an area gaining great importance is that of Artificial Intelligence. This is used in several scientific articles, and it is being more and more applied, with the advent of large volumes of data, the Big Data. Within Artificial Intelligence, there is the branch of Machine Learning, and in this, Deep Learning. Machine Learning has been widely used in Finance, as well as Artificial Intelligence and its history is old. Machine Learning has become an essential technique in the area of Finance, due to the growing number of data used, such as historical records by year, month, day, hours and even seconds. Thus, the purpose of this work is to make use of models from the area of Time Series and Artificial Intelligence and compare their performance.

Keywords

Finance, Time Series, Machine Learning, Financial Market.

Objective: Identify which model best fits the resource data, looking at the RMSE and its accuracy, to finally predict the best time series. Analyzing the nature of the series used in the study, the bibliographic survey of the models, their construction and application, to finally compare the results, with the possible choice of one or more models for each series approached.

Financial Resources: Share is defined as the smallest portion of capital stock of companies or even corporations. It can be considered an equity title, granting the shareholders all rights and duties of a partner, within the limit of the shares held [1]. Another type of financial resource that has been gaining strength is cryptocurrencies. With the advent of technology and the most varied forms of payment and receipt of money, the so-called cryptocurrencies were created.

Time Series: Monitoring resources in the financial market can be made daily, weekly, etc. Understanding how an asset will behave tomorrow, or after a certain cycle, is of great relevance for investors. This idea of monitoring observations in a certain period of time is called time series [2]. Regarding the definition of time series, stochastic processes can be defined as a set of observations ordered by time [4]. As for the models created by Box and Jenkins (1970), such linear models were designed to capture autocorrelation and create models with one part showing how much the past affects the present and another

part, demonstrating how much errors can affect the present. These, ARIMA, SARIMA ... When they were created, the objective of autoregressive models with conditional heteroscedasticity was to estimate the variability of inflation. Examples of these models are the GARCH.

Artificial intelligence: Artificial intelligence (AI) is divided into two more parts: Machine Learning and Deep Learning. As the names show, the first defines machine learning and the second, deep learning, in the field of neural networks. In this work, the machine learning model called Support Vector Regression (SVR) was used, whose objective is to trace hyperplanes in order to stipulate an optimal function for the data. The Artificial Neural Network (ANN), in the field of Deep Learning, is a structure based on a biological neuron, called multilayer perceptron, with 3 layers, one of them being hidden [1].

Database: Database was composed by 6 different types of actions: MSFT: Microsoft; SBUX: Starbucks; IBM: IBM company; AAPL: Apple; GSPC: Active; AMZN: Amazon. And, three different types of cryptocurrencies, namely: BTC-USD: Bitcoin; ETH-USD: Ethereum; and LTC-USD: Litecoin.

Results: The results analyzed show that the AI models had a better performance. In the asset series, the SVR was the best, while in the Cryptocurrencies series, the ANN stood out as the best fit.

References

- [1] BTG Pactual, Carteira de investimentos: o que é, como montar e diversificar. Available at www.btgpactualdigital.com Accessed: 06/01/2020.
- [2] P. J. Brockwell, A. D. Richard and V. C. Matthew. Introduction to time series and forecasting. Vol. 2. New York: Springer, 2002.
- [3] A. Carvalho, K. Faceli, A. Lorena and J. Gama . Inteligência Artificial - uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC., 2011.
- [4] P. A. Morettin and C. M. Toloi. Análise de séries temporais: modelos lineares univariados. Editora Blucher, 2018.

Image analysis using neural convolutional network in health

Andressa Contarato Rodrigues¹ and Gleici da Silva Castro Perdoná¹

¹Universidade de São Paulo, Brazil

E-mail addresses: *andressacontarato@gmail.com; pgleici@fmrp.usp.br*

Artificial Intelligence, previously considered only as a theoretical area with little applicability, has been developing with the advent of Big Data and powerful computers. One of the techniques within the Artificial Intelligence area is Deep Learning, which contains studies in the area of so-called Neural Networks. A Neural Network is based on a biological neuron, which becomes a basic unit of the human brain, being a special cell in the transmission of instructions. One of the neural networks that has been designed with great efficiency in its application in several areas of knowledge are the Convolutional Neural Networks - CNN [2].

Keywords

Artificial Intelligence, Deep Learning, Convolutional Neural Network, Health, Respiratory Diseases.

Objective: Although, according to the WHO [3], the disease called pneumonia is a type of acute respiratory infection that mainly affects the lungs. The aim of this work is to detect and differentiate a normal patient from one who has respiratory disease. Pneumonia can, in the current scenario, be confused with Covid-19 and influence the patient's condition.

Methodology: The structure of a CNN is a neural network that uses layers with a convolution filter [2]. These neural networks showed results both in the field of computer vision and in several problems of Natural Language Processing (NLP). The neural network starts with the input, inside it there are the convolution layers, each convolutional layer tries to understand the basic patterns. In the first, the network tries to learn patterns and borders of the images, in the second, it tries to understand the shape/color and so on.

Database: Initially, we will consider the Chest-X Ray Images base present on the Kaggle platform [4], to develop the structure of the CNN. The base is composed of a set of images, these from chest radiography (anteroposterior) that were selected from retrospective cohorts of pediatric patients from one to five years old at the Guangzhou Women and Children Medical Center.

Results: Partial results were obtained, with 91.3% accuracy in detecting and identifying pneumonia compared with x-ray images of patients without respiratory diseases. And it is on process to improve.

References

- [1] Katti et al. Inteligência Artificial: Uma abordagem de aprendizado de máquina. 2011.
- [2] Yoon. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882, 2014.
- [3] World Health Organization. Available at https://www.who.int/health-topics/coronavirus/#tab=tab_1. Accessed: 05/07/2020.
- [4] Kaggle, Rio de Janeiro. Available at: <https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia>. Accessed: 08/10/2020.

Impacts caused by COVID-19 in the Brazilian educational sector: An application of exploratory factor analysis

Tuany Esthefany Barcellos de Carvalho Silva¹, Isabela da Costa Granja¹, Daiane Rodrigues dos Santos², Marco Aurélio Sanfins¹ and Pablo Silva Machado Bispo dos Santos¹

¹Universidade Federal Fluminense, Brazil

²Universidade Candido Mendes, Brazil

E-mail addresses: *tuanybarcellos@id.uff.br; isabelacg@id.uff.br; daianesanoseco@gmail.com; marcosanfins@id.uff.br; psmbasantos@gmail.com*

The year 2020 was marked by the beginning of a pandemic caused by the new coronavirus (COVID-19). This virus with easy and fast contagion impacted several sectors around the world, as well as countless countries, and Brazil seeks to minimize such impacts. Due to the ease of contamination, restrictive measures needed to be adopted, such as social distancing. The educational sector was severely affected, and it was extremely necessary to implement measures that would allow the school year to continue. That said, applying factor analysis, this work aims to measure the impacts caused by COVID-19 in the educational scope. The data were collected through a virtual questionnaire, which made it possible to analyse the effects of such scenario on the lives of education professionals. The results were satisfactory, showing factors that directly influence the performance of professionals during this challenging period.

Keywords

Coronavírus, Education, Educational Impact.

At the end of 2019, a new virus called Coronavirus (COVID-19) spread across Wuhan (China) a disease with fast and easy contagion that became the latest global threat (SOUTO, 2020) [1]. On December 31, the World Health Organization (WHO) was notified of the occurrence of an outbreak of pneumonia in the city of Wuhan caused by the new coronavirus (SARS-COV-2). A disease that challenges health professionals, researchers and government leaders, something unknown that quickly impacted the Chinese population, despite the adoption of some restrictive measures in 2020. The disease spread to many countries, and as of February of the same year, there were already approximately 80 thousand cases disease and 2,838 deaths in 53 countries (CRODA, 2020) [2]. It did not take long for the disease to take on overwhelming proportions in the world, arriving in Brazil, with its first case confirmed on February 26, 2020(CRODA, 2020)[2]. The country was on the alert and measures needed to be taken to minimize contamination of the population, thus requiring the implementation of social distancing. This resulted in the closure of industries, businesses, schools, leisure areas and everything not considered essential, in order to try to reduce the number of cases. Such restrictive measures directly impacted several sectors such as economy and education, among others. The educational sector was extremely affected, requiring the implementation of a plan that would allow the continuation of the school year, but the time for planning was short, and immediate measures needed to be applied. As an initial proposal, the July vacations were brought forward,

but long-term measures were necessary. After this period of early recess, education professionals together with government leaders proposed the use of technology, already used in distance learning (DE) , as a tool to help students and teachers, making possible through remote classes the continuity of the school period. The use of technology enabled the continuity of classes online, becoming essential to mitigate the consequences of the pandemic in the educational field. The scenario is still challenging for students and teachers, more functions were added to the roles of professionals, as it is necessary to rethink and innovate teaching methods, in order to guarantee a quality education that meets the academic demand. That said, the present work aims to measure, through descriptive statistics and statistical methods, such as factor analysis [3] and chi-square test, the impact of the pandemic caused by the new coronavirus on the lives of educators. This study allows us to observe which factors most impacted the lives of professionals of education and what are the effects on the way of teaching. The analysis is conducted by a total sample of 107 education professionals. The research was developed through a virtual questionnaire, the data collection period took place from April to August 2020, and the sample inclusion criteria was to be a professional in the educational sector. For a better analysis of the collected data, they were coded and categorized. With the application of factor analysis, it was possible to identify that students and professionals are striving to maintain a good academic performance, but factors such as the lack of technological resources and an increase in the student workload is an additional obstacle for the already complex and challenging scenario.

References

- [1] J. H. R. Croda and L. P. Garcia. Resposta imediata da vigilância em saúde à epidemia da covid-19. Editorial, Epidemiol. Serv. Saúde, Brasília, 2020.
- [2] B. F. Damásio. Uso da análise fatorial exploratória em psicologia. Universidade Federal do Rio Grande do Sul, 11, 2012.
- [3] X. M. Souto. Covid-19: Aspectos gerais e implicações globais. Revista de Educação, Ciência e Tecnologia de Almenara/ MG, 2020.

Organized Session 3

Statistical Theory and Applications

Organizer: Carla Santos

Statistical modelling of the performance of Portuguese granites exposed to salt mist: the cases of RA and SPI granites

Isabel Borges¹, Cristina Dias², Carla Santos³ and Maria Varadinov⁴

¹Polytechnic Institute of Portalegre and Interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

²Polytechnic Institute of Portalegre and Center of Mathematics and Applications (CMA), Caparica, Portugal

³Polytechnic Institute of Beja and Center of Mathematics and Applications (CMA), Caparica, Portugal

⁴Polytechnic Institute of Portalegre and Interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

E-mail addresses: *m.i.borges@ipportalegre.pt ; cpsd@ipportalegre.pt; carla.santos@ipbeja.pt*

The development of methodologies and models of degradation to predict the useful life of natural stones used as building materials or as ornamental stones assumes an unquestionable importance for the optimization of its choice, application criteria and maintenance in the short and long-term period of time. The analysis of the behavior of granites exposed to salt mist artificial weathering, simulated by accelerated aging tests, allows us to obtain data that when analyzed with the proper techniques of statistical prevision modelling, provide us a precious information about their expected performance over time.

Keywords

Aggressive Environments, Laboratorial Data, Natural Stones, Polynomial Regression Analysis.

Material and methods

Despite the common held belief that granite is not as prone to weathering as other natural stones, even granite is susceptible to it when exposed to aggressive environments, such as, for instance, soluble salts present in coastal areas. According to several authors salt mist is widely recognized as a promoter of alteration and degradation of natural stone present in coastal areas, leading to notable damage also in the stone-built heritage. The influence of this natural phenomena could spread to zones up to about 20 km from the shoreline, being able to accelerate the weathering process by a factor of 1.59. In order to modelling the degradation by salt mist of two Portuguese granites, widely used as construction materials and also as ornamental stones, fourteen samples of each litology were subject to 150 cycles of a salt mist atmosphere, according to the EU standards, and the variation in their masses recorded.

Results and discussion

A polynomial regression analysis was applied to the laboratorial data obtained and it was possible to verify that the pattern of the degradation curve for the two granites is quite

similar, with a "S" shape and both present very high values of coefficient of determination, meaning a very good fit of the laboratorial data to this model.

References

- [1] M. I. Borges. *Estudio de la acción de ambientes agresivos en rocas graníticas del Alentejo*. Tesis Doctoral, Universidad de Extremadura, Spain, 2019.
- [2] T. Kovács. *Durability of crystalline monumental stones in terms of their petrophysical characteristics*. Tesi Dottorato di Ricerca, Università di Bologna, Italy, 2009.
- [3] A. Silva, J. de Brito and P. Gaspar. Service life prediction model applied to natural stone wall claddings (directly adhered to the substrate). *Construction and Building Materials* **25(9)**: 3674-3684, 2011.
- [4] M. I. Santos and A. M. Porta Nova. Validation of nonlinear simulation metamodels. In *M. H. Hamza, editor, Proceedings of Applied Simulation and Modeling* **33**: 421–425, 2001.

Multivariate collective risk models. Inference and special case

Ana Cantarinha^{1,2}, Elsa Moreira^{3,4} and João T. Mexia^{3,4}

¹University of Évora, Portugal

²Research Center for Mathematics and Applications - CIMA, Portugal

³Nova University of Lisbon, Portugal

⁴Center for Mathematics and Applications - CMA, Portugal

E-mail addresses: *aicantarinha@gmail.com; efnm@fct.unl.pt*

Univariate collective models have played an important role in Actuarial Mathematics. The inference about these models is usually made for the totals of claims. We now present a multivariate version of these models that may be of interest, as a special case, for application in forest fires. The inference in this case is now made for the total burnt area and the number of fires.

Keywords

Collective Models, Asymptotic Distributions, Confidence Intervals, Risk Theory.

References

- [1] N. L. Bowers, H. U. Gerber, J. C. Hickman, D. A. Jones and C. J. Nesbet. *Actuarial Mathematics*. The Society of Actuaries, 1986.
- [2] A. Khuri. *Advanced calculus with applications to statistics*. John Wiley & Sons, 2003.
- [3] J. T. Mexia. *Variance free models*, Trabalhos de Investigação, N^o Departamento de Matemática, Faculdade de Ciências e Tecnologia. Universidade Nova de Lisboa, 1990.
- [4] S. Marques, J. G. Borges, J. Garcia-Gonzalo, F. Moreira, J. M. B. Carreiras, M. M. Oliveira, A. Cantarinha, B. Botequim and J. M. C. Pereira. Characterization of wildfires in Portugal. *European Journal of Forest Research*, **130(5)**: 775–784, 2011.

Analysis of adaptability and production stability of common wheat genotypes

Cristina Dias¹, Carla Santos², Isabel Borges³ and Maria Varadinov⁴

¹Polytechnic Institute of Portalegre and Center of Mathematics and Applications (CMA), Caparica, Portugal

²Polytechnic Institute of Beja and Center of Mathematics and Applications (CMA), Caparica, Portugal

³Polytechnic Institute of Portalegre and Interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

⁴Polytechnic Institute of Portalegre and Interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

E-mail addresses: *cpsd@ipportalegre.pt; carla.santos@ipbeja.pt; m.i.borges@ipportalegre.pt*

Corn, along with rice and wheat, is one of the three most important crops in the world, in terms of area under cultivation and total production. It is a versatile crop, being grown on all continents, with a wide number of cultivars adapted to local conditions. This work had the objectives of (i) comparing the values of the genotype-environment interaction (GA) obtained using the model of the main additive effects and multiplicative interaction (AMMI) and in the analysis of the linear regression (RL),(ii) comparing the production stability of common wheat genotypes.

Keywords

Cultivars, Genotype-Environment Interaction, AMMI Model, Linear Regression Analysis.

Material and methods

Twenty-two genotypes were evaluated in different environments (combinations of location and year) in Elvas, Beja, Herdade da Abobada (Serpa), Portugal, based on the analysis of data obtained in field trials that took place in the period from 2015 to 2019. The experimental design used was randomized blocks, with two repetitions.

Results and discussion

The sum of squares (SQ) of the regressions only explained 21,6% of the SQ of the GA interaction, while the first component (CP1) of the analysis of the main components explained 46.3%. The SQ of CP1 was greater than twice the SQ of all combined regressions (joint, genotypic and environmental). Therefore, the AMMI analysis was more efficient in describing the GA interaction than the RL. Cultivars 8, 11, 14, 16, 19 are the most stable, with the highest yield, revealing considerable adaptability to the region. Cultivars 7, 13, 20 and 22 are also quite stable with high medium yield. The cultivars, 12 and 21 are highly stable, with production above average, revealing wide adaptability to the region, the remaining cultivars are more unstable with production below average.

Acknowledgements: This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

References

- [1] L. Gusmão. An adequate design for regression analysis of yield trials. *Theor. Appl. Genet* **71**: 314–319, 1985.
- [2] H. D. Patterson and E. R. Williams. A new class of resolvable incomplete block designs. *Biometrika* **63**: 83–92, 1976.
- [3] J. T. Mexia, D. G. Pereira and J. Baeta. L2 environmental indexes. *Biom. Lett.* **36**: 137–143, 1999.
- [4] D. G. Pereira and J. T. Mexia. Reproducibility of joint regression analysis. *Colloquium Biometryczne* **33**: 279–299, 2003.

The likelihood ratio test of independence for random sample sizes - power studies and computational issues

Nadab Jorge¹, Carlos A. Coelho², Filipe J. Marques² and
Célia Nunes³

¹Centro de Matemática e Aplicações, Universidade da Beira Interior, Covilhã, Portugal

²Centro de Matemática e Aplicações (CMA) and Departamento de Matemática, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

³Centro de Matemática e Aplicações and Departamento de Matemática, Universidade da Beira Interior, Covilhã, Portugal

E-mail addresses: *nadab.jorge@ubi.pt; cmac@fct.unl.pt; fjm@fct.unl.pt; celian@ubi.pt*

In this work the test of independence of two groups of variables is addressed in the case where the sample size is not known in advance. This assumption lead us to assume the sample size as a realization of a random variable N .

We consider the sampling scheme in which N follows a Poisson distribution. For the case of two groups with p_1 and p_2 variables, it is shown that when either p_1 or p_2 (or both) are even the exact distribution of the test statistic corresponds to an infinite mixture of Exponentiated Generalized Integer Gamma distributions. In this case a computational module is made available for the cumulative distribution function of the test statistic. When both p_1 and p_2 are odd, the exact distribution may be represented as an infinite mixture of products of independent Beta random variables which density and cumulative distribution functions do not have a manageable closed form. Thus, a computational approach for the evaluation of the cumulative distribution function is given based on a numerical inversion formula developed for Laplace transforms. Simulation studies are provided in order to assess the power of the test, allowing us to analyze the properties of the testing procedure when using the random sample size approach.

Keywords

Laplace Transform Inversion, Series Expansions, Characteristic Function, Mixtures, Simulations, Power Studies.

Acknowledgements: This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through projects UIDB/00212/2020 and UIDB /00297/2020.

References

- [1] B. C. Arnold, C. A. Coelho and F. J. Marques. The distribution of the product of powers of independent uniform random variables - a simple but useful tool to address and better understand the structure of some distributions, *J. Multiv. Analysis* **113**: 19–36, 2013.

- [2] C. A. Coelho. The Generalized Integer Gamma Distribution - A Basis for Distributions in Multivariate Statistics. *J. Mult. Anal.* **64**: 86–102, 1998.
- [3] C. Nunes, G. Capristano, D. Ferreira, S. S. Ferreira and J. T. Mexia, Exact critical values for one-way fixed effects models with random sample sizes. *Journal of Computational and Applied Mathematics* bf 354: 112–122, 2019.
- [4] F. J. Marques, C. A. Coelho and B. C. Arnold, A general near-exact distribution theory for the most common likelihood ratio test statistics used in Multivariate Analysis, *Test* **20**: 180–203, 2011.

On the optimization of unbiased linear estimators in models with orthogonal block structure

Carla Santos^{1,2}, Célia Nunes³, Cristina Dias^{2,4} and João Tiago Mexia²

¹Polytechnic Institute of Beja, Portugal

²Center of Mathematics and its Applications -FCT- New University of Lisbon, Portugal

³Department of Mathematics and Center of Mathematics and Applications, University of Beira Interior, Portugal

⁴Polytechnic Institute of Portalegre, Portugal

E-mail addresses: *carla.santos@ipbeja.pt; celian@ubi.pt; cpsd@ipportalegre.pt; jtm@fct.unl.pt*

In a special class of models with orthogonal block structure, a necessary and sufficient condition for its least squares estimators to be best linear unbiased estimators is that the variance-covariance matrix commutes with the orthogonal projection matrix on the space spanned by the mean vector. As an alternative to the requirements for this purpose, established when this particular class of linear models was introduced, to ensure commutability we resort to U-matrices and use the fundamental partition of the observations vector.

Keywords

U-matrices, Best Linear Unbiased Estimators, Mixed Models, Models With Commutative Orthogonal Block Structure

Models with orthogonal block structure, OBS, are linear mixed models whose covariance matrices are the positive definite linear combinations of known pairwise orthogonal orthogonal projection matrices (POOPM) that add up to the identity matrix [1]. OBS have estimators with good behaviour for estimable vectors and variance components, however we can achieve best linear unbiased estimators, for estimable vectors, considering a particular class of OBS, those of models with commutative orthogonal block structure (COBS). COBS are based on the commutativity between the orthogonal projection matrix (OPM) on the space spanned by the mean vector, and the POOPM, belonging to the principal basis of the commutative Jordan algebra of symmetric matrices, associated to the model. Regarding estimation in COBS, a necessary and sufficient condition for its least squares estimators to be best linear unbiased estimators is that the OPM on the space spanned by the mean vector commutes with the variance-covariance matrix, which is ensured by the commutativity condition mentioned above or, alternatively, resorting to U-matrices and using the fundamental partition of the observations vector, constituted by the sub-vectors corresponding to the different sets of the levels of the fixed effects factors, [2].

Acknowledgements: This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações) and UIDB/00212/2020

References

- [1] J. A. Nelder. The analysis of randomized experiments with orthogonal block structure I. Block structure and the null analysis of variance, *Proceedings of the Royal Society, Series A* **283**: 147–162, 1965.
- [2] C. Santos, C. Nunes, C. Dias and J. T. Mexia. Models with commutative orthogonal block structure: a general condition for commutativity. *Journal of Applied Statistics*, 2020, DOI: 10.1080/02664763.2020.1765322
- [3] J. Seely. Quadratic subspaces and completeness. *Ann. Math. Statist.* **42**: 710–721, 1970.
- [4] R. Zmyślony. A characterization of best linear unbiased estimators in the general linear model. *Mathematical Statistics and Probability Theory* **2**: 365–373, 1978.

Multivariate APC model in the analyses of the logistics activities within agri-food supply chains

Maria Varadinov¹, Cristina Dias², Isabel Borges³ and Carla Santos⁴

¹Polytechnic Institute of Portalegre, and interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

²Polytechnic Institute of Portalegre and Center of Mathematics and Applications (CMA), Caparica, Portugal

³Polytechnic Institute of Portalegre and interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

⁴Polytechnic Institute of Beja and Center of Mathematics and Applications (CMA), Caparica, Portugal

E-mail addresses: *mjvaradinov@ipportalegre.pt; cpsd@ipportalegre.pt; m.i.borges@ipportalegre.pt*

The Reverse Logistics is associated with a growing awareness of people and institutions to environmental problems, the worrying shortage of raw materials globally and the increase of "green" products, that are environmentally friendly. So, there is the need to manage a RL flow to the traditional forward flow (between the point of final consumption and the point of origin). The goal is to conduct a study to analyze the practices and characteristics of RL within agri-food enterprises in Portugal.

Keywords

Agri-Food Supply Chains, Reverse Logistics, End-Of-Use Products, Methodologies.

Material and methods

The goal is to conduct a study, using descriptive research methods, to analyze the practices and characteristics of RL within agri-food enterprises in Portugal in the various agri-food sectors (e.g., agrarian, production, transformation, recycling), through an online questionnaire with closed questions. A quantitative study is developed, using a webpage and electronic mailing to promote the online questionnaire to the companies representative of agri-food sector registered in the INE database and that operate and reside in Portugal.

Results and discussion

As a result, this study can contribute to the knowledge and the perception by the Portuguese agri-food managers to the RL processes and it can help with the criteria that influence decision making for RL implementation, either as driver or as barrier to the development of RL systems in agri-food companies.

Acknowledgements: Research partially funded by FCT- Fundação para a Ciência e a Tecnologia, Portugal, through the project FCT-UID/ECI/04028/2020.

References

- [1] S. Agrawal, R. Singh and Q. Murtaza, A literature review and perspectives in RL, *Resources, Conservation and Recycling Journal* **97**: 76–92, 2015.
- [2] K. Govindan, S. Azevedo, H. Carvalho and V. Cruz-Machado. Impact of supply chain management practices on sustainability. *Journal of Cleaner Production* **30**: 1–14, 2014.
- [3] J. L. Miranda, M. J. Varadinov and S. Rubio. Green Logistics, RL and Agro-Industries: Overview of scientific articles and international programs. In *Green Supply Chains: Applications in Agroindustries*. Ariel, C. and Castro, W. (Eds.). Universidad Nacional de Colombia, Manizales 97–106, 2014.
- [4] M. J. Varadinov. *A Gestão da Logística Inversa nas Empresas Portuguesas*. Ph.D. Thesis, Badajoz, Extremadura University, 2016.

Posters included in the Organized Session 3

Coefficient of variation: properties, bounds and monotony

Carla Santos^{1,2}, Cristina Dias^{2,3}, Isabel Borges³, Maria Varadinov³
and João Tiago Mexia⁴

¹Polytechnic Institute of Beja, Portugal

²Center of Mathematics and its Applications -FCT- New University of Lisbon, Portugal

³Polytechnic Institute of Portalegre, Portugal

⁴Department of Mathematics and Center of Mathematics and its Applications -FCT-
New University of Lisbon, Portugal

E-mail addresses: *carla.santos@ipbeja.pt; cpsd@ipportalegre.pt;*
m.i.borges@ipportalegre.pt; jtm@fct.unl.pt

The widespread use of the standard deviation as dispersion measure, in most investigations that use statistical methods, has disregarded the coefficient of variation, not taking advantage of its potentialities as measure of risk sensitivity, to represent the reliability of trials or in the assessment of the accuracy of experiments. In this work, we revisit the coefficient of variation definition and properties and investigate the case of samples with binomial distribution.

Keywords

Coefficient of Variation, Dispersion, Binomial Distribution.

