SMTDA 2026-

Hybrid

9th Stochastic Modeling Techniques and Data Analysis International Conference
with Demographics 2026 Workshop

Venue: Hellenic Mediterranean University

Department of Management and Technology

Lakonia Campus

Agios Nikolaos, Crete, Greece

8 – 11 June, 2026

Book of Abstracts

Preface

Dear Guests, Dear Colleagues,

We welcome you to the 9th Stochastic Modeling Techniques and Data Analysis International Conference and Demographics 2026 Workshop in Agios Nikolaos.

As the previous years, a considerable number of high-quality research papers are accepted for oral or poster presentations.

Important speakers present onsite and virtually their work promoting scientific development and research.

We have submissions from 33 countries: Algérie, Belgium, Brasil, Bulgaria, Canada, Czech Republic, France, Greece, Hungary, India, Italy, Japan, Latvia, Lithuania, Luxembourg, Malta, Mexico, Northern Ireland, P. R.of China, Poland, Portugal, Russia, Serbia, Slovakia, Spain, Sweden, Taiwan, United Kingdom, Ukraine, United Arab Emirates. USA, Uzbekistan.

I'd like to thank the Committees of the Conference which worked hard for this success.

I'd also like to thank the Region of Crete, our Sponsor.

Dr Christos H. Skiadas
ex- vice Rector
Technical University of Crete
Chair of ISAST

Plenary Talks

Insurance coverage for a SIR epidemic

Claude Lefèvre

Department of Mathematics, Université Libre de Bruxelles.

claude.lefevre@ulb.be

Abstract. SIR epidemic processes are well-established models for describing the spread of an infectious disease with permanent removals (such as COVID-19). First, we will present this class of models with the novelty that the epidemic faresponse to the progress of the epidemic. Next, our objective will be to define a premium strategy for an insurance company that has to cover a population against such an epidemic risk. Typically, the company will reimburse medical expenses to the infected individuals during their period of treatment and may also pay a lump-sum to the removed individuals in case of death or disability. To this end, the company will collect premiums from the susceptible individuals as long as they remain healthy. Using the classical equivalence principle in life insurance, we want to obtain the level of premium to apply by the insurance based on the entire duration of the epidemic. To get the expected benefit outgo, we will determine the expected total infectivity duration and the expected total number of removed cases. To get the expected premium income, we will determine the expected total susceptibility duration. The proposed methodology will be illustrated by a few numerical examples.

Structural Properties and Role of Load-Sharing Models in Applied Probability and Social Choices

Fabio L. Spizzichino, Fondazione Sapienza, Rome, Italy

Abstract: The term Time-Homogeneous Load-Sharing (THLS) Models designates a remarkable class of jointly absolutely continuous probability distributions for random vectors $X \equiv (X_1, \ldots, X_m)$. The random variables $X_1, \ldots, X_m$ are non-negative and typically represent lifetimes of devices which are embedded in a same environmental situation and share some common loss or benefit. Furthermore such devices start working simultaneously and the possibility of a complete observation is assumed about the subsequent failures of them (Longitudinal Observation).

Such models have a long tradition in the Reliability field and are defined as follows: for $1 \leq i_1 \neq \ldots \neq \neq i_h \neq j \neq < n$ and for positive constants $\mu_j(\emptyset), \mu_j(i_1, \ldots, i_h)$, X is jointly distributed according to a THLS model with parameters $\mu_j(\emptyset), \mu_j(i_1, \ldots, i_h)$, if

$$\lambda_j(t|\emptyset) = \mu_j(\emptyset) \;;\; \lambda_j^{h+1}(t|x_1, \ldots, x_h; i_1, \ldots, i_h) = \mu_j(i_1, \ldots, i_h)$$

where $\lambda_j(t|\emptyset)$ and $\lambda_j^{h+1}(t|x_1, \ldots, x_h; i_1, \ldots, i_h)$ are the multivariate conditional hazard rate functions. See e.g. [Shaked, Shantikumar (2015)], [S. (2019)], and references cited therein, for more detailed reviews and complete lists of bibliographic references.

Besides traditional applications in Reliability theory, both the modeling value and the conceptual interest of THLS models can be found in several different fields of applied probability, and beyond.

As it will be pointed out in a first part of this talk, the THLS models can be considered as simple generalizations of the joint distributions of independent (non-identical distributed) exponential lifetimes.

The discussion in the second part will be based on such a background and on results recently obtained in [De Santis, S. (2023a)], [De Santis, S. (2023b)]: it will be shown how (and way) THLS models can have an important role in several different studies, even far away from the reliability problems. Specifically I will point out the interest - and the meaning- of such models in the fields of Social Choices, and of Voting Theory in particular.

Explanatory distributions for Business and Healthcare process durations

Sally McClean

Ulster University, School of Computing, Northern Ireland, UK,

I plan to discuss aspects of Explanatory AI, Root Causes, real-time processes alongside relevant duration distributions such as: Phase-type distributions, Mixed gamma distributions, and Mixed Weibull distributions. These topics will be explored with regard to software, business, smart homes, and hospital billing duration data.

INVITED TALKS

A flexible semiparametric step-stress model based on piecewise monotone hazard function

Aman Prakash1; Raj Kamal Maurya1, Debashis Samanta2 and Debasis Kundu3

1Department of Mathematics, National Institute of Technology, Surat, India 2Department of Mathematics and Statistics, Aliah University, Kolkata, West Bengal, India 3Department of Mathematics and Statistics, Indian Institute of Technology Kanpur, India

Abstract: This work proposes a flexible semiparametric framework for a simple step-stress accelerated life-testing experiment in the presence of competing risks data. The stress is assumed to change from an initial level to a higher level at a pre-specified time. Instead of assuming a single parametric form for the hazard function at each stress level, it is assumed that the underlying lifetime distribution has a piecewise monotone hazard function. Within each sub-interval, the cause-specific hazard rates are assumed to follow a piecewise monotone hazard function. This framework provides a good approximation to a broad class of hazard functions and allows the failure rate to exhibit increasing, decreasing, or constant patterns over time, depending on the underlying failure mechanism. It is a natural generalization of the piecewise constant hazard function. The piecewise constant hazard function can be obtained as a special case. We develop both classical and Bayesian inference procedures for the proposed framework. In the classical framework, maximum likelihood estimation is used to estimate model parameters, and their asymptotic confidence intervals are constructed. Bayesian inference is carried out by assuming a flexible prior distribution. We obtained the Bayes estimates and the associated credible intervals for the model parameters under the squared-error loss function. The performance of the proposed estimators has been examined through an extensive simulation study. A real-life dataset is analyzed to illustrate the practical applicability and effectiveness of the proposed methodology. In practice, the locations of the cut points are often unknown and play a crucial role in model performance. So, we address this issue by determining optimal cut points using

a non-homogeneous Poisson process on the real dataset. Keywords: Accelerated life testing, Piecewise monotone hazard function, Competi

ng risk, Step stress models, Maximum likelihood estimation, Bayesian estimation

A hybrid nonparametric framework for outlier detection in functional time series

David Solano1,2, Marc Saez1,2, Maria A. Barceló1,2

1 Research Group on Statistics, Econometrics and Health (GRECS), University of Girona, Spain

2 Professor Dr, Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Spain

Dr., University of Girona, Carrer de la Universitat de Girona 10, Campus de Montilivi, Girona 17003,

Abstract. Outlier detection in functional time series is challenging due to temporal dependence and thecoexistence of magnitude, shape, and partially contaminated anomalies. Existing methods often assume independence or rely on model-based approaches, such as the Standard Smoothed Bootstrap on Residuals (SmBoR), which may perform poorly under model misspecification. Model-free alternatives, based on the Moving Block Bootstrap (MBBo), improve robustness but may exhibit modest true positive rates for magnitude anomalies. This work proposes a fully model free pipeline with two components. First, the Directional Outlyingness (DirOut) framework is extended by recalibrating its cutoff via MBBo, improving the detection of shape and partial outliers while controlling false positives. Second, a Sliding Window Functional Boxplot (SWOD) is introduced to exploit local temporal neighborhoods and detect magnitude anomalies that global summarie s may miss. Simulations show that SWOD achieves high detection rates for magnitude outliers, while MBBo calibrated DirOut attains close to perfect detection for shape and partial anomalies, outperforming SmBoR. The approach is further validated on a real temperature dataset, demonstrating its practical effectiveness.

Keywords: Functional time series; Outlier detection; Directional outlyingness; Functional depth; Moving block bootstrap; Sliding window methods; Magnitude and shape anomalies; Derivative-based features

A methodological approach to modeling the advanced ages reporting in censuses in countries undergoing consolidated demographic transition

Neir Antunes Paes

(PPGMDS/UFPB)Brazil.

João Batista Carvalho

Unidade Acadêmica de Estatística (CCT/UFCG)Brazil

Rozane Pereira de Sousa

(PPGMDS/UFPB)Brazil.

Abstract. The demographic transition is characterized by the shift from a scenario of high birth and death rates to one marked by lower levels. Countries that have already completed this transition are known for their advanced aging process and usually have a consistent civil registration system, extensive historical series, and fewer errors in age reporting. However, problems persist in reporting advanced ages in these countries. The objective was to propose a methodological approach for modeling the advanced age structure of the population in countries undergoing advanced demographic transition. Population data for ages 65 to 110 by sex were extracted from the Human Mortality Database (HMD), an international reference database known for the standardization and consistency of its historical series. Twenty-four countries were selected, most of which had data from the latest available census. The age structure of the population wa s modeled by applying the Gamma, Gompertz, logarithmic, and mean proportion distributions to the data for each sex. Diagnostic analyses were performed, and the Mean Square Error (MSE) was used for residual analysis. The results indicated good performance of the Gompertz model for ages 65 to 84 and the Gamma model for ages 85 and above. For confirmatory purposes, Brass' logit system, which measures the level and pattern of adjustment, was applied to reshape the observed data for countries with values adjusted by the best-performing models (Gompertz and Gamma). The countries were stratified based on EQM quartiles composed of six countries each, whose adherence to the model was classified as follows: highest (below Quartile 2),

intermediate (from Quartile 2 to 3), and lowest (above Quartile 3). It was concluded that the models performed well in adjusting for the extreme ages of the census populations for both sexes in the selected countries, which are in consolidated stage of demograph ic transition. The methodological strategy ! proposed in this study has the potential to satisfactorily model population structures in countries with deficiencies in the reporting of advanced ages in censuses and at less advanced stages of demographic transition.

A New Technique of "-Heaping" vis-à-vis Turner in Age Data

Barun Kumar Mukhopadhyay

Abstract. The fundamental objective of my paper is to provide a new index of age heaping that builds on the foundation work of Turner (1958) , S.H published in the Proceedings of the Social Statistics Section of the American Statistical Association in 1958. The method is very simple on digit preference error on age reporting error, an approach has been made using perfect ranking of ten digits from life table stationary population of any country and compared with the actual ranking obtained for either developed or developing country using Spearman rank-difference correlation coefficient. The present paper is an important exercise in the academic field where scientists researchers, students (young demographers) regularly corroborate with each other. The data used is single year age distribution quite available for almost all the countries of the world in the network system. However, in the present attempt data on a few countries of both the developed and developing regions are used. As there are almost one to one matching between the ranking of actual data of countries with their life table counterpart, a new nomenclature of "digit non-heaping" has been proposed in the paper. And it is well known that in software package on digit preference error, the calculation on King (1915)", Myers' Blended Method (1940)" though is given for its wider applicability but in addition there are other methods given for instance Bachi (1952), Ramachandran (1965) which are not used in practical purpose. Hence in the literature of a subject on quality of age reporting every attempt (method) should be included at least for the coming generations to work on subject of their choice.

Keywords: Turner (1958), age data, Bach (1952), Myers (1940)

A Statistical Framework for Analyzing Electricity Consumption in Construction Projects

Jana Kalická[1], Dominika Sónak Ballová[2], and Štefan Krištofič[3]

[1] *Institute of Computer Science and Mathematics, Faculty of Electrical Engineering and Information Technology, Slovak University of Technology Bratislava, Ilkovicova 3, 841 04 Bratislava, Slovakia*

[2] *Department of Mathematics and Descriptive Geometry, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05 Bratislava, Slovakia*

[3] *Department of Building Technology, Faculty of Civil Engineering, Slovak University of Technology in Bratislava, Radlinského 11, 810 05, Bratislava, Slovakia*

Abstract. This paper presents an advanced statistical analysis of electrical energy consumption at construction sites, with an emphasis on methodological contribution and model-based interpretation of consumption patterns. The study incorporates multiple explanatory factors, including seasonal effects, workforce size, and the functional categorization of electricity use into on-site construction activities and temporary residential facilities for construction personnel. A comprehensive time series framework is employed to decompose electricity consumption into trend, seasonal, and irregular components, enabling a detailed examination of temporal dynamics. Several forecasting and explanatory models, including regression models with seasonal dummy variables and autoregressive components, are developed and systematically compared. Model performance is evaluated using statistical validation techniques such as goodness-of-fit measures and residual diagnostics. Furthermore, interaction effects between workforce size and seasonal conditions are examined to capture non-linear and context-dependent influences on energy demand. The proposed approach contributes a structured statistical methodology for analyzing construction-site energy consumption, offering improved interpretability and predictive accuracy. The results provide a robust basis for energy demand forecasting and support data-driven decision-making aimed at enhancing energy efficiency and sustainability in construction project management.