The coefficient of variation of a distribution is a useful descriptive measure for describing variability or dispersion. Although, in many research fields, the most common measure of dispersion is the standard deviation, in agriculture, finance and others, the coefficient of variation has emerged in a prominent place as measure of risk sensitivity, to represent the reliability of trials or in the assessment of the accuracy of experiments. This relevance comes from the advantage of being a unit-free measure, allowing to compare data from different distributions or data in different scales. The sample coefficient of variation is usually used as an estimator of the population coefficient of variation, however, in many situations, it is interesting to consider interval estimation. Due to its practical interest, point estimation and confidence intervals for coefficient of variation has been well addressed in the literature. Theoretical investigation of properties of the sample coefficient of variation has provided limits for this measure, especially in a Normal distribution. In this work we consider the case of samples with binomial distribution in which the coefficient of variation is monotone decreasing both with n and p . For finding n we obtain bound for that coefficient.

Acknowledgements: This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações).

References

- [1] S. Banik and B. M. Golam Kibria. Estimating the Population Coefficient of Variation by Confidence Intervals, *Communications in Statistics - Simulation and Computation* **40**:(8): 1236–1261, 2011.
- [2] R. Mahmoudvand and T. A. Oliveira. On the Application of Sample Coefficient of Variation for Managing Loan Portfolio Risks. In: Oliveira T., Kitsos C., Oliveira A., Grilo L. (eds) *Recent Studies on Risk Analysis and Statistical Modeling. Contributions to Statistics*. Springer, 2018.
- [3] M. Smithson. On relative dispersion: A new solution for some old problems. *Qual Quant* **16**: 261–271, 1982.
- [4] M. R. Spiegel and L. J. Stephens *Schaum's outline of theory and problems of statistics*. McGraw-Hill, 2008.

Basic models with orthogonal structure

Cristina Dias¹, Carla Santos² and João Tiago Mexia³

¹Polytechnic Institute of Portalegre and Center of Mathematics and Applications (CMA), Caparica, Portugal

²Polytechnic Institute of Beja and Center of Mathematics and Applications (CMA), Caparica, Portugal

³New University of Lisbon and Center of Mathematics and Applications (CMA), Caparica, Portugal

E-mail addresses: *cpsd@ipportalegre.pt; carla.santos@ipbeja.pt; jtm@fct.unl.pt*

Today several techniques exist for treating a series of studies on the Joint Analysis of Tables. The presented formulation allows us to make inference about the series of studies, since the results presented in this work can be applied to matrices of Hilbert-Schmidt products that are very important in first step of the Statis methodology. The models we consider are based in the spectral decomposition of the mean matrices. These models gave the basis to perform inference for isolated matrices and for structured matrix families. In these families, the matrices, all of the same order, correspond to the treatments of base models. An application to legislative elections held in mainland Portugal is presented. Our results point towards to the existence of a common structure of degree one.

Keywords

Treatments, Base Design, ANOVA, Statis Methodology.

Acknowledgements: This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

References

- [1] C. Dias. *Modelos e Famílias de Modelos para Matrizes Estocásticas Simétricas*. Ph.D. Thesis, Évora University, 2013.
- [2] C. Dias, S. Santos, M. Varadinov and J. T. Mexia. ANOVA Like Analysis for Structured Families of Stochastic Matrices. *AIP Conf. Proc.* **1790**: 1400061–1400063, 2016.
- [3] H. Scheffé, *The Analysis of Variance*. New York: John Wiley and Sons, 1959.
- [4] M. M. Oliveira and J. T. Mexia. AIDS in Portugal: endemic versus epidemic forecasting scenarios for mortality. *International Journal of Forecasting* **20**: 131–137, 2004.

Comparing COVID-19 case fatality ratios through Marascuilo Procedure

Carla Santos^{1,2}, Cristina Dias^{2,3} and João Tiago Mexia⁴

¹Polytechnic Institute of Beja, Portugal

²Center of Mathematics and its Applications -FCT- New University of Lisbon, Portugal

³Polytechnic Institute of Portalegre, Portugal

⁴Department of Mathematics and Center of Mathematics and its Applications -FCT- New University of Lisbon, Portugal

E-mail addresses: *carla.santos@ipbeja.pt; cpsd@ipportalegre.pt; jtm@fct.unl.pt*

To understand the severity of COVID-19, identify populations at risk and guide policy, it is critical to measure associated mortality. This information can be obtained through case fatality ratio (CFR). To assess the differences between the COVID-19 CFR of European Union countries, we used World Health Organization data on COVID-19 infections and deaths, until August 26. Rejecting the null hypothesis of equality of CFR, we proceeded to the comparisons between all pairs of countries by carrying out Marascuilo's multiple comparison procedure. We found evidence of significant differences in CFR between different countries in Europe.

Keywords

CFR, COVID-19, Marascuilo's procedure, European Union countries, Severity.

The appropriate metric for determining the true severity of a disease is the IFR (Infection fatality ratio), however, the accuracy of its calculation has enormous limitations and, for the moment, for most countries in the world the information available only allows to know the CFR (case fatality ratio). Although reliable values of the CFR can only be obtained at the end of a pandemic, and the value of the CFR tends to vary considerably in the course of that pandemic and be overestimated at certain times [4], the knowledge of the CFR constitutes an important contribution to the knowledge of the characteristics of the pandemic [2]. Alluding to the imprecise nature of CFR values, while an epidemic is going on, [1], used simulation and COVID-19 data available on April 5th, for United States of America, Italy, Spain, Germany, France, United Kingdom, The Netherlands and Korea, arguing that the differences found between CFR are due to different evolutionary stages of the pandemic in these countries and that, at the end of the pandemic, the CFR will tend to be similar. More than eight months after the beginning of the dissemination of COVID-19, it is important to evaluate the CFR values. In this study we look at the countries of the European Union, aiming to compare the CFR in these countries. The data used in this study were the cumulative frequencies of infections and deaths by COVID-19, until August 26th, which are part of the WHO's records, published on <https://covid19.who.int/region/euro>. Rejecting the null hypothesis of equality of proportions by chi-square test it is concluded that not all population proportions are equal. Focusing on differences in proportions, we carried out Marascuilo's multiple comparison procedure [3], which enables us to make comparisons between all pairs of countries. We found evidence of significant differences in CFR between different countries in Europe.

Acknowledgements: This work was partially supported by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the project UIDB/00297/2020 (Centro de Matemática e Aplicações).

References

- [1] F. M. Granozio. On the problem of comparing COVID-19 fatality rates, 2020. Available in: <https://arxiv.org/ftp/arxiv/papers/2004/2004.03377.pdf>
- [2] M. A. Khafaie and F. Rahim Cross-country comparison of case fatality rates of COVID-19/SARS-COV-2. *Osong Public Health and Research Perspectives* **11**: 2, 74, 2020.
- [3] L. A. Marascuilo, M. McSweeney. Nonparametric post hoc comparisons for trend, *Psychological Bulletin* **67**: 401–412, 1967.
- [4] WHO, Estimating mortality from COVID-19, 2020. Available in: <https://www.who.int/news-room/commentaries/detail/estimating-mortality-from-covid-19>

The need to manage a RL flow to the traditional forward flow

Maria Varadinov¹ and Carla Santos²

¹Polytechnic Institute of Portalegre and Interdisciplinary Coordination for Research and Innovation (C3i), Portalegre, Portugal

²Polytechnic Institute of Beja and Center of Mathematics and Applications (CMA), Caparica, Portugal

E-mail addresses: *mjvaradinov@ipportalegre.pt; carla.santos@ipbeja.pt*

A survey using descriptive and empiric methods is outlined. It aims at consulting on the implementation of RL practices in European agri-food companies, including the current mechanisms influencing RL activities (e.g., economic, social, and legislative factors). The study will allow a comparative analysis with the existing results and studies conducted in other countries. The expected results will indicate the relevance given by agri-food companies 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 agri-food SC.

Keywords

Agri-food companies, Supply Chains, Reverse Logistics, End-Of-Use Products, RL processes.

Material and methods

In this study is outlined 1) a specific analysis to highlight the corporate events in Portuguese agri-food companies; 2) We also propose to check other studies (of the same scope) developed in other countries; and the proxy variables (e.g., as economic, social and legislative factors) to be identified by their level of importance, 3) We also propose for correlation, and checking, the relevance given by agri-food managers to RL as an important tool for decision making, and 4) It is also intended to verify if RL can take part of the companies's management (in the near future; and in an autonomous/permanent way). The statistical analysis thus involves measures of descriptive statistics: both absolute frequencies and relative frequencies; averages and the associated standard deviations. Statistical analysis is prepared with SPSS-Statistical Package for the Social Sciences for MS Windows.

Results and discussion

The study results can contribute either for decision aiding or for the integration of RL at management level. In this way, allowing the implementation of the RL system, realizing competitive advantages for agri-food companies, as well as giving the possibility to consider RL in the long term.

Acknowledgements: Research partially funded by FCT- Fundação para a Ciência e a Tecnologia, Portugal, through the project FCT-UID/ECI/04028/2020.

References

- [1] S. Agrawal, R. Singh and Q. Murtaza, A literature review and perspectives in RL, *Resources, Conservation and Recycling Journal* **97**: 76–92, 2015.
- [2] K. Govindan, S. Azevedo, H. Carvalho and V. Cruz-Machado. Impact of supply chain management practices on sustainability. *Journal of Cleaner Production* **30**: 1–14, 2014.
- [3] J. L. Miranda, M. J. Varadinov and S. Rubio. Green Logistics, RL and Agro-Industries: Overview of scientific articles and international programs. In *Green Supply Chains: Applications in Agroindustries*. Ariel, C. and Castro, W. (Eds.). Universidad Nacional de Colombia, Manizales, 97–106, 2014.
- [4] M. J. Varadinov. *A Gestão da Logística Inversa nas Empresas Portuguesas*. Ph.D. Thesis, Badajoz, Extremadura University, 2016.

Organized Session 4

Computational Models and Machine Learning Approaches for solving Real-world Applications

Organizer: Padmanabhan Seshaiyer

Efficient image segmentation and machine learning algorithms for improved malaria detection from blood smear images

Viraj Boreda¹, Kirthi Kumar², Soraya Ngarnim³, Nathalia Peixoto³
and Padmanabhan Seshaiyer³

¹Thomas Jefferson High School for Science and Technology, Virginia, USA

²University of California, Berkeley, California, USA

³George Mason University, Virginia, USA

E-mail addresses: *virajboreda514@gmail.com; kirthikumar50@gmail.com;*
sngarnim@masonlive.gmu.edu; npeixoto@gmu.edu; pseshaiy@gmu.edu

Malaria is one of the deadliest endemic diseases in developing countries with millions of cases recorded each year. Having widespread access to efficient malaria detection would help reduce malaria cases in these countries. In recent years, several computational algorithms have been applied to detect malaria from a variety of approaches, including analysis of blood smear images and presence of protein biomarkers in urine and saliva. For blood smear images, recent literature employs a combination of image segmentation techniques and machine learning classifiers. While multiple techniques have been suggested, and advances are being made, there is still work that needs to be done to identify the best combination of these algorithms.

The aim of this study was to perform a comparative analysis of four different image segmentation techniques and five machine learning classifiers on a dataset of roughly 27,000 blood smear images, half of which were infected with malaria. The focus of the work involved identifying combinations of image segmentation and machine learning algorithms that resulted in the highest accuracy in classifying a blood smear image as infected or uninfected, leading to improved malaria detection. For our computational algorithms, the Python programming language with the libraries OpenCV and Scikit-learn were used for the implementation of the computational algorithms developed in this work. The resulting confusion matrices helped to identify efficient combinations of image segmentation techniques and machine learning classifiers. We will discuss the accuracy of the various machine learning algorithms employed in this work and make suitable recommendations.

Keywords

Malaria Detection, ML Classifier, Image Segmentation and Blood Smear.

Acknowledgements: This work was supported in part by the George Mason University's 2020 Aspiring Summer Scientist Internship Program.

Deep learning with neural networks for airborne spread of COVID-19 in enclosed spaces

Udbhav Muthakana¹, Maziar Raissi² and Padmanabhan Seshaiyer³

¹Thomas Jefferson High School for Science and Technology, Virginia, USA

²University of Colorado Boulder, Colorado, USA

³George Mason University, Virginia, USA

E-mail addresses: *umuthakana@gmail.com; Maziar.Raissi@colorado.edu; pseshaiy@gmu.edu*

The COVID-19 pandemic has caused more than 16 million cases and 600,000 deaths worldwide in little over half a year. Most epidemic models studying the virus have focused on the macro scale, evaluating the progression of infected cases across large regions. Alarming, there is a severe lack of coronavirus-specific literature that models the medium to long term progression of infections in enclosed spaces like schools, aircraft, and hospitals. This work develops a general framework but focuses on hospitals, imagining a two-week scenario of infectious spread across several connected hospital wards. The airborne infectivity of COVID-19 has led to important public policy questions about safety measures in enclosed spaces like schools, aircraft, and hospitals.

In this work, we introduce a novel framework that combines the Wells-Riley airborne infection model, the SEIR epidemiological model, and an infectious concentration transport model. We make use of two fundamental models including the *SEIR compartmental model*, which follows a population to evolve through four stages of infection, and the *Wells-Riley model* of indoor airborne infection, which relates probability of infection to parameters describing the air in the enclosed space. We apply it to a benchmark application, and then perform parameter estimation with Physics Informed Neural Networks via Deep Learning. We show that the technique performs well with limited and noisy data. We also study the performance of method on the number of layers as well as the level of noise in the data. The novelty of this work is in extending standard epidemiological models for parameters of airborne infection models to include an infectious concentration transport model. Our computational results show that the method presented in this work is efficient and reliable for supervised learning while respecting constraints defined by 15 coupled ODEs.

Keywords

Deep Learning, COVID-19, Modeling, Airborne, Parameter Estimation.

Acknowledgements: This work is supported in part by the National Science Foundation DMS 2031027 and DMS 2031029.

Computational mathematics for solving real-world problems arising from COVID-19

Padmanabhan Seshaiyer

George Mason University, Virginia, USA

E-mail address: *pseshaiy@gmu.edu*

Computational mathematics, which comprises of modeling, analysis and simulation has served as a foundation for solving most multidisciplinary problems in science and engineering. These real-world problems often involve complex dynamic interactions of multiple physical processes which presents a significant challenge, both in representing the physics involved and in handling the resulting coupled behavior. If the desire to control the system is added, then the complexity increases even further. Hence, to capture the complete nature of the solution, a coupled multidisciplinary approach is essential. In this work, we will present examples of real-world challenges related to infectious diseases such as COVID-19 modeled via coupled system of differential equations and present methodologies to solve them using deep learning frameworks

Through our research in this field, we have come to appreciate that developing and solving a mathematical model from a real-world system requires a combined theoretical and computational approach. Theory is needed to guide the performance and interpretation of the mathematical model while computation is necessary to synthesize the results. The focus of the work will therefore be to discuss how research and education programs can be developed around computational mathematics that can not only help solve several challenging multidisciplinary applications such as those related to COVID-19 but also will help to train the next generation STEM (Science, Technology, Engineering and Mathematics) workforce to solve real-world challenges. Specifically, we will describe how research focus can be integrated with education programs where the primary goal will be to engage students to apply these well-developed research concepts in engineering, computer science, and mathematics. Incorporating concepts developed herein, into new or existing inter-disciplinary computational mathematical courses and also by mentoring students at the graduate, undergraduate and high school level as well as teachers through workshops, seminars and other enrichment activities will also be discussed.

Keywords

Mathematical Modeling, Computational Mathematics, COVID-19.

Acknowledgements: This work is supported in part by the National Science Foundation DMS 2031027 and DMS 2031029.

Computational techniques using machine learning for rehabilitation and relapse-detecting technology to assist recovering Opioid addiction patients

Diego Valencia¹, Kendall Johnson², Holly Matto³ and
Padmanabhan Seshaiyer³

¹Thomas Jefferson High School for Science and Technology, Virginia, USA

²U.S. Army Combat Capabilities Development Command, Virginia, USA

³Mathematical Sciences, George Mason University, Virginia, USA

E-mail addresses: *dav222575@gmail.com; kjohns21@masonlive.gmu.edu; hmatto@gmu.edu; pseshaiy@gmu.edu*

Opioid addiction rates in the United States have long been on the rise, but the expansion of rehabilitation programs have been falling behind. In 2013, an estimated 20.4 million people out of 22.7 million who required treatment for illicit drug addiction failed to receive proper treatment at specialty facilities. Without proper rehabilitation, patients can often fall back into a cycle of substance abuse. Considering the societal stigma of addiction and seeking treatment therefore, technological rehabilitation provides a more conservative alternative for these patients in terms of cost, privacy, and convenience.

The goal of this mathematical modeling project is to develop a discrete wearable device that will determine whether an opioid rehabilitation patient is having a response to a stressor, which indicates the potential for a relapse event. In this project, we created a prototype with a heart rate sensor and Raspberry Pi Zero with the Python module scikit-learn to measure heart rate and use deep learning to perform data analysis that would evaluate Heart Rate Variability (HRV) and make complex predictions. This project has helped us to provide insight on how technology can be used for opioid addiction treatment, and has also examined how heart rate (HR) values can be used to evaluate HRV, a more accurate determinant of emotional responses to stimuli. We hope to further develop and deploy our prototype with the user community to validate the mathematical model and computational algorithms. The deep learning algorithms developed were validated and used to perform data analysis to evaluate heart rate variability and for complex predictions.

Keywords

Opioid Epidemic, Machine Learning, Wearable Technology, Rehabilitation.

Acknowledgements: This work was supported in part by George Mason University's Aspiring Scientists Summer Internship Program. We also thank graduate mentors Andy Hoang and Nithin Ellanki, and faculty member Dr. Nathalia Peixoto for their useful advise.

Organized Session 5

Statistical Modeling

Organizer: Milan Stehlík

Generalized means: progresses and challenges in statistics of extremes

M. Ivette Gomes^{1,2}, Frederico Caeiro^{3,4} and
Lígia Henriques-Rodrigues^{5,6}

¹Universidade de Lisboa, Portugal

²Centro de Estatística e Aplicações (CEAUL), Portugal

³Universidade Nova de Lisboa, Portugal

⁴Centro de Matemática e Aplicações (CMA), Portugal

⁵Universidade de Évora, Portugal

⁶Centro de Investigação em Matemática e Aplicações (CIMA), Portugal

E-mail addresses: *ivette.gomes@fc.ul.pt; fac@fct.unl.pt; ligiahr@uevora.pt*

Extreme value theory (EVT) helps us to potentially control disastrous events. Floods, fires, hurricanes and other extreme occurrences have provided impetus for new developments of EVT. *Generalized means* (GM) have recently been used with success in the estimation of a positive *extreme value index* (EVI). Due to the specificity of the *Weibull tail coefficient* (WTC), and its deep link to a positive EVI, we shall now make use of GM in the estimation of the WTC.

Keywords

Extreme Value Theory, Semi-Parametric Estimation, Weibull Tail-Coefficient.

The *Weibull tail-coefficient* (WTC) is the parameter θ in a right-tail function of the type

$$\overline{F}(x) := 1 - F(x) =: e^{-H(x)}, \quad H \in RV_{1/\theta}, \quad \theta \in \mathbb{R}^+, \quad (1)$$

where the notation RV_β stands for the class of *regularly varying* functions at infinity with an *index of regular variation* equal to β , i.e. positive measurable functions g such that $\lim_{t \rightarrow \infty} g(tx)/g(t) = x^\beta$, for all $x > 0$. We are thus working with models with a regularly varying cumulative hazard function, $H(x) = -\ln(1 - F(x))$.

With the notation $U(t) := F^\leftarrow(1 - 1/t)$, $t \geq 1$, $F^\leftarrow(t) := \inf\{x : F(x) \geq y\}$, the generalized inverse function of F , Equation (1) is equivalent to saying that

$$U(e^t) = H^\leftarrow(t) \in RV_\theta \iff U(t) =: (\ln t)^\theta L(\ln t),$$

with $L \in RV_0$, a slowly varying function.

Moreover, a model F is said to have a heavy right-tail if and only if there exists a positive real ξ such that $\overline{F} \in RV_{-1/\xi} \iff U \in RV_\xi$. The parameter ξ is the well-known *extreme value index* (EVI), the primary parameter in statistics of extremes.

Most of the recent WTC-estimators are proportional to the class of Hill EVI-estimators ([4]), defined by

$$H_{k,n} := \frac{1}{k} \sum_{i=1}^k V_{i,k}, \quad V_{i,k} := \ln X_{n-i+1:n} - \ln X_{n-k:n}, \quad k = 1, 2, \dots, n-1, \quad (2)$$

where $X_{i:n}, 1 \leq i \leq n$ denotes the i -th ascending order statistic associated with a random sample $X_i, 1 \leq i \leq n$. Just as an example, we here refer one of the estimators in [3],

$$\hat{\theta}_{k,n} := \frac{\ln(n/k)}{k} \sum_{i=1}^k V_{i,k} = \ln(n/k) H_{k,n}, \quad (3)$$

with $H_{k,n}$ defined in (2).

The interesting performance of EVI-estimators based on *generalized means* (GM), leads us to consider a simple generalization of the Hill EVI-estimators, in (2), studied in [1], among others, under a second-order framework. Indeed, since the H EVI-estimators are the logarithm of the *geometric mean* (or *mean-of-order-0*) of $U_{ik} := \exp(V_{i,k}) = X_{n-i+1:n}/X_{n-k:n}$, $1 \leq i \leq k$, with $V_{i,k}$ given in (2), we can more generally consider the mean-of-order- p of $U_{i,k}$, being led to the *mean-of-order- p* (MO_p) EVI-estimators (see also [2], where these estimators are dealt with under a third-order framework). More generally than $\hat{\theta}_{k,n}$, in (3), we can base the WTC-estimation on the MO_p EVI-estimators, or even on other GM EVI-estimators. Consistency and asymptotic normality of the estimators under study is proved. The performance of the new estimators for finite samples is illustrated through a Monte-Carlo simulation.

Acknowledgements: Research partially supported by National Funds through **FCT** - Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/0006/2020 (CEA/UL), UIDB/MAT/0297/2020 (CMA/UNL) and UIDB/MAT/04674/2020 (CIMA).

References

- [1] M. F. Brilhante, M. I. Gomes and D. Pestana. A simple generalisation of the Hill estimator. *Comput. Statist. and Data Analysis* **57(1)**: 518–535, 2013.
- [2] F. Caeiro, M. I. Gomes, J. Beirlant and T. de Wet. Mean-of-order p reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19(4)**:, 561–589, 2016.
- [3] L. Gardes and S. Girard. Comparison of Weibull tail-coefficient estimators. *Revstat—Statist. J.* **4**: 163–188, 2006.
- [4] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174, 1975.

Comparison of different estimators for a reflective SEM: burnout syndrome in Portuguese workers

Luís M. Grilo

Instituto Politécnico de Tomar (IPT) & Universidade Aberta, Portugal
Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal
Centro de Investigação em Cidades Inteligentes (Ci2), IPT, Portugal
CIICESI, ESTG, Politécnico do Porto, Portugal

E-mail address: *lgrilo@ipt.pt*

Structural equation modeling (SEM) is considered a family of statistical techniques that investigates the relationships between observed/manifest and latent variables. Mostly SEM methods require manifest variables to be quantitative (measured on interval or ratio scales), but in many studies they are qualitative and measured in an ordinal scale, such as variables that operationalize latent variables (i.e. not directly observable) as stress and burnout syndrome. However, if this scale is well presented (symmetrical and with equidistant categories), it can approach an interval-level measurement, and the corresponding variables can be used in SEM, which is a common situation in the specialized literature. In this study, a theoretical reflective SEM was proposed and a model was estimated using the consistent Partial Least Squares (PLSc) estimator. This model was also estimated using the estimators: maximum likelihood estimator (ML), the robust ML (MLM) and the weighted least square mean and variance adjusted (WLSMV), used for SEM with categorical variables. The results obtained were compared showing strong similarities.

Keywords

Likert Scale, Non-Normal Distribution, Primary Data, Stress, WLSMV Estimator.

Chronic stress damages the immune system, which is a serious problem because we need it strong to fight diseases and also unknown bacteria and viruses (like COVID-19, since there are people who have no symptoms) [1]. This condition can even lead to burnout syndrome which can be considered as a limit situation of depression. The Copenhagen Psychosocial Questionnaire (COPSOQ) has been used in an increasing number of countries, namely to assess workers' burnout syndrome [3]. The manifest variables available in the validated COPSOQ are expressed on a 5-point Likert scale. In this study, the primary data collected in a Portuguese company violates the multivariate normal distribution and almost all variables are not approximately symmetric and/or mesokurtic. To apply SEM, two types of methods are available, covariance based (CB) and variance based (VB). Although they complement each other, they differ statistically and have different objectives and requirements. Among CB-and-VB estimators, PLSc (which corrects bias to consistently estimate SEM with common factors), ML, MLM (robust standard error estimate) and WLSMV (based on polychoric correlations) were used to assess which latent variables have the greatest impact on stress and burnout [2–6]. Based on the outputs produced by SmartPLS, AMOS and R (Lavaan package) software algorithms, the measurement and structural (sub)models were accessed. The estimated values for the path coefficients and

the coefficients of determination are almost equal, with the endogenous latent variable burnout having a $R^2 \approx 85\%$.

Acknowledgements: This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

References

- [1] D. M. Davis. *The Beautiful Cure: Harnessing Your Body's Natural Defences*. Vintage Publishing, 2018.
- [2] L. M. Grilo, A. Mubayi, K. Dinkel, B. Amdouni, J. Ren and M. Bhakta. Evaluation of Academic Burnout in College Students. Application of SEM with PLS approach. *AIP Conf. Proc.* 2186, 090008, 2019; <https://doi.org/10.1063/1.5138004>.
- [3] L. M. Grilo, H. L. Grilo and E. Martire. SEM using PLS Approach to Assess Workers Burnout State. *AIP Conf. Proc.* 2040, 110008-1-110008-5, 2018; <https://doi.org/10.1063/1.5079172>.
- [4] J. F. Hair, G. T. M. Hult, C. M. Ringle and M. Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, (2nd Ed., Thousand Oakes, CA: Sage), 2017.
- [5] C. M. Ringle, S. Wende and J.-M. Becker. “SmartPLS 3.” Boenningstedt: SmartPLS GmbH, 2015, www.smartpls.com.
- [6] F. Schuberth, J. Henseler and T. K. Dijkstra. Partial least squares path modeling using ordinal categorical indicators. *Qual Quant* **52**: 9–35, 2018.

Ratios of order statistics and regularly varying tails

Pavlina Jordanova¹, Venelin Todorov^{2,3} and Milan Stehlík^{4,5}

¹Konstantin Preslavsky University of Shumen, Bulgaria

²Institute of Mathematics and Informatics, Bulgarian Academy of Sciences

³Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

⁴Johannes Kepler University, Austria

⁵Universidad de Valparaíso, Chile

E-mail addresses: *pavlina_kj@abv.bg; mlnstehlik@gmail.com; venelin@parallel.bas.bg*

Six distribution sensitive estimators of the index of regular variation are considered. Given that the observed random variable is Pareto the first of them is unbiased, consistent, asymptotically efficient and asymptotically normal. The second one is asymptotically equivalent to the first one. It has faster rate of convergence, however, it is biased. Analogous results for the case when the observed random variable is Log-logistically distributed follow. In cases when it is Fréchet or Hill-Horror distributed the last two estimators are consistent, however, their asymptotic unbiasedness and asymptotic efficiency are still open problems. The talk is based on two papers [3] and [4]. A very extensive study on these and similar results could be found in [2].

Keywords

Point Estimators, Order Statistics, Heavy Tails.

Given that the observed random variable has a distribution in the max-domain of attraction of a fixed Generalized Extreme Value Distribution in 1995 [1] proposed three semi-parametric estimators of the tail index. By using the second-order regular variation parameter [1] investigated the most appropriate choices of the scale parameter in the sequence (in our talk it is s) which minimizes the asymptotic mean square error, bias and variance of the estimator. Then, the rate of convergence was investigated and the asymptotic normality and strong consistency of the estimators were proved.

In [3] we found two very simple, unbiased and asymptotically efficient estimators of the index of regular variation given that the observed random variable X is Pareto distributed with parameter $\alpha > 0$. More precisely if X_1, X_2, \dots, X_n are n independent observations on a random variable X with order statistics $X_{(1,n)} \leq X_{(2,n)} \leq \dots \leq X_{(n,n)}$, the first

estimators is $Q_{i,s} := \frac{\log \frac{X_{(is,n)}}{X_{(i,n)}}}{H_{is-1} - H_{i-1}}$, where H_i is the i -th harmonic number. When we made

the asymptotic we obtained Fraga's [1] estimator $Q_{i,s}^* := \frac{\log \frac{X_{(is,n)}}{X_{(i,n)}}}{\log(s)}$. It is clear that both are asymptotically equivalent, however, the first one is unbiased. This approach allowed us to find the exact explicit formula for the probability density function and variance of the estimators and to compute small sample confidence intervals, for example when the sample size $n = 89$.

In [4] we considered similar four new distribution sensitive estimators $Q_{i,s}^{LL} := \frac{Q_{i,s}}{2}$, $Q_{i,s}^{LL*} := \frac{Q_{i,s}^*}{2}$, $Q_{i,s}^{Fr*} := -\frac{\log \frac{X_{(is,n)}}{X_{(i,n)}}}{\log[1 - \frac{\log(s)}{\log(s+1)}]}$, and $Q_{i,s}^{HH*} := \frac{\log \frac{X_{(is,n)}}{X_{(i,n)}} + \log[1 - \frac{\log(s)}{\log(s+1)}]}{\log(s)}$, given

that the observed random variable is Log-logistic for the first two of them, Fréchet distributed in the third case or Hill-Horror distributed for the last one. The transformations of order statistics are chosen in such a way that we can obtain unbiased or at least asymptotically unbiased or strongly consistent estimators. Moreover, first two of them are asymptotically normal and asymptotically efficient. Note that for one and the same α they are not asymptotically equivalent to the corresponding estimators in Pareto case.

We have obtained the exact formulae for the probability densities of the estimators and although their formulae look different we have observed that the plots of these densities, at least in case $\alpha = 1$, are approximately symmetric with respect to the center of the distribution and have properties which are similar to the corresponding normal one. The last conclusions are very useful when we compute small sample confidence intervals of the parameter. They are consequences of the asymptotic normality of the estimators which is useful only for working with large samples. In the Fréchet and Hill-Horror cases the asymptotic unbiasedness and asymptotic efficiency are still open problems.

We show that the proposed estimators could be useful for achieving a better precision in the estimation of the quantiles outside the range of the data.

Acknowledgements: This work has received funding from Bulgarian National Science Funds under the bilateral projects Bulgaria - Austria, 2016-2019, Feasible statistical modelling for extremes in ecology and finance, Contract number 01/8, 23/08/2017 and WTZ Project BG 09/2017.

References

- [1] M. I. F. Alves. Estimation of the tail parameter in the domain of attraction of an extremal distribution. *Journal of statistical planning and inference* **45(1–2)**: 143–173, Elsevier, 1995.
- [2] P. Jordanova. *Probabilities for p-outside values and heavy tails*, Shumen University Publishing House, 2020.
- [3] P. Jordanova and M. Stehlík. *Logarithm of ratios of two order statistics and regularly varying tails*, AIP Conference Proceedings **2164(1)**: 020002-1 – 020002-11, 2019.
- [4] P. Jordanova and M. Stehlík. *Distribution sensitive estimators of the index of regular variation based on ratios of order statistics*. Submitted, 2020.

Modelling and prediction of COVID-19 outbreaks

M. Stehlík^{1,2}

¹Department of Applied Statistics & Linz Institute of Technology, Johannes Kepler University in Linz, Austria

²Institute of Statistics, Universidad de Valparaíso, Valparaíso, Chile

E-mail address: *Milan.Stehlik@jku.at*

Keywords

COVID-19 outbreaks, Exponential Growth Model, SIR model, Sensitivity.

We formulate the ill-posedness of inverse problems of estimation and prediction for COVID-19 outbreaks from statistical and mathematical perspectives. These leave us with a plenty of possible statistical regularizations, thus generating plethora of sub-problems. We can mention the as examples stability and sensitivity of peak estimation, starting point of exponential growth curve, or estimation of parameters of SIR model. We also illustrate that several country-specific covariates, like social structure, or air pollution, can play crucial way in regularization of the estimators. We will illustrate this on example of Chile, where start of exponential growth, grounded on microbiological-epidemiological model was severely underestimated. Moreover, in a specific country, one can define several social groups which can contribute in a heterogeneous way to whole country epidemiological curves. Moreover, each parameter has its own specific sensitivity, and naturally, the more sensitive parameter deserves a special attention. E.g. in SIR (Susceptible-Infected-Removed) model, parameter β is more sensitive than parameter γ . In simple exponential epidemic growth model, b parameter is more sensitive than a parameter. We provide sensitivity and illustrate it on the country specific data. We also discuss on statistical quality of COVID-19 incidence prediction, where we justify an exponential curve considering the microbial growth in ideal conditions for epidemic. We model number of infected in Iowa State, USA, Hubei Province in China, New York State, USA. All empirical data justifies an exponential growth curve for initial prediction during epidemics. We also discuss peculiarities of COVID-19 prediction in Chile and Slovak Republic. For more details see [1].

Acknowledgements: Author acknowledge support of WTZ Project BG 09/2017 (Austria-Bulgaria)

References

- [1] M. Stehlík, J. Kiselak, A. Dinamarca, Y. Li and Y. Ying. On COVID-19 outbreaks predictions: issues on stability, parameter sensitivity and precision. *Stochastic Analysis and Applications*, 2020, DOI: 10.1080/07362994.2020.1802291.

Contributed Talks

The impact of entrepreneurial framework conditions on entrepreneurship early activities: an European longitudinal study

Ana Borges¹, Aldina Correia¹ and Eliana Costa e Silva¹

¹CHICESI, ESTG, Politécnico do Porto

E-mail addresses: *aib@estg.ipp.pt; aic@estg.ipp.pt; eos@estg.ipp.pt*

In this work we study how the Entrepreneurial Framework Conditions indicators from Global Entrepreneurship Monitor, are associated with Entrepreneurial Behaviour and Attitudes. Data from 34 European Countries between 2001 and 2019 is modeled by using the General Method of Moments estimator.

Keywords

Entrepreneurship Intentions, Entrepreneurial Activity, GEM, longitudinal data, GMM model.

Global Entrepreneurship Monitor (GEM) is a large-scale database for internationally comparative entrepreneurship that includes information about many aspects of entrepreneurship activities of a large number of countries. It is based on collecting primary data through an Adult Population Survey (APS) of at least 2000 randomly selected adults (18-64 years of age) in each economy. Additionally, national teams collect experts' opinions about components of the entrepreneurship ecosystem through a National Expert Survey (NES) [3]. Entrepreneurship Intentions are defined as the entrepreneurial orientation, vocational aspirations and the desire to own a business [7]. In GEM, this indicator is measured as the percentage of adult-age population who are latent entrepreneurs and who intend to start a business within three years. The early stage Entrepreneurial Activity (TEA) index indicates how many of adult-age population are actively involved in the early-stage entrepreneurial process, either actively trying to start new firms or as owner-managers of new start-ups that are less than 42 months old. This study intends to contribute to the understanding of this emerging concept that plays a vital role on economic development, job creation and innovation ([6], [5]). We aim to understand which GEM Framework Conditions have impact on entrepreneurial intentions (EI) and on TEA. The database used in this work contains the GEM indicators, corresponding to the Entrepreneurial Behaviour and Attitudes (EBA) and Entrepreneurial Framework Conditions (EFC). In particular, we analyze how twelve EFC's indicators relate to EI and TEA through a longitudinal study of an unbalance data on 34 European countries between 2001 and 2019. Entrepreneurship is heterogeneous across countries, according to past GEM data analyses. Hence, it is important to take into account intra-country specificity when conducting this type of analyses, since company performance can variate regarding the economic level of the country. We control for possible endogeneity and unobserved heterogeneity by using a General Method of Moments (GMM) estimator proposed by [1]. This model analyzes autoregressive-distributed lagged models from unbalanced panels with many cross-sectional units observed for relatively few time periods. Similarly to the work of [4] we control for per-capita GDP, to capture overall economic and social context, since country's wealth has a significant effect on the character of its entrepreneurial activity [8]. As instrumental variable we consider the size of country's population since it is likely to impact the supply of individuals that

can be active in the labor force, thus impacting entrepreneurship rates, as explained by [2]. For model diagnostic the Arellano-Bond tests and Hansen J-test were applied. Results show that only a reduce number of EFC's are associated with Entrepreneurship intentions and Early Stage Entrepreneurial Activity.

Acknowledgements: This work has been partially supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

References

- [1] M. Arellano and O. Bover. Another look at the instrumental variable estimation of error-components models, *Journal of econometrics* **68(1)**: 29–51, 1995.
- [2] D. M. Hechavarría and A. E. Ingram. Entrepreneurial ecosystem conditions and generated national-level entrepreneurial activity: a 14-year panel study of GEM, *Small Business Economics* **53(2)**: 431–458, 2019.
- [3] M. Herrington and P. Kew. GEM 2016/17 global report. *Global Entrepreneurship Research Association*, 2017.
- [4] J. Levie and E. Autio. Entrepreneurial framework conditions and national-level entrepreneurial activity: Seven-year panel study, *Third Global Entrepreneurship Research Conference*, 1–39, 2007.
- [5] M. Pruett, R. Shinnar, B. Toney, F. Llopis and J. Fox. Explaining entrepreneurial intentions of university students: a cross-cultural study, *International Journal of Entrepreneurial Behavior & Research*, 2009.
- [6] R. Remeikiene, D. Dumciuvienė, G. Startienė and others. Explaining entrepreneurial intention of university students: The role of entrepreneurial education, *Active Citizenship by Knowledge Management & Innovation: Proceedings of the Management, Knowledge and Learning International Conference 2013*, 299–307, 2013.
- [7] E. R. Thompson. Individual entrepreneurial intent: Construct clarification and development of an internationally reliable metric, *Entrepreneurship theory and practice*, **33(3)**: 669–694, 2009.
- [8] A. Van Stel, M. Carree and R. Thurik. The effect of entrepreneurial activity on national economic growth, *Small business economics*, **24(3)**: 311–321, 2005.

Harvesting policies with stepwise effort in random environments

Nuno M. Brites^{1,2} and Carlos A. Braumann^{3,4}

¹ISEG/UL - Universidade de Lisboa, Department of Mathematics, Portugal

²REM - Research in Economics and Mathematics, CEMAPRE, Portugal

³Departamento de Matemática, Escola de Ciências e Tecnologia, Universidade de Évora, Portugal

⁴Centro de Investigação em Matemática e Aplicações, Instituto de Investigação e Formação Avançada, Universidade de Évora

E-mail addresses: *nbrites@iseg.ulisboa.pt; braumann@uevora.pt*

In a randomly varying environment, we can describe the evolution of a fished population size using stochastic differential equations. For the logistic and the Gompertz stochastic growth models, we have previously (see [2,3]) compared the profit performance of two harvesting policies, one with variable harvesting effort, called optimal policy, and the other with constant harvesting effort, called optimal sustainable policy. The former is characterized by fast and abrupt variations of the harvesting effort associated with the frequent variations in population size due to the random environmental fluctuations. This type of policy is inapplicable due, for instance, to the logistics of the fisheries being incompatible with abrupt and frequent changes in the harvesting effort. It also poses social problems during the periods of no or low harvesting effort. Furthermore, this type of policy requires the knowledge of the population size at each instant and estimating population size is an inaccurate, lengthy and expensive task. The optimal sustainable policy considers the constant application of the same harvesting effort and leads to population sustainability, as well as to the existence of a stationary probability density for the population size (see [1]). This policy has the advantage of being easily applicable and there is no need to estimate the population size at every instant. The performance of the two policies was compared in terms of the profit over a finite time horizon. Using data based on a real fished population, we show that there is only a slight reduction in profit by using the optimal sustainable policy (based on constant effort) instead of the inapplicable optimal policy (based on variable effort).

Since the optimal variable effort policy is not applicable, we present here *stepwise policies* (see [4]), which are sub-optimal policies where the harvesting effort is determined at the beginning of each year (or of each biennium) and kept constant throughout that year (or biennium). These policies are not optimal and pose social problems common to the optimal policies, but have the advantage of being applicable, since the changes in harvesting effort are much less frequent and are compatible with the fishing activity. Furthermore, even though estimating population size is still required, that needs to be done less frequently. We present the comparison in terms of profit of these stepwise policies with the two previously mentioned policies.

Keywords

Fisheries Management, Stochastic Differential Equations, Profit Optimization, Stepwise Effort.

Acknowledgements: Nuno M. Brites was partially supported by the Project CEMAPRE/REM - UIDB/05069/2020 - financed by FCT/MCTES through national funds. Carlos A. Braumann belongs to the research centre Centro de Investigação em Matemática e Aplicações, Universidade de Évora, supported by FCT (Fundação para a Ciência e a Tecnologia, Portugal), project UID/04674/2020.

References

- [1] C. A. Braumann. Variable effort fishing models in random environments. *Mathematical Biosciences* **156**: 1–19, 1999.
- [2] N. M. Brites and C. A. Braumann. Fisheries management in random environments: Comparison of harvesting policies for the logistic model. *Fisheries Research* **195**: 238–246, 2017.
- [3] N. M. Brites and C. A. Braumann. Fisheries management in randomly varying environments: Comparison of constant, variable and penalized efforts policies for the Gompertz model. *Fisheries Research* **216**: 196–203, 2019.
- [4] N. M. Brites and C. A. Braumann. Harvesting in a random varying environment: Optimal, stepwise and sustainable policies for the gompertz model. *Statistics, Optimization and Information Computing* **7(3)**: 533–544, 2019.

A cluster analysis of the business behavior and attitudes indicators defined by global entrepreneurship monitor

Mariana Cardante^{1,2} and Aldina Correia^{1,2}

¹ CIICESI – Center for Research and Innovation in Business Sciences and Information Systems,

² Escola Superior de Tecnologia e Gestão, Politécnico do Porto, Portugal

E-mail addresses: *8140395@estg.ipp.pt; aic@estg.ipp.pt; amb@estg.ipp.pt; vbarga@estg.ipp.pt*

Entrepreneurship has been pointed out as a key contributor to sustained economic growth and development as it not only creates employment, but increases spending in markets, knowledge transfers, employment and innovation, according to Quadrini and Vincenzo in [5]. Global Entrepreneurship Monitor (GEM) has been tracking rates of entrepreneurship across multiple phases of entrepreneurial activity, assessed the characteristics, motivations and ambitions of entrepreneurs and explored the attitudes societies have towards this activity in a wide range of countries, according to 2017/2018 Global Report. GEM data has been used in numerous studies on entrepreneurship [1,3,4] as well as studies that highlight the important role of GEM in the area [2]. This research focuses on fifteen indicators of Business Behavior and Attitudes defined by GEM in forty countries in the year of 2018. The main objective is to apply the multivariate statistical technique called clustering to identify homogeneous groups of countries, in order to discover patterns in the indicators of entrepreneurship in groups of countries.

The results obtained were quite satisfactory: twelve of the fifteen indicators allowed a good distinction between the two clusters formed, according to the parametric ANOVA test. The second cluster has three times fewer countries than the first cluster and has nine of the fifteen indicators above the average, in contrast to the first cluster, which only has eight. The results also show that on the one hand, in the first cluster the indicator that is significantly above the average is Business Services Sector and in the second cluster is the Entrepreneurial Intentions. On the other hand, the Entrepreneurial Intentions and Income Level indicators are significantly below the average for the both clusters.

Keywords

Entrepreneurial Activity , Entrepreneurship Indicators , Global Entrepreneurship Monitor, Hierarchical Cluster Analysis , K-means Cluster Analysis , Multivariate Statistics.

Acknowledgements: This work has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) under the project UID/GES/04728/2020.

References

- [1] S. Z. Ahmad and S. R. Xavier. Entrepreneurial environments and growth: evidence from malaysia gem data. *Journal of Chinese Entrepreneurship*, 2012.
- [2] N. Bosma. The global entrepreneurship monitor (gem) and its impact on entrepreneurship research. *Foundations and Trends in Entrepreneurship* **9(2)**: 2013.
- [3] A. C. Martínez, J. Levie, D. J. Kelley, R. J Sæmundsson and T. Schøtt. Global entrepreneurship monitor special report: A global perspective on entrepreneurship education and training. 2010.
- [4] M. Herrington, J. Kew, P. Kew and G. E. Monitor. *Tracking entrepreneurship in South Africa: A GEM perspective*. Graduate School of Business, University of Cape Town South Africa, 2010.
- [5] V. Quadrini. The importance of entrepreneurship for wealth concentration and mobility. *Review of income and Wealth* **45(1)**: 1–19, 1999.

The role of framework conditions in income level of countries

Aldina Correia^{1,2}, Vítor Braga^{1,2} and Sofia Gonçalves^{1,2}

¹CHICESI – Center for Research and Innovation in Business Sciences and Information Systems,

²Escola Superior de Tecnologia e Gestão, Politécnico do Porto, Portugal

E-mail addresses: *aic@estg.ipp.pt; vbraga@estg.ipp.pt; 8130365@estg.ipp.pt*

In this work GEM 2018 data indicators are considered, in particular is studied Entrepreneurial Framework Conditions (EFC) significance in Income Level of countries, using multivariate discriminant analysis and factor analysis.

The results shown that it appears that many indicators probably do not contribute to discriminate Income Level, and also multicollinearity problems are observed. Thus a Factor Analysis for EFC is considered and it was observed that Factor 1, including Physical and services infrastructure, Internal market openness, Financing for entrepreneurs, R & D transfer, Governmental programs, Commercial and professional infrastructure, is significant to discriminate Level Income, being possible correctly classify 69,8% of original grouped cases.

Keywords

Entrepreneurship, GEM Framework Conditions, Global Entrepreneurship Monitor, Multivariate Discriminant Analysis, Factor Analysis.

Entrepreneurship is highly recognized as an important international subject, often defined as largely responsible for new jobs, new projects and boosting economies, as described by Wach et al. (2016) [12].

Valliere (2010) [11] defined Entrepreneurial Framework Conditions as the environmental conditions that encourage and support entrepreneurial activity at the national level. In this work the 2028 GEM Framework Conditions indicators from GEM are considered. GEM (Global Entrepreneurship Monitor) is an investigation program about the entrepreneurship around the world. Datasets from this project are considered by several authors, [1,4,5,8–10], and it allows describe, analyze and compare the entrepreneurship for the different economies. GEM project began in 1999 and is conceived to analyse the phenomenon of entrepreneurship conducting two types of surveys: the Adult Population Survey (APS) and the National Expert Survey (NES). The GEM Framework Conditions from GEM includes Entrepreneurship Behavior and Attitudes (EBA) and Entrepreneurial Framework Conditions (EFC). EFC includes: Financing For Entrepreneurs, Governmental Support And Policies, Taxes And Bureaucracy, Governmental Programs, Basic School Entrepreneurial Education And Training, Post School Entrepreneurial Education And Training, R and D Transfer, Commercial And Professional Infrastructure, Internal Market Dynamics, Internal Market Openness, Physical And Services Infrastructure, Cultural And Social Norms and Business Services Sector.

In particular this study focuses on the role of EFC indicators in Income Level of countries.

The approach presented in this work is part of a more general work about the effects of other GEM Framework Conditions in Entrepreneurship Intentions (EI) and Income Level (EL) of countries, [3,6,7].

Bosma and Kelley (2019) [2] underline the importance of Income Level, because different regions of the world has different economies. Thus the significance of framework conditions in Income Level is studied using multivariate discriminant analysis and GEM 2018 data indicators are considered. Observing the tests, for a significance level of 5%, it appears that many variables probably do not contribute to discriminate Income Level, and also multicollinearity problems are observed. Thus a Factor Analysis for EFC is considered and it was observed that Factor 1, including Physical and services infrastructure, Internal market openness, Financing for entrepreneurs, R & D transfer, Governmental programs, Commercial and professional infrastructure, is significant to discriminate Level Income, being possible correctly classify 69,8% of original grouped cases.

In future a cluster analysis would be also a possibility in order to identify countries with similar Entrepreneurship Intentions. Other possibility can be to do the same study using a longitudinal approach, both in regression and cluster analysis.

Acknowledgements: This work has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e a Tecnologia under the project UID/GES/04728/2020

References

- [1] V. Braga, M. Queirós, A. Correia and A. Braga, High-Growth Business Creation and Management: A multivariate quantitative approach using GEM data. *Journal of the Knowledge Economy* **9(2)**: 424–445, 2018.
- [2] N. Bosma and D. Kelley. *GEM 2018/2019 global report*, 2019.
- [3] A. Correia, V. Braga and S. Gonçalves. Country income levels and entrepreneur's perceptions. 4th Contemporary Issues in Economy & Technology Conference – *CIET 2020*, 2020.
- [4] A. Correia, E. Costa e Silva, I. C. Lopes and A. Braga. MANOVA for distinguishing experts' perceptions about entrepreneurship using NES data from GEM. *AIP Conference Proceedings* **1790(1)**: 140002, 2016.
- [5] A. Correia, E. Costa e Silva, I. C. Lopes, A. Braga and V. Braga. Experts' perceptions on the entrepreneurial framework conditions. *AIP Conference Proceedings* **1906(1)**: 110004, 2017.
- [6] S. Gonçalves, A. Correia and V. Braga. The role of Entrepreneurial Intentions in Income Level, using GEM Framework Conditions. 7th Global Conference on Business Management and Social Sciences *7th GCBMS 2020*, 2020.
- [7] S. Gonçalves, A. Correia and V. Braga. Entrepreneurship Intentions: the role of framework conditions, using GEM. *ESTG Masters*, 2020.
- [8] E. Costa e Silva, A. Correia and F. Duarte. How Portuguese experts' perceptions on the entrepreneurial framework conditions have changed over the years: A benchmarking analysis. *AIP Conference Proceedings*, 2040(1), 110005, 2018.
- [9] M. F. Pilar, M. Marques and A. Correia. New and growing firms' entrepreneurs' perceptions and their discriminant power in EDL countries. *Global Business and Economics Review* **21(3-4)**: 474–499, 2019.
- [10] C. Sampaio, A. Correia, V. Braga and A. Braga. Environmental variables and entrepreneurship – a study with GEM NES global individual level data. *International Journal of Knowledge Based Development*, 2040, 4476, 2016.
- [11] D. Valliere. Reconceptualizing entrepreneurial framework conditions. *International Entrepreneurship and Management Journal* **6(1)**: 97–112, 2010.
- [12] D. Wach, U. Stephan and J. Gorgievski. More than money: Developing an integrative multifactorial measure of entrepreneurial success. *International Small Business Journal* **34(8)**: 1098–1121, 2016.

A novel predator-prey strategy for optimization problems: preliminary study

Sérgio Cavaleiro Costa^{1,2}, Fernando M. Janeiro^{1,3} and Isabel Malico^{1,2}

¹University of Évora, Portugal

²LAETA-IDMEC, Portugal

³Instituto de Telecomunicações, Portugal

E-mail addresses: *smcac@uevora.pt; fmtj@uevora.pt; imbm@uevora.pt*

Several evolutionary algorithms are described in the literature, but few try to mimic the predator-prey behavior. A novel Predator-Prey Algorithm is presented here, where several species may prey each other and reproduce, pushing the overall population to the best solutions of the cost function. This algorithm not only proved to be capable of finding solutions, but a niching behavior emerged from the algorithm itself, which is very desirable in multiobjective problems.

Keywords

Optimization, Evolutionary Algorithms, Predator-prey Model, Benchmark Functions.

Evolutionary strategies play a crucial role in optimization algorithms. Both because of the complexity of the problems or of the unavailability of a better alternative, approaches like Genetic Algorithms, Particle Swarm Algorithms, Bee Swarm Algorithms, Firefly Algorithms, and Ant Colony Algorithms, among many others, are often used [1]. All of them are nature-inspired, make use of swarm intelligence, and usually are capable of finding a global solution.

Of the different evolutionary strategies available, few try to mimic the predator-prey behavior. In one of the most relevant works on the subject, a set of agents evolve based on the mutation of the worst candidate from a prey to a predator, which then pushes the overall population in a descent direction [2]. In the approach of Li et al. [3], a lattice is used to support the PPA operations in a multi-objective problem. Another example of the use of a PPA strategy in a multi-objective problem is presented in the work of Laumanns et al. [4].