Keywords: Electrical energy consumption, Time series analysis, Regression modeling, Seasonal effects.

A unified model for SPC and maintenance in high investigation cost processes with multiple quality disruptions

Konstantina Tsiota and Konstantinos A. Tasias

Department of Mechanical Engineering, University of Western Macedonia, 50150 Kozani, Greece.

Abstract. In modern manufacturing systems, integrating process quality control and equipment maintenance is crucial for maintaining operational efficiency and reliability. Traditional Statistical Process Control (SPC) methods often initiate investigations in response to alarms, which can be impractical in environments with high investigation costs or frequent disruptions. This study introduces a unified framework that integrates SPC with preventive maintenance to address these challenges. Designed for manufacturing environments characterized by high investigation costs and multiple quality shifts, this framework leverages quality monitoring data to guide both quality policy and maintenance decisions. A probabilistic decision-making model dynamically balances quality control and maintenance efforts, optimizing cost efficiency and reducing downtime. A detailed numerical analysis demonstrates the model's effectiveness in improving cost efficiency, reducing downtimes, and optimizing decision-making in complex industrial settings.

Keywords: Statistical Process Control (SPC), Multiple assignable causes, Markov chain, Preventive Maintenance (PM), Age replacement

ALMAB-DC: A Unified Stochastic Framework Integrating Active Learning, Multi-Armed Bandits, and Distributed Computing

Foo, Hui-Men and Yuan-chin Ivan Chang

Institute of Statistical Science

Academia Sinica

Taipei, Taiwan

Abstract. We propose ALMAB-DC, a unified stochastic framework integrating active learning, multi-armed bandits, and distributed computing for scalable statistical learning and optimization. The framework formulates sequential data acquisition and model updating as an adaptive stochastic decision process, where sampling and

computational resources are allocated according to uncertainty-guided criteria. Multi-armed bandit mechanisms provide principled exploration–exploitation trade-offs, while active learning enhances statistical efficiency through informative sample selection. The distributed architecture enables coordinated learning across multiple computational nodes, allowing efficient scaling to large datasets. We investigate the stochastic and computational properties of the proposed framework, including convergence behavior and adaptive resource allocation efficiency. The proposed methodology provides a principled foundation for integrating distributed statistical inference and adaptive learning in modern large-scale data analysis environments.
Keywords: Multi-Armed Bandits; Active Learning; Distributed Computing; Stochastic Optimization; Sequential Decision-Making; Adaptive Sampling; Distributed Statistical Inference; Stochastic Modeling

Alternative Computations of Whole-Word Variability in Child Developmental Speech

Mojtaba Taghvafard, Elena Babatsouli

University of Louisiana at Lafayette

Abstract. A bulk of research in child phonological acquisition demands quantification (Chandlee, 2025; Sonderegger & Márton Sóskuthy, 2025). This quantitative analysis can be a simple count of different variables such as mean length of utterance (Brown, 1973), percentage of consonants correct (Shriberg & Kwiatkowski, 1982), whole-word measures (Ingram & Ingram, 2001; Ingram, 2002), mean length of utterance in words (Parker & Brorson, 2005), measure for cluster proximity (Babatsouli & Sotiropoulos, 2018), etc. Ingram (2002) introduced four measures of whole-word productions, including phonological mean length of utterance (pMLU), the proportion of whole-word proximity (PWP), the proportion of whole-word correctness (PWC), and the proportion of whole-word variation (PWV). Among these, PWV, which divides the number of distinct forms by the total number of productions, is under-represented in studies. The range of PWV scores is from 0 to 1. When the child produces one form across different productions of a targeted word, the value of PWV will differ compared to values of PWV involving variable productions that also have different numbers of repetitions. Ingram (2002) provided a solution to this problem by dividing zero by the

number of productions. Nevertheless, Ingram's PWV can be computed in other ways that can not only solve the problem of comparing different tasks but also include the frequency and dominance of productions in the mathematical computation of PWV. We have outlined the following three methods to compute PWV. The simplest way is PWV = (K-1)/(N-1), where k=the number of distinct forms and N=the number of productions. This simple formula can correct the "minimum zero problem." However, this still does not account for the frequency of produced variable forms. To this end, normalized Shannon Entropy can be used because the formula is sensitive to both the number of productions and the frequency of each production. It is mathematically more complex than the first formula. Furthermore, the normalized Simpson Diversity Index can be used to calculate PWV, which is similar to Shannon's Entropy but less sensitive to rarer forms. Gierut's data (Gierut, 2015) has been used to compute whole-word variability using the mentioned metrics. Our results show that Ingram's PWV is consistently higher than the simple PWV, meaning that Ingram's PWV gives larger variability scores, and that the simple PWV accounts for the "minimum zero problem". It is found that turning to Shannon's entropy or Simpson's method is more advantageous, because the frequency of productions and dominant productions are weighted. In conclusion, the findings here have applications in educational and clinical settings to quantitatively assess PWV in child phonological data, thus providing clinicians with useful information to guide interventions.

Keywords: quantitative measures, phonological analysis, proportion of whole-word variability, distribution of productions, frequency

Analytical Reformulation and Fixed-Point Yield Inversion for Fixed-Income Securities

Fausto Acernese, *Department of Physics "E.R. Caianiello," University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy*

Valeria D'Amato, *MEMOTEF Department, Faculty of Economics, Sapienza University of Rome, via Del Castro Laurenziano, 00161, Rome, Italy*

Rocco Romano, *Department of Physics "E.R. Caianiello," University of Salerno, Via Giovanni Paolo II, 132, 84084 Fisciano (SA), Italy*

Abstract.The pricing equation that defines the yield of a fixed-rate bond is well known to be nonlinear and to require numerical inversion, most commonly performed via

Newton–Raphson methods. In this paper, we show that the nonlinear price–yield relationship can be rewritten in a simplified and fully differentiable form that preserves the original pricing identity while isolating the key source of nonlinearity. This analytical reformulation leads to an exact fixed-point representation of the yield and gives rise to a derivative-free iterative scheme with a closed-form update.

The resulting mapping exhibits a simple analytical structure and satisfies contraction conditions over a broad and economically relevant range of bond prices. As a consequence, yield computation becomes globally convergent, numerically stable, and faster than standard root-finding methods. The number of iterations required can be characterised analytically and remains very small even for deep-discount, premium, or low-duration bonds. The analytical transparency of the reformulated equation enhances computational efficiency in large-scale fixed-income applications, where yields must be computed repeatedly across large cross-sections of instruments and scenarios.

Overall, the proposed approach transforms yield computation into a numerically efficient and analytically interpretable procedure, offering clear advantages for pricing engines, RFQ workflows, and asset–liability management systems in modern fixed-income markets. The framework can be extended to floating-rate bonds, and the paper concludes by deriving the analytical pricing expression and an exact fixed-point formulation for the corresponding discount margin.

Analyzing ordinal categorical data related to dance using multilevel modelling

Liberato Camilleri[1] and Milena Pellicano[1]

[1]*Department of Statistics and Operations Research, University of Malta, Malta*

Multilevel models, also known as hierarchical linear models, extend generalized linear regression models by relaxing the assumption of independence among the responses. Multilevel models are explicitly designed to analyse clustered data structures where observations at micro level are nested in macro level structures, which in turn may be nested in clusters of higher order. Moreover, multilevel models can accommodate both fixed and random variables, where the equations defining the hierarchical linear models contain an error term at each level of nesting. This makes it possible to study the variation at different levels of the hierarchy. When parameters are estimated, multilevel models takes into account the hierarchical nested structure of the data such that the intra-class correlation describe the proportion of the total variance due to within

cluster variance. This dissertation mainly focuses on 2-level and 3-level models together with the numerical approximation techniques used in order to obtain the parameter estimates. The ordinary Gaussian quadrature and the adaptive Gaussian quadrature are the estimation techniques used to approximate the marginal likelihood, which is then maximized by using the modified Newton-Raphson algorithm.

Mixed effects multilevel models have several research applications since hierarchical structured data arises in many research fields. In this dissertation, the dataset comes from the Czech television dance competition Stardance. The data follow a three-level structure: individual performances (level 1) are nested within gala nights (level 2), which are in turn nested within seasons (level 3). Each performance was judged by four judges, with total scores ranging from 0 to 40. The distribution of scores is left-skewed, making analysis challenging since most standard distributions are symmetric or right-skewed. To address this, the scores were grouped into five categories (≤ 20, 21–25, 26–30, 31–35, 36–40), yielding an ordinal categorical dependent variable.

Keywords: Multilevel models, Adaptive Gaussian quadrature, Intraclass correlation

Analyzing Time-Varying Ranking Data

Vladimír Holý

Department of Econometrics,

Prague University of Economics and Business,

Winston Churchill Square 1938/4, 130 67 Prague 3, Czech Republic

Abstract. We present a score-driven framework for modeling and analyzing time-varying ranking data. Rankings are modeled using the Plackett–Luce distribution with latent worth parameters that evolve dynamically according to an updating mechanism based on the score, i.e., the gradient of the log-likelihood. This approach allows the dynamics to be driven directly by the information contained in past rankings, enabling estimation using the maximum likelihood method. The framework accommodates partial rankings and time-varying covariates. The methodology is implemented in the open-source R package gasmodel, which provides tools for estimation, forecasting, and simulation of score-driven ranking models. We illustrate the practical relevance of the framework through applications in sports analytics and operations research. Overall, the proposed score-driven ranking model provides a unified and versatile tool for the statistical analysis of evolving rankings across a wide range of empirical applications.

Keywords: Time-Varying Rankings, Plackett-Luce Distribution, Score-Driven Model.

Asymptotic Relation between the Bayesian Predictive Distributions for Random Sampling

Nasrollah Saebi

Kingston University,

Faculty of Engineering, Computing and the Environment,

School of Computer Science and Mathematics,

Penrhyn Road, London KT1 2EE - UK.

Abstract: We are concerned with problems that arise from the observation of sequential occurrences of a random point process over a known time period (0, T), studied in a Bayesian context. We initially proceed to derive the predictive distribution and its moments with regard to the final number of occurrences at time point T, conditioned on the number of observations sampled at an intermediate time point, and studied for different point processes. Since we shall be concerned with inferences in a Bayesian framework, for each point process, with reference to the concept of conjugacy, a suitable choice of a prior distribution is introduced. The choice of prior will be of considerable importance as it ensures mathematical tractability and provides an informative interpretation of the posterior distribution.

In this paper, we first assume a stochastic Poisson sampling with a gamma prior distribution for the rate involved, and derive and discuss the predictive distribution. Then, a binomial process is defined, and by using a beta prior density for the success rate, the predictive distribution for the future number of occurrences is identified. The asymptotic relationship between the predictive distributions arising from these sampling methods, conditioned on the assumed prior knowledge about the involved rates, is determined and discussed.

Bayesian Linear Regression for Modeling and Predicting Short-Distance Swimming Performance

Konstantinos Koutras[3], Sonia Malefaki[2], and John Garofalakis[3]

[1] *Hellenic Center of Marine Research, Anavyssos, Greece*

[2] *Department of Mechanical Engineering and Aeronautics, University of Patras, Greece*

[3] *Department of Computer Engineering and Informatics, University of Patras, Greece*

Abstract. Modern competitive swimming in short-distance events is highly demanding, requiring coaches to engage with increasingly complex datasets to evaluate athletes' performance and identify key determinants. In recent years, Bayesian statistical methods have gained prominence in addressing such challenges, particularly in the presence of historical data, owing to their flexibility in modeling uncertainty and generating probabilistic predictions. The present study investigates the application of Bayesian linear regression to performance data from 50m and 100m freestyle events. A bootstrapping-based approach is applied for the estimation of prior distributions. Variable selection is performed using projection predictive methods, with model comparison based on the Expected Log Predictive Density (ELPD). Shannon entropy is employed to quantify posterior uncertainty and distributional dispersion. Posterior inference is conducted via Markov Chain Monte Carlo (MCMC) algorithms, enabling comprehensive quantification of uncertainty in both model estimates and predictions. The proposed bootstrapping framework provides more efficient posterior mean estimation, requiring fewer iterations and yielding narrower credible intervals. Finally, the application of the models to out-of-sample data provides further insights into the variables that have the greatest impact on short-distance swimming performance, while facilitating robust predictive inference.

Keywords: Bayesian Linear Regression, Bootstrapping, Shannon Entropy, Variable Selection, Swimming Athletes, Swimming Performance, Freestyle Stroke.

Big Data Analytics and Human Capital as Joint Information-Processing Capabilities: A Theory-Building Framework for Economic Performance under Uncertainty

Ourania Kitsou,

Department of Management Science and Technology, Hellenic Mediterranean University, Crete, Greece.

Abstract: Organizations are increasingly navigating environments characterized by considerable uncertainty, complexity, and volatility, where economic success depends not only on information access but also on the organization's ability to process, interpret, and act on it effectively. Big Data analytics capabilities greatly improve the

amount, speed, and accuracy of information available, while human capital plays a vital role in interpreting and integrating this data into strategic and operational insights. The proposed framework, based on Organizational Information Processing Theory (OIPT), views these two elements as complementary, creating a combined information-processing capability. Strengthening this capability enhances decision quality in uncertain conditions, resulting in superior short-term economic performance (e.g., efficiency, revenue optimization) and long-term outcomes (e.g., learning, adaptability, resilience). Although exemplified in the tourism sector due to high demand volatility, the framework is applicable to a broad range of service and knowledge- intensive industries. The paper further examines how this theory can guide stochastic and dynamic modeling, with Big Data and human capital acting as complementary factors in decision-making under uncertainty.

Keywords: Big Data Analytics; Human Capital; Organizational Information Processing Theory; Economic Performance; Uncertainty; Decision-Making; Tourism Sector

Big Data Analytics for Digital Accounting: Bankruptcy Prediction Models for Financial Risk Management

Georgios Kampiotis, Georgios L. Thanasas, Georgia Kontogeorga

Abstract: Digital transformation has significantly influenced the accounting profession by introducing advanced technologies such as artificial intelligence (AI) and machine learning (ML) into financial analysis and risk management. This study investigates the application of bankruptcy prediction models in digital accounting environments, focusing on their ability to support corporate risk evaluation and strategic financial planning. Drawing on the US Company Bankruptcy Prediction Dataset, the research assesses the performance of traditional statistical methods in comparison with modern machine learning techniques, including Random Forest, Support Vector Machines, Gradient Boosting, and CatBoost. The results suggest that AI-based approaches provide more accurate predictions than conventional models, thereby improving the early detection of companies facing financial distress. The study also highlights the relevance of these predictive tools for auditors, financial analysts, and corporate executives, as

they offer data-driven insights that can enhance decision-making, strengthen risk assessment processes, and support more proactive financial management.

Keywords: Machine Learning, Predictive Analytics, Data-Driven Decision-Making, Strategic Planning.

Can Greece increase productivity while its population declines?

Ioanna Atsalaki, George S. Atsalakis

Technical University of Crete

Abstract. The transition from the macroeconomic cycle of 1968–2024 to the new megacycle of 2024–2080 is accompanied by profound economic, technological, and demographic changes that are reshaping both the global and national development landscape. Greece, having emerged from a prolonged financial crisis, is now called upon to redefine its growth model within an uncertain, multipolar polycentric, and continuously transforming environment. The article argues that the long-term trajectory of an economy is not predetermined but is shaped by collective will, institutional maturity, and a society's ability to understand its strengths and weaknesses.

The analysis focuses on Greece's demographic challenges — population decline and aging — and how these factors negatively affect productivity, innovation, and labor market dynamics. At the same time, it highlights that technological advances, such as artificial intelligence and automation, can act as counterbalancing forces, provided they are accompanied by policies that strengthen education, research, and youth participation in economic activity.

Finally, the article proposes the formulation of a "Grand Strategy" for the country, based on four key pillars: increasing productivity, revitalizing the regions, enhancing fertility, and achieving institutional consensus. The success of this strategy depends on Greece's ability to unite social and political forces, harness domestic knowledge and talent, and seize the opportunities of the new megacycle — ensuring sustainable growth and social prosperity for the decades to come.

Cluster-Based Classification of Czech Districts by Reproductive Behaviour

Kristýna Dvořáková[1], Filip Hon[2]

[1] *Department of Demography, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic*

[2] *Department of Demography, Faculty of Informatics and Statistics, University of Economics, Prague, Czech Republic*

Abstract. This article focuses on the classification of Czech districts according to their reproductive behaviour using cluster analysis. The method consists of performing cluster analysis for the years 2015, 2019, and 2023 and examining shifts in cluster membership over time. Districts were grouped into seven clusters using Ward's method combined with principal component analysis for data transformation. Based on the monitored demographic characteristics, the Ústí nad Labem Region and the Zlín Region emerged as the most internally homogeneous. While the Ústí nad Labem Region is characterised by a low mean age of mothers at childbirth, a high proportion of children born outside marriage, elevated total abortion rates, a substantial share of induced abortions, and a higher proportion of third and higher-order births, the Zlín Region displays the opposite pattern.

Keywords: Reproductive behaviour, Cluster analysis, Principal component analysis, Regional demography.

Conflict-Related Attention and Market Dynamics: GARCH and Granger Evidence from Israel case.

Papanikolaou Nikolaos

Hellenic Mediterranean University, Department of Accounting and Finance

Papanikolaou Anastasios

University of the Aegean, Department of Financial and Management Engineering

Pantos Themistoclis

Lincoln University, Finance Management and Investments Department

Vasileiou Evangelos

Hellenic Mediterranean University, Department of Accounting and Finance

Abstract. This paper examines the impact of the 2023-2025 Israel-Hamas war on the Tel Aviv Stock Exchange TA-35 index using weekly data, emphasizing how different

conflict-related information channels are reflected in asset prices. Using Google Trends data, we construct four conflict related attention measures: global and within Israel English language searches for the phrase “War in Israel” and Hebrew language searches for the word “Hamas” (חמאס). These indicators are used to explain the fluctuations in the TA-35 index.

Our results indicate that Israeli market returns co-move strongly with war-related public attention in Hebrew both worldwide and domestic. In contrast, English searches exhibit a weaker relationship and are marginally significant (at the 10% level).

However, Granger causality test paints a different picture about predictability. Neither global nor domestic Hebrew searches Granger cause TA-35 returns. Although English search attention indicators contain short-horizon predictive information: global “War in Israel” Granger causes TA-35 returns at 10% level, while the domestic relevant searches Granger cause returns strongly, with no evidence of reverse causality.

These findings imply that Hebrew search attention primarily captures market relevant information that is incorporated quickly, whereas war intensity measures reflect predictability. This distinction sharpens the paper’s contribution by showing that the market response during conflict depends not only on the volume of attention, but on the linguistic and geographic origin.

Design and Evaluation of a Strategic Digital Decision-Support Application for Smart Hospitals

Georgia Thanasa, Ioanna Giannoukou, Christos Klavdianos

Abstract. The digital transformation of hospitals is increasingly driven by advanced technologies such as artificial intelligence, the Internet of Things, big data analytics and cloud computing. Despite their growing adoption, healthcare organizations often lack strategic tools that translate digital investments into managerial insight and informed decision-making. Existing digital health systems predominantly focus on clinical or operational functions, leaving a gap at the strategic management level. This study addresses this gap by designing and evaluating a web-based strategic digital decision-support application for smart hospitals. The proposed application integrates key dimensions of digital maturity, operational performance, sustainability and patient-centric outcomes into a unified decision-support environment. Through interactive dashboards and scenario-based simulation, hospital managers can explore the strategic

implications of alternative digital investment portfolios and assess their potential impact before implementation. Following a design and evaluation approach, the application is implemented as a fully functional web prototype and assessed through scenario-based validation and expert feedback. The study contributes to the literature on smart hospitals and digital health by moving beyond conceptual frameworks and empirical correlations, offering a practical and replicable digital artefact that bridges digital technologies and strategic healthcare management.

Keywords: Digital maturity, Management decision-making, Information systems design, Multi-criteria decision analysis, Healthcare strategy

Directed information in graphical models of causality

Wojciech Niemiro

Professor

University of Warsaw, Institute of Applied Mathematics and Mechanics, Banacha 2, 02-097 Warszawa, Poland, wniem@mimuw.edu.pl

Abstract. I will speak of using the tools of information theory to quantify the strength of causal relations. In my talk I focus on models based on possibly cyclic directed graphs, with nodes corresponding to time series, as in [3]. In this setup the so called 'transfer entropy', introduced by Thomas Schreiber in [5] serves as a natural measure of Granger causality. It quantifies to what extent the past of X helps predict the future of Y, given the past of Y.

Graphical models allow to rigorously define the notion of intervention, ie. an actual or imagined controlled experiment in which values of some variables are set by the experimenter. Measuring the effect of intervention is quite different from measuring the predictive power. Several approaches based on information theory have been proposed in the literature. I will compare different definitions of 'interventional directed information', in particular those due to Ay and Polani [1] and Simoes et al. [4]. I will point out some problems and rather controversial interpretations.

Keywords: Causality, Directed Graphs, Information, Cyclic, Entropy Transfer

Distance Transform and AIC-based Model Selection for Tree Detection

Smaragda Markaki[1] and Costas Panagiotakis[2]

1 Department of Management Science and Technology, Hellenic Mediterranean University, 72100, Agios Nikolaos, Crete, Greece

2 Department of Management Science and Technology, Hellenic Mediterranean University, 72100, Agios Nikolaos, Crete, Greece

This work introduces two innovative approaches, the Decremental Circle Fitting Algorithm (DCFA) and its evolution, the Distance Transform Circle Detection (DTCD) method, for tree detection and counting from UAV-derived imagery and 3D point clouds. The framework bypasses the need for labor-intensive training data by treating tree detection as a statistical model selection problem. Initially, canopy regions are enhanced using a vegetation index (RGBVI) and segmented into binary masks. The DCFA pipeline models crowns by refining circle hypotheses through a Gaussian Mixture Model Expectation Maximization (GMM-EM) algorithm. The DTCD pipeline advances this by utilizing the Euclidean Distance Transform (EDT) to identify candidate tree centers at local maxima of the distance map.

The core innovation lies in the application of the Akaike Information Criterion (AIC) to determine the optimal number of trees. By iteratively evaluating candidate circles, the algorithm selects the model that maximizes canopy coverage while minimizing redundancy, effectively preventing over-segmentation. For complex urban landscapes, the DTCD-PC extension incorporates 3D vertical filtering to isolate tree structures from spectrally similar artificial objects.

Experimental results on the Acacia-6 and the urban Agios Nikolaos-3 datasets demonstrate high accuracy, with F1-scores up to 0.915. Furthermore, the framework offers exceptional computational efficiency, performing significantly faster than state-of-the-art supervised and unsupervised techniques.

Keywords: Tree Detection, Akaike Information Criterion, Distance Transform, Circle Estimation, UAV, Point Clouds.

Energy security and public support for renewable energy: Evidence from the UK

Andreas Markoulakis, Eleanya Nduka

University of Warwick, Department of Economics, University of Warwick, Coventry CV4 7AL, United Kingdom.

Abstract: We investigate how different facets of energy security, e.g., energy vulnerability (domestic energy supply, import dependency, technology development on energy sources), energy affordability (higher prices) and energy reliability (power cuts frequency) impact the support for different sources of renewable energy —offshore and onshore wind power, biomass energy and solar power. Our results show that there is a common pattern for energy vulnerability since as concerns decline, the probability of support for each renewable source also declines, but the rate of decline is larger for biomass and onshore wind. Energy imports dependency and affordability reveal a distinction between the wind power sources and the other sources since both offshore and onshore wind power are affected less by energy imports concerns or affordability concerns. Energy reliability is the only facet that leads to a rise in the probability of support for offshore wind. The above results are critical for policy appraisal purposes to inform policymakers on the differences between energy security facets and renewable energy sources when designing future energy policies towards net zero strategies.

Keywords: Energy security; renewable energy; offshore wind; onshore wind; solar; biomass.

Entropy Decomposition in Mortality Modeling

Apostolos Bozikas

Department of Statistics & Insurance Science, University of Piraeus, Piraeus, Greece

Abstract. In this work, entropy measures are used to analyze uncertainty in survival outcomes within mortality models. Instead of looking only at overall entropy measures, we focus on how different age-specific components contribute to total survival uncertainty. We develop a component-wise entropy decomposition that separates the contribution of individual model components within a broad class of mortality models.

The proposed framework offers a useful tool for understanding and managing uncertainty in longevity risk for practical applications.

Keywords: Entropy, Life expectancy, Mortality Model

Evolution of Private Health Expenditure (Out - of - Pocket Money) in Greece & how it affects Inequality and Poverty

Constantinos Stathopoulos, George Matalliotakis

Abstract. The change over time of informal payments in the Greek health system shows particularly high percentages for health by households, while at the same time during the economic recession there was a large reduction in public spending on health. These factors contributed to the emergence of financial risk and financial burden on households, with an increase in the percentage of poor households and their further impoverishment. The consequences of the deterioration of various socio-economic indicators and the increase of catastrophic health expenditures, show that the mechanisms of financial protection of citizens against emergency issues in Greece are incomplete.

The aim of this study was to use as recently as possible available data on household consumption expenditure, which came from the Family Budget Survey of ELSTAT for the years 208-2017, with theultimategoal of measuring the impact of catastrophic health expenditure in households as well as measuring the degree of further impoverishment in relation to a number of variables. The analysis included the application of the probability model (probit) in order to calculate the probability of catastrophic costs and poor households as well as the probability of their further impoverishment in relation to various demographic and socio-economic data. The statistical analysis was performed using the statistical packages SPSS and STATA.

According to the results of the statistical analysis, the advent of the financial crisis and the beginning of the memoranda resulted in the reduction of public health expenditures, resulting in increased financial risk and financial burden on households, as well as their level of impoverishment.