In this work, a novel strategy is proposed. It is based on the assumption that the search space is the habitat of the species and that the fitter the species get (i.e. closer to a solution), the more sedentary it gets. A species is a set of agents with particular characteristics such as sight range, nominal locomotion speed, longevity, and feeding rate. The position of an agent is the set of parameters to be optimized. The sight range is a sphere centered on the agent position and is used to identify predators from which to run away and other members of the species to approach because they are breeding candidates. Longevity refers to the number of iterations an agent of a given species is alive. To balance the decision between hunting and reproduction, the probability for an agent to hunt is

$$P(\text{Hunting}) = \text{Hunger} \times \text{Feeding Rate} \times \text{Prey Availability}.$$

This probability is directly related to the hunger, prey availability and feeding rate, defined as the rate a species must eat. As already said, the agent becomes more sedentary, the more its location is closer to the solution. The displacement of the agent (step) is given

by the multiplication of the nominal locomotion speed of the species and a function $f(\mathbf{x})$, which penalizes the speed.

$$\text{step} = \text{nominal speed} \times 2 \times \arctan\left(\frac{f(\mathbf{x}) - f(\mathbf{x}_{\text{best}})}{\pi}\right) \times \mathbf{d}$$

where \mathbf{x} is the position of the agent, $f(\mathbf{x}_{\text{best}})$ the best overall solution found so far, and \mathbf{d} the direction vector.

Four benchmark functions were used to confirm the convergence of the method: Goldstein Price, Egghold, Ackley, and Shubert. To proceed with the evaluation, 3 species were defined, all with the same characteristics, i.e., 30 agents each, sight range of 0.3, a feeding rate of 0.1, and a nominal locomotion speed of 0.1. For all the Benchmark functions the proposed PPA was able to converge to the absolute minimum. The next step will be to apply the proposed method to practical optimization problems.

Acknowledgements: This work was financed by the Fundação para a Ciência e Tecnologia through the PhD scholarship BD/139113/2018 and by the projects UID/50022/2020 and UID/EEA/50008/2019.

References

- [1] A. K. Kar. Bio inspired computing - A review of algorithms and scope of applications. *Expert Systems with Applications* **59**: 20–32, 2016.
- [2] S. L. Tilahun and H. C. Ong. Prey-predator algorithm: A new metaheuristic algorithm for optimization problems. *International Journal of Information Technology and Decision Making* **14(6)**: 1331–1352, 2015.
- [3] X. Li. A Real-Coded Predator-Prey Genetic Algorithm for Multiobjective Optimization. In *International Conference on Evolutionary Multi-Criterion Optimization*, 207–221, 2003.
- [4] M. Laumanns, G. Rudolph and H. P. Schwefel. A spatial predator-prey approach to multi-objective optimization: A preliminary study. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, **1498** LNCS, 241–249, 1998.

Clusterwise regression for interval data

Nikhil Suresh¹, Paula Brito^{1,2} and Sónia Dias^{2,3}

¹Faculdade de Economia, Universidade do Porto, Portugal

²LIAAD - INESC TEC, Portugal

³ESTG, Instituto Politécnico de Viana do Castelo, Portugal

E-mail addresses: *up201801861@fep.up.pt; mpbrito@fep.up.pt; sdias@estg.ipv.pt*

In classical data analysis, data is usually represented as an array where rows represent individuals and columns represent the variables (or attributes). It is possible to represent the data in a two-dimensional array of n rows and p columns because only one single value, numerical or categorical, is recorded for each variable and for each individual. However, when data is grouped to a higher level, the classical solution which consists in using the mean, median or mode to represent each group leads to a loss of information, especially the variability present in each group. In situations as such, Symbolic Data Analysis [3,2] provides a framework to represent data with inherent variability by using special variables. Among these representations, the focus in this work is on interval-valued data. A combination of existing dynamic clustering techniques [1] and the Interval Distributional (ID) regression model [4] for interval-valued data is proposed. A uniform distribution is assumed within each observed interval, which is then represented by the corresponding quantile function. The error between predicted and observed intervals is evaluated using the Mallows Distance. The proposed approach consists in clustering the data, and for each cluster, fit a regression model. The data is then re-partitioned, by verifying which regression model fits each observation best. This algorithm is iterated until the best clusters and corresponding regression models fit the data, locally optimizing a criterion that measures the fit between the model associated with each cluster and its members. The process is replicated for different initial partitions. The obtained solutions may be evaluated by different criteria, e.g. weighted Coefficient of Determination [4], (adapted) Silhouette Coefficient [3] or Root Mean Square Errors. The final clusters can then be used to predict target intervals for new observations. The developed model is applied on real interval-valued data to illustrate the behaviour and success of the proposed method.

Keywords

Data With Variability, Interval-Valued Variables, Symbolic Data Analysis, Interval Distributional Regression Model.

Acknowledgements: This work is financed by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia, within project UIDB/50014/2020.

References

- [1] E. Diday and J. J. Simon. Clustering Analysis. *In: K. S. Fu (Ed.), Digital Pattern Recognition. Springer, Heidelberg.* 47–94, 1976.
- [2] H.-H Bock and E. Diday (Eds.) *Analysis of Symbolic Data. Exploratory methods for extracting statistical information from complex data.* Springer, Heidelberg, 2000.
- [3] P. Brito. Symbolic Data Analysis: Another Look at the Interaction of Data Mining and Statistics. *WIREs Data Mining and Knowledge Discovery.* **4(4)**: 281–295, 2014.
- [4] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* *20*. 53–65, 1987.
- [5] S. Dias and P. Brito. Off the beaten track: a new linear model for interval data. *European Journal of Operational Research.* **258(3)**: 1118–1130, 2017.

A wavelet based neural network scheme (NNS) for supervised and unsupervised learning

Manuel L. Esquível¹ and Nadezhda P. Krasii²

¹Department of Mathematics and CMA FCT NOVA, New University of Lisbon, Campus de Caparica, 2829-516 Monte de Caparica, Portugal.

²Don State Technical University, Gagarin Square 1, Rostov-on-Don 344000, Russian Federation.

E-mail addresses: *mle@fct.unl.pt; krasnad@yandex.ru*

We introduce a scheme for supervised and unsupervised learning based on successive decomposition of random inputs by means of wavelet basis. To each successive layer corresponds a different wavelet basis with more null moments than its predecessor. The random wavelet coefficients are supposed to be Gaussian distributed and constitute the inner *representation of the world* in the scheme. We consider two types of operations in the scheme: firstly, a input signal treatment phase – the *awake* phase – and, secondly, a reorganizing phase of the random wavelet coefficients obtained in the previous *awake* phase – the *asleep* phase. The next *awake* phase input treatment will include a feedback from the previous *asleep* phase. We show that in the case of a constant average value of the inputs in each successive *awake* phases, the mean value of the feedback converges to a multiple of the constant average value of the inputs.

Keywords

Neural Networks, Supervised Learning, Unsupervised Learning, Wavelets.

$(\Omega, \mathcal{F}, \mathbb{P})$ is a complete probability space. The first *awake* phase for a NNS with two levels is given by a *realization* indexed by $\omega \in \Omega$ of a single input $f(\omega)$ – for $f : \mathbb{R} \mapsto \mathbb{R}$ – decomposed by a wavelet $L^2(\mathbb{R})$ basis $(\psi_{ij}^1)_{i,j}$ with n_1 null moments in the first layer of the NNS, as:

$$\forall x \in \mathbb{R}, f(\omega)(x) = \sum_{i,j} \langle f(\omega), \psi_{ij}^1 \rangle \psi_{ij}(x) = \sum_{i,j} \hat{f}_{ij}^{1,1}(\omega) \psi_{ij}^1(x).$$

Coefficients $\{\hat{f}_{ij}^{1,1}(\omega) : \omega \in \Omega\}$ define $\mathbb{X}_{ij}^{1,1} \sim \mathcal{N}(\mu_{ij}^{1,1}, (\sigma^2)_{ij}^{1,1})$. A *cut-off* threshold $\delta^1 > 0$ and define:

$$(\mathbb{X}_{ij}^{1,1})^{\delta^1} = \begin{cases} \mathbb{X}_{ij}^{1,1} & \text{if } |\mathbb{E}[\mathbb{X}_{ij}^{1,1}]| = |\mu_{ij}^{1,1}| \geq \delta^1 \\ 0 & \text{if not} \end{cases}.$$

If $|\mu_{ij}^{1,1}| \geq \delta^1$ the reconstructed input for those indices is in the set $I_1 := \{ij : \mu_{ij} \geq \delta^1\}$ is:

$$f_{\delta^1}(\omega)(x) = \sum_{i,j \in I_1} \langle f(\omega), \psi_{ij}^1 \rangle \psi_{ij}(x) = \sum_{i,j \in I_1} \hat{f}_{ij}^{1,1}(\omega) \psi_{ij}^1(x)$$

and its decomposition by another wavelet $L^2(\mathbb{R})$ basis $(\psi_{ij}^2)_{i,j}$ with $n_2 > n_1$ null moments in the second layer of the NNS, is:

$$\forall x \in \mathbb{R}, f_{\delta^1}(\omega)(x) = \sum_{i,j} \langle f(\omega), \psi_{ij}^2 \rangle \psi_{ij}(x) = \sum_{i,j} \hat{f}_{ij}^{2,1}(\omega) \psi_{ij}^2(x).$$

Coefficients $\{\hat{f}_{ij}^{2,1}(\omega) : \omega \in \Omega\}$ define $\mathbb{X}_{ij}^{2,1} \sim \mathcal{N}(\mu_{ij}^{2,1}, (\sigma^2)_{ij}^{2,1})$. First *asleep* phase for a NNS with two levels is the structuring given by $(\mathbb{X}_{ij}^1)_{i,j}$ and $(\mathbb{X}_{ij}^1)_{i,j}$ obtained in the first *awake* phase. Joint distributions of $(\mathbb{X}_{ij}^{1,1})_{i,j}$ and $(\mathbb{X}_{ij}^{2,1})_{i,j}$ are Gaussian and $\mathbb{X}^{1,1} \equiv (\mathbb{X}_{ij}^{1,1})_{i,j} \sim \mathcal{N}(\mu^{1,1}, \Sigma^{1,1})$ and $\mathbb{X}^{2,1} \equiv (\mathbb{X}_{ij}^{2,1})_{i,j} \sim \mathcal{N}(\mu^{2,1}, \Sigma^{2,1})$. Joint distribution of $\mathbb{X}^{1,1}$ and $\mathbb{X}^{2,1}$ is Gaussian $\mathcal{N}(\nu_1, \Phi_1)$. Second *awake* phase is similar as the first but the inputs wavelet analysis will be modified by a *feedback* influence of the second level upon the first, given by $\mathbb{E}[\mathbb{X}^{1,1} | \mathbb{X}^{2,1}]$. After 2 steps in the training phase, the first level recursion is given by:

$${}^*\mathbb{X}_{ij}^{1,3} := \mathbb{X}_{ij}^{1,3} + \alpha \mathbb{E}[\mathbb{X}_{ij}^{1,2} | {}^*\mathbb{X}_{ij}^{2,2}] + \alpha^2 \mathbb{E}[\mathbb{E}[\mathbb{X}_{ij}^{1,1} | {}^*\mathbb{X}_{ij}^{2,1}] | {}^*\mathbb{X}_{ij}^{2,2}].$$

Theorem 1 (Stabilization of the feedback restructuring) Suppose that all the inputs f are Hölder continuous with exponents bounded by $0 < \rho < 1$, and that for some constant c , $\forall x, y \in \mathbb{R}$, $|f(x) - f(y)| \leq c|x - y|^\rho$. Consider the feedback recursion given by ${}^*\mathbb{X}_{ij}^{1,n} = \mathbb{X}_{ij}^{1,n} + Z_{n,i,j}^\alpha$ with $0 < \alpha < 1$ and

$$\begin{aligned} Z_{n,i,j}^\alpha &:= \alpha \mathbb{E}[\mathbb{X}_{ij}^{1,n-1} | {}^*\mathbb{X}_{ij}^{2,n-1}] + \dots \\ &\dots + \alpha^{n-1} \mathbb{E}[\mathbb{E}[\dots \mathbb{E}[\mathbb{X}_{ij}^{1,1} | {}^*\mathbb{X}_{ij}^{2,1}] | \dots] \dots | {}^*\mathbb{X}_{ij}^{2,n-1}]. \end{aligned} \quad (1)$$

Suppose $\mathbb{X}_{ij}^{1,k} \sim \mathcal{N}(\mu_{ij}^{1,k}, (\sigma^2)_{ij}^{1,k})$ and, for constants A_{ij} and B_{ij} , $\forall k \geq 1$, $|\mu_{ij}^{1,k}| \leq A_{ij}$ and $|(\sigma^2)_{ij}^{1,k}| \leq B_{ij}$. Then, the sequence $(|\mathbb{E}[Z_{n,i,j}^\alpha]|)_{n \geq 2}$ **has a subsequence that converges** to some constant $z_{\infty,i,j}^\alpha$ and such that for some $M_{i,j} = M_{ij}(A_{ij}, B_{ij}, \rho) > 0$ we have:

$$z_{\infty,i,j}^\alpha \leq \frac{\alpha}{1 - \alpha} M_{i,j}.$$

Acknowledgements: CMA project UID/MAT/00297/2020; RFBR grant 19-01-00451.

References

- [1] M. Huang and B. Cui. A novel learning algorithm for wavelet neural networks. In: *Advances in Natural Computation. ICNC 2005. Lecture Notes in Computer Science* **3610**: 1–7, Springer, Berlin, Heidelberg, 2005.
- [2] Z. Zhang, Y. Shi, H. Toda and T. Akiduki. A study of a new wavelet neural network for deep learning. In: *2017 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, 127–131 (July 2017). DOI: 10.1109/ICWAPR.2017.8076676.

Bivariate multilevel modelling of Portuguese student achievement

Susana Faria¹ and Carla Salgado²

¹Department of Mathematics, Centre of Molecular and Environmental Biology,
University of Minho, Portugal

²Department of Mathematics, University of Minho, Guimarães, Portugal

E-mail address: *sfaria@math.uminho.pt*

The purpose of this study was to investigate the relationship between Portuguese students' mathematics and science test scores and the characteristics of students and schools themselves. The results showed that the index of the socioeconomic status of the student, being a male student and the average index of the socioeconomic status of the student in school, positively influence the students' performance in mathematics and science. On the other hand, the grade repetition had a negative influence on the performance of the Portuguese student in mathematics and science.

Keywords

Multilevel Models, Bivariate Models, PISA Data.

Multilevel regression models are well-suited to the analysis of educational data because they recognize the hierarchical structure of the data with students nested in schools. These models simultaneously investigate relationships within and between hierarchical levels of grouped data, thereby making it more efficient at accounting for variance among variables at different levels than other existing analyses [2].

This study using data from the Programme for International Student Assessment (PISA) [1] investigates the factors from both student and school perspectives, that impact the mathematics and science achievement of 15-year-old Portuguese students.

To conduct the analysis, a bivariate two-level linear model was used to model the variation in mathematics and science achievements as a function of student - (level 1) and school - (level 2) factors. Using this bivariate model it is possible to compare the associations between students' characteristics or school resources and their performances in mathematics and science achievements [3].

Acknowledgements: This work was supported by the strategic programme UID-BIA-04050-2019 funded by national funds through the FCT.

References

- [1] OECD. *What is PISA? in PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematic and Financial Literacy*, OECD Publishing, Paris, 2016.
- [2] H. Woltman, A. Feldstain, J. C. MacKay and M. Rocchi. An introduction to hierarchical linear modeling. *Tutorials in Quantitative Methods for Psychology* **8**(1): 52–69, 2012.
- [3] C. Masci, F. Ieva, T. Agasisti and A. M. Paganoni. Bivariate multilevel models for the analysis of mathematics and reading pupils' achievements, *Journal of Applied Statistics*, **44**(7): 1296–1317, 2017.

General theory and construction of k-dimensional survival functions

Jerzy K. Filus

Dept. of Mathematics and Computer Science, Oakton College, USA

E-mail address: *jkfilus98@gmail.com*

The theory and methods of construction of survival functions of any dimension k is mainly based on earlier developed general theory of bivariate probability distributions. So, given k ($k = 3, 4, \dots$) univariate survival functions, we present the method of construction of classes of k -variate probability distributions whose marginals are those, originally given, univariates. This is a nice fact that in many cases the obtained k -variate distributions can be, in a sense, reduced to the bivariate distributions earlier analyzed. Namely, the joiner of the k -variate model can be obtained as a simple sum of the formerly investigated bivariate joiners. The analytic conditions for such k -variate joiners, to be proper once, will be presented. As a special case we analyze the situation when $k = 3$. The analytical "situation" turns out to be very interesting. Namely, the two possible **modes of stochastic dependences** emerge. For the first mode three random variables are pairwise dependent but then, as we will show, they must be "3 - independent" (the notion defined throughout). Consequently, the 3-dependence totally exclude any bi-dependence. Similarly, in higher dimensions, any r -dependence (r not greater than k) exclude occurring any $(r - 1)$ - dependence among the r random variables which are subjected to the r - dependence. This fact significantly simplifies the stochastic dependences structure, the phenomena, probably, **unknown** in the current **literature**. Some analytical conditions for existence of the so defined models will also be presented. Remarkably, this kind of stochastic modeling for higher than two or three dimensions by means of the joiners seems evidently be easier, as well as more accurate in applications, than that by means of copulas.

Theoretical aspects and constructions of bivariate probability distributions, most general case

Lidia Z. Filus

Dept. of Mathematics, Northeastern Illinois University, USA

E-mail address: *L-Filus@neiu.edu*

Given, initially, any two marginal probability distributions, the general form of (all) bivariate distributions (in the form of two-dimensional survival functions) as the universal form via **joiner** representation will be presented. In addition, methods of the construction of any such bivariate model is described. Generality of the model's form, which is universal, allows to consider it as competitive to the **copula** methodology. Also, in many cases our method of the construction is evidently simpler and more efficient than the construction by the use of copulas. Especially important fact is that, in most of practical problems, the association between a modeled *physical* reality and a joiner, which determines the stochastic model (probability distribution), is much more straightforward than when the copula methodology is applied. On the other hand, unlike in the case of copulas, not every joiner fits to a given pair of marginal distributions. Therefore, methods of determining classes of joiners (so the models) that fit to a given pair of univariate distributions will be presented. In most of the important cases, such classes (or subclasses) of the models turn out to be easy determined. In the so defined "continuous cases", a "*canonical*" class of bivariate survival functions, given two marginal hazard rates is automatically given. In all the, mentioned above, constructions the (only) initial data required are two **marginal** survival functions. Other form of bivariate models, that also will be presented, is given by the **baseline** distributions representation which, in general, are different than the marginal. These two baseline distributions (as the input data) are very often met in practical situations. The relationship between the baseline and the marginal representations of the same bivariate model will be presented too.

Factors influencing exports and season sales tend in a footwear industry

Raquel Francisco¹, Eliana Costa e Silva¹, Sandra P. Sousa² and
Vitor Braga¹

¹CHICESI, ESTG, Politécnico do Porto

²NIPE, Universidade do Minho

E-mail addresses: *8080214@estg.ipp.pt; eos@estg.ipp.pt; sis@estg.ipp.pt; vbrega@estg.ipp.pt*

Exporting involves sunk costs and the increase in exports is an important component for companies. The footwear industry is no exception. Exporting 95% of its production, it is one of the sectors that most positively contributes to the trade balance. This study aims to analyze the determining factors of exports of a shoe exporting company, as well as the trend of seasonality of sales. For this real data of the company from January 2017 to June 2020 is considered. As for the econometric study, 42-month sample, a time series was formed considering several theoretically relevant explanatory variables found in the literature, namely productivity, Equity, number of workers, wages and lagged sales variable.

To explain the evolution of exports, two econometric models are presented. Model 1 explores the determining factors that influence exports. The results show that the variables productivity, number of workers, equity and sales lag have an impact on the value of exports. However, the wage variable was not statistically significant. In model 2, the fixed effects of time were added in order to control time specific effects and to analyse the seasonality trend. The results showed that the sales lag is no longer significant. However, the time series exhibits seasonality, shown be statistical coefficient for the months relative to the production of the winter collection.

Keywords

Seasonality, Determinants of Export, Econometric Study.

In a growing global world and with an increase in international trade, it is important to increase the competitive capacity of economies and companies in order to guarantee their success in an increasingly demanding foreign market. The Portuguese footwear industry is one of the most competitive in the world, exporting its products to the five continents.

In the present econometric study, the value of the company's exports (values in Euros) for the 42-month period from January 2017 to June 2020 is used as the dependent variable. Based on the literature review, several explanatory variables were considered, namely Productivity (Prod) more productive companies are better able to explore export markets, this plays an important role in companies and intensifies exports[4]; Number of workers (NTr) the more workers a company has, the more likely it is to be an exporter and therefore have higher levels of production, resources and capacity [1]; Wages (Rem) studies like that of Srinivasan and Archana [2] that argue that higher wages cause an increase in production costs and thus decrease the company's production and competitiveness; Equity (CP) companies with higher equity are more likely to export [3], because it has a greater capacity to finance itself, showing that the company is financially healthy; Sales

lag (LagExp), companies that export in a previous period are more likely to export in the future than those that previously do not export [1]; Fixed time effect (Dtempo) in order to analyze the seasonal trend, a fixed time effect was added to the model, in order to control macroeconomic shocks and other unobservable effects that may exist and were not taken into account.

Thus, the econometric models developed were as follows:

$$EXP_t = \beta_1 + \beta_2 Prod_t + \beta_3 NTr_t + \beta_4 Rem_t + \beta_5 CP_t + \beta_6 LagExp_t + u_t \quad (4)$$

$$EXP_t = \beta_1 + \beta_2 Prod_t + \beta_3 NTr_t + \beta_4 Rem_t + \beta_5 CP_t + \beta_6 LagExp_t + \beta_7 Dtempo + u_t \quad (5)$$

Regarding the model 4, it was found that the explanatory variables Prod, NTr, CP and LagExp present statistically significant coefficients. Regarding the coefficient obtained for the explanatory variable productivity, the increase of one percentage point in productivity causes a positive impact on the value of exports, keeping everything else constant. The results obtained for the explanatory variable NTr show that the increase of one employee in the company has a positive impact in the value of exports, ceteris paribus. However, the coefficient of the variable CP has a negative impact on the dependent variable, that is, the increase of in equity causes a decrease in the value of exports. Lagged exports are very significant for the estimated model, suggesting that exports from previous years increase the likelihood of increasing the value of exports today.

The fixed effect of time added to the model 5 in to control the specific time effects, it also allows us to draw some conclusions regarding the seasonality trend of sales. It is concluded that there are seven months (from March to September) whose coefficient are statistically significant. Note that these months refer to the production of the winter collection.

References

- [1] S. Girma, D. Greenaway and R. Kneller. *Does Exporting Lead to Better Performance? A Microeconometric Analysis of Matched Firms*, Leverhulme Centre for research on globalisation and economic policy, 2002.
- [2] A. Sterlacchini. The determinants of export performance: A firm-level study of Italian manufacturing, *Review of World Economics* **137(3)**: 450–472, 2001.
- [3] M. J. Roberts and J. R. Tybout. The decision to export in Colombia: An empirical model of entry with sunk costs. *The American Economic Review*, 545–564, 1997.
- [4] D. Greenaway and R. Kneller. Exporting and productivity in the United Kingdom. *Oxford Review of Economic Policy* **20(3)**: 358–371, 2004.

Ontological and probabilistic aspects of assessing predictors quality

Sergey Frenkel

Federal Research Center "Computer Science and Control" of Russian Academy of Science

E-mail address: *fsergei51@gmail.com*

At present, an impressive range of programs is known that implement the functions of predictors of future values of various sequentially presented data sets (for example, predictors from known cloud resources (MS AWS Azure Machine Learning, Google Cloud Machine Learning, etc)). The quality of prediction by one or another predictor depends along with the prediction algorithm also on the properties of the predicted data and prediction goals. For example, for relatively highly correlated numerical data (time series), prediction of absolute values is performed fairly well by regression (ARIMA) predictors. However, if the predicted values are included (from the user's point of view) in rather complex ontology described with data of various types, then the choice of the predictor is not so unambiguous. The paper analyzes the possibility of evaluating the properties of popular predictors as predictors of binary sequences or binary classifiers, regarding data corresponding to a specific user ontology.

Keywords

Data Prediction, Probabilistic Models, Ontological Models.

Let x_1, x_2, \dots, x_n be some (as a sequence), $x_t \in A = \{A_1, \dots, A_k\}$, $t = 1, \dots, n$, A be some alphabet, for example, $\{0, 1\}$. At the moment t , the prediction function f of previous (known) values is used, i.e. it computes the following value $b_{t+1} = f\{x_1, x_2, \dots, x_t\}$, where b_{t+1} is an estimation of x_{t+1} from the observed sequence x_1, x_2, \dots, x_t of length t , b_{t+1} may not coincide with x_{t+1} . The algorithm for calculating functions f can be based either on probabilistic or on semantic-ontological models [1]. There are known approaches to forecasting and classification based on the use of semantic and ontological models supplemented by some probabilistic estimates [2]. Let knowledge about data (semantics) be presented as an ontology, i.e. triplet $O = \{T, Rel, F\}$, where T is the set of concepts of the subject area described by the ontology, Rel is many relationships between concepts, F is an interpretation functions set on entities and/or relations of ontology. For example, let the sequentially observed data set is an alternation of traffic intensity values, and let the data set is a network traffic. Then, for example, $A_1 = L$ (small change in the volume of traffic), and $A_2 = H$ (high value of change), or as Low, and as High and, respectively, can be represented as binary sequence $\{Low(0), High(1)\}$. So, we can consider the concepts of the network traffic ontology as "traffic intensity", and the order of alternation of intensity patterns as a relation between the concepts. Then the user can try to develop a model to predict the concepts that appear in the ontology based on "historical data", which which predicts the appearance of the concept of ontology O at time $t + 1$ using only the data available at time t (in fact, the transition of the ontology state is predicted at the moment

t). For this, estimates can be done according to the "historical" data of the probabilities of the presence of ontology concepts. If we consider the relationship in an ontology as ordering rule (see example outlined above) then a prediction pattern (0 or 1) matches a sequence of events within an event sequence if 1) the events within the prediction pattern match events within the event sequence (say, in the majority branches of decision tree), 2) the ordering constraints expressed in the prediction pattern are obeyed, and 3) the events involved in the match occur within the pattern duration. This information enables prediction rules to be constructed, for example, such as the rule: if 3 (or more) 0 events and 4 (or more) 1 events occur within a sequence window (a 4-bit window, for example), then predict the target event. However, the only way to do this is to draw on the theory of a given ontology, which consists of axioms and inference rules. The analysis of many modern predictors, e.g. Random Forest Algorithm-based, shows they actually use of the ontology of the predicted area with the probabilities of certain concept. But all knowledge about the applied domain of predicted (classified) data is reduced to knowledge about the features and their possible meanings in the predictors mathematical models. Therefore, the important question is how to assess the predictive power of the prescript used online to predict the future value of the incoming data. For this, it is possible to interpret the results of predictions obtained to this moment as the result of the successful use of information contained in the ontology, in which the predicted data is interpreted. In other words, we assume that there is ontology (in the above sense) that can be better or worse predicted by a specific predictor, and, accordingly, there are more or less "light" ontology for this predictor. Let us pose the question, which representation of the ontology will be the simplest from the point of view of the algorithm implementation. The answer is obvious if there are only two features that take a binary value of 0 or 1, with probabilities of values p_0 and p_1 . Obviously, in this case, predicting the ontology concept at time $t + 1$ from known data at previous times is equivalent to predicting the Bernoulli sequence.