Exploring Multidimensional Digital Burden: A Data-Driven Analysis of Functional, Economic and Psychosocial Dimensions

Angeliki Kazani[1] and Apostolos Linardis[2]

[1]*Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

[2] *National Centre for Social Research, Greece, Athens, Greece*

Abstract. This study presents a data-driven methodological framework for the empirical analysis of functional, economic, and psychosocial dimensions of digital burden in interactions with digital services in Greece. Each dimension is operationalised separately using survey-based indicators, allowing for a focused examination of their individual structure and empirical behaviour. Psychosocial Likert items are reverse-coded to ensure conceptual consistency, and dimension-specific measures are constructed using mean-based aggregation. To support multivariate analysis, the three dimensions are subsequently standardised (z-scores) and examined using Principal Component Analysis (PCA) to explore their dimensional coherence and shared structure. To further investigate how different forms of digital burden manifest across the population, k-means clustering is then applied to the three standardised dimensions, enabling the identification of distinct user profiles characterised by differentiated burden patterns. In addition, binary logistic regression is employed to examine how key sociodemographic variables shape the likelihood of belonging to higher-burden profiles. The proposed approach adopts a statistically robust and flexible analytical workflow for examining multidimensional digital burden, identifying population subgroups, and supporting evidence-based research on digital inequalities. The analysis is based on a large-scale survey dataset (N=5,010) conducted by the National Centre for Social Research (EKKE) as part of the project "Towards a Just Green and Digital Transition" (JustReDI).

Keywords: Digital Burden, PCA, clustering, data-driven modelling, digital inequality, user profiling, sociodemographic factors

Extreme Decline in the Birth Rate in the Czech Republic

Kornélia Svačinová[1], Elena Říhová[2] and Petr Kasal[3]

[1] *Kornélia Svačinová, Department of Quantitative Methods, Škoda Auto University, Na Karmeli 1457, Mladá Boleslav, Czech Republic (E-mail: k.csefalvaiova@seznam.cz)*

[2] *Elena Říhová, Department of Quantitative Methods, Škoda Auto University, Na Karmeli 1457, Mladá Boleslav, Czech Republic (E-mail: elena.rihova@savs.cz)*

[3] *Petr Kasal, Department of Quantitative Methods, Škoda Auto University, Na Karmeli 1457, Mladá Boleslav, Czech Republic (E-mail: petr.kasal@savs.cz)*

Abstract. Research have shown that the Czech Republic is experiencing its historically lowest birth rate. Declining fertility and an ageing population have been a trend in many countries in recent years. Aim: Submitted paper focuses on the severe decline in birth rates in the Czech Republic. Data: Czech Statistical Office, Eurostat, OECD. Methods: Literature review was conducted. Analysis of birth rate indicators was calculated. Results: 84.3 thousand children were born in the Czech Republic in 2024 representing the third consecutive year of a substantial year-on-year decrease. A major change is not expected in the near future.

Keywords: Birth Rate, Fertility, Decline, Czech Republic.

From Quantification to Right Confirmation: Visual Modeling and Rights Protection Path of Intensive Motherhood

Nana Guo [1], Hong Mi [2]

[1] *School of Humanities and Law, Northeast Forestry University, Harbin 150040, P.R.China (Associate Professor)*

[2] *School of Economics and Management, Northeast Forestry University, Harbin 150040, P.R.China; Research Base of Population Big Data and Policy Simulation (Workshop), Zhejiang University, Hangzhou 310058, P.R.ChinaSchool of Humanities and Law, Northeast Forestry University, Harbin, China(Leading Talent and Chair Professor, Corresponding author, E-mail: spsswork@163.com)*

Abstract. The maternal practice of contemporary women is undergoing a profound cognitive transformation: on the one hand, women have achieved the return of subjectivity and the enhancement of self-awareness; on the other hand, the concept of

parenting has shifted from the traditional “continuing the family line” to the ethical responsibility of “nurturing life”. Driven by this dual transformation, “intensive motherhood” has become an independent choice for contemporary women, distinguishing them from previous generations. However, in sharp paradox with this, the phenomenon of “the value vacuum of maternal labor” still persists stubbornly. This gives rise to the core dilemma that this study aims to address—the “measurement paradox”, that is, the content and intensity of contemporary women's motherhood have increased significantly, yet the recognition of their social value and economic compensation remain in a state of vacuum. Instead of granting motherhood the higher evaluation and compensation it deserves, this reality imposes more obstacles to women's right to personal development, seriously inhibiting fertility willingness and affecting the sustainable development of the population. The systematic neglect of the value of motherhood stems from the characteristics of parenting labor, such as long-term continuity, high intensity, private scenarios, high emotional and psychological costs, and blurred boundaries with housework. Coupled with the profound impact of traditional gender division of labor concepts, its value is difficult to be effectively quantified and has never been integrated into the social evaluation and compensation system. Although the development of gender equality theory has brought widespread attention to the predicament of motherhood, it has not yet been translated into practical compensation and rights protection.

Keywords: Intensive Motherhood; Visualization Model; Maternal Labor Value; Rights Protection; Gender Equality

Gender differences in digital competence: A probabilistic analysis approach

Angeliki Kazani[1], Glykeria Stamatopoulou[2], Nafsika Moschovakou[1] and Maria Symeonaki[1]

[1]*Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

[2] *National Centre for Social Research, Greece, Athens, Greece*

Abstract. Gnder differences in digital competence are often examined by analysing the average scores between male and female students. While informative, this comparison

provides limited insight into the uncertainty underlying the gender gap. This study adopts a probabilistic modelling approach to explore gender differences in computer and information literacy by analysing the distribution of the gender gap, rather than a single point estimate. The analysis is based on student-level data for Greece drawn from the International Computer and Information Literacy Study (ICILS), a large-scale assessment that examines students' ability to use computers and digital information for inquiry, creation and communication in everyday contexts. A composite indicator of digital competence is constructed by combining multiple plausible values into a single continuous measure. Separate performance distributions are estimated for male and female students. Monte Carlo simulation is then used to generate repeated random draws for every group based on these distributions. In each simulation, the gender gap is calculated as the difference between male and female scores, producing a range of possible gender differences rather than a single fixed value. The resulting simulated distribution allows examination of both the typical magnitude of the gender gap and its variability across outcomes. The results indicate a small average advantage for female students. At the same time, there is a significant overlap between the performance distributions of male and female students in a non-negligible percentage of simulations. In conclusion, the findings suggest that gender differences in digital competence are not uniform but vary across individuals and contexts. Treating the gender gap as a variable rather than a fixed quantity highlights the role of uncertainty in educational achievement.

Keywords: probabilistic modelling, gender differences, digital competence, Monte Carlo simulation, large-scale survey, ICILS

Gender reassignment in translation: French masculine nouns to Greek neuter and feminine

Maria-Sofia Sotiropoulou

Linguistics & Cognitive Science Department and Romance Languages & Literatures Department

Pomona College, USA

Abstract. One of the difficulties in learning/acquiring a language, native or other, which has grammatical gender is to identify gender correctly. Studying this area contributes

to the understanding of language processing in the monolingual/bilingual mind and offers guidelines to teaching/learning methods. The present paper visits grammatical gender across two languages, French and Greek, the former having two grammatical genders, feminine and masculine, and the latter having three genders, feminine, masculine, and neuter. When translating nouns from French to Greek, gender in Greek either remains the same as in French or it is reassigned to other genders. Here, the full data of French masculine nouns as given in the French-Greek Dictionary of Lust & Pandelodimos (2022) is utilized to compute reassignment to Greek neuter and feminine nouns. First, it is found that the most frequent French masculine nouns have word-initial c, followed by p, a, s, while the least frequent nouns have word-initial x, followed by y, w, u, and that the frequency of masculine nouns is proportional to the frequency rank in the alphabet according to the Zipf-Mandelbrot law, 1/(rank+2.7), with Pearson correlation coefficient, r=0.9608. Next, the distribution of the frequency of reassigned nouns in the alphabet is significantly similar to the distribution of the source in the alphabet as attested by the chi-squared test (p=0.768), that is, the more nouns in the alphabet, the more get similarly reassigned in the alphabet. Total reassignment is interestingly at 70%, of which 61% become neuter and 39% feminine. The neuter dominance over feminine in reassignment is attributed to the dominance of concrete nouns over abstract in French masculine, 70% concrete (of which 68% are reassigned) vs 30% abstract (of which 74% are reassigned), and to the dominance of reassigned concrete to neuter over feminine, 73% vs 27%. French abstract masculine nouns are reassigned primarily to Greek feminine at 65%. In conclusion, the grammatical gender reassignment of French masculine nouns to Greek is major and, therefore, special attention needs to be paid to French-Greek child and adult bilingual studies as well as to teaching and learning methods.

Keywords: grammatical gender, reassignment, bilingual, French/Greek, computation

Graphical predictive model of multimorbidity in allergic diseases

Konrad Furmańczyk[1,4], Wojciech Niemiro[2,3], Marta Zalewska[4]

[1] *Institute of Information Technology, Warsaw University of Life Sciences, Warsaw, Poland*

[2] *Institute of Applied Mathematics University of Warsaw, Poland*

[3] *Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland*

[4] *Department of Prevention of Environmental Hazards, Allergology and Immunology, Medical University of Warsaw, Poland*

Abstract. Multimorbidity is common in allergic diseases. We construct a graphical model based on data gathered in the big epidemiological survey ECAP [2]. The goal is to better understand causal relations between several allergic diseases and also to provide medical doctors with a tool to assist in making correct diagnosis. The variables included in the model are partitioned into three groups: the diseases Y, the symptoms S and the covariates X, all considered as nodes of a directed acyclic graph (DAG), similarly as in [1]. The arrows of the graph lead from X to Y to S and also among some variables Y. We used expert medical knowledge and suggestions produced by AI to choose the structure of the DAG within Y, which reflects comorbidities. For every node Y and S, we fitted a regression model (logistic or linear) in the training phase. In the predictive regime, we input values of variables S and X for new patients and compute the MAP (maximum a posteriori) estimate for the presence of diseases (configuration of Y variables). We conducted experiments in which we partitioned the ECAP data into training and testing sets. The results are promising and confirm the model usefulness in predicting the structure of co-occurring allergic diseases.

Improved Insurer's Capital Efficiency of non-life underwriting risk using copula approach and hypothesis tests

Ilze Zarina-Cirule, Irina Voronova Gaida Pettere and Ilze Zarina-Cirule
Riga Latvia

Abstract. Maintaining adequate capital buffers and the ability to absorb losses during economic downturns are central to financial stability management and the protection of shareholders' interests. In non-life insurance, loss distributions are typically heavy-tailed and exhibit nonlinear dependence structures, which complicates risk aggregation. Although copula models provide a flexible framework for capturing such dependencies, their application in insurance practice remains limited. This study investigates the selection of appropriate copula models for aggregating premium and reserve risk in non-life insurance. We estimate regulatory capital requirements using the Value-at-

Risk (VaR) measure at the 99.5% confidence level, as mandated under the European Union solvency framework. Several copula families are evaluated and compared using statistical goodness-of-fit and hypothesis testing procedures to identify the most suitable dependence structure for capital aggregation. The proposed methodology is illustrated through a case study based on Baltic non-life insurer data, highlighting its practical relevance for internal capital modeling and risk management.

Keywords: Value-at-Risk; copula models; non-life insurance; reserve risk; premium risk; internal capital modeling; dependence modeling; hypothesis testing; financial stability management

Inference for Poisson Change-point Analysis

James Freeman

Alliance Manchester Business School, University of Manchester,

Abstract: Over the years, a diverse repertoire of procedures has evolved for analysing Poisson change-point data. Adding to this collection, a new approach is presented here based on a Poisson regression formulation.

Tested across a range of contrasting datasets the procedure is shown to perform comparably to mainstream alternatives but with the advantage of being able to distinguish qualitatively between abrupt and progressive types of change.

Keywords: Change-point, Deviance, Goodness of fit, Poisson regression

Influence of Chemical Fertilizer Application Technical Efficiency on Land Output Efficiency

Xiuyi Shi and Hong Mi

Northeast Forestry University, Harbin, China

Abstract The agricultural resources are being increasingly overdrawn for the agricultural development of China, and the excessive application of chemical fertilizer has affected the land output efficiency. In this paper, with the selection of appropriate input-output indexes, the technical efficiency of chemical fertilizer application and the efficiency of land output in various provinces of China were calculated by using the stochastic frontier analysis model (SFA). Then the two are set as explanatory variables

and explained variables respectively. With some related explanatory variable, the influence of chemical fertilizer application technical efficiency on land output efficiency is analyzed by using Tobit model. The results showed that the average land output efficiency of agricultural operators in China from 2018 to 2020 was 0.7731, and the technical efficiency of chemical fertilizer application was 0.3499, both of which ha ve the potential to be increased. During the same period, the efficiency of land output increased slowly, while the technical efficiency of chemical fertilizer application decreased slowly. The technical efficiency of chemical fertilizer application, productive labor force input, planting area of crops, and the amount of agricultural subsidy had obvious positive effects on the efficiency of land output. However, non-agricultural employment rate and chemical fertilizer market price have obvious negative effect on land output efficiency. Based on the above conclusions, this paper put forward some suggestions as establishing a perfect technical guidance system of chemical fertilizer application, constantly strengthening the subsidy policy of new chemical fertilizer application, strengthening the training of specialists in the field of chemical fertilizer application, and perfecting the mechanism of agricultural administration, law enforcement, supervision, and inspection.