It is known that the optimal according to the Hamming loss criterion ("incorrectly predicted the next bit-lost everything" predictor of the Bernoulli sequence is an algorithm which, for the Bernoulli distribution parameter $p_1 \geq 0.5$, predicts the future value $x_{t+1} = 1$, and $x_{t+1} = 0$ for $p_1 \leq 0.5$ [3]. Suppose we use a predictor f , to predict the value of a binary time series at time $t + 1$, knowing the predicted values at times $\{t - k, \dots, t\}$, where k is the window size (training window).

Let we computed the success rate of the predictions $SR_f = \frac{N_f}{k}$ in the window $\{t - k, \dots, t\}$, N_f is the number of correct predictions by the predictor f (more strictly, biased Laplace estimate to estimate the number of correct predictions in Bernoulli's trials can be used, that is $\frac{N_f + 1}{k + 2}$). Then, if SR_f does not differ statistically significantly from the success rate calculated under the assumption that the predictor u is the optimal predictor for the Bernoulli sequence, then we can conclude that the predictor is ineffective for the ontology. In this case, we can also talk about the difference some predictor binary sequence fp predictor of the simplest ontology as by the estimated value SR_f , and by the structure errors, that is, according to the error ratio "0 instead of 1" and "1 instead of 0".

Acknowledgements: Research partially supported by the Russian Foundation for Basic Research under grants RFBR 18-07-00669 and 18-29-03100.

References

- [1] Y.-B. Kang, J. Z. Pan, S. Krishnaswamy, W. Sawangphol and Y.-F. Li. How long will it take? accurate prediction of ontology reasoning performance, *AAAI*, 2014.
- [2] M. Marzocchia and T. Jordan. Testing for ontological errors in probabilistic forecasting models of natural system, *PNAS* 111,33, pp. 11973–11978.

- [3] N. Merhav and M. J. Weinberger, On universal simulation of information sources using training data, *IEEE Trans. Inform. Theory* 50, 1, 2004, pp. 50.

The role of perceptual variables and country-level culture on international entrepreneurship

Carlos Gomes^{1,2,3}, Aldina Correia^{1,2,3} and Vítor Braga^{1,2,3}

¹CIICESI – Center for Research and Innovation in Business Sciences and Information Systems

²Escola Superior de Tecnologia e Gestão

³Politécnico do Porto, Portugal

E-mail addresses: *8140122@estg.ipp.pt; aic@estg.ipp.pt; vbraga@estg.ipp.pt*

International Entrepreneurship is a largely investigated theme in the last decades [3–8]. Based on that growing interest in this subject, in this work [1] approach is considered and updated, using the international entrepreneurship current data, obtained through GEM (Global Entrepreneurship Monitor). First an "International Entrepreneurship" variable was created. This variable was then chosen to be studied as dependent in this research, which intends to make an analysis about the influence of three groups of variables on the international entrepreneurial choices: Demographic and economic variables; Perceptual variables and Country-effects variables.

In a first phase of literary review, a set of opinions about the variables that influence the choice to entrepreneurship and international entrepreneurship is made, making it possible to create the hypothesis of study. Moving on to a second phase, with the objective of understand which variables presented by the literature are significant to explain the dependent variable, are performed logistic regression models, using the software SPSS. The primordial objective of this study is to understand the effect of perceptual variables and individual characteristics in the International Entrepreneurship.

Keywords

International Entrepreneurship, Perceptual variables, Country-level.

There are a lot of policies and stimulus for entrepreneurship, however they do not produce the same effects on different people, being so important to study the role of the perceptual variables in entrepreneurship. [2], thus is also intended to fill this gap in the literature. In addition to the demo-economic variables, when making decisions with respect to their employment, individuals also consider a set of subjective perceptions about entrepreneurship that they form based on knowing other entrepreneurs, confidence in one's skills and abilities, risk propensity, and alertness to unexploited opportunities. After all, entrepreneurship is about people [1]. Evidence across many countries indicates that subjective perceptions about one's own skills, the existence of opportunities and knowing other entrepreneurs are all highly and positively correlated to the decisions to start businesses and internationalize. The fear of failure, instead, is strongly but negatively correlated to being a international entrepreneur. Thus, in addition to the demo-economic and personal characteristics, we include the different individuals' environment as a further component of entrepreneurial behavior and consider the possibility of country specific effects.

There are several entrepreneurs' characteristics associated with motivations and perceptions which can be identified in early internationalization. Some of these motivations

are related to the entrepreneurs' needs and personality, while others depict the competitive landscape of the venture's environment. Identifying entrepreneurs' motivations can be crucial for understanding how resources and strategic decisions are managed. Our paper contributes to the research in this field providing a more accurate analysis of what variables are statistically significant in the creation of international entrepreneurs, including three different groups of variables and using data that is well suited for the study, because they provide information about individuals who are in the process of starting a new business. Our results find that both perceptual variables and country-effects are significant and correlated to the international entrepreneurship and improve the model of study.

Acknowledgements: This research has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e Tecnologia under the project UID/GES/04728/2020.

References

- [1] P. Arenius and M. Minniti. Perceptual variables and nascent entrepreneurship. *Small business economics*, **24**(3): 233–247, 2005.
- [2] M. Entrialgo and V. Iglesias. Entrepreneurial intentions among university students: The moderating role of creativity. *European Management Review*, 2020.
- [3] M. V. Jones, N. Coviello and Y. K. Tang. International entrepreneurship research (1989–2009): a domain ontology and thematic analysis. *Journal of business venturing*, **26**(6): 632–659, 2011.
- [4] P. P. McDougall. International versus domestic entrepreneurship: New venture strategic behavior and industry structure. *Journal of Business Venturing*, **4**(6): 387–400, 1989.
- [5] P. P. McDougall, B. M. Oviatt and R. C. Shrader. A comparison of international and domestic new ventures. *Journal of international entrepreneurship*, **1**(1): 59–82, 2003.
- [6] A. R. Reuber, G. A. Knight, P. W. Liesch and L. Zhou. International entrepreneurship: The pursuit of entrepreneurial opportunities across national borders, 2018.
- [7] A. Tabares, Y. Chandra, C. Alvarez and M. Escobar-Sierra. Opportunity-related behaviors in international entrepreneurship research: a multilevel analysis of antecedents, processes, and outcomes. *International Entrepreneurship and Management Journal*, 1–48, 2020.
- [8] S. Terjesen, J. Hessels and D. Li. Comparative international entrepreneurship: A review and research agenda. *Journal of Management*, **42**(1): 299–344, 2016.

TBATS models and linear regression models approaches for forecasting daily weather time series

A. Manuela Gonçalves¹, Marco Costa² and Cláudia Costa³

¹University of Minho, Department of Mathematics and Center of Mathematics, Portugal

²University of Aveiro, Águeda School of Technology and Management, Center for Research and Development in Mathematics and Applications, Portugal

³University of Minho, Department of Mathematics, Portugal

E-mail addresses: *mneves@math.uminho.pt; marco@ua.pt; claudiacosta_mf@hotmail.com*

The main aim of this work is to estimate and forecast daily weather variables in real time in the field of agriculture irrigation at a particular site (a farm). Weather time series have strong correlation and high-frequency seasonal patterns. How to best model and forecast these patterns has been a long-standing issue in time series analysis. In this study we perform a state space modelling framework using exponential smoothing, the TBATS (Trigonometric Seasonal, Box-Cox Transformation, ARMA errors, Trend and Seasonal Components) models and linear regression models with correlated errors.

Keywords

Time Series Analysis, Forecasting, TBATS, Regression With Correlated Errors, Weather Variables.

Predicting and forecasting weather has always been a difficult field of research analysis with a very slow progress rate over the years. Weather data consider the noises and outliers and therefore investigation in this area may not be accurate. The main aim of this work is to estimate and forecast daily weather variables in real time at a particular site (a farm located in Carrazeda de Ansiães, in the district of Bragança in the North of Portugal), in the field of agriculture irrigation. Like many other environmental time series, weather time series have strong correlation and complex/high-frequency seasonal patterns. How to best model and forecast these patterns has been a long-standing issue in time series analysis. This way, we perform a state space modelling framework using exponential smoothing by incorporating Box-Cox transformations, ARMA residuals, Trend and Seasonality - the TBATS (Trigonometric Seasonal, Box-Cox Transformation, ARMA errors, Trend and Seasonal Components) models [1,3]. In particular, this study presents a comparison of the modelling and forecasting performances of forecasting methods of the TBATS models and the classical approach of the linear regression models with correlated errors ([2,3]) for weather time series.

Acknowledgements: This work was partially supported by the Portuguese FCT (FCT - Fundação para a Ciência e a Tecnologia) Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM. This work was also supported by the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

References

- [1] A. M. De Livera, R. J. Hyndman and R. D. Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of American Statistical Association*, **106(496)**: 1513-1527, 2011.
- [2] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*. 2nd edition. OTexts: Melbourne, Australia, 2018. <https://otexts.com/fpp2/>
- [3] T. Alpuim and A. El-Shaarawi. On the efficiency of regression analysis with ar(p) errors. *Journal of Applied Statistics* **35(7)**: 717-737, 2008.
- [4] T. Alpuim and A. El-Shaarawi. Modelling monthly temperature data in Lisbon and Prague. *Environmetrics* **20(7)**: 835-852, 2009.

Individual growth modelling with stochastic differential equations

Gonalo Jacinto¹, Patr cia A. Filipe¹ and Carlos A. Braumann¹

¹Departamento de Matem tica, Escola de Ci ncias e Tecnologia, Universidade de  vora
Centro de Investiga  o em Matem tica e Aplica  es, Instituto de Investiga  o e
Forma  o Avan ada, Universidade de  vora, Portugal.

E-mail addresses: *gjcj@uevora.pt; pasf@uevora.pt; braumann@uevora.pt*

To describe individual growth dynamics, regression methods are inappropriate. So, we use stochastic differential equation (SDE) models that, in the most general form, can be written as $dY(t) = \beta(\alpha - Y(t))dt + \sigma dW(t)$, where $Y(t) = h(X(t))$ with $X(t)$ being the weight of an animal at age t and h an appropriate strictly increasing C^1 function. The parameters are $\alpha = h(A)$, where A is the maturity weight of the animal, β the rate of approach to maturity and σ the intensity parameter of the random fluctuations. $W(t)$ is a standard Wiener process. We have previously (see [2], [4], [3]) studied estimation, prediction and optimization issues using cattle weight data from females of Mertolengo cattle breed.

In this new project, we have adjusted and extended the methodologies and applied them to the weight data of males of Mertolengo cattle breed and Alentejana cattle breed. Since model parameters may vary from animal to animal and that variability can be partially explained by their genetic differences, we introduce the extension of the study to SDE mixed models in order to incorporate the effect of environmental variability and genetic effects. For this, we consider that the parameters of the model may now depend on the genetic values of the animal.

The estimation of parameters for this type of models may present some difficulties, in particular, the assumption of approximate normality of the maximum likelihood estimators may fail when sample sizes are not large enough. So, we also obtain the parametric bootstrap estimators, which can be used to compare with the maximum likelihood estimators, correct their possible biases and improve their confidence intervals (see [1] for a similar approach).

The SDE methodology is much more appropriate, and therefore more precise. The farmer's expected profit as a function of the animal's selling age can now be determined on an individual animal basis by considering the animal's genetic values, thus improving profit optimization.

Keywords

Application To Cattle Data, Bootstrap Estimation, Genetic Values, Maximum Likelihood Estimation, Mixed Models.

Acknowledgements: The authors belong to the research centre Centro de Investiga  o em Matem tica e Aplica  es, Universidade de  vora, supported by FCT (Fund  o para a Ci ncia e a Tecnologia, Portugal, project UID/04674/2020). This work was developed within the project *GO Bov+* (PDR2020-101-031130), financed by PDR2020-1.0.1-FEADER-031128.

References

- [1] P. A. Filipe and C. A. Braumann. Modelling individual animal growth in random environments, *Proceedings of the 23rd International Workshop on Statistical Modelling*, 232–237, 2008.
- [2] P. A. Filipe, C. A. Braumann, N. M. Brites and C. J. Roquete. Modelling Animal Growth in Random Environments: An Application Using Nonparametric Estimation, *Biometrical Journal* **52(5)**: 653–666, 2010.
- [3] P. A. Filipe, C. A. Braumann and C. Carlos. Profit optimization for cattle growing in a randomly fluctuating environment, *Optimization*, **64(6)**: 1393–1407, 2015.
- [4] P. A. Filipe, C. A. Braumann and C. J. Roquete. Multiphasic individual growth models in random environments, *Methodology and Computing in Applied Probability* **14(1)**: 49–56, 2012.

A preliminary study on the transportation of non-urgent patients by a fire station

Emanuel Lopes¹ and Eliana Costa e Silva¹

¹CHICESI, ESTG, Politécnico do Porto

E-mail addresses: *8150105@estg.ipp.pt; eos@estg.ipp.pt*

Abstract

Non-urgent patients requiring transportation services carry out by institutions have to be picked from their homes and transported to the health care appointments on time. Using the Vehicle Routing Problem (VRP) solver developed by [2], the present work intends to study its performance in solving real transportation of non-urgent patients performed by a fire station. The differences between the real distance travelled by the vehicles and the routes obtained using the solver on five instances are analysed.

Keywords

Vehicle Routing Problem; Transport; Statistics.

The transport of non-urgent patients is part of the daily life of the Portuguese fire-fighters. This type of organization seeks to optimize the allocation of resources to reduce its expenses and maximize the availability of resources. In this work, the needs of a fire station that intends to minimize the number of route and the distances taken by the ambulances transporting non-urgent patients is addressed. This problem can be seen as a Vehicle Routing Problem with Time Windows (VRPTW) in which each vehicle leaves the fire station, travels to the residence of non-urgent patients and, after reaching its maximum capacity in terms of maximum number of passengers or maximum time, transports the patients to the clinic where treatments are carried out. There is a vast literature on VRP (see e.g. [3]). Furthermore there are specific studies of practical problems such as for e.g. Bektas [1], that seeks to optimize the routes taken by school buses, and the transportation of non-urgent patients by the Portuguese Red Cross [4]. Erdoğan [2] presents an open source tool Spreadsheet Solver for VRP and demonstrates its application in case studies of two real-world, namely to the healthcare and tourism sectors. In the present work, five days of transport of non-urgent patients performed by the fire station are analysed concerning the transport of 213 patients. The methodology proposed in [2] is used in five real instances for minimizing the distance traveled by ambulances that carry out this type of scheduled transportation. The statistics of the obtained preliminary results are presented. For each instance the differences between the real distance travelled by the vehicles and the routes obtained using the solver will be analysed. Furthermore, the existence of the time violations and waiting times will also be examine. Some limitations have already been found, such as violation of the time windows, therefore in the future extensions of the proposed methodology will be carried out.

Acknowledgements: This work has been partially supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDP/04728/2020.

References

- [1] T. Bektaş and S. Elmastaş. Solving school bus routing problems through integer programming. *Journal of the Operational Research Society* **58(12)**: 1599–1604, 2007.
- [2] G. Erdoğan. An open source Spreadsheet Solver for Vehicle Routing Problems. *Computers and Operations Research* **84**: 62–72, 2017.
- [3] G. Laporte. Fifty years of vehicle routing. *Transportation Science* **43(4)**: 408–416, 2009.
- [4] M. D. M. R. Loureiro. Optimização de rotas de transporte de doentes programados: O caso da cruz vermelha portuguesa amadora - sintra. Master’s thesis, IST- Universidade Técnica de Lisboa, Lisboa, 2010.

Stochastic model to analyze the changes in the fisheries catch and species composition due to varying oceanic temperature along the Colombian Pacific coast

Erika Johanna Martínez Salinas¹, Viswanathan Arunachalam¹ and John Josephraj²

¹Universidad Nacional de Colombia, Sede Bogotá Colombia

²Universidad Nacional de Colombia, Sede Tumaco Colombia

E-mail addresses: ejmartinezsa@unal.edu.co; varunachalam@unal.edu.co; jojselvaraj@unal.edu.co

The fishing activity in the Colombian Pacific coast together, with the forestry, form the primary income of the inhabitants of the sector. According to the UN intergovernmental panel, Colombia along with countries such as Ecuador, the Southeast United States, among others, make up the group of most vulnerable countries in the world affecting fishing resources due to global warming. The consequence of climate change is likely to increase the sea surface temperature, modifying the availability of the fishery resources. Based on these claims, different research groups have posted the following question: Is there a real change in environmental variables due to climate change that will affect the fisheries species catch and its composition in the Pacific Colombian coast?

In this work, a stochastic model is proposed in order to analyze the behavior over time, between the observed changes in sea surface temperature and its relationship with the catch and abundance of the Goldfish species, in a specific area of the Colombian Pacific. Statistical analysis and estimation methods such as *maximum likelihood* and the *Euler-Maruyama method* are applied in order to estimate the model parameters and estimate values of sea surface temperature, species abundance and catch per unit of effort for the species under study, between the years 2000-2012. Finally, the simulation results of the exposed model are presented and a sensitivity analysis is developed for the estimated parameters.

Keywords

Stochastic Differential Equations, Ornstein Uhlenbeck Process, Mean Reversion, Maximum Likelihood Estimation, Euler-Maruyama Method.

Deep learning method to identify fire ignitions

João Mendes^{1,2}, Thadeu Brito^{1,2}, Ana I. Pereira^{1,2} and José Lima^{1,2}

¹Research Center in Digitalization and Intelligent Robotics (CeDRI)

²Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253 Bragança, Portugal

E-mail addresses: *a32299@alunos.ipb.pt; brito@ipb.pt; apereira@ipb.pt; jllima@ipb.pt*

The SAFe project aims to create and implement a set of innovative operations that minimize the time of forest ignitions identification contributing to the development of the Trás-os-Montes region. Thus, it is intended to locate a set of sensors in the forest, data information will be collected, and the artificial intelligence algorithm Deep Learning will be applied to achieve the intended end. Numerical results demonstrated the approach reliability.

Keywords

Artificial Intelligence. Learning algorithms. Forest Ignitions.

The forest fires are a continuous worldwide problem since a sufficiently effective solution has not yet been found. One of the major problems of this theme in our country is the fact that the Portuguese forest has a very tight vegetation, thus limiting the use of image sensors, which according to [1] are the best option for the detection of forest ignitions. Once this surveillance option is excluded and considering that the first 20 minutes are essential to minimize the damage caused by the fire, innovative approaches must be study and implemented [2].

For that, SAFe project presents a solution that implements a set of innovative operations that minimize the identification time of forest ignitions and contribute to the development of the Trás-os-Montes region. Thus, it is intended to locate a set of sensor modules in the forest, data information will be collected from them, and artificial intelligence algorithms will be applied to send alerts to the surveillance agents.

The prototype of the wireless sensor module is composed by the central element based on a microcontroller which will send the data through a wide area network communication. The nodes, through LoRaWAN, connect to the gateways that send the acquired data to the application server where the processing will be done. The level of data transmission will also be a simpler process as there is no need to transmit images, producing only data in numerical format. The data generated will be in massive quantities since the sensors collect with an interval of two minutes between them [3]. This system will allow to obtain an immediate control thus facilitating the task of processing and later alerting within the time window necessary to minimize the damages caused by the fire. Taking into account the amount of data on a daily basis, it is possible to apply artificial intelligence, more precisely, the method of deep learning. This system will become more intelligent each hour, due to constant training with values of daily values which will consequently result in a more credible system day after day. The prototype of the node will consist of several elements, namely, a microcontroller, through LoRaWAN communication, sends data from five infrared sensors, air temperature and humidity. This type of sensors will be a more economical solution both in terms of money and energy consumption when compared to image sensors [1].

In this work, the data obtained from five infrared sensors were analysed. The Deep learning algorithm was implemented using the predefined MatLab functions and it was used to predict the behaviour of the sensors data. It is possible to verify, in 1, the final alert, combined from the individual alerts of each of the five sensors, identifying the forest ignition.

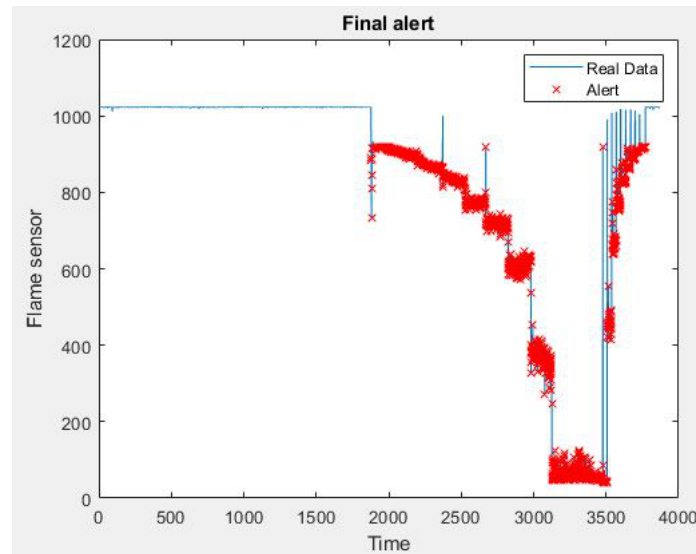


Fig. 1. Alert chart.

Future work will involve the use of more types of sensors and not just flame sensors, in order to combat any false alert given by them. Thus making the software even more reliable.

References

- [1] F. Félix and L. Lourenço. O tempo de resposta do ataque inicial a incêndios florestais nos espaços mais sensíveis de Portugal: o exemplo prático da serra da Lousã. *Territorium: Revista Portuguesa de riscos*, 187–211, 2017.
- [2] N. Verma and D. Singh. Analysis of cost-effective sensors: Data Fusion approach used for Forest Fire Application. *Materials today: proceedings* **24**: 3584–3588, 2020.
- [3] T. Brito, A. Pereira, J. Lima, J. Castro, and A. Valente. Optimal Sensors Positioning to Detect Forest Fire Ignitions. *Proceedings of the 9th International Conference on Operations Research and Enterprise Systems - Volume 1: ND2A*, 411–418, 2020.

How does organizational ambidexterity influence the firms' internationalization speed? The contribution of network clustering

Telma Mendes^{1,2}, Vítor Braga^{1,2} and Carina Silva^{1,2}

¹CHICESI – Center for Research and Innovation in Business Sciences and Information Systems,

²Escola Superior de Tecnologia e Gestão, Instituto Politécnico do Porto, Portugal

E-mail addresses: *8140224@estg.ipp.pt; vbraga@estg.ipp.pt; ccs@estg.ipp.pt*

Firms' in industrial clusters need a certain combination of organizational and collective skills to grow rapidly on foreign markets. This study aims to evaluate the role played by network clustering and organizational ambidexterity on the clustered firms internationalization speed. Adopting a paradoxical perspective, we argue that there are two opposite orientations (exploitation and exploration) differentially linked to the output variable (speed). We test the developed hypotheses against 1467 Portuguese's manufacturing firms obtained by the Community Innovation Survey (CIS). Our results suggest that network clustering increases the number of international markets explored by agglomerated firms (scope) and, contrary to was verified in the exploitative capability, the adoption of explorative activities strengthens this relationship. This study contributes to shape ambidextrous orientations and to consider internationalization speed as a springboard for the strategic intent to continuously and recursively outperform global competition.

Keywords

Industrial Clusters, Networks, Geographical Diversity, Exploitation, Exploration.

The speed of internationalization is receiving considerable attention from the literature, playing an important role in the global strategy research [1]. However, theoretical and practical knowledge on how clustered firms address their speed is still scant [2,3]. Drawing on International Business (IB) literature, the aim of this study is to shed light on how networks embedded in industrial clusters help agglomerated firms to grow faster internationally, exploring the influence of each organizational orientation on this relationship. For accomplish the research purpose, we employed the Partial Least Squares Structural Equation Modelling (PLS-SEM) in order to maximize the explanation of variance (R²) for internationalization speed in a latent model. After evaluating the selection criteria, we choose this approach since its assumptions are less restrictive and yield in littler identification problems as estimates are more robust against violations of normality, minimum sample size and maximum model complexity. Moreover, in recent stimulation studies, PLS-SEM has proven to provide greater statistical power (i.e., lower levels of type-II error) [4]. Therefore, the estimated SEM model allows to attain a better understanding of the variables with potential effects on firms' internationalization speed. The total explained variance of the target construct (speed) is $R^2 = 13.2\%$. Our findings suggest that network clustering (national and international ties) positively influence agglomerated firms willingness to pursue explorative strategies. The two network dimensions can produce synergistic

effects when it comes to international exploration. Furthermore, beyond network clustering has a significant impact on explorative capabilities, it also explains the firms' presence in different geographical regions. The outcomes also indicate that a higher orientation to exploration positively relates with the enlargement of territorial scope. Hence, if affiliated firms intend to achieve a long-term growth, they should allocate resources to exploration in a higher extent than those devoted to exploitation. This trade-off is extremely important in international markets since the adoption of ambidextrous strategies can produce different results in terms of businesses' development.

Acknowledgements: CIICESI, ESTG, Polytechnic of Porto. This research has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e Tecnologia (Portuguese Foundation for Science and Technology) under the project UID/GES/04728/2020.

References

- [1] M. Hilmersson, M. Johanson, H. Lundberg and S. Papaioannou. Time, temporality, and internationalization: The relationship among point in time of, time to, and speed of international expansion, *Journal of International Marketing* **25(1)**: 22–45, 2017.
- [2] B. Jankowska and M. Götz. Internationalization intensity of clusters and their impact on firm internationalization: the case of Poland, *European Planning Studies* **25(6)**: 958–977, 2017.
- [3] J. Novotná and L. Novotný. Industrial clusters in a post-socialist country: The case of the wine industry in Slovakia, *Moravian Geographical Reports* **27(2)**: 62–78, 2019.
- [4] W. Reinartz, M. Haenlein and J. Henseler. An empirical comparison of the efficacy of covariance-based and variance-based SEM, *International Journal of Research in Marketing* **26(4)**: 332–344, 2009.