Integrating Multi-Criteria Decision Analysis into FX Forecasting: A Conceptual Framework

Fotis Papatheofanous and Christos Floros

Department of Accounting and Finance, Hellenic Mediterranean University,

Abstract: This paper extends prior research on the persistent *exchange rate disconnect puzzle* by developing a conceptual framework that integrates Multi-Criteria Decision Analysis (MCDA) into foreign exchange forecasting (Herrera et al., 2025). Traditional macroeconomic models frequently display weak out-of-sample predictive performance and limited capacity to synthesize heterogeneous macro-financial information, constraining both accuracy and interpretability (Pfahler, 2021). The proposed framework bridges quantamental macro analysis with decision theory, addressing structural instability, parameter uncertainty, and regime dependence that characterize exchange-rate dynamics (Φεppαρα et al., 2023).

MCDA methodologies enable the systematic combination of macroeconomic fundamentals, financial variables, and qualitative signals through structured weighting, ranking, and scenario evaluation under uncertainty (Peng et al., 2023). This approach enhances the interpretation of forecast errors by linking them to key macroeconomic drivers and improves practical relevance for investors and policymakers engaged in currency-risk management and hedging strategy design (Abouzaid & Boussedra, 2025; Neghab et al., 2024). Furthermore, the capacity of MCDA to incorporate both quantitative and qualitative information supports applications in portfolio optimization, derivative pricing, and risk management within volatile FX environments (Alaminos et al., 2023; Rubaszek et al., 2022). Overall, the framework represents a methodological advancement toward robust, interpretable, and decision-oriented exchange-rate forecasting adaptable to evolving economic conditions (Peng et al., 2024).

Keywords: FX forecasting; Multicriteria decision analysis; Quantamental macro; Exchange- rate modeling; Decision theory

Investigating Active Ageing index: From macrolevel to individual scores

D. Chaidemenaki G. Verropoulou

Department of Statistics and Insurance Science

University of Piraeus

Abstract. This study investigates the determinants of active ageing, good mental health and productivity among older adults in Europe using SHARE longitudinal microdata. Building on – and contrasting with – the macro-level Active Ageing Index (AAI), I construct a new individual-level active ageing index that integrates domains of employment, social participation, health, independent living and mental well-being. The index is derived from the panel data of wave 9, using carefully justified decisions on item selection, scaling, missing-data treatment and aggregation. Particular emphasis is placed on the methodological choices behind composite indicator construction: the distinction between inputs and outputs, explicit versus implicit weights, and the trade-off between normative weighting schemes and data-driven approaches such as PCA. The findings are expected to show that macro-level indices conceal substantial within-country heterogeneity and policy-relevant gaps by gender, education and income, and

that the individual-level index provides a more defensible basis for studying inequality and policy effects.

Keywords: active ageing, composite index, SHARE dataset, weighted indicators

Investigating the performance of an overall index of political trust to representative institutions cross-nationally

Aggeliki Yfanti[1], Anastasia Charalampi [2] and Catherine Michalopoulou[3]

[1] *Research Centre for Greek Society, Academy of Athens, Athens, Greece*

[2] *Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

[3] *Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

Abstract. The purpose of this paper is to explore the performance of an Overall Index of Political Trust (OIPT), constructed by averaging political trust to the national government, politicians and political parties. The index was developed within the framework of the research program "TRUEDEM: Trust in European Democracies 2023-2025". The analysis was based on the 2022 European Social Survey datasets of France, the Netherlands, Portugal and Slovenia. Conceptualizing political trust as a unidimensional construct, the common structures were validated, and Confirmatory factor analyses were conducted on the full samples, suggesting perfect model fit for each country. Their psychometric properties were assessed, providing reliable and valid constructs for each case. Future research could extend this analysis by examining the performance of the OIPT in additional countries from the same ESS Round and overtime.

Keywords: Overall index of political trust, Confirmatory factor analysis, Reliability, Validity, European Social Survey.

Large-Scale Customer Conversion Prediction in Digital Marketing Using PySpark and Distributed Machine Learning

Leonidas Theodorakopoulos, Christos Klavdianos, Alexandra Theodoropoulou.

Abstract: This paper investigates large-scale customer conversion prediction in digital marketing using PySpark and distributed machine learning, with the goal of supporting more accurate targeting and campaign decision-making under high-volume behavioral data. A large dataset of customer interactions and conversion outcomes is processed in a scalable pipeline that includes data cleaning, missing-value handling, and feature normalization. Multiple neural network architectures are implemented and compared alongside established machine learning baselines within a unified evaluation framework, enabling a practical assessment of predictive performance under distributed training and inference. Model performance is evaluated using accuracy, precision, recall, F1-score, and threshold-sensitive analysis through ROC curves and confusion matrices. To improve interpretability and make the outputs actionable, the study also conducts feature importance analysis to identify the most influential behavioral and campaign-related predictors of conversion, linking model insights to controllable marketing levers.

Keywords: Behavioral data, Marketing Decision-making, Neural networks, Costumers insights.

Macroprudential Policy and Downside Risk: Regime-Dependent Effects of Capital Regulation∗

Vivien Czofa

Central Bank of Hungary,

Hungary Tibor Szendrei

National Institute of Economic and Social Research, UK.

Katalin Varga†

Central Bank of Hungary, Hungary.

Abstract . This paper employs a Threshold Bayesian Vector Autoregression (TBVAR) to estimate the regime-dependent macroeconomic effects of capital regulation in Hungary. Using the Factor-based Index of Systemic Stress (FISS) as the threshold variable, the model identifies normal and stress regimes consistent with the

occasionally binding constraints literature. The TBVAR offers a practical multivariate alternative to Growth-at-Risk for data-constrained economies. Generalised impulse responses reveal a pronounced asym metry: releasing regulatory capital by one percentage point during stress raises GDP growth by approximately one percentage point at the peak, with effects persisting for roughly twenty months, while the cost of accumulating capital in the normal regime is economically negligible. These findings are robust to alternative Cholesky orderings, sam ple periods, and credit variable definitions, providing direct empirical support for the countercyclical operation of the capital buffer. Keywords: Macroprudential Policy, Downside Risk, Capital Regulation, Threshold VAR.

Mean Value Theorems for Random Variableswith applications to Probability Models

Georgios Psarrakos[1]

[1] Department of Statistics & Insurance Science, University of Piraeus, Piraeus, Greece

Abstract. The mean value theorem (MVT) is very popular in calculus, and states that if g is a function continuous in [x, y] and differentiable in (x, y), then there exists a point u in the interval (x, y) such that

g(y) - g(x) = g'(u) \, (y - x).

A probabilistic analogue of the mean value theorem (PMVT) for nonnegative continuous/discrete random variables was introduced and studied by Di Crescenzo (1999), providing also applications to various contexts of reliability theory. In this presentation, we provide further extensions of MVT, giving also some applications to probability models and actuarial science.

Keywords: Mean value theorem, Stochastic order, Variability order, Mixture of distributions, Deductible value, Delta method.

Measuring Digital Poverty in Greece: A Multidimensional Approach with data drawn from a large-scale survey

Glykeria Stamatopoulou[1], Angeliki Kazani[2] and Maria Symeonaki[2]

[1] *National Centre for Social Research, Greece, Athens, Greece*

[2] *Department of Social Policy, Panteion University of Social and Political Sciences, Athens, Greece*

Digital poverty is increasingly understood as a form of social exclusion, rather than merely a lack of access to digital technologies. The term has been previously linked to a lack of access to digital resources (first-level digital divide), the lack of digital skills (second-level digital divide), or the effective use of ICT to functionally utilise digital technologies in a constructive way (third-level digital divide), and it may characterise individuals regardless of their economic poverty status. Within this framework, this study investigates digital poverty by adopting a multidimensional concept that focuses on functional rather than purely infrastructural aspects. The analysis draws on data from a large-scale survey conducted in Greece in 2025, under the flagship initiative JUSTReDI «Resilience, Inclusion and Development Towards a Just Green and Digital Transition of Greek Regions». A digital poverty index is constructed based on four dimensions, i.e. recent use of the internet, self-reported digital skills, the ability to complete online public service procedures, and experiences of exclusion due to mandatory digitalisation. Descriptive statistics and regression analysis are employed to examine both the trends and the social distribution of digital poverty in Greece. The findings reveal variations across specific socio-demographic groups and indicate that digital

is closely associated with attitudinal factors related to digital engagement and perceived barriers to digital services.

Keywords: digital poverty, exclusion, digitalisation, Greece, Large-scale survey

Measuring the literacy in sustainable finance of European students by means of a computerized adaptive assessment tool

Jang Schiltz, Emilie Takla[1], Maurice Dumrose, Daniel Engler, Christian Klein[2], and Ramon Hörler, Othar Kordsachia, Sabrina Urban[3]

[1] *University of Luxembourg, Department of Finance, 6, rue Richard Coudenhove-Kalergi, L-1359 Luxembourg, Luxembourg*

[2] *University of Kassel, Chair of Sustainable Finance, Henschelstrasse 4, D-34127 Kassel, Germany*

[3] *University of Liechtenstein, Sustainable Finance and Investments, Fürst-Franz-Joseph-Strasse, FL-9490 Vaduz, Liechtenstein*

Abstract. We conducted a comprehensive cross-country survey of students at European universities. The survey measures multiple dimensions of sustainable finance literacy, including knowledge of environmental, social and governance (ESG) factors, sustainable investment instruments and strategies, relevant regulatory frameworks and awareness of concepts such as greenwashing and the ESG rating system. We also assess general financial literacy, factual knowledge about ESG issues, personal factors and preferences (including risk preferences, time preferences, and environmental attitudes), and actual investment behavior.

We then used the results of this survey to develop a finance literacy assessment tool. This tool is actually a computerized adapted test, based on item response theory.

More precisely, we used 27 items coming from the survey questions and the answers of 171 participants in the survey to calibrate the item parameters (discrimination, difficulty, and guessing) by means of python's IRT modeling tools using the classical 3PL model of item response theory. We then we implemented an interactive assessment tool that can be run directly in R. The tool presents items from the item bank, collects the answers from the console, selects subsequent items by using the Maximum Fisher Information criterion, and re-estimates the ability score θ after each item. Since there are currently only relatively few items in the item bank of the tool, each test is constituted by exactly 10 questions, but this can easily be changed in a future extension of this work and alternative stopping rules can be implemented if desired.

The standardized assessment tool emerging from this research provides an actionable resource for universities worldwide to integrate sustainable finance education across different disciplines and can easily by enlarged to a larger item bank.

Keywords: Sustainable finance literacy assessment tool, item response theory, computerized adaptive testing

Mining Skill Checklists and Constructing Career Sequences Based on Massive Dynamic Data

Yiming Zhang[1], Zhongshi Jiang[2], Zhong Wang[3]

[1] *Distinguished Professor, Shandong University, China*

[2] *Ph.D. Candidate, School of Public Affairs, Zhejiang University, China*

[3] *Ph.D.·Wurzweiler School of Social Work, Yeshiva University, US*

Abstract. President Xi Jinping, in his speech at the 14th collective study session of the Political Bureau of the CPC Central Committee on May 27, 2024, proposed to "actively explore and cultivate new career sequences." On September 15, 2024, the *Opinions of the CPC Central Committee and the State Council on Implementing the Employment-First Strategy to Promote High-Quality and Full Employment* pointed out the need to establish and improve the national qualifications framework. The document calls for promoting the two-way mutual recognition of vocational qualifications/skill levels with corresponding professional titles and academic credentials, advancing the implementation of the "Academic Certificate + Several Vocational Skill Certificates" system, actively mining and cultivating new career sequences, and timely releasing new professions.

Furthermore, the Ministry of Human Resources and Social Security and the Ministry of Finance jointly issued the *Notice on Implementing the "Skills Illuminate the Future" Training Action*, requiring the promotion of a "four-in-one" project-based training model comprising "Job Demand + Skills Training + Skills Evaluation + Employment Services" to empower laborers for high-quality and full employment. Simultaneously, as various regions promote high-quality economic development and industrial upgrading, new industries, new business formats, and new professions are constantly emerging. There is a strong realistic demand to construct career sequences, training systems, and talent introduction mechanisms that are adaptable to industrial development. Meanwhile, the accumulation of massive dynamic data and the rapid iteration of knowledge graph and natural language processing (NLP) technologies provide strong technical support for the construction of occupational graphs.

Model of Lee-Carter type for fertility – analyzing baby bust in Poland

Wojciech Otto

Faculty of Economic Sciences

University of Warsaw

Abstract. Lee-Carter (or Renshaw-Haberman) model is just a special case of more general bilinear models that can be applied whenever we observe a population of objects that experience some event during her lives. In the most typical application to mortality tables the event is just death. Another case is birth.

Quite interesting conclusions could be pulled out by combining a collection of models:

- a model explaining probability of birth by a female of age x in year t
- a model explaining probability of first (second, …) birth by a female of age x in year t.

The research explores Polish data on number of females by age, and on first births, on second births,…etc. by age of a mother. Data cover the period 1971-2024. Getting a good picture of baby bust on the basis of these data raises several methodological issues. On the other hand it brings quite sound answers to important questions.

Most important methodological issues concern:

- extracting kohort effects (dealing with calendar effects is easy)
- coping with disturbances:
    - heteroscedasticity
    - autocorrelation
    - non-normal distribution (heavy tails)
- preventing too strong influence of marginal age groups on estimates

Most important empirical findings concern answering the following questions:

- do models restricted to calendar effects only (or to cohort effects only) explain observed changes sufficiently?
- does smaller number of children per female come from:
    - smaller number of children per mother (female having at least one child), or rather rising ratio of females having no children at all?

    - smaller number of children per mother having two or more children, or rather rising ratio of females having at most one child?
- How fast average age at birth (at first birth, at second birth) increases?
- To what extent co-occurrence of decline in fertility with rising age at birth clouds the picture of baby bust

The speech will cover a selection of the above issues.

Keywords:
Lee-Carter Model, Renshaw-Haberman model, calendar and kohort effects, fertility

Modeling the spread of several rumors using competing Markov chains

Jelena Jocković[1] and Bojana Todić[2]

[1] *Faculty of Mathematics, University of Belgrade, Serbia*

[2] *Faculty of Mathematics, University of Belgrade, Serbia*

Abstract. Several recent models for asynchronous rumor spread in discrete time are based on Markov chains , and provide exact and/or asymptotic properties of the expected waiting time for the rumor to spread through the whole population of the given size, or its subpopulation. We build on these results by introducing the possibility of several rumors spreading at the same time, as well as the possibility of failed transmission (the receiver who doesn't believe the rumor, or even reassures the spreader). The model is based on the results related to identically distributed Markov chain competing for transitions, obtained in .

Keywords: Asynchronous rumor spread, Competing Markov chains, Waiting time, Failed rumor transmission.

Multidimensional Policy Regimes across Countries - A Comparative Clustering Analysis

Samir K. Safi[1] and Afreen Arif[2]

[1] *Department of Statistics and Business Analytics, United Arab Emirates University, Al Ain, United Arab Emirates.*

[2] *Amity Business School, Amity University, Dubai, United Arab Emirates.*

Abstract. This paper uses k-means, hierarchical, and hybrid clustering methods on macroeconomic indicators to categorize the countries into five different groups according to the development patterns, financial structure, and financial institutions. The hybrid model is very stable and interpretable, showing changes in the global. To the public policy, this group of findings implies that structural similarities can be mischaracterized by these traditional income-based or loan-based categories. Evidence-based clustering will provide policymakers with a more detailed structure of identifying economic analogues, building specific development plans, and better resource distribution to make international economic and institutional policies more precise and profound.
Keywords: Cluster analysis, social-economic similarity, hybrid clustering, bootstrapped stability, adjusted Rand index, policy classification.

Nonparametric Estimation for Compound Renewal Processes in Weighted Cadlag Spaces

Anna Kouroukli1 and Konstadinos Politis

Department of Statistics and Insurance Science, University of Piraeus, 80 M. Karaoli & A. Dimitriou, 18534 Piraeus, Greece,

Abstract. Based on classical renewal–reward models and prior nonparametric work on compound distributions and decompounding (e.g., (1; 2; 3; 4)), we study the compound renewal process N(t) Z(t) = i=1 Xi, where N(t) is the renewal counting process driven by i.i.d. interarrival times Yi > 0 (a.s.) and i.i.d. jumps Xi. The process {Z(t),t ≥ 0} is typically used to represent the aggregate claim paid by an insurance company through time. For each fixed t ≥ 0, we define the distribution function Gt of Z(t) through the functional representation ∞ Φt(F,H)(x) = k=0 pk(t, H)F∗k(x), pk(t, H) = Pr{N(t) = k} = H∗k(t)−H∗(k+1)(t), where F and H denote the marginal distribution functions of the jumps and interarrival times, respectively. We propose a nonparamet ric plug-in estimator Gt,n = Φt(Fn, Hn), constructed from the empirical distribution functions of the observed jumps and interarrivals. To control the infinite series and obtain uni form bounds, we work in weighted c`adl`ag spaces Dββ and derive mo ment conditions ensuring Φt(F,H) ∈ Dββ, using Cram´er–Chernoff bounds for the renewal counts,

uniform domination bounds for the infinite series defining Φt, and dominated-convergence arguments for series in Banach spaces. By employing a refined Glivenko–Cantelli re sult and the Functional Delta Method, we establish strong consistency and asymptotic normality of Gt,n, and show that bootstrap techniques provide valid confidence bounds

Nonparametric Testing of Copula Equality via Young Diagram Shapes

Garcia, Jesus E

Gonzalez-Lopez, Veronica

UNICAMP, Statistics, Campinas, Brazil

Abstract. We propose a nonparametric two-sample procedure for testing whether two bivariate pop ulations share the same copula, and therefore the same dependence structure up to strictly increasing marginal transformations. Given a sample of size n from (X,Y), we form the permutation that maps the ranks of the X observations to the corresponding ranks of the Y observations. Applying the Robinson–Schensted correspondence to this permutation yields a Standard Young Tableau and its shape, that is, the partition λ = (λ1,λ2,...) of n describing the row lengths of the associated Young diagram. For continuous margins, the distribution of λ depends only on the copula, which turns copula equality into a two-sample problem on Young diagram shapes. We build test statistics from the empirical distribution of shape-based summaries, including weighted ℓ1 and quadratic discrepancies computed on a low-dimensional feature vector extracted from λ. Critical values are obtained by an exact permutation calibration: we pool the two samples and randomly permute the group labels to approximate the null distribution. Simulation experiments indicate good power across a broad range of alternatives, including settings with zero linear correlation, while maintaining invariance under monotone marginal transformations. We also illustrate the method on real data, where the resulting Young diagram profiles provide an interpretable signature of dependence.

Keywords: Copula equality, Young diagram, Permutation test

On the modelling of Mortality Improvement Rates

Peter Hatzopoulos

Department of Statistics and Insurance Science, University of Piraeus, 80 M. Karaoli & A. Dimitriou, 18534 Piraeus, Greece

Abstract. Recent actuarial research has increasingly shifted attention from modelling the logarithm of central mortality rates to modelling mortality improvement rates, defined as the logarithmic ratios of successive mortality rates, as the primary response variable. This alternative formulation offers important advantages. Inference based on large reference populations can be transferred readily to subpopulations with different baseline mortality levels, while mortality reduction factors provide a transparent and interpretable framework for expressing assumptions about future mortality dynamics.

In this work, we develop a comprehensive mortality modelling framework in which the predictor structure is specified in terms of mortality improvement rates rather than log-mortality levels. The proposed model follows a general age–period–cohort (APC) specification, allowing for age-specific average improvement rates as well as time-varying deviations and cohort-related effects. Parameter uncertainty is assessed using semiparametric bootstrap procedures under the assumption of normally distributed errors.

The framework is applied to national population data from the United Kingdom, Sweden, Greece, and the United States. Empirical results are used to compare the proposed approach with conventional log-mortality models in terms of goodness-of-fit and out-of-sample forecasting performance. The analysis highlights the practical advantages of modelling mortality improvement rates, both for statistical inference and for the communication of mortality trend assumptions in actuarial applications.

Keywords: *mortality improvement rates; reduction factors; Poisson log-bilinear models; parameter uncertainty; age–period–cohort models; bootstrapping; longevity risk.*

Optimal annuitization post-retirement with labor income

Gao X; Hyndman C; Pirvu TA; Jevtić P

McMaster University, 1280 Main St W, Hamilton ON L8S4K1, Canada

Abstract. Evidence shows that the labor participation rate of retirement age cohorts is non-negligible, and it is a widespread phenomenon globally. In the United States, the labor force participation rate for workers age 75 and older is projected to be over 10 percent by 2026 as reported by the Bureau of Labor Statistics. The prevalence of post-retirement work changes existing considerations of optimal annuitization, a research question further complicated by novel factors such as post-retirement labor rates, wage rates, and capacity or willingness to work. To our knowledge, this poses a practical and theoretical problem not previously investigated in actuarial literature. In this paper, we study the problem of post-retirement annuitization with extra labor income in the framework of stochastic control, optimal stopping, and expected utility maximization. The utility functions are of the Cobb-Douglas type. The martingale methodol ogy and duality techniques are employed to obtain closed-form solutions for the dual and primal problems. The effect of labor income is investigated by exploiting the explicit solutions and Monte-Carlo simulation. The latter reveals that the optimal annuitization time is strongly linear with respect to the initial wealth, with or without labor income. When it comes to optimal annuitization, we find that the wage and labor rates may play opposite roles. However, their impact is mediated by the leverage ratio.

Keywords: Optimal annuitization, retirement, labor income

Portfolio Optimization using Stochastic Controls and Technical Investment Strategies

Jean-Paul Murara[1]

1 Department of Business and Mathematics (IEM), Mälardalen University, Universitetsplan 1, Västerås, Box 883, 721 23 Västerås, Sweden

Abstract. In this article, we perform a comparative study of different investment strategies. We start by modifying the Merton investment-consumption problem and then suggest a new investment strategy based on our newly derived optimal controls. Here we investigate the impact of constant and non-constant volatility on the optimal wealth. We also optimize our portfolio considering investment strategies that use RSI,

TSI, and SMA as indicators. Numerical analyses are conducted to compare the performances of our developed strategies, with different realistic assumptions.

Keywords: Stochastic Optimal control, Portfolio Optimization, RSI, TSI, SMA.

Predicting Sustainable Tourism Choices Using AI-Driven Marketing Signals: A Machine Learning Approach

Ioanna Kalliampakou, Christos Klavdianos, Hera Antonopoulou

Abstract. This study explores how AI-driven marketing signals can be used to predict sustainable tourism choices through a machine learning approach. As sustainability becomes a central concern in tourism decision-making, understanding the digital signals that influence tourists' preferences is increasingly important for destinations, hospitality firms, and policymakers. The research focuses on marketing-related indicators such as personalized recommendations, social media engagement, online reviews, eco-label visibility, sentiment patterns, website interaction data, and targeted sustainability messages. By applying supervised machine learning techniques, the study aims to classify and predict tourists' likelihood of choosing sustainable tourism options, including eco-friendly accommodation, low-impact transport, community-based tourism, and environmentally responsible travel activities.

The proposed approach integrates behavioral, attitudinal, and digital marketing data to identify the most influential predictors of sustainable tourism choices. Machine learning models such as Random Forest, Gradient Boosting, Support Vector Machines, and Logistic Regression may be compared in terms of predictive accuracy, interpretability, and practical usefulness. The findings are expected to contribute to both academic knowledge and managerial practice by showing how artificial intelligence can improve sustainable tourism marketing strategies. Specifically, the study highlights the potential of AI-based predictive analytics to support more targeted, ethical, and effective communication with environmentally conscious tourists. At the same time, it emphasizes the importance of transparency, data privacy, and responsible use of AI in tourism marketing.

Overall, this research demonstrates that AI-driven marketing signals can provide valuable insights into sustainable tourism behavior, enabling tourism organizations to design data-informed strategies that encourage more responsible travel choices.

Keywords: Sustainable tourism; AI-driven marketing; machine learning; predictive analytics; tourist behavior; digital marketing signals; sustainable consumer choices; tourism decision-making; artificial intelligence; eco-friendly travel.

Predicting two-stage screening/rescreening and randomization pipeline in multicentre clinical trials accounting for future sites initiation

Volodymyr Anisimov

Center for Design & Analysis, Amgen, London, UK

Abstract. We consider a multicentre clinical trial at an interim stage where some sites are active and additional sites are scheduled to initiate in the future. Patients arriving at a site undergo an initial screening of fixed duration $\Delta 1$. After screening, a patient is randomized with probability p1. If the patient fails initial screening, they may enter a rescreening stage with probability r, lasting $\Delta 2$, and then be randomized with probability p2. Forecasting time to a randomization target should account for patients currently in screening/rescreening and predict how many patients out of those will be randomized, since stopping screening too late can cause over-enrollment.

We develop a novel analytic methodology to predict the randomization process along with screening/rescreening while explicitly incorporating future site initiations; related work has considered another logistic of screening/pre-screening without modelling future site openings.

Patient arrivals are modelled using a Poisson–gamma (PG) framework capturing between-site heterogeneity. PG parameters are estimated from interim data in active sites, and the rates for not-yet-open sites are evaluated using country-level aggregation of posterior rates in active sites. The randomization process in active sites is modelled as a PG process using some aggregated characteristics of screening-rescreening stages. The randomization process in the new sites is modelled as an evolving hierarchical process where each newly opened site generates a delayed arrival process, feeding into the screening-to-randomization pipeline. The global process is then formed as a sum of delayed site-level hierarchic processes propagated through the screening-rescreening pipeline. For closed-form prediction we approximate this process by a single PG process basing on the previous results on the approximation of the sums of PG

processes by a PG process with aggregated parameters. The methodology also supports prediction of a stopping time for new arrivals for screening to reduce over-enrollment. Simulation studies and case examples show strong potential for real-time trial monitoring.