New approach of total least square algorithm for nonlinear models

Oumaima Mesbahi¹, Mouhaydine Tlemçani¹, Fernando M. Janeiro²,
Abdeloawahed Hajjaji³ and Khalid Kandoussi³

¹Instituto de Ciências da Terra, Universidade de Évora, Portugal

²Instituto de Telecomunicações, Universidade de Évora, Évora, Portugal

³LabSIPE, University of Chouaib Doukkali, Morocco

E-mail addresses: *oumaimames@gmail.com; tlem@uevora.pt; fmtj@uevora.pt; hajjaji.a@ucd.ac.ma; kkandouss@gmail.com*

Many optimization mathematical problems [1] are based on the minimization of a cost function formed by a sum of quadratic terms representing the quantification of the errors between a mathematical model and a measured set of experimental data [2], which is usually accompanied by measurement uncertainties. The objective is mainly to achieve the best estimation of numerical model parameters from this data. Many optimization studies adopt this methodology; however they focus only on using the ordinary least squares function presented in (1). This equation is defined as a sum of quadratic terms of the differences between a mathematical function as a set of data which are only represented in Y-axis. In engineering [3], [4], it is considered the easiest method for the application of the least squares approach, due to its uncomplicated calculations. Nonetheless, the ordinary least squares method does not consider the measurement uncertainties contained by the variable in X-axis, since it only focuses on calculating the vertical distances. This could lead to non-accurate and non-precise model parameters estimation. This work presents a new approach to solve the total least squares problem represented in equation (2) which considers the measurement uncertainties contained in both x and y variables. This approach is also applicable in the case of the fitting of an implicit non-linear function. A simulation study of the five parameters photovoltaic cell model is presented along with a comparative study with the ordinary least squares approach.

$$S = \frac{1}{N} \sum_{k=1}^N (y_k - \hat{y}_k)^2 \quad (6)$$

$$S = \frac{1}{N} \sum_{k=1}^N [(y_k - \hat{y}_k)^2 + (x_k - \hat{x}_k)^2] \quad (7)$$

Keywords

Least Squares Approach, Ordinary Least Squares, Simulation.

References

- [1] A. Antoniou and W. S. Lu. Practical optimization: algorithms and engineering applications. Springer Science and Business Media, 2007.

- [2] S. Van Huffel and J. Vandewalle. The total least squares problem: computational aspects and analysis, vol. 9. Siam, 1991.
- [3] P. M. Ramos, F. M. Janeiro, M. Tlemçani and A. C. Serra. Recent Developments on Impedance Measurements With DSP-Based Ellipse-Fitting Algorithms, *IEEE Trans. Instrum. Meas.* bf 58(5):, 1680–1689, 2009.
- [4] O. Mesbahi, M. Tlemçani, F. M. Janeiro, A. Hajjaji and K. Kandoussi. Estimation of photovoltaic panel parameters by a numerical heuristic searching algorithm, *2019 8th International Conference on Renewable Energy Research and Applications (ICR-ERA)*, 401–406, 2019.

Short-term forecast models for the maximum temperature time series

Fernanda Pereira¹, A. Manuela Gonçalves², Marco Costa³ and Sofia O. Lopes⁴

¹University of Minho, Department of Mathematics, Portugal

²University of Minho, Department of Mathematics and Center of Mathematics, Portugal

³University of Aveiro, Águeda School of Technology and Management, Center for Research and Development in Mathematics and Applications, Portugal

⁴University of Minho, Department of Mathematics, Center of Physics and SYSTEC, Portugal

E-mail addresses: *fernandaccp@outlook.pt; mneves@math.uminho.pt; marco@ua.pt; sofialopes@math.uminho.pt*

This work presents a comparative study of two models to estimate the maximum temperature in real time: the linear state space models (associated to the Kalman filter) and linear regression models. The goal is to improve short-term forecast of the maximum temperature, obtained from the website *weatherstack.com*, in the context of increasing the efficiency of daily water use in irrigation systems.

Keywords

Time Series Analysis, State Space Model, Kalman Filter, Calibration, Meteorological Variables.

The challenge of the To Chair project is to study the behaviour of humidity in the soil by mathematical/statistical modelling in order to find optimal solutions to improve the efficiency of daily water use in irrigation systems [1].

The procedure proposed is based on the state space modelling ([3]) that provides a very flexible tool for analysing dynamic phenomena and evolving systems, and has significantly contributed to extending the classical domains of application of statistical time series analysis [2].

In this study, the objective is to calibrate the short-term forecasts in real time, obtained from the *weatherstack* website through the state space modelling in order to improve the accuracy of the initial h -steps-ahead, considering a six days temporal window of forecasts of the maximum temperature. For this purpose, two calibration models were adjusted considering models with a deterministic calibration factor (linear regression models) and a stochastic calibration factor (state space models associated to the Kalman filter).

The linear regression model has been the most applied approach when a predict model is needed. However, statistical models with fixed effects are unlikely to yield a good predictive accuracy, particularly in situations where the predictor and predictand relationship changes over time [3].

This methodology is applied to a data set that includes observations of daily maximum temperature in a farm, Senhora da Ribeira, located in the municipality of Carraceda de Ansiães in the North of Portugal, between February 20 and October 11, 2019.

Acknowledgements: This work has received funding from FEDER/COMPETE/NORTE 2020/POCI/FCT funds through grants UID/EEA/-00147/20 13/UID/IEEA/00147/006933-SYSTEC, project and To CHAIR - POCI-01-0145-FEDER-028247. This work

was also partially supported by the Portuguese FCT Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM and the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020.

References

- [1] C. Costa, A. M. Gonçalves, M. Costa and S. Lopes. Forecasting Temperature Time Series for Irrigation Planning Problems. In: Proceedings of the 34th International Workshop on Statistical Modelling (IWSM 2019), 2019. <http://hdl.handle.net/10773/26437>
- [2] R. J. Hyndman and G. Athanasopoulos. *Forecasting: principles and practice*, 2nd edition. OTexts: Melbourne, Australia, 2018. <https://otexts.com/fpp2/>
- [3] A. M. Gonçalves, O. Baturin and M. Costa. Time Series Analysis by State Space Models Applied to a Water Quality Data in Portugal. *AIP Conference Proceedings* Vol. **1978**, 470101-1-470101-4, 2018. <https://doi.org/10.1063/1.5044171>
- [4] A. M. Gonçalves and M. Costa. Predicting seasonal and hydro-meteorological impact in environmental variables modelling via Kalman filtering. *Stoch Environ Res Risk Assess* **27**(5), 1021–1038, 2013. <https://doi.org/10.1007/s00477-012-0640-7>

Modelling complex relationships in dentistry - The GAMLSS approach

J. A. Lobo Pereira^{1,2}, Davide Carvalho¹, Miguel Gonçalves Pinto¹,
Luzia Mendes¹ and Teresa Oliveira^{2,3}

¹University of Porto, Portugal

²Universidade Aberta Lisboa, Portugal

³Research Center CEAUL, Portugal

E-mail addresses: *jpereira@fmd.up.pt; davideccarvalho@netcabo.pt; mpinto@fmd.up.pt; luziamgoncalves@gmail.com; Teresa.Oliveira@uab.pt*

To explain and compare the variability of PPD between diabetic and non diabetic subjects, GAMLSS models were fitted to data of 79 diabetic and 79 non diabetic individuals. The results suggest a U-shaped relationship between HDL levels and PPD more pronounced in the diabetic, together with a significant impact of age and HDL on PPD.

Keywords

Periodontitis, Diabetic, GAMLSS models.

Introduction: Periodontitis is a chronic inflammatory destructive process in the supporting tissues of the teeth associated with gram negative bacteria present in a biofilm on the surface of the teeth, and is one of the main causes of tooth loss in adulthood. Promoting a low level endotoxemia [1], appears to sustain a low-grade systemic inflammation associated to dyslipidemia, yet whether periodontitis causes an increase in levels of plasma lipids or whether hyperlipidemia is a risk factor for periodontal infection. The probing pocket depth (PPD) is a clinical parameter for periodontal destruction/inflammation.

Generalized additive models for location, scale and shape (GAMLSS), are an extension of the classical GAM approach, introduced by Hastie and Tibshirani [2]. Besides modelling the conditional mean (location) of the response variable distribution also, model the variance (scale) and the shape parameters (skewness and kurtosis) which may depend on explanatory variables. Under the GAMLSS framework, we assume that independent observations y_i of a random variable Y for $i = 1, 2, 3, \dots, n$ have probability distribution function $f_Y(i_i | \theta^i)$ with $\theta^i = (\theta_1^i, \dots, \theta_p^i)$ vector of p parameters of Y being μ , σ , v and τ for location, scale, skewness, and kurtosis, respectively. The aim of this work is to assess non-linear relationships between mean PPD and HDL, age in diabetic and nondiabetic subjects.

Material and methods: Data on lipid profile, and socio-demographic from 79 diabetic and 79 non-diabetic individuals matched by age and gender.

The dependency of PPD on HDL levels (mg/dl) and age (years) in both groups was modeled using exGAUS (convolution of normal and exponential distributions) and Gamma distributions truncated between 0 and 6 (TexGAUS[0, 6] and TGA[0,6]) in the diabetic and non-diabetic groups, respectively.

The data was processed with *R* software [3] using the GAMLSS Package [2] .

Results and Conclusion: The best fits of diabetic TexGAUS[0, 6] and non diabetic TGA[0, 6] models were: $\hat{\mu}=cs(HDL)+cs(Age)$, $\log(\hat{\sigma}) = Age$, and $\log(\hat{\nu})$; and $\hat{\mu}=cs(HDL)+cs(Age)$, and $\log(\hat{\sigma}) = Age$, respectively (tables 1 and 2). Where cs are cubic-splines .

Table1: Non-Diabetic GAMLSS fit parameters with TGA[0, 6] distribution

Parameter	Variable	$\hat{\beta}$	Std. Error	p-value	χ^2 p-value*
μ	Intercept	0.663	0.131	3.03e-06	-
	HDL	-0.007	0.002	8.84e-05	0.0001
	Age	0.010	0.001	2.58e-10	9.659e-08
σ	Intercept	-2.304	0.288	1.75e-11	-
	Age	0.013	0.005	0.0163	0.0201

Total df=9.001; GAIC=103.634; Global deviance = 85.631; $R^2=0.51$;

Table2: Diabetic GAMLSS fit parameters with TexGAUS[0, 6] distribution

Parameter	Variable	$\hat{\beta}$	Std. Error	p-value	χ^2 p-value*
μ	Intercept	1.553	0.460	0.0010	-
	HDL	-0.007	0.005	0.2040	0.0501
	Age	0.018	0.005	0.0003	0.0033
σ	Intercept	-3.484	1.982	0.0833	-
	Age	0.035	0.032	0.2809	0.07541
ν	Intercept	-0.418	0.156	0.0093	-

Total df = 10.000; GAIC= 150.508; Global deviance = 130.508; $R^2=0.24$;

HDL – High Density Lipoprotein; k=1 – Penalty in GAIC; * – test of significance of the parametric terms;

Our results suggest a U-shaped relationship between PPD and HDL, similar to that founded by Madsen et al [4] between HDL and risk of infectious disease. The gamlss modeling framework revealed versatile enough to undercover non linear relationships between PPD and the independent variables.

References

- [1] O. Andrukhov, H. Haririan, K. Bertl, W. D. Rausch, H. P. Bantleon, A. Moritz, et al. Nitric oxide production, systemic inflammation and lipid metabolism in periodontitis patients: possible gender aspect. *J Clin Periodontol*; **40(10)**: 916–23, 2013.
- [2] T. J. Hastie and R. J. Tibshirani, *Generalized additive models*. CRC Press, 1990.
- [3] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. 2018.
- [4] C. M. Madsen and A. Varbo. A. Tybjaerg-Hansen, R. Frikke-Schmidt, B. G. Nordestgaard, U-shaped relationship of HDL and risk of infectious disease: two prospective population-based cohort studies. *European Heart Journal*, **20**, 1–11, 2017.

Modelling spreading process of a wildfire in heterogeneous orography, fuel distribution and environmental conditions - a complex networks approach

Sara Perestrelo¹, Maria Grácio^{1,3}, Nuno Ribeiro^{1,4} and Luís Lopes^{2,3}

¹Universidade de Évora, Portugal

²ISEL – Instituto Superior de Engenharia de Lisboa, Portugal

³CIMA – Centro de Investigação em Matemática e Aplicações, Portugal

⁴MED – Mediterranean Institute for Agriculture, Environment and Development, Portugal

E-mail addresses: *belperes@uevora.pt; mgracio@uevora.pt; nmcar@evora.pt; luismariolopes@gmail.com*

Forest fires are phenomena that represent a great danger for the population and have severe environmental consequences. The greatest efforts in regard to direct confrontation require a large investment in terms of both financial and human resources, partly due to the unpredictability of fire behaviour and the fuel distribution that feeds its growth and spread. To deal with this unpredictability in a more efficient manner, our work focuses around the establishment of an optimal fire-break structure whose purpose is to block the spread along the landscape, at a minimal cost. It's a preventative approach, which acts as a complement to direct confrontation and contributes to the reduction of material, economic and human losses. For the establishment of this optimal fire-break structure it is necessary to model fire spreading process and, for that, we use the multilayer network model, within the area of complex networks. We aim the construction of a network of networks that allows us to simulate the fire spreading process both at a local scale and at a more global scale.

Keywords

Forest Fires, Spreading Process, Graph Theory, Geographic Information Systems (GIS), Multilayer Network Model.

We perform computational simulations that test fire spreading at the national scale. Records of fires occurred in previous years provide us the area and perimeter of the burnt area, which we use for calibration. These simulations start with an ignition point and require input parameters. Once calibrated, we intend to study properties of this network, such as the connectivity and robustness, among others. Finally, to establish the fire-break structure, we sequentially eliminate several combinations of edges and evaluate the respective effect in the neutralization of the spreading process. Computational tools: ArcGIS, a GIS software for the construction and visualization of maps; Python 3, a programming language that is the basis of the ArcGIS software and that serves us for data processing, simulations and graph construction.

Acknowledgements: This is one of the subprojects within the project CILIFO (Forest Fires Fight and Investigation Center), supported by the international European fund Interreg, which acts in three POCTEP euroregions, Alentejo, Algarve and Andaluzia. Research partially sponsored by national funds through the Fundação Nacional para a Ciência e Tecnologia, Portugal-FCT, under the project UIDB/04674/2020 (CIMA).



Fig. 1. Map of a section of Portalegre, Portugal. Each closed line forms a polygon. Green polygons: different fuel types; brown polygons: burnt area of a 2017 fire. From the burned area we construct a network, by associating each polygon to a node.

References

- [1] C. Grácio, S. Fernandes and L. M. Lopes. Strong generalized synchronization with a particular relationship R between the coupled systems, *Journal of Physics: Conf. Series* 990, 2018.
- [2] L. Russo, P. Russo and C. Siettos. A Complex Network Theory Approach for the Spatial Distribution of Fire Breaks in Heterogeneous Forest Landscapes for the Control of Wildland Fires. *PLoS ONE* 11(10): e0163226. doi: 10.1371/journal.pone.0163226, 2016.
- [3] M. Domenico, C. Granell, M. Porter and A. Arenas. The physics of spreading processes in multilayer networks. *Nature Physics* vol 12. doi: 10.1038/NPHYS3865, 2016.
- [4] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno and M. Porter. Multilayer Networks, arXiv:1309.7233v4 [physics.soc-ph], 2014.
- [5] L. Shekhtman, M. Danziger, D. Vaknin and S. Havlin. Robustness of spatial networks and networks of networks, *C. R. Physique* 19, 233–243, <https://doi.org/10.1016/j.crhy.2018.09.005>, 2018.
- [6] R. Seidl, J. Muller, T. Hothorn, C. Bassler, M. Heurich and M. Kautz. Small beetle, large-scale drivers: how regional and landscape factors affect outbreaks of the European spruce bark beetle, *Journal of Applied Ecology* 2016, 53, 530–540 doi: 10.1111/1365-2664.12540, 2016.

A stochastic diffusion process based on brody curve with exogenous factors

Oussama Rida¹, Ahmed Nafidi¹ and Boujemaa Achchab¹

¹Hassan First University of Settat, École Nationale des Sciences Appliquées, LAMSAD, Morocco

E-mail addresses: *oussamar@gmail.com; nafidi@hotmail.com; achchab@yahoo.fr*

In this paper we consider for the diffusion process based on brody curve [4,2] with exogenous factors [2]. After building the model, an exhaustive study of its main characteristics is presented, so how to get the estimations of the parameters of the process and theirs characteristics functions is presented in this paper. Finally, the ability of the new process to model economic data are shown by means of an application to simulated data.

Keywords

Diffusion Process, Maximum Likelihood, Optimization Methods and Economic Data.

References

- [1] L. M. Ricciardi. *Diffusion Processes and Related Topics in Biology*, Springer Verlag, 1977.
- [2] R. Gutiérrez, R. Gutiérrez-Sánchez and A. Nafidi. *Electricity consumption in Morocco: stochastic Gompertz diffusion analysis with exogenous factors*, *Applied Energy* **83**(10): 1139–1151, 2006.
- [3] S. Brody. *Bioenergetics and Growth*, Reinhold Publishing Corporation: New York, 1945.

Modeling analysis and numerical estimation of a stochastic epidemic SEIR model with environmental stochasticity

Andrés Ríos-Gutiérrez¹, Viswanathan Arunachalam¹ and
Anuj Mubayi²

¹Universidad Nacional de Colombia, Colombia

²PRECISIONheor, Los Angeles, USA

E-mail addresses: *asriosg@unal.edu.co; varunachalam@unal.edu.co;*
anujmubayi@yahoo.com

Studying the spread and transient behavior of an infectious disease outbreak over time is vital for all public health departments. It allows health departments to identify disease control thresholds and design timely and effective intervention measures such as implementing vaccine- and treatment-related policies. In the last decade, we have seen many global outbreaks are directly and indirectly directly transmitted infectious diseases such as Zika, Chikungunya, Dengue, Leishmaniasis, influenza H1N1, SARS, Ebola, and more recently, Coronavirus SARS CoV-2.

The present study develops and analyzes a stochastic epidemic SEIR model (Susceptible-Exposed-Infectious-Recovered) with environmental stochasticity based on random perturbations [2,5,6]. The random perturbations in the model allow us to capture the impact of variations due to external factors such as changes in environmental conditions that may impact an infectious disease. The stochastic epidemic model is analyzed and four fundamental aspects of its dynamics; 1) existence and uniqueness of the solution, 2) eventual extinction of the infection states, 3) persistence of infection in the mean, and 4) existence of the stationary distribution of the infectious state [3,4]. The mathematical analysis suggests sufficient and necessary conditions for existence and stability via methods from probability theory. The numerical simulations of different scenarios in the parameters' region are studied. The model parameters are estimated by using the Euler-Maruyama approximation and maximum likelihood estimators [1]. Finally, we illustrate numerical simulations using data from the year 2003-2004 SARS outbreak in China.

Keywords

Stochastic perturbations, Infectious diseases, Brownian Motion, Random perturbations

Acknowledgements: This work has received funding from UNIVERSIDAD NACIONAL DE COLOMBIA under the project "Modelamiento estadístico y estocástico de las dinámicas epidemiológicas de enfermedades tropicales en Colombia".

References

- [1] P. Aguirre, E. González-Olivares and S. Torres. Stochastic predator-prey model with Allee effect on prey. *Nonlinear Analysis: Real World Applications* **14(1)**: 768–779, 2013.

- [2] A. Gray, D. Greenhalgh, L. Hu, X. Mao and J. Pan. A stochastic differential equation SIS epidemic model. *SIAM Journal on Applied Mathematics* **71(3)**: 876–902, 2011.
- [3] Y. Lin, D. Jiang and P. Xia. Long-time behavior of a stochastic SIR model. *Applied Mathematics and Computation* **236(1–9)**: 2014.
- [4] R. Ullah, G. Zaman and S. Islam. Stability analysis of a general *SIR* epidemic model. *VFAST Transactions on Mathematics* **1(1)**: 2013.
- [5] P. J. Witbooi. An *SEIRS* epidemic model with stochastic transmission. *Advances in Difference Equations* **2017(1)**: 109, 2017.
- [6] Y. Zhou, W. Zhang and S. Yuan. Survival and stationary distribution of a *SIR* epidemic model with stochastic perturbations. *Applied Mathematics and Computation* **244**: 118–131, 2014.

Detection analysis of breaks in water consumption patterns: a simulation study

Marta Santos¹, Ana Borges¹, Davide Carneiro¹ and Flora Ferreira²

¹CIICESI, ESTG, Politécnico do Porto

²Center of Mathematics, University of Minho

E-mail addresses: *8150180@estg.ipp.pt; aib@estg.ipp.pt; dcarneiro@estg.ipp.pt; fferreira@math.uminho.pt*

Breaks in water consumption records can represent apparent losses, that is, volumes of water that are consumed but not billed. These losses may be associated with water theft by its consumers, billing anomalies, data handling errors, and metering errors. The study is based on simulation data and described herein evaluates the utilization of changepoint methods for the detection of different breaks in water consumption time series, which are being incorporated in a software for water meter surveillance. The simulated data were based on real data, applying different probabilistic distributions with different decreasing breaks on the mean in order to validate the effectiveness of the methods considered. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used as performance measurement criteria. The results indicate that the breakpoint method implemented in *strucchange* R package presents a higher performance in the detection of structural breaks.

Keywords

Water Losses, Apparent Losses, Structural Breaks, Probability Distribution.

Apparent losses have a major economic impact on the water company's revenue and, although they occur in small quantities, they are difficult to quantify [1]. In order to try to minimize this type of loss, the Águas do Norte company seeks solutions to support the water meter replacement process, minimizing unnecessary losses and expenses inherent in their replacement. To assist the company in detecting the meters malfunction, a software will be created incorporating an algorithm with several methods that detect an abnormal decrease in water consumption time series, that we intend to validate in this study. This software will give an alert for the company at the moment a meter is working abnormally, thus helping in the decision process of replacing the meters.

The methods used for analysis in this study were the *breakpoint* method incorporated in the package **strucchange**, the *binary segmentation* and the *PELT* method of the package **changepoint**.

As the data provided were few and short in time horizon, and we had no guarantees that in that period there was a malfunction of the meter, we simulated data from the actual data provided, using the R software package **fitdistrplus** version 1.1-1. Parameter estimates were obtained after adjusting the data to each probabilistic distribution based on the maximum likelihood estimate [2]. The data were simulated based on probabilistic distributions such as Weibull, Log-Logistic, Gamma and Log-normal suggested in the literature for the simulation of data in water consumption by [3] and [4]. The simulated datasets, corresponding to daily consumption data and 914 days were simulated. To validate the effectiveness of the methods used, we introduced decreasing breaks on the mean

for each dataset simulated of 5%, 10%, 25%, and 50% magnitudes, varying the parameters of each distribution every 6 months. For each distribution, 100 datasets were created for each type of break, with a total of 3600 synthetic datasets. Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) were used as performance measurement criteria, with the criterion that the number of points detected is the same, or close to the simulated points. In order to validate the performance of the methods under analysis, we determined the ratio between the number of true break points and the number of estimated break points, which is multiplied by 100 to obtain the percentage. For the methods with a performance of 100% we calculated the RMSE and the MAE to quantify the distance between the estimated break from the methods under evaluation, and the simulated break. The results show that the RMSE and the MAE are lower for the simulated breaks with a decreasing magnitude of 50% on the mean, with the exception of the log-logistic distribution in which the 25% magnitude breaks are the ones with the lowest RMSE and MAE. In conclusion, although most of the methods were successful in identifying breaks in water consumption related to water meter malfunction, the method that obtains the best results is *strucchange*, with the lowest RMSE and MAE in all break magnitudes and water consumption distributions simulated.

Acknowledgements: This work has been partially supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDB/04728/2020 and UIDB/00013/2020.

References

- [1] F. J. Arregui, J. Soriano, E. Cabrera Jr. and R. Cobacho. Nine steps towards a better water meter management. *Water Science and Technology*, **65(7)**: 1273–1280, 2012.
- [2] M. L. Delignette-Muller, C. Dutang, et al. *fitdistrplus*: An r package for fitting distributions. *Journal of statistical software*, **64(4)**: 1–34, 2015.
- [3] P. Kossieris and C. Makropoulos. Exploring the statistical and distributional properties of residential water demand at fine time scales. *Water*, **10(10)**: 1481, 2018.
- [4] S. Surendran and K. Tota-Maharaj. Effectiveness of log-logistic distribution to model water-consumption data. *Journal of Water Supply: Research and Technology-AQUA*, **67(4)**: 375–383, 2018.

Trends identification in medical care

Inês Sena¹ and Ana I. Pereira¹

¹Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, Portugal

Email addresses: *a33970@alunos.ipb.pt; apereira@ipb.pt*

Daily the health professionals make numerous diagnoses that can be wrong considering different reasons. To avoid errors in medical diagnosis, an application was developed to support health professionals in diagnostic procedures. The application is based on training, through the classification algorithm, Support Vector Machine, and testing a learning procedure with the health data set to identify clusters and patterns in human symptoms. The application begins with a questionnaire, where health professionals fill key questions with the patient medical results. The application will indicate a suggested diagnosis with a percentage associated. Until now, the application considers diagnosis associated with heart disease, dementia, and breast cancer. On average, the application will indicate a correct diagnosis in 84% of the cases.

Keywords

Artificial Intelligence, Support Vector Machine, Classification Algorithm.

Currently, there is greater availability of data, a greater variety of data analysis techniques, and an increase in the information processing capacity, which allows for a paradigm shift in several areas. Derived from this, Artificial Intelligence (AI), Machine Learning, and Deep Learning are evolving in the medical industry, to support physicians in different domains and applications, such as the study of disease transmission and risk identification of disease, among others [3].

A domain supported by these computer sciences is medical diagnosis, through the identification or prediction of diseases [3]. Daily, health professionals are sought out by patients, which causes them to make numerous diagnoses that can be wrong for several reasons [1].

In order to support healthcare professionals in the diagnosis, an application called ProSmartHealth was developed, which manages to provide diagnostic suggestions, so far, of three diseases: dementia, heart disease, and breast cancer.

The ProSmartHealth is based on three model codes performed to identify each disease individually. It is necessary to have a set of labeled data. at least of 100, in which each data needs a set of parameters that represents it. The application performs depends largely on the number of input parameters taken into account.

The data set was obtained through the University of Wisconsin, the Cleveland database, and the Open Access Series of Imaging Studies (OASIS). Table 1 shows the data sets used to train and test each disease identification model.