Keywords: Clinical Trials, Patient Enrolment, Forecasting, Screening/Rescreening, Randomization, Poisson-gamma Enrolment Model

Principal Component Quantile Sampling: A Structure-Preserving Strategy for Efficient Data Reduction and Statistical Learning

Foo, Hui-Men and Yuan-chin Ivan Chang

Institute of Statistical Science

Academia Sinica

Taipei, Taiwan

Abstract. Modern statistical analysis increasingly encounters massive and high-dimensional datasets, where computational and storage constraints necessitate effective data reduction strategies. Conventional sampling methods, such as simple random sampling, often fail to preserve intrinsic geometric and distributional structures, particularly under aggressive reduction rates. We propose Principal Component Quantile Sampling (PCA-QS), a novel framework that selects representative observations based on quantile positions along principal component directions. By leveraging the geometric and probabilistic structure revealed through principal component analysis, the proposed approach ensures systematic and representative coverage of the data distribution. We investigate its statistical and geometric properties and demonstrate through simulations and applications that PCA-QS substantially improves preservation of distributional characteristics, statistical inference accuracy, and downstream learning performance compared with conventional random sampling. The proposed method provides a principled and computationally efficient solution for structure-preserving data reduction in modern statistical learning and stochastic modeling applications.

Keywords: Principal Component Analysis; Quantile Sampling; Structure-Preserving Sampling; Data Reduction; High-Dimensional Data Analysis; Statistical Learning; Multivariate Analysis; Stochastic Modeling

Quantifying the Impact of Sustainability on ROE: Insights from Non-Linear & Quadratic Regression Models"

Spyridon D. Lampropoulos, Georgios L. Thanasas, Sotiris Kotsiantis, Nicos Sykianakis

Abstract.In this study, we examine the relationship between Return on Equity (ROE) and key financial indicators, including Return on Assets (ROA), Return on Invested Capital (ROIC), and Sustainability Rank. The main research hypothesis is that companies with higher sustainability scores are likely to demonstrate superior ROE, driven by more efficient capital management and long-term value creation. The quadratic regression model was finally chosen for its ability to capture non-linear relationships while maintaining ease of interpretation, although advanced ensemble models like Extra Trees Regressor (ET), Random Forest Regressor (RF), and XGBoost showed high predictive accuracy. ROE is positively affected by ROIC, ROA, and Sustainability Rank in a moderate but significant way. These findings indicate that companies focusing on ESG have superior financial performance, demonstrating the importance of incorporating sustainability into financial planning.

Keywords: ROE, ROA, ROIC, ESG, Sustainability Rank

Quantum Circuit based Workflow for Multivariate Connectivity Estimation and Classification

Vishwambhar Pathak

BIT Mesra Jaipur Campus, Jaipur, India, Email: v.pathak@bitmesra.ac.in

Nisheeth Joshi

Banasthali Vidyapith, Rajasthan, India Email: jnisheeth@banasthali.in

Prabhat Mahanti

University of New Brunswick, Saint John, Canada, Email: pmahanti@unb.ca

Abstract. Multivariate connectivity estimation and classification play a critical role in the analysis of complex networked systems such as brain functional networks, communication infrastructures, and other spatio-temporal graph-structured data. Classical machine learning techniques often face limitations when modelling high-

dimensional connectivity patterns due to non-linearity, combinatorial complexity, and the need for aggressive dimensionality reduction. To address these challenges, this paper proposes a quantum circuit–based hybrid workflow for multivariate connectivity estimation and classification, integrating quantum feature encoding, quantum kernel evaluation, and classical learning components.

The proposed framework represents connectivity patterns as graphs and encodes graph-derived features into quantum states using low-depth variational quantum circuits based on angle embedding. To systematically evaluate the expressivity and robustness of the approach, extensive experimentation is conducted on synthetic graph datasets, including Erdős–Rényi random graphs and cycle graphs, which exhibit fundamentally different structural properties. These synthetic benchmarks enable controlled analysis of the model's ability to discriminate between random and structured connectivity patterns.

A swap-test-based quantum circuit is employed to estimate state fidelities, forming a quantum kernel that captures similarity between graph connectivity instances in a high-dimensional Hilbert space. This kernel is then integrated into a hybrid Quantum-Kernel Support Vector Classifier (QSVC), enabling efficient classification with a reduced number of trainable parameters. The hybrid workflow is systematically compared against classical kernel-based methods, including radial basis function (RBF) SVC, in terms of classification accuracy, robustness, and kernel separability.

Experimental results on synthetic graph datasets demonstrate that the quantum-enhanced kernel provides improved discrimination of structurally distinct connectivity patterns while maintaining shallow circuit depth compatible with noisy intermediate-scale quantum (NISQ) devices. Latent space visualization and kernel analysis further reveal enhanced geometric separation between graph classes, contributing to improved interpretability. The proposed workflow establishes a scalable and modular foundation for applying quantum machine learning to multivariate connectivity analysis and provides a principled pathway toward future deployment on real quantum hardware.

Keywords: Quantum Machine Learning, Quantum Kernel Methods, Graph Connectivity Analysis, Hybrid Quantum–Classical Learning, Multivariate Network Classification, Variational Quantum Circuits

Research on Aging Risk Identification and Policy Simulation of China's Families Who Lost Their Only Child and Families with Disabled Only Child

——From the Perspective of the Three-Dimensional Framework of "Population Dynamics - Risk Mechanism - Security Adaptation"

Guoyong Ma*[Ph.D., Dean and Professor, School of Economics and Management, Northeast Forestry University]*

Hong Mi ***[Ph.D****., Corresponding Author (E-mail: spsswork@163.com), Chengdong Distinguished Professor, School of Economics and Management, Northeast Forestry University; Director (Adjunct), Research Base of Population Big Data and Policy Simulation (Workshop), Zhejiang University]*

Abstract: Aiming at the multiple risks and security deficiencies faced by China's families who lost their only child and families with disabled only child in the aging process, this study constructs a three-dimensional analytical framework of "Population Dynamics - Risk Mechanism - Security Adaptation" from an interdisciplinary perspective. Breaking through the limitations of traditional single-discipline research, this study innovatively integrates the accurate prediction method of "three-parameter + dual-gender" in demography, the social vulnerability theory in disaster risk management, and the risk loss distribution analysis in actuarial science, aiming to systematically identify the risk structure of this group and explore the path of policy optimization.

In terms of methodology, the original three-parameter model life table and dual-gender model are adopted, which significantly improve the prediction accuracy and characterization ability of the life expectancy evolution, gender differences and risk dynamics of this group. By transforming the status of "losing the only child / having a disabled only child" into quantifiable and traceable multi-dimensional dynamic parameters, this study realizes a refined analysis from macro-scale prediction to micro-risk characteristics.

Based on the above framework and methodology, the study further conducts multi-policy scenario simulation to simulate the implementation effects of intervention measures such as economic subsidies, integration of medical and elderly care services, and psychological support, providing visualized and verifiable decision-making references for policy formulation. The results show that establishing a differentiated

and targeted support system is the key to alleviating the economic, health, care and psychological risks of this group.

This study not only provides an interdisciplinary methodological paradigm for research on similar vulnerable groups, but also offers empirical evidence and policy implications for China to improve the elderly care security system for special families during the 15th Five-Year Plan period; meanwhile, it provides a Chinese example for carrying out population simulation research in global South-South cooperation.

Robust Estimation of Value-at-Risk under Weibull Loss Distributions: A Simulation and Empirical Study

Derya Karagoz and Ugur Karabey

Department of Statistics & OR, Faculty of Science, Phsics and MAths Bulding, 504 , University of Malta, Msida MSD 2080, Malta karagozzderya@gmail.com

Abstract. Modelling insurance losses accurately is essential for effective risk management and solvency control. Owing to its flexible tail behaviour, the Weibull distribution is commonly applied in non-life insurance modelling. Value-at-Risk (VaR) remains one of the most widely used risk measures in actuarial and financial applications; however, classical VaR estimation methods are sensitive to outliers and model deviations.

This study proposes robust estimation approaches for VaR under the Weibull loss distribution. The statistical performance of the proposed methods is assessed via extensive Monte Carlo simulation studies. Additionally, the practical applicability of the robust VaR estimators is demonstrated using real loss data from a Turkish insurance company. The results indicate that robust methods outperform classical approaches in terms of stability and reliability, especially when the underlying data exhibit contamination or heavy-tailed behaviour.

Keywords: Robust Value-at-Risk, Weibull Distribution, Monte Carlo Simulation, Non-life Insurance

Statistica l Modeling and Assessment of Regression Models Using Data from Computer Modeling

Fatima Sapundzhi[1], Irina Naskinova[2], Mikhail Kolev[2], Meglena Lazarova[3,4], and Michail Todorov[5]

[1]Dept of Communication and Computer Engineering, Faculty of Engineering, Neofit Rilski South-West University of Blagoevgrad, 66 Iv. Mihaylov str., 2700 Blagoevgrad, Bulgaria

[2]Dept of Mathematics, University of Architecture, Civil Engineering and Geodesy, 1 H. Smirnenski Blvd, 1164 Sofia, Bulgaria,

[3]Faculty of Applied Mathematics and Informatics, Technical University of Sofia, 8 Kl. Ohridski Blvd., 1000 Sofia, Bulgaria

[4]Dept of Mathematics, Informatics and Physics, Prof. Dr Assen Zlatarov State University of Burgas, 1 Prof. Yakimov Blvd., 8000 Burgas, Bulgaria,

[5]Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Akad.G. Bonchev bl. 8, 1113 Sofia, Bulgaria,

Abstract. This study examines a problem for statistical modelling and assessment of the predictive ability of regression models illustrated by computer data of biological compounds. The relationship between the quantitative predictors and a continuous dependent variable is analysed through the investigation of stability and interpretability of regression estimates in the presence of a correlation between the predictors. Linear regression and Partial Least Squares regression are used. The parameter estimation is performed by using the least squares method. The preliminary analysis includes a standardization of variables and examination of the correlation structure. The generalizability of the models is assessed by using the cross-validation scheme. In order to determine the quality of the given models, root mean square error, and the cross-validated indicator we evaluate the determination's coefficient.

The obtained results show the presence of statistically significant dependencies and stability of the regression estimates, emphasizing the applicability of classical and latent regression approaches as an analytical tool in the context of applied stochastic models and data analysis.

Keywords: Biological compounds; cross-validation scheme; regression model

Acknowledgments

The research work of Meglena Lazarova is supported by Project No 2025-FNSE-02 “Research of Natural and Anthropogenic Phenomena” financed by Scientific Research Fund of Ruse University.

Studying the Impact of Degradation of Data through Ecological Network Analysis Indices in Linear Inverse Modeling for Marine Trophic Systems

Valerie Girardin[1] Jacques Brehelin[1] Theo Grente2 Nathalie Niquil[3]

1 Laboratoire de Math´ematiques Nicolas Oresme, Universit´e de Caen Normandie, CNRS-6139, 14032 Caen, France

2 Groupe de Recherche en Informatique, Image, et Instrumentation de Caen, Universit´e de Caen Normandie, CNRS-6072, 14032 Caen, France

3 Laboratoire Morphodynamique Continentale et Cotiere, Universite de Caen Normandie, CNRS-6143, 14032 Caen, France

Abstract. The Linear Inverse Modeling (LIM) approach of food webs relies on the principle of steady states for the biomasses of all internal compartments. More con straints are added from field measurements of mass transfers and from the literature on similar ecosystems. All these linear equations and inequalities define a bounded multidimensional polyhedron, called a polytope. Further, the Ecological Network Analysis functions (ENA)–such as Quadratic Energy, McArthur Index, Overhead quantify interesting aspects of the systems. The deterministic approach [1] uses Se quential Quadratic Programming to find the optimum solution of a given ENA. The stochastic approach [3] provides a Markov Chain Monte Carlo sampling of the poly tope, thus yielding simulated ENA values. However, both measures and numerical data issued from other models may be inaccurate. In the continuity of [2], for studying the impact on ENA of this inherent uncertainty of data, we relax the constraints, by transforming some equations into inequalities. This process, which we call degradation, allows us to quantify some pertinent error bounds on the measures, as well as to characterize the most important flows in the system in terms of ENA. This is illustrated by degradation scenarios from seven sites of the Sylt-Rømø Bight ecosystem; see [3].

Keywords: ENA, LIM, MCMC.

The effect decomposition and mechanism of green innovation empowering carbon emission performance

Xiuting Cai

Northeast Forestry University

No. 26, Hexing Road, Xiangfang District, Harbin City, Heilongjiang Province, Harbin, 150040China

Abstract. Green innovation is an important path to reduce environmental pollution and achieve carbon emission reduction targets. Based on the panel data of 30 provinces (municipalities) in China from 2011 to 2022, this paper uses the super-efficiency SBM model to calculate the green innovation efficiency and carbon emission performance respectively. On the basis of spatial correlation tests, it uses the spatial econometric model to explore the total effect, direct effect and spatial spillover effect of green innovation empowering carbon emission performance by region. The results show that: (1) Both green innovation and carbon emission performance present a spatial distribution characteristic where the eastern region is higher than the western region. During the study period, the green innovation and carbon emission performance of most regions showed an upward trend; (2) Green innovation has a significant positive promoting effec t on carbon emission performance. In addition, the level of economic development, the degree of opening up to the outside world, the level of science and technology, and the level of urbanization have a significant positive impact on carbon emission performance, while the energy structure and the level of financial development have a significant negative impact on it. (3) Heterogeneity analysis indicates that green innovation in the eastern region has a significant positive direct effect and spatial spillover effect on carbon emission performance. Green innovation in the central region has a significant negative direct effect on carbon emission performance, while the spatial spillover effect is not significant. Green innovation in the western region has a significant positive direct effect and a significant negative spatial spillover effect on carbon emission performance.

Keywords: Green innovation; Carbon emission performance; Effect decomposition; Super-efficiency SBM model; Spatial econometric model

The Inverse Shortest Path Problem for Earthquakes' Motion

Jeremy Farrugia[1], Mark Anthony Caruana[2],

[1] *Department of Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta.*

[2] *Department of Statistics and Operations Research, Faculty of Science, University of Malta, Msida, Malta*

Abstract. According to Fermat's principle, seismic waves follow paths of least travel time. Thus, shortest path algorithms such as Dijkstra's can be used to determine these paths. Conversely, inferring the parameters of a mathematical program from an observed optimal path defines the inverse shortest path problem, an area within inverse optimisation.

With its wide range of applicability, inverse optimisation has attracted considerable interest. One of the earliest topics in this field was the inverse shortest path problem, with Burton and Toint (1992) laying its foundations. Since then, this problem has been explored across several domains, with various mathematical formulations and algorithms proposed.

In this paper we examine the inverse shortest path and apply it to a seismological case study. Three algorithms are employed to solve the problem: the column generation algorithm, a quadratic programming algorithm, and a deep inverse optimisation algorithm using a modern deep learning framework. The aim is to estimate the weight vector using these algorithms, thereby reconstructing the mathematical program that defines the shortest paths taken by seismic waves. To the best of the author's knowledge, this is the first study to apply the inverse shortest path problem to local seismic data from the Maltese Islands and Sicily.

Keywords Inverse: Shortest path problem, column generation, quadratic programming, deep inverse optimization, deep learning.

Using the robust Cox regression model to analyze student dropouts in tertiary education

Liberato Camilleri[1], Derya Karagoz[1] and Mireille Fenech[1]

[1]*Department of Statistics and Operations Research, University of Malta, Malta*

Abstract. This paper applies the Cox Proportional Hazards (CPH) model to analyze student dropouts in tertiary education. Despite that the CPH model has been used extensively to analyze time-to-event data in educational settings, it is vulnerable to outliers. Consequently, for atypical student paths, the result may yield imprecise approximations and incorrect inferences. This study employs the Robust Cox Regression Estimator to overcome this issue. The considered approach is based on a smooth modification of the partial likelihood which is a reliable estimation technique originally proposed by Bednarski (1993). Such a method is specifically designed to provide better and unbiased estimation in the presence of data contamination and so this robust methodology will be used to assess dropout trends among University of Malta (UOM) undergraduate students in the Faculty of Science.

Using institutional data from the cohort that applied for their respective course in 2022, the study examines the long-term effects of gender, admission scores, and subject combinations on dropout rates. The analysis compares results obtained from traditional Cox regression analysis with its robust alternative approach to assess the stability of hazard ratio estimates when potential outliers are down-weighed. Methodologically, the thesis demonstrates that reducing observations with high leverage in the partial likelihood function can yield more accurate parameter estimates without a notable loss of efficiency.

The end of the first year of tertiary education is the time when dropouts are most prevalent. Other findings also show that academic preparedness and particular subject combinations have a significant impact on dropout rates. Moreover, while pointing out situations where classical estimates might be unduly impacted by atypical cases, the robust approach also identifies a number of covariates whose effects hold true across various estimation frameworks. These findings are essential to university administration to allocate resources and devise plans to reduce student dropouts from university courses

Keywords: Robust Cox regression estimator, Partial likelihood, Dropout rates

Visual Attention and Cognitive Effort in Financial Statement Analysis: Insights from an Experimental Study

Georgios Antonopoulos, Alexia Gazeta, Maria Rigou, Georgios L. Thanasas

Abstract: Financial statement analysis is a core activity in accounting and finance, requiring users to integrate numerical, visual, and contextual information in order to assess a firm's financial condition. Despite its importance, limited empirical evidence exists on how individuals actually process and interpret financial information at a cognitive and affective level. This study presents an experimental investigation of financial data interpretation using eye-tracking and neurophysiological measures. Participants—including undergraduate accounting students and experienced accounting professionals—were asked to evaluate a firm's financial condition based on charts, tables, financial ratios, and balance sheet information displaying challenging or potentially problematic financial indicators. Visual attention patterns, indicators of cognitive effort, and basic affective responses were recorded while participants analyzed the financial information and formed their assessments. The study adopts a behavioral perspective to explore how financial information is visually explored, which elements attract attention, and how cognitive load and emotional responses may influence judgment. The findings provide managerial and educational insights into how financial statements are interpreted in practice, highlighting differences in information processing strategies and the potential implications for financial reporting, disclosure design, and accounting education. By combining behavioral data with traditional financial analysis tasks, the study contributes to the growing literature on behavioral accounting and supports the development of more effective and user-centered financial communication.

Keywords: Behavioral accounting, Financial statement analysis, decision-making, Visual attention, Cognitive effort, Eye-tracking

Waiting time until absorption in the coupon collector problem with penalty coupon

Bojana Todić[1] and Jelena Jocković[2]

[1] *Faculty of Mathematics, University of Belgrade, Serbia*

[2] *Faculty of Mathematics, University of Belgrade, Serbia*

Abstract. Coupon collector problem with penalty coupon (CCPPC) is recently introduced generalization of the classical problem ([1]). It is obtained by adding a, so called, penalty coupon, which interferes with collecting standard coupons, to the coupon set. Therefore, the process of completing a full collection of standard coupons is slower than in the case of classical coupon collector problem, and can possibly fail. We provide several results related to the expected waiting time until a full collection of standard coupons is obtained (or has failed), and compare the results across several variants of the coupon collector problem with additional coupons ([2],[3],[4]). We also discuss the connection of CCPPC to random walk with two absorbing barriers. Keywords: Coupon collector problem with penalty coupon, Markov chains, Waiting time until absorption, Random walk with absorbing barriers.

What are the options for ensuring the financial stability of the Czech pension system?

Dr. Tomáš Fiala, Jitka Langhamrová, Jana Vrabcová

Prague University of Economics and Business, nám. W.Chuchilla 4, Praha 3, 13067, Czech Republic
fiala@vse.cz

Abstract. The baby boom in Czechia peaked in 1974, so after 2040 there will be a gradual and significant increase in the ratio of people of post-productive and productive age. A number of demographic and economic measures can contribute to the financial stability of a pension system operating on a PAYG basis: increasing the birth rate, increasing net migration, raising the retirement age, increasing contributions to the pension system, or reducing pension levels.

This paper presents model calculations of the level of these individual measures that would be necessary to ensure the financial sustainability of the pension system after 2030. The calculations are based on the current forecast of demographic developments

in Czechia published by the Czech Statistical Office in 2023 and adjusted according to current demographic developments

Key Words: Population ageing, old-age pension system, old-age dependency ratio, fertility, migration.

Accurate forecasting of regional carbon emissions in China is essential for the government to formulate dual-carbon policies

Hongmin Li
Harbin, China, lhm@nefu.edu.cn

Abstract

Accurate forecasting of regional carbon emissions in China is essential for the government to formulate dual-carbon policies. The regional carbon emissions data are complicated and the influencing factors are difficult to determine. Grey forecasting models have advantages in solving the problem of insufficient data and information. However, existing grey models for carbon emission forecasting have shortcomings such as insufficient feature expression and parameters to be optimized. Therefore, this study designs a novel intelligent multivariate grey modeling mechanism that incorporates multi-stage feature selection and the honey badger algorithm to accurately identify and quantify feature effects. Firstly, a multi-stage feature selection model is proposed to gradually identify the driving factors based on different weights. Subsequently, the driving terms were designed and optimized to different degrees to reduce the unce rtainty caused by the fusion of factors from different domains. Finally, a new meta-heuristic algorithm is applied to optimize the background value coefficients of the grey model. The results show that the combined error of the new model is only 0.93% compared to other candidate models. The proposed model can be used as an effective information analysis tool for high quality and sustainable development of regional economy.

Evaluating Test Battery Behavior with Contaminated Numerical Sequences

Dra. **Elena Almaraz Luengo**
Department of Statistics and Operational Research, Faculty of Mathematical Science, Complutense University of Madrid, Spain

Madrid, Spain

ealmaraz@ucm.es

Abstract. In many scientific and engineering fields, assessing the statistical quality of numerical sequences is essential, as these sequences are often used in subsequent applications such as computational simulation or cryptography. This assessment is typically performed through statistical hypothesis tests, which in practical settings are organized into collections commonly referred to as test suites or batteries.
In this work, we address the problem of sensitivity analysis for a statistical test battery when evaluating deliberately contaminated sequences, and we present a case study to illustrate the methodology.

Geometry of Euclidean matrices and multifractal analysis

Leonidas Sakalauskas, Neringa Urbonaite

Vytautas Magnus University, Vilnius, Lithuania

Many processes in environmental protection, economics, and technology are fractal. An important fractal parameter is the Hurst coefficient. The paper presents a method for multivariate and multivariate data analysis based on the geometry of fractal Euclidean distance matrices. The fractal Brownian vector field is defined only through the fractal Euclidean distance matrix, avoiding the correlation functions of positive definitions usually used for long-distance definitions. The model created in this way is a multi-year and multi-output generalization of the well-known Wiener process. The maximum likelihood method for estimating model parameters and the kriging method for

predicting data are presented. Applications of the method to the analysis of heavy metals and climate data are discussed.

Graphical predictive model of multimorbidity in allergic diseases

Konrad Furmańczyk, Wojciech Niemiro, Marta Zalewska

Department of Prevention of Environmental Hazards, Allergology and Immunology, Medical University of Warsaw, Poland
marta.zalewska@wum.edu.pl

Abstract: Multimorbidity is common in allergic diseases. We construct a graphical model based on data gathered in the big epidemiological survey ECAP [2]. The goal is to better understand causal relations between several allergic diseases and also to provide medical doctors with a tool to assist in making correct diagnosis. The variables included in the model are partitioned into three groups: the diseases Y, the symptoms S and the covariates X, all considered as nodes of a directed acyclic graph (DAG), similarly as in [1]. The arrows of the graph lead from X to Y to S and also among some variables Y. We used expert medical knowledge and suggestions produced by AI to choose the structure of the DAG within Y, which reflects comorbidities. For every node Y and S, we fitted a regression model (logistic or linear) in the training phase. In the predictive regime, we input values of variables S and X for new patients and compute the M AP (maximum a posteriori) estimate for the presence of diseases (configuration of Y variables). We conducted experiments in which we partitioned the ECAP data into training and testing sets. The results are promising and confirm the model usefulness in predicting the structure of co-occurring allergic diseases.

Key Words: multimorbidity, DAG, logistic regression, maximum a posteriori

References:

1. Furmańczyk K. Niemiro W. Chrzanowska M. Zalewska M. Network Model with Application to Allergy Diseases. International
Conference on Computational Science (ICCS). 2024, 14835, 105-112 2.
2. Samolinski B, Raciborski F, Lipiec A i wsp. (2014), Epidemiologia Chorób Alergicznych
w Polsce (ECAP), Alergol Pol 2014; 1: 10-8

An innovative statistical method for co-localization in super-resolution microscope images

Professor Hui Zhang

Northwestern University, 680 North Lake Shore Drive, Suite 1400, Chicago IL, 60611 USA

hzhang@northwestern.edu

Abstract: Spatial data is common in many scientific disciplines. For example, single-molecule localization microscopy, such as stochastic optical reconstruction microscopy, provides super-resolution images to help scientists investigate the co-localization of proteins and hence their interactions inside cells, which are key events in living cells. However, there are few accurate methods for analyzing co-localization in super-resolution images. The current methods and software are prone to produce false-positive errors and are restricted to only 2-dimensional images. We will propose a novel statistical method to effectively address the problems of unbiased and robust quantification and comparison of protein co-localization for multiple 2- and 3-dimensional image datasets. This method significantly improves the analysis of protein co-localization using super-resolution image data, as shown by its excellent performance in simulation studies and an analysis of light chain 3-lysosomal-associated membrane protein 1 protein co-localization in cell autophagy. Moreover, this method is directly applicable to co-localization analyses in other disciplines, such as diagnostic imaging, epidemiology, environmental science, and ecology.

Key Words: Co-localization, Non-parametric, Robust

CONTENTS