For the tests, it was divide the database into two databases: training and testing. The training database has 85% of the numerical data, and testing database has 15% of the numerical data.

The models are trained and tested in 100 interactions through the classification algorithm, Support Vector Machine, the type of classification to be used depends on the number of diagnostics possibles of obtain.

Table 1. Indicates the data set used for each disease.

Disease	Patient	Parameters	Data collection
Heart disease	303	14	Banco de dados de Cleveland
Breast cancer	569	10	Universidade de Wisconsin
Dementia	150	5	Open Access Series of Imaging Studies

To train the model for heart disease and breast cancer it was used the linear binary classification combined with the SMO (Minimum Sequential Optimization) algorithm, since there are only two possible diagnoses, presence or absence, and malignant or benign nodule, respectively. In the end, an accuracy of 82.22% was obtained for heart disease and 84.71% for breast cancer.

The model trained for dementia was carried out through “one vs one” method of the ECOC of the multi-class classification because it gets three different diagnoses, not demented, converted, and demented. A final precision of 86.36% had obtained.

The application begins with a questionnaire associated to the parameters corresponding to the chosen disease, as soon as questions are answered with the patient’s medical results, the application will indicate a diagnostic suggestion associated with a percentage rate.

In the future, the ProSmartHealth application may be applied to a set of data with symptoms that covers a greater number of diseases, which will allow a suggestion of diagnosis with the percentage of contracts each of these diseases that can be observed with the data entered in the questionnaire.

References

- [1] P. Kim. *Matlab, Deep Learning - With Machine Learning, Neural Networks and Artificial Intelligence*. APRESS, 2017.
- [2] M. C. dos Santos, A. Grilo, G. Andrade, T. Guimarães and A. Gomesäk. Comunicação em saúde e a segurança do doente: problemas e desafio. *Revista Portuguesa de saúde pública* **10**: 43–57, 2010.
- [3] N. Deng, Y. Tian and C. Zhang. *Support Vector Machine - Optimization Based Theory, Algorithms, and Extensions*. Chapman & Hall Book/CRC Press, 2013.
- [4] G. Battineni, N. Chintalapudi and F. Amenta. Machine Learning in medicine: Performance calculation of dementia prediction by support vector machine (SVM). *Informatics in Medicine Unlocked, Published by Elsevier Ltd.* **16(8)**:, 2019.

European economies clustering based on the entrepreneurial framework conditions

Eliana Costa e Silva¹, Aldina Correia¹ and Ana Borges¹

¹CIICESI, ESTG, Politécnico do Porto

E-mail addresses: *eos@estg.ipp.pt, aic@estg.ipp.pt, aib@estg.ipp.pt*

In this work, the twelve Entrepreneurial Framework Conditions (EFCs) indicators of the entrepreneurial ecosystem, defined by the Global Entrepreneurship Monitor (GEM) project, are used to cluster the European economies between 2010 and 2019. Several multivariate cluster analysis techniques are used to group the European economies based on the experts' perceptions on the EFCs of their home country with the aim of establishing a comparative analysis between the reality of the European countries since 2010.

Keywords

Internal Validation Measures, Unsupervised learning, Global Entrepreneurship Monitor Project.

Entrepreneurship has a crucial role in the development and well-being of the society, since entrepreneurs create jobs, while drive and shape innovation. Thus it is critical for the economic growth of a country and for increasing its competitiveness. With the aim of studying the entrepreneurial dynamics in the world's economies, the Global Entrepreneurship Monitor (GEM) project has gathered information on entrepreneurship since 1997 [4]. The large amount of data that has been gathered was been studied by a large number of researchers (see e.g [2]).

In this work, the twelve indicators of the entrepreneurial ecosystem of GEM, i.e. the Entrepreneurial Framework Conditions (EFCs), are used to cluster the European economies since 2010.

Clustering is an unsupervised learning technique that will assign each economy into a cluster based on some similarity measure, so that the within-cluster variations are small when compared with the variation between clusters. In this work four clustering algorithms are considered, namely: the Hierarchical Agglomerative Clustering (HAC) algorithm, with Euclidean distance and Ward method; the iterative method K-means; the Partitioning Around Medoids (PAM); and the Fanny algorithm which performs fuzzy clustering.

For assessing the quality of the clustering connectivity, Dunn's index and silhouette width are used [3]. These are functions of three very often used internal validation measures (IVM): connectivity - which indicates the extension to which observations are placed in the same cluster as their nearest neighbors; compactness - which evaluates cluster homogeneity recurring to the intra-cluster variance; and separation - which quantifies the degree of separation between clusters, that is typically by measured by the distance between the centroids of the clusters [1].

The results show that the best results for the majority of the years are obtained using HAC and the best number of clusters is two. More precisely, in terms of connectivity, the best results are obtained using HAC with the European economies divided into two clusters, for all years from 2010 to 2019. The same is observed in terms of silhouette, except for 2013 - for which the best results are obtained using K-Means with two clusters;

for 2019, for which the best results is obtained using PAM also with two clusters; and for 2015 with three clusters using HAC. Furthermore, HAC yields the best result in terms of Dunn's index for 2011, 2012, 2013 and 2018, with two clusters for 2011 and 2012 and five for the remaining years.

Acknowledgements: This work has been partially supported by national funds through FCT - Fundação para a Ciência e Tecnologia through project UIDB/04728/2020.

References

- [1] R. C. De Amorim and C. Hennig. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences* **324**: 126–145, 2015.
- [2] D. M. Hechavarría and A. E. Ingram. Entrepreneurial ecosystem conditions and gendered national-level entrepreneurial activity: a 14-year panel study of GEM. *Small Business Economics* **53**(2): 431–458, 2019.
- [3] J. Ledolter. Data mining and business analytics with R. *John Wiley & Sons*, 2013.
- [4] S. Singer, M. Herrington and E. Menipaz. *Global entrepreneurship monitor: Global report*, 2017/18, 2018.

A new RFM approach for customer segmentation using a SAF-T based business intelligence system

Rosa Silveira¹, Aldina Correia¹, Bruno Oliveira¹ and Telmo Matos¹

¹CIICESI, Escola Superior de Tecnologia e Gestão, Politécnico do Porto, Portugal

E-mail addresses: *8180717@estg.ipp.pt; aic@estg.ipp.pt; bmo@estg.ipp.pt; tsm@estg.ipp.pt*

In 2018, 70% of the Portuguese companies produced thousands of Standard Audit Files for Tax purpose (SAF-T Portuguese Version). These files contain valuable information that can represent an important tool for analytical procedures to support decision-making processes. For this purpose, a Business Intelligence System based on SAF-T files was created. Next, Statistical and Data Mining techniques were employed to create knowledge. In this particular case, a new approach of Recency (R), Frequency (F) and Monetary (M) criterion was implemented. Other attributes were suggested for customer segmentation as Age (A), Item average (I), Debt (D) and Customer Life Value (CLV) index. Next, various Clustering techniques were used to perform the customer segmentation in four different groups. Finally, the k-means classification was cross-validated by Linear Discriminant Analysis (LDA). Results show that accuracy is superior to 98% when R and CLV are not considered.

Keywords

Intelligent Control Systems, Data Warehousing, SAF-T (PT), Clustering, Linear Discriminant Analysis.

According to the Portuguese Institute of Statistics (INE), in 2018, 99% of the Portuguese companies were Small or Medium-sized Enterprises (SMEs) and 96% of these were Micro-companies, that is, companies that employ less than ten individuals and bill less than two million euros per year. From another perspective, since 2017, the Portuguese Tax Authority (PTA) requires all companies that have organized accounting (approx.70% of all) to monthly submit the Standard Audit File for Tax Purposes (SAF-T (PT)) for validation.

Generally, those companies have a disorganized or immature digital level and have significant difficulties for implementing intelligent control systems or Business Intelligence (BI) systems once resources are limited in a technical or human perspective or simply because they do not have data management policies that allow them to access to consistent and current data. Usually, these premises are the cause of the non-return in a BI investment. In fact, one of the biggest challenges for the implementation is in the data itself. Due to its lack of maturity, the implementation of the BI system can be compromised and the risk of error in the decision-making process increased.

To minimize this problem the SAF-T (PT) were considered, once they are standard documents for representing companies' accounting data regardless of the operational software, used for capturing business events. Although simple, these documents provide the basis for using a standard data representation to create a skeleton of data storage systems that can be further enriched to support more business requirements. In this way data standardization, its historical and evolutionary perspective can be used to provide knowledge to the company [1]. This was our motivation for this particular study that

aims to achieve Customer Segmentation based on SAF-T (PT) attributes and facilitate the decision making process from the user's perspective, once owners do not need to have technical background to interpret the data provided by the systems' interface.

To address this challenge, firstly, the data warehouse was created based on SAF-T (PT) attributes and data was simulated for companies. Then, Statistical and Data Mining techniques were applied to perform customer segmentation using a new approach of the Recency (R), Frequency (F) and Monetary (M) criterion. In the pre-processing phase product lines and invoices were aggregated, variables were transformed and normalized and a company randomly chosen. The result is a consistent data set with 52 customers and 7 attributes where: (R) Recency is the period after the last purchase; (F) Frequency, the count of purchases in a given period; (M) Monetary is the amount spent in that period; (A) Age, the period since the first purchase (age of relationship); (I) Item is the average number of item purchased by a customer; (D) Debt is the credit amount in the same period; and, finally, (C) CLV is the Customer Life Value Index that measures the importance of a particular client for the company. The CLV index was created based on a weighted criterion of the values of the previous attributes where weights from 1 to 5 were applied, considering the quintiles of each distribution and the importance of each to the company.

In the modelling phase, various Clustering techniques were performed. First, an exploratory analysis based on three hierarchical clustering methods was used - Single, Ward and Complete linkage - to identify groups and define a range of solutions for the number of clusters (k). Second, the k-means clustering was implemented and k was optimized considering the literature and based on measures as the Dunn Index, the Silhouette, the Connectivity and finally the R-squared. The value of k clusters was then established at k=4. Finally the clustering results were validated by the Linear Discriminant Analysis (LDA) and assumptions were validated as performed by Marôco [2]. Results show that 98% of the cases were correctly classified and the percentage is 96.2 %, when using cross-validation.

Acknowledgements: This research has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e Tecnologia under the project UID/GES/04728/2020.

References

- [1] R. Kimball and M. Ross. *The data warehouse toolkit: the complete guide to dimensional modeling*. John Wiley & Sons, 2011.
- [2] J. Marôco. *Análise Estatística com o SPSS Statistics*. 7th Ed., Report Number, Lda, 2018.

A new kind of stability for the Bessel equation: the σ -semi-Hyers-Ulam stability

A. M. Simões^{1,2} and L. P. Castro²

¹CMA-UBI – Center of Mathematics and Applications, Department of Mathematics,
University of Beira Interior, Covilhã, Portugal

²CIDMA – Center for Research and Development in Mathematics and Applications,
Department of Mathematics, University of Aveiro, Portugal

E-mail addresses: *asimoes@ubi.pt; castro@ua.pt*

The main aim of the work is to seek adequate conditions to derive a different kind of stability for the Bessel equation and for the modified Bessel equation considering a perturbation of the trivial solution. We identify the σ -semi-Hyers-Ulam stability as in-between the two well known stabilities of Hyers-Ulam-Rassias and Hyers-Ulam. Sufficient conditions are obtained in order to guarantee Hyers-Ulam-Rassias, σ -semi-Hyers-Ulam and Hyers-Ulam stabilities for those equations. The results are obtained mostly based on convenient majorations and the use of integral techniques.

Keywords

Bessel Equation, Modified Bessel Equation, Hyers-Ulam Stability, σ -semi-Hyers-Ulam Stability, Hyers-Ulam-Rassias Stability.

We will analyse Hyers-Ulam, Hyers-Ulam-Rassias and the so-called σ -semi-Hyers-Ulam stabilities for the Bessel equation

$$x^2 y''(x) + xy'(x) + (x^2 - \alpha^2)y(x) = 0, \quad (8)$$

and for the modified Bessel equation

$$x^2 y''(x) + xy'(x) - (x^2 + \alpha^2)y(x) = 0,$$

by using initial conditions

$$y(a) = y'(a) = 0,$$

with $y \in C^2([a, b])$, for $x \in [a, b]$ where a and b are fixed positive real numbers and $\alpha \in \mathbb{R}$ or α is a pure imaginary number.

We define the new stability, introduced in [5] (see also [4]), in the following way:

Let σ be a non-decreasing function defined on $[a, b]$. If for each function y satisfying

$$|x^2 y''(x) + xy'(x) + (x^2 - \alpha^2)y(x)| \leq \theta, \quad x \in [a, b],$$

where $\theta \geq 0$, there is a solution y_0 of the Bessel equation (8) and a constant $C > 0$ independent of y and y_0 such that

$$|y(x) - y_0(x)| \leq C \sigma(x), \quad x \in [a, b]$$

then we say that the Bessel equation (8) has the σ -semi-Hyers-Ulam stability.

Acknowledgements: This work is supported by The Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020 and by the Center of Mathematics and Applications of University of Beira Interior (CMA-UBI) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDP/00212/2020 and UIDB/00212/2020.

References

- [1] N. B.-Belluot, J. Brzdęk and K. Ciepliński. On some recent developments in Ulam's type stability. *Abstract and Applied Analysis* **2012**: 41 pp, 2012.
- [2] E. Biçer and C. Tunç. On the Hyers-Ulam Stability of Laguerre and Bessel Equations by Laplace Transform Method. *Nonlinear Dynamics and Systems Theory* **17(4)**: 340–346, 2017.
- [3] L. P. Castro and A. M. Simões. Hyers-Ulam and Hyers-Ulam-Rassias Stability for a Class of Integro-Differential Equations. *Mathematical Methods in Engineering: Theoretical Aspects*, K. Tas, D. Baleanu and J.A. Tenreiro Machado (eds.), Springer, 81-94, 2019.
- [4] L. P. Castro and A. M. Simões. Different types of Hyers-Ulam-Rassias stabilities for a class of integro-differential equations. *Filomat* **31(17)**: 5379–5390, 2017.
- [5] L. P. Castro and A. M. Simões. Hyers-Ulam-Rassias stability of nonlinear integral equations through the Bielecki metric. *Mathematical Methods in the Applied Sciences* **41(17)**: 7367–7383, 2018.
- [6] Y. J. Cho, C. Park, T. M. Rassias and R. Saadati. *Stability of Functional Equations in Banach Algebras*. Springer International Publishing, Heidelberg, 2015.
- [7] G. L. Forti. Hyers-Ulam stability of functional equations in several variables. *Aequationes Mathematicae* **50**: 143–190, 1995.
- [8] D. H. Hyers. On the stability of linear functional equation. *Proceedings of the National Academy of Sciences USA* **27(4)**: 222–224, 1941.
- [9] S.-M. Jung. *Hyers-Ulam-Rassias Stability of Functional Equations in Mathematical Analysis*. Hadronic Press, Palm Harbor, 2001.
- [10] Th. M. Rassias. On the stability of the linear mapping in Banach spaces. *Proceedings of the American Mathematical Society* **72**: 297–300, 1978.

Cooperation, innovation and environmental sustainability: Portuguese companies research

Marta Teixeira^{1,2}, Aldina Correia^{1,2}, Alexandra Braga^{1,2} and Vítor Braga^{1,2}

¹CIICESI – Center for Research and Innovation in Business Sciences and Information Systems,

²Escola Superior de Tecnologia e Gestão, Politécnico do Porto, Portugal

E-mail addresses: *8140395@estg.ipp.pt; aic@estg.ipp.pt; amb@estg.ipp.pt; vbrega@estg.ipp.pt*

The increase in globalization has led to greater competitiveness, so that companies are successful they need to increase their competitive advantages, that is, develop new strategies to remain competitive in the market. Therefore, innovation is essential for the performance of firms [1]. In the innovation process, cooperation plays an important role. It helps to release internal restrictions on innovation, facilitating access to knowledge sources that facilitate the entire innovation process [2]. According to Freel and Harrison in [3], product innovations are influenced by cooperation with customers and institutions, while process innovations are driven by cooperation with suppliers and universities. Cooperation then serves as a mechanism to maximize the company's value because the greater the collaboration with partners, the greater the chance of obtaining more innovative products [4]. Taking into account the constant degradation of the environment, it is necessary that companies adopt major innovations in an environmentally sustainable way to be able to respond to the growing consumer demand for sustainable products and services [5]. Eco-innovation refers to innovation directed towards sustainability [6], being a type of innovation that causes new products that use clean energy, are less polluting and have less impact on the environment [7]. Eco-innovation is a way of addressing future environmental problems, taking into account the reduction of energy / resources / waste / consumption, through sustainable economic activities [8]. In addition to the concern for the environment, companies can adopt eco-innovation practices to improve their company's reputation, achieve cost savings, respond to market demand, enter new markets, act correctly or simply, to meet regulatory requirements [7,5]. The main objective of this study is to understand the influence of cooperation on innovation and the factors that contribute to decision making by companies to choose innovations with environmental benefits. Using the CIS 2014 database and the MANOVA multivariate techniques and Linear Regression, we found that there is a relationship between innovation and cooperation. Since the more innovation there is, the greater the existing cooperation in companies. The results also show that the factors that most contribute to the adoption of innovations with environmental benefits are essentially the current or expected demand in the market for environmental innovations, the improvement of the company's reputation and the high costs of energy, water or materials. This study presents several contributions, both from a theoretical and practical perspective. In theoretical terms, this study responds to a gap in the literature because there is no study using the three variables (eco-innovation, cooperation and innovation) simultaneously. In practical terms, this study aims to help companies realize the advantages that cooperation provides in the innovation process, as well as the benefits that eco-innovation practices provide to companies.

Keywords

Eco-innovation; MANOVA; Multiple Linear Regression.

Acknowledgements: This work has received funding from FEDER funds through P2020 program and from national funds through FCT - Fundação para a Ciência e a Tecnologia under the project UID/GES/04728/2020.

References

- [1] R. Belderbos and B. Lokshin. Cooperative r & d and firm performance. *Research policy* **33**(10): 1477–1492, 2004.
- [2] J. Carrillo-Hermosilla, P. Del Río and T. Könnölä. Diversity of eco-innovations: Reflections from selected case studies. *Journal of cleaner production* **18**(10-11): 1073–1083, 2010.
- [3] M. Freel and R. Harrison. Innovation and cooperation in the small firm sector: Evidence from "northern britain". *Regional Studies* **40**(4): 289–305, 2006.
- [4] T. Hellström. Dimensions of environmentally sustainable innovation: the structure of ecoinnovation concepts. *Sustainable development* **15**(3): 148–159, 2007.
- [5] J. Hojnik, M. Ruzzier and T. Manolova. Internationalization and economic performance: The mediating role of eco-innovation. *Journal of Cleaner Production* **171**: 1312–1323, 2018.
- [6] A. Ingram. Innovation and the Family Firm: Leadership, Mindsets, Practices and Tensions. PhD thesis, University of Cincinnati, 2011.
- [7] L. Miotti and F. Sachwald. Co-operative r&d: why and with whom?: An integrated framework of analysis. *Research policy* **32**(8): 1481–1499, 2003.
- [8] E. Severo, J. de Guimarães and E. Dorion. Cleaner production and environmental management as sustainable product innovation antecedents: A survey in brazilian industries. *Journal of Cleaner Production* **142**: 87–97, 2017.

Using a multidimensional statistical approach to select experts opinion in a context of naval operations

M. Filomena Teodoro^{1,2}, Mário J. Simões Marques¹, Isabel Nunes^{3,4}
and Gabriel Calhamonas⁴

¹Center for Computational and Stochastic Mathematics (CEMAT, IST, Lisbon University, Portugal

²Center of Naval Research(CINAV), Portuguese Naval Academy, Portuguese Navy Portugal

³UNIDEMI, Department of Mechanical and Industrial Engineering, Faculty of Sciences and Technology, New University of Lisbon, Portugal

⁴FCT, Faculty of Sciences and Technology, New University of Lisbon, Portugal

E-mail addresses: *maria.alves.teodoro@marinha.pt; mteodoro64@gmail.com*

The Portuguese Navy supported THEMIS project to build a decision support system (DSS) to optimize short time decisions under a disaster context, allowing to improve the performance of tasks execution with a reduction of costs.

In [4,3], the authors have considered the facilities and high qualified staff of Portuguese Navy and proposed a variant of the Delphi method, a method that is exceptionally useful where the judgments of individuals are considered as an important information source. They proposed a system that prioritize certain teams for specific incidents taking into account the importance of each team that acts in case of emergency.

In the present work we propose the application of multidimensional statistical techniques to create an importance index of experts opinion complementing the method used in [4,3].

In this manuscript we re-weight the experts experience that contributes to DSS. The results are obviously distinct the ones presented in [4,3].

Keywords

Decision Support System, Catastrophe, Hierarchical Classification, Discriminant Analysis, Multidimensional Scaling.

Acknowledgements: The author was supported by Portuguese funds through the *Center of Naval Research* (CINAV), Portuguese Naval Academy, Portugal and *The Portuguese Foundation for Science and Technology* (FCT), through the *Center for Computational and Stochastic Mathematics* (CEMAT), University of Lisbon, Portugal, project UID/Multi/04621/2019.

References

- [1] I. Borg and et. al. *Applied Multidimensional Scaling*. 2nd ed. Springer Science & Business Media, 2013. Available at DOI:10.1007/978-3-642-31848-1

-
- [2] J. Kruskal and M. Wish. *Multidimensional Scaling*. 2nd Ed., Thousand Oakes, CA: Sage, 1978. Available at DOI:<https://dx.doi.org/10.4135/9781412985130>
 - [3] I. L. Nunes, G. Calhamonas, M. Simões-Marques and M. F. Teodoro. Building a decision support system to handle teams in disaster situations - a preliminary approach, In: A. Madureira, A. Abraham, N. Gandhi, M. Varela, M. (eds.) *Hybrid Intelligent Systems. HIS 2018*. Advances in Intelligent Systems and Computing, AISC Vol. 923, pp. 551–559. Springer, Cham, 2020. Available at https://doi.org/10.1007/978-3-030-14347-3_54.
 - [4] M. Simões-Marques and et al. Applying a Variation of Delphi Method for Knowledge Elicitation in the Context of an Intelligent System Design. In: Nunes, I. L. (Eds) *Proceedings of Advances in Human Factors and System Interactions, AHFE 2019 Conference on Human Factors and System Interactions, Advances in Intelligent Systems and Computing, 24-28 July*, Washington D.C. USA , AISC, Vol. 959, pp. 386–398. Springer, Cham, 2020. Available at https://doi.org/10.1007/978-3-030-20040-4_35.

Contributed Poster

Minimize the production of scrap in the extrusion process

Fátima De Almeida¹, A. Filipe Ferrás¹, Aldina Correia¹ and
Eliaana Costa e Silva¹

¹CIICESI, ESTG, Politécnico do Porto

E-mail addresses: *mfalmeida.f@gmail.com; afmf@estg.ipp.pt; aic@estg.ipp.pt; eos@estg.ipp.pt*

In this work, an empirical study of a Portuguese company in the industrial sector is presented. The main objective of this work is to model the aluminium extrusion process, with the aim of minimizing the production of scrap.

Keywords

Aluminium Extrusion, Scrap, Sustainability, Extrusion Variables, Multiple Linear Regression.

In the discussion on environmental policies, the notion of eco-efficiency is often used. Eco-efficiency is defined as the delivery of products and services with competitive values, while reducing the ecological impacts and satisfying human needs. In the area of Environmental Policies, the Sustainability and the Eco-efficiency issues are becoming increasingly important, both for safeguarding the environment, through the awareness of power and society, as for the competitiveness of the companies [2,3].

In an environment of great competitiveness in which the companies operate, the improvements in the productive process is one of the differentiating factors for guaranteeing a strong competition. Also it is essential rationalizing energy consumption and natural resources [1,2,4]. For this it is necessary to understand the main operations and dynamics of the company.

This work presents an empirical study of a Portuguese company in the industrial sector. The problematic here presented is based on the company's growing concern to reduce the amount of scrap produced. From the literature research that was performed, a strong dependence relation between the different variables in the extrusion process was found. The main objective of this work is to model the aluminium extrusion process, with the aim of minimizing the production of scrap. For this several variables involved in the process are taking into account. Using statistical techniques, in particular multiple linear regressions, it was possible to identify the importance of the variables under study for the scrap production.

References

- [1] P. Groche, S. Wohletz, M. Brenneis, C. Pabst and F. Resch. Joining by forming a review on joint mechanisms, applications and future trends. *Journal of Materials Processing Technology* **214(10)**: 1972–1994, 2014.
- [2] E. H. Moors. Technology strategies for sustainable metals production systems: a case study of primary aluminium production in the Netherlands and Norway. *Journal of Cleaner Production* **14(12-13)**: 1121–1138, 2006.

- [3] S. Schmidheiny and L. Timberlake. *Changing course: A global business perspective on development and the environment*. Vol. 1, MIT press, 1992.
- [4] K. Singh and I. A. Sultan. Modelling and evaluation of KPIs for the assessment of sustainable manufacturing: An extrusion process case study. *Materials Today: Proceedings* **5(2)**: 3825–3834, 2018.

Structured families of multivariate collective models

Ana Cantarinha^{1,2}, Elsa Moreira^{3,4} and João T. Mexia^{3,4}

¹University of Évora, Portugal

²Research Center for Mathematics and Applications - CIMA, Portugal

³Nova University of Lisbon, Portugal

⁴Center for Mathematics and Applications - CMA, Portugal

E-mail addresses: *aicantarinha@gmail.com; efnm@fct.unl.pt*

Univariate collective models have played an important role in Actuarial Mathematics. We extend the theory to structured families of multivariate collective models.

We develop an application to forest fires in Portugal, where in these families we consider that for each treatment of a linear orthogonal model there is a multivariate collective model, which will able us to study of the influence of several factors on these models parameters.

Keywords

Collective Models, Risk Theory, Structured Families.

References

- [1] M. L. Centeno. *Teoria do Risco na Actividade Seguradora*. Celta Editora - Coleção Económicas Oeiras, 2003.
- [2] J. Grandell. *Aspects of Risk Theory*. Springer Series in Statistics, 1992.
- [3] J. T. Mexia. Best linear unbiased estimates, duality of F tests and Scheffé's multiple comparison method in the presence of controlled heterocedasticity. *Journal Comp. Stat. Data Analysis*. **10(3)**: 271–281, 1990.
- [4] H. Scheffé. *The Analysis of Variance*. John Wiley & Sons, 1959.

One-dimensional theories to study three-dimensional problems: theoretical and numerical aspects

Fernando Carapau^{1,2} and Paulo Correia^{1,2}

¹University of Évora, Portugal

²Research Center CIMA-UÉ, Portugal

E-mail addresses: *flc@uevora.pt; pcorreia@uevora.pt*

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

Cosserat Theory, One-dimensional Model, Mathematical Fluid Dynamics.

Acknowledgements: This work has been partially supported by Centro de Investigação em Matemática e Aplicações da Universidade de Évora (CIMA-UE) through the grant UIDB/04674/2020 of FCT-Fundação para a Ciência e a Tecnologia.

References

- [1] 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): 190–196 (1987).
- [2] C. Truesdell and W. Noll. *The non-linear field theories of mechanics*, 2nd edition, Springer, New York, 1992.
- [3] R. L. Fosdick and K. R. Rajagopal. Thermodynamics and stability of fluids of third grade, *Proc. R. Soc. Lond. A.*, **339**: 351–377 (1980).
- [4] 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): 927–941 (1993).

Extreme value index estimation under non-regularity and through generalized means

M. Ivette Gomes^{1,2}, Lgia Henriques-Rodrigues^{3,4} and Dinis Pestana^{1,2}

¹Universidade de Lisboa, Portugal

²Centro de Estatística e Aplicações (CEAUL), Portugal

³Universidade de vora, Portugal

⁴Centro de Investigao em Matemática e Aplicaes (CIMA), Portugal

E-mail addresses: *ivette.gomes@fc.ul.pt; ligiahr@uevora.pt; dinis.pestana@fc.ul.pt*

An average of statistics S_i can be regarded as the logarithm of the geometric mean, or the power mean of order $p = 0$, of $\exp(S_i)$, $1 \leq i \leq k$. Instead of such a geometric mean, we can more generally consider the power *mean of order* p (MO_p) of those statistics, $p \in \mathbb{R}$, and to build a class of MO_p -estimators. The Hill estimators, average of the log-excesses, are the most popular *extreme value index* (EVI) estimators for heavy right tails. We here consider associated MO_p EVI-estimators, under the so-called non-regular framework.

Keywords

Generalized Means, Non-Regular Frameworks, Semi-Parametric Estimation, Statistics of Extremes.

The *extreme value index* (EVI), denoted by ξ , is the main parameter in *extreme value theory* (EVT), measuring the heaviness of the *right-tail function* (RTF), $\bar{F}(x) := 1 - F(x)$. The heavier the right-tail, the larger ξ is. For Pareto-type models, with a positive EVI, the classical EVI-estimators are the Hill (H) estimators ([4]), which are the averages of the log-excesses, $V_{ik} := \ln X_{n-i+1:n} - \ln X_{n-k:n}$, $1 \leq i \leq k < n$, where $X_{i:n}$, $1 \leq i \leq n$, denotes the ascending *order statistics* (OSs) associated with a sample X_i , $1 \leq i \leq n$. Consequently,

$$H(k) := \frac{1}{k} \sum_{i=1}^k V_{ik} =: \sum_{i=1}^k \ln U_{ik}^{1/k} = \ln \left(\prod_{i=1}^k U_{ik} \right)^{1/k}, \quad U_{ik} = \frac{X_{n-i+1:n}}{X_{n-k:n}}, \quad 1 \leq i \leq k < n,$$

i.e. the Hill estimator is the logarithm of the *geometric mean* (or *mean-of-order-0*) of the statistics U_{ik} . More generally, the *mean-of-order- p* (MO_p) of U_{ik} was recently considered, together with the class of EVI-functionals,

$$H_p(k) \equiv \text{MO}_p(k) := \begin{cases} \left(1 - \left(\frac{1}{k} \sum_{i=1}^k U_{ik}^p \right)^{-1} \right) / p, & \text{if } p \neq 0 \\ H(k), & \text{if } p = 0, \end{cases} \quad (9)$$

a highly flexible class of functionals, dependent on a *tuning* parameter $p \in \mathbb{R}$ (see, [1,2], among others).

In the area of statistical EVT and whenever working with large values, a *cumulative distribution function* (CDF), F , is often said to be *heavy-tailed* whenever the RTF $\bar{F} \in \mathcal{R}_{-1/\xi}$, $\xi > 0$, where \mathcal{R}_α denotes the class of regularly varying functions with an index of

regular variation α . Then, F is in the max domain of attraction of the EV_ξ CDF, and the notation $F \in \mathcal{D}_{\mathcal{M}}(\text{EV}_\xi)_{\xi>0} =: \mathcal{D}_{\mathcal{M}}^+$ is used. Reciprocally, if $F \in \mathcal{D}_{\mathcal{M}}^+$, we necessarily have $\overline{F} \in \mathcal{R}_{-1/\xi}$, the so-called *first-order condition* (FOC). Also, with the notation $F^{\leftarrow}(t) := \inf\{x : F(x) \geq t\}$ for the generalised inverse function of F , the reciprocal tail quantile function $U(t) := F^{\leftarrow}(1 - 1/t)$ is in \mathcal{R}_ξ . The *second-order parameter* ρ (≤ 0) controls the rate of convergence in the FOC, which is ruled by $A(t)$, such that $|A| \in \mathcal{R}_\rho$.

To have consistency of the H_p EVI-estimators, in all $\mathcal{D}_{\mathcal{M}}^+$, we not only have to work with intermediate values of k , i.e. a sequence of integers $k = k_n$, $1 \leq k < n$, such that $k = k_n \rightarrow \infty$ and $k_n = o(n)$, as $n \rightarrow \infty$, but also to have $p < 1/\xi$, in (9). Under the aforementioned *second-order condition* (SOC), ruled by the function A , and with $\sigma_p(\xi) := \xi(1 - p\xi)/\sqrt{1 - 2p\xi}$, $b_p(\xi|\rho) := 1 - p\xi/(1 - p\xi - \rho)$,

$$H_p(k) \stackrel{d}{=} \xi + \sigma_p(\xi) \mathcal{N}_\infty^{(p)}(0, 1)/\sqrt{k} + b_p(\xi|\rho) A(n/k) + o_p(A(n/k))$$

holds for all $p < 1/(2\xi)$ and $\rho \leq 0$, with $\mathcal{N}_\infty^{(p)}(0, 1)$ asymptotically standard normal.

But the simulated mean values, at optimal levels, of $H_p(k)$, $p = 1/\xi$, provide very interesting results, for a large variety of underlying models and different values of ξ . Such a behaviour was theoretically investigated in [3]. We here consider large-scale simulations, under non-regular frameworks, i.e. for $p > 1/(2\xi)$, and illustrations with simulated random samples from known heavy-tailed models and for real data in the field of environment.

Acknowledgements: Research partially supported by National Funds through **FCT**—Fundação para a Ciência e a Tecnologia, projects UIDB/MAT/0006/2020 (CEA/UL) and UIDB/MAT/04674/2020 (CIMA).

References

- [1] M. F. Brilhante, M. I. Gomes and D. Pestana. A simple generalisation of the Hill estimator. *Comput. Statist. and Data Analysis* **57(1)**: 518–535, 2013.
- [2] F. Caeiro, M. I. Gomes, J. Beirlant and T. de Wet. Mean-of-order p reduced-bias extreme value index estimation under a third-order framework. *Extremes* **19(4)**: 561–589, 2016.
- [3] M. I. Gomes, L. Henriques-Rodrigues and D. Pestana. *Non-regular Frameworks and the Mean-of-order- p Extreme Value Index Estimation*. Preprint, <https://doi.org/10.13140/RG.2.2.28347.64804>, 2020.
- [4] B. M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* **3**: 1163–1174, 1975.

Comparison of time series models' performance applied to economic data forecasting

Susana Lima¹, A. Manuela Gonçalves² and Marco Costa³

¹University of Minho, MEtRICs Research Center, Portugal

²University of Minho, Department of Mathematics and Center of Mathematics, Portugal

³University of Aveiro, Águeda School of Technology and Management and Center for Research and Development in Mathematics and Applications, Portugal

E-mail addresses: *susanamrlima@gmail.com; mneves@math.uminho.pt; marco@ua.pt*

In Economics, particularly in the retail segment, sales forecasting supports most of the strategic planning decisions and its accuracy is crucial to ensure profitability. The main purpose of this work is to compare the accuracy of various proposed forecasting methods applied to a monthly retail sales time series in Portugal from 2000 to 2018 in order to identify the most appropriate methodologies.

Keywords

Retail Sales, Time Series, Forecast, Accuracy, Evaluation Measures.

Like many other economic time series, retail sales present strong trend and seasonal patterns. How to best model and forecast these patterns has been a long-standing issue in time series analysis. This study compares the forecasting performance of economic time series based on ARMA models and their extensions [2], on classical decomposition time series associated with multiple regression models with correlated errors [1], and on exponential smoothing methods as the Holt-Winters method [4]. These methods are applied because of their ability to model trend and seasonal fluctuations present in economic data, particularly in retail sales data.

In this work, these models' formulations are performed to the same economic dataset: monthly indexes of turnover (total, TOVT) of retail trade in Portugal, collected in the EUROSTAT retail databases [3,2]. The collected data were divided into two sets: training data (in-sample data) and testing data (out-of-sample data) in order to test the accuracy of the suggested forecasting models. The selected training period was from January 2000 to December 2016 (the first 204 observations), and the test period was from January 2017 to February 2018 (the last 14 observations). To evaluate the predictive capacity of the three methodologies adopted, several evaluation measures are used, namely MSE, RMSE, and U-Theil statistic.

Acknowledgements This work is supported by the Portuguese Foundation for Science and Technology (FCT- Fundação para a Ciência e a Tecnologia) Projects UIDB/00013/2020 and UIDP/00013/2020 of CMAT-UM, and is also supported by the Center for Research and Development in Mathematics and Applications (CIDMA) through the Portuguese Foundation for Science and Technology (FCT - Fundação para a Ciência e a Tecnologia), references UIDB/04106/2020 and UIDP/04106/2020. Susana Lima was financially supported by UMINHO/BI/145/2020.

References

- [1] T. Alpuim and A. El-Shaarawi. Modeling monthly temperature data in Lisbon and Prague. *Environmetrics* **20(7)**: 835–852, 2009.
- [2] G. Box, G. Jenkins, G. Reinsel and G. Ljung. *Time series analysis: forecasting and control*. 5th Ed., New Jersey: John Wiley and Sons, 2016.
- [3] EUROSTAT databases. <https://ec.europa.eu/eurostat/data/database> (accessed in 04.May.2018).
- [4] R. Hyndman, A. B. Koehler, J. K. Ord and R. D. Snyder. *Forecasting with Exponential Smoothing. The State Space Approach*. 1st Ed., Berlin, Germany: Springer, 2008.
- [5] S. Lima, A. M. Gonçalves and M. Costa. Time series forecasting using Holt-Winters exponential smoothing: An application to economic data. *AIP Conference Proceedings* **2186**, 090003-1-090003-4, 2019.

Geographically weighted panel logistic regression to model COVID-19 data

Conceição Leal^{1,2}, Teresa Oliveira^{1,2} and Amílcar Oliveira^{1,2}

¹Departamento de Ciências e Tecnologia, Universidade Aberta, Portugal

²Centro de Estatística e Aplicações, Universidade de Lisboa, Portugal

E-mail addresses: *conceicao.leal2010@gmail.com; teresa.oliveira@uab.pt; amilcar.oliveira@uab.pt*

In this paper, we aim to present a review of the theoretical foundations of the Geographically Temporal Weighted Regression Model, considering a panel data approach, suitable for modeling COVID-19 data at the local level - Municipalities of Mainland Portugal.

Keywords

Geographically and Temporally Weighted Regression, COVID-19

With the number of cases of infections by COVID-19 growing, modeling the transmission dynamics, identifying possible factors and estimating its development are crucial to providing decisional support for public health departments and healthcare policy makers.

There is some literature reporting statistical analyzes and future scenarios of the COVID-19 epidemic based on logistic regression models, without considering the spatial component or temporal correlation.

Spatial regression models can effectively estimate the influence of independent factors on target variables by differentiating the spatial dependence by including the lag and error components of independent features [2,4]. Geographically Weighed Regression (GWR) [1] is a local form of linear regression to model the spatially varying association between dependent and independent variables. And it is also the most frequently used method to study spatial nonstationary. With the Geographically Temporal Weighted Regression Approach [3], the temporal correlation of the data can also be integrated in the model. Because one of the most important properties of epidemics is their spatial spread, which is conditioned by factors that vary spatially and temporal, and that can be spatially and temporal correlated, the use of models that accommodate spatial and temporal effects allows results to be more consistent with the actual data.

In this paper, we will review the theoretical foundations of the Geographical Time Weighted Regression Model considering a panel data approach, and discuss its application to data on confirmed cases of COVID-19 in the Municipalities of Mainland Portugal.

References

- [1] C. Brunsdon, A. S. Fotheringham and M. E. Charlton. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis* **28**(4): 281–298, 1996.

-
- [2] H. Rahman, N. M. Zafri, F. Ashik and Waliullah. Gis-Based Spatial Modeling to Identify Factors Affecting COVID-19 Incidence Rates in Bangladesh. <https://ssrn.com/abstract=3674984orhttp://dx.doi.org/10.2139/ssrn.3674984>. 2020 (Accessed 31.08.2020).
 - [3] J. Liu, Y. Yang, S. Xu, Y. Zhao, Y. Wang and F. Zhang. A geographically temporal weighted regression approach with travel distance for house price estimation. *Entropy* **18(8)**: 303, 2016.
 - [4] S. Sannigrahi, F. Pilla, B. Basu, A. S. Basu and A. Molter. Examining the association between socio-demographic composition and COVID-19 fatalities in the European region using spatial regression approach. *Sustainable Cities and Society*. Vol **62**: 2020.

Control charts for attribute control based on life distributions with applications on e-learning classes monitoring

Teresa A. Oliveira^{1,2} and Amílcar Oliveira^{1,2}

¹Universidade Aberta, Portugal

²Centro de Estatística e Aplicações, Faculdade de Ciências, Universidade de Lisboa, Portugal

E-mail addresses: *teresa.oliveira@uab.pt; amilcar.oliveira@uab.pt*

Charts for attribute control will be explored, considering: (i) the number of defective items, named np-charts; (ii) means of the proportion (p) of defective items, named p-charts; (iii) u-charts; (iv) c-charts. An update and some recent ideas on the subject based on lifetime distributions will be presented, as well as the advantages and disadvantages of each kind of these charts. Implementations in software R will be discussed through examples in different contexts.

Keywords

Life Distributions, Control Charts, R Software.

Acknowledgements: This research was supported partially by National Funds through Fundação para a Ciência e a Tecnologia of the Portuguese government, project grant UIDB/00006/2020 of the CEAUL, Faculdade de Ciências, Universidade de Lisboa, Portugal

References

- [1] Z. W. Birnbaum and S. C. Saunders. A new family of life distributions, *Journal of Applied Probability* **6**: 319–327, 1969.
- [2] V. Leiva, G. Soto, E. Cabrera and G. Cabrera. New control charts based on the Birnbaum-Saunders distribution and their implementation, *Colombian Journal of Statistics* **34**: 147–176, 2011.
- [3] V. Leiva and T. A. Oliveira. p-charts for Attribute Control, *Wiley StatsRef: Statistics Reference Online*, 1–8, 2015.
- [4] V. Leiva and T. A. Oliveira, A. np-charts for Attribute Control, *Wiley StatsRef: Statistics Reference Online*, 1–6, 2015.

Mechanical behavior of the skin: A Statistical Approach

M. Filomena Teodoro^{1,2}

¹Center for Computational and Stochastic Mathematics CEMAT/IST/Lisbon

University, Portugal

²Center of Naval Research(CINAV), Portuguese Naval Academy, Portuguese Navy
Portugal

E-mail addresses: *maria.alves.teodoro@marinha.pt; mteodoro64@gmail.com*

The objective of this study is to verify if it is possible to reproduce the mechanical behavior of the skin subject to indentation tests. An indentation test allows to relate the pressure force with the maximum deformation which is got when the maximum pain is reached, taking into account the moment when the pain threshold occurs (onset of pain).

An analysis of variance [4] was applied to the experimental data and verifies that the limits of load application in safety and comfort remain stable for a group of individuals, following the work presented in [1,2].

The approximation to the experimental data produced adequate indicators, in accordance with [3]. The behavior of strength and deformation at the pain threshold and at maximum bearable pain were modeled, having distinct profiles between men and women.

Keywords

Skin, Pain, Indentation Tests, ANOVA.

Acknowledgements: The author was supported by Portuguese funds through the *Center of Naval Research* (CINAV), Portuguese Naval Academy, Portugal and *The Portuguese Foundation for Science and Technology* (FCT), through the *Center for Computational and Stochastic Mathematics* (CEMAT), University of Lisbon, Portugal, project UID/Multi/04621/2019. The author appreciates the availability of data by colleagues Paula Silva and Célio Figueiredo-Pina.

References

- [1] P. Silva. Computational Modelling of a Wearable Ankle-Foot Orthosis For Locomotion Analysis and Comfort Evaluation. PhD Thesis, DEM, Instituto Superior Técnico, Lisbon University, Lisbon, 2012.
- [2] M. Rodrigues, V. Bernardo, P. Silva and C. Figueiredo-Pina. Influência da Velocidade de Penetração No Limiar da Dor à Compressão. *Proc. 5º Cong. Nac. Biomecânica*, Espinho, Portugal, Fevereiro, 2013.
- [3] W. C. Mann. Smart Technology For Aging, Disability And Independence: The State Of The Science. John Wiley & Sons, 2005. Available at <https://doi.org/10.1002/0471743941.ch1>
- [4] A. C. Tamhane and D. D. Dunlop. Statistics and Data Analysis: from Elementary to Intermediate. Prentice Hall, New Jersey, 2000.

The global satisfaction of Portuguese expatriate workers: application of PLS-SEM

Luís M. Grilo^{1,2,3,4}, Graciete Honrado¹ and Ricardo Vitorino¹

¹Instituto Politécnico de Tomar (IPT) & Universidade Aberta, Portugal

²Centro de Matemática e Aplicações (CMA), FCT, UNL, Portugal

³Centro de Investigação em Cidades Inteligentes (Ci2), IPT, Portugal

⁴CIICESI, ESTG, Politécnico do Porto, Portugal

E-mail addresses: *lgrilo@ipt.pt; graciete.honrado@ipt.pt; ricardo.wr.vitorino@gmail.com*

Portuguese expatriates (on their own initiative or at the decision of the companies they work for) face various difficulties in their personal and professional lives in the destination countries [1,6]. To assess this situation, a survey was conducted through a questionnaire validated in previous studies on the topic. The answers to the questions were expressed on a Likert-type scale with seven categories and generated a dataset of 207 observations. In order to build a model based on the perceptions of workers, implicit in their responses, on global satisfaction with expatriation, the multivariate statistical technique Structural Equation Modeling (SEM) was used. Based on the specialized literature and empirical knowledge, a reflective theoretical model, that reflects a set of structural relationships, was proposed. The estimated model was obtained using the consistent Partial Least Squares (PLSc), which provides a correction for estimates and relaxes the demands on data, like the strong assumption of multivariate normality [2–5]. The latent exogenous construct *satisfaction with the destination country* has a direct effect on the constructs *personal satisfaction* and *global satisfaction*. The latent construct *job satisfaction* has a direct effect on *personal satisfaction* and an indirect effect on the endogenous *global satisfaction*, through the mediator construct *personal satisfaction*. These findings allow a better understanding of the relationship between the variables under study and the impact they have on the global satisfaction of Portuguese expatriate workers.

Keywords

Bootstrapp, Consistent PLS, Job Satisfaction, Latent Constructs, Primary Data.

Acknowledgements: This work is funded by national funds through the FCT - Fundação para a Ciência e a Tecnologia, I.P., under the scope of the project UIDB/00297/2020 (Center for Mathematics and Applications).

References

- [1] Y. Baruch, Y. Altman and R. L. Tung. Career mobility in a global era: Advances in managing expatriation and repatriation. *The Academy of Management Annals* **10**(1): 841–889, 2016.
- [2] L. M. Grilo, A. Mubayi, K. Dinkel, B. Amdouni, J. Ren and M. Bhakta. Evaluation of Academic Burnout in College Students. Application of SEM with PLS approach. *AIP Conf. Proc.* 2186, 090008, 2019; <https://doi.org/10.1063/1.5138004>.

- [3] L. M. Grilo, H. L. Grilo and E. Martire. SEM using PLS Approach to Assess Workers Burnout State. *AIP Conf. Proc.* 2040, 110008-1-110008-5, 2018; <https://doi.org/10.1063/1.5079172>.
- [4] J. F. Hair, G. T. M. Hult, C. M. Ringle and M. Sarstedt. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*, (2nd Ed., Thousand Oaks, CA: Sage), 2017.
- [5] C. M. Ringle, S. Wende and J.-M. Becker. “SmartPLS 3.” Boenningstedt: SmartPLS GmbH, 2015, www.smartpls.com.
- [6] V. Vaiman, A. Haslberger and M. C. Vance. Recognizing the important role of self-initiated expatriates in effective global talent management. *Human Resource Management Review* **25**: 280–286, 2015.

Index of authors

- Achchab, Boujemaa, *108*
Akoto, Isaac, *15, 18, 19*
Almeida, Fátima De, *127*
Antunes, Patrícia, *16*
Arunachalam, Viswanathan, *95, 109*
- Banerjee, Malay, *1*
Barbeiro, Sílvia, *2*
Boreda, Viraj, *51*
Borges, Ana, *65, 111, 115*
Borges, Isabel, *29, 32, 38, 43*
Braga, Alexandra, *121*
Braga, Vítor, *71, 82, 87, 98, 121*
Braumann, Carlos A., *67, 91*
Brites, Nuno M., *67*
Brito, Paula, *75*
Brito, Thadeu, *96*
- Caeiro, Frederico, *56*
Calhamonas, Gabriel, *123*
Cantarinha, Ana, *31, 129*
Carapau, Fernando, *130*
Cardante, Mariana, *69*
Carneiro, Davide, *111*
Carvalho, Davide, *104*
Castro, L. P., *119*
Coelho, Carlos A., *34*
Correia, Aldina, *65, 69, 71, 87, 115, 117, 121, 127*
Correia, Paulo, *130*
Costa, Cláudia, *89*
Costa, Marco, *89, 133*
Costa, Sérgio Cavaleiro, *73*
- Dias, Cristina, *29, 32, 36, 38, 43, 45, 46*
Dias, Sónia, *75*
- Esquível, Manuel L., *77*
- Faria, Susana, *79*
Ferrás, A. Filipe, *127*
Ferreira, Dário, *15–19*
Ferreira, Flora, *111*
Ferreira, Sandra, *16–19*
Filipe, Patrícia A., *91*
Filus, Jerzy K., *80*
Filus, Lidia Z., *81*
Francisco, Raquel, *82*
Frenkel, Sergey, *84*
- Garção, Eugénio, *21*
Ghosh, Indranil, *4*
Gomes, Carlos, *87*
Gomes, M. Ivette, *56, 131*
Gonçalves, A. Manuela, *89, 133*
Gonçalves, Sofia, *71*
Grácio, Maria, *106*
Granja, Isabela da Costa, *25*
Grilo, Luís M., *21, 58, 139*
Guerreiro, Gracinda R., *15*
- Hajjaji, Abdelowahed, *100*
Honrado, Graciete, *139*
- Jacinto, Gonçalo, *91*
Janeiro, Fernando M., *73, 100*
Johnson, Kendall, *54*
Jordanova, Pavlina, *60*
Jorge, Nadab, *34*
Josephra, John, *95*
- Kandoussi, Khalid, *100*
Krasii, Nadezhda P., *77*
Kumar, Kirthi, *51*
- Leal, Conceição, *135*
Lima, José, *96*
Lima, Susana, *133*
Lopes, Emanuel, *93*
Lopes, Luís, *106*
- Malico, Isabel, *73*
Marques, Filipe J., *34*
Marques, Mário J. Simões, *123*
Martínez, Erika Johanna, *95*
Matos, Telmo, *117*
Matto, Holly, *54*
Mendes, João, *96*
Mendes, Luzia, *104*
Mendes, Telma, *98*
Mesbahi, Oumaima, *100*
Mexia, João T., *15–19, 21, 31, 36, 43, 45, 46, 129*
Milan, Stehlík, *62*
Moreira, Elsa, *31, 129*
Mubayi, Anuj, *109*
Muthakana, Udbhav, *52*
- Nafidi, Ahmed, *108*
Ngarnim, Soraya, *51*

- Nunes, Célia, *17–19, 34, 36*
Nunes, Isabel, *123*
- Oliveira, Amílcar, *135, 137*
Oliveira, Bruno, *117*
Oliveira, Manuela, *21*
Oliveira, Teresa, *104, 135, 137*
- Peixoto, Nathalia, *51*
Perdoná, Gleici da Silva Castro, *24*
Pereira, Ana I., *96, 113*
Pereira, Fernanda, *102*
Pereira, J. A. Lobo, *104*
Perestrelo, Sara, *106*
Pestana, Dinis, *131*
Pinto, Miguel Gonçalves, *104*
- Ríos-Gutiérrez, Andrés, *109*
Raissi, Maziar, *52*
Ribeiro, Nuno, *106*
Rida, Oussama, *108*
Ringle, Christian M., *9*
Rodrigues, Andressa Contarato, *22, 24*
Rodrigues, Lúcia Henriques, *56, 131*
Rodrigues, Pedro Ivo, *22*
- Salgado, Carla, *79*
- Sanfins, Marco Aurélio, *22, 25*
Santos, Carla, *29, 32, 36, 38, 43, 45, 46, 48*
Santos, Daiane Rodrigues dos, *25*
Santos, Marta, *111*
Santos, Pablo Silva Machado Bispo dos, *25*
Sena, Inês, *113*
Sequeira, Adélia, *5*
Seshaiyer, Padmanabhan, *51–54*
Silva, Carina, *98*
Silva, Eliana Costa e, *65, 82, 93, 115, 127*
Silva, Tuany Esthefany Barcellos de
Carvalho, *25*
Silveira, Rosa, *117*
Simões, A. M., *119*
Sousa, Sandra P., *82*
Stehlík, Milan, *60*
Suresh, Nikhil, *75*
- Teixeira, Marta, *121*
Teodoro, M. Filomena, *123, 138*
Tlemçani, Mouhaydine, *100*
Todorov, Venelin, *60*
- Valencia, Diego, *54*
Varadinov, Maria, *29, 32, 38, 43, 48*
Vitorino, Ricardo, *139*