From the Editor

It is with great pleasure that we present our readers with the September issue consisting of 12 articles arranged, as usual, in three sections: Research papers, Other articles, and Research communicates and letters. A wide spectrum of topics is discussed in these papers by 25 authors from a large group of countries: Czechia, Poland, India, Cameroon, Brazil, Jordan, Saudi Arabia, Nigeria, Iran, and Botswana.

Research articles

The issue starts with the paper by **Kamila Hasilová**, **Ivana Horová**, **David Vališ**, and **Stanislav Zámecník,** entitled *A comprehensive exploration of complete cross-validation for circular data.* The aim of the article is to propose a novel circular-specific method that is based on a crossvalidation procedure with a von Mises density used as a kernel function. Using simulated data as well as real-world circular data sets, the authors evaluate and validate the proposed method and compare it with the existing methods. This method extends the estimate of the MISE and has better theoretical properties than the LSCV. From the presented results and outcomes it was concluded that the CCV is applicable to various data types with a respective success rate. From the data-driven method of bandwidth selection, the authors focused only on the cross-validation methods which target the MISE to have the consistent group of methods to compare.

Adam Szulc's paper *Reconstruction of the social cash transfers system in Poland and household wellbeing: 2015–2018 evidence* examines the impact of changes in the social benefits system on the wellbeing, poverty, and economic activity in Poland. The core element of those changes was a programme of large cash transfers, referred to as Family 500+, introduced in 2016. It was intended to support families with children, especially the least affluent ones, and to foster fertility. The impact of the transfers is examined through the observation of changes in the monetary and multidimensional wellbeing of households. The study also analyzed the changes in recipients' economic activity using estimates of regression models and treatment effects. The Family 500+ programme proved to be successful as an anti-poverty tool and also resulted in the increase in the average wellbeing for the whole population.

In the paper *On autoregressive processes with Lindley-distributed innovations: modeling and simulation* **K. U. Nitha** and **S. D. Krishnarani** develop an autoregressive

process of order one, assuming that the innovation random variable has a Lindley distribution. The key properties of the process under investigation embrace five distinct estimation techniques employed to estimate the respective parameters. Parametric and non-parametric estimating techniques are effectively employed. The authors explored a first order autoregressive model with the Lindley error distribution and its properties. The stationarity of the process is tested using a unit root test. The application of the proposed process to the analysis of time series data is demonstrated using real data sets. Based on some important statistical measures, the analysis of the data sets reveals that the proposed model fits well, and the errors are independent and Lindley-distributed. The stationary series of additive autoregressive models could feature non-Gaussian errors and marginal.

In the next article, *Comparing logistic regression and neural networks for predicting skewed credit score: a LIME-based explainability approach,* **Jane Wangui Wanjohi**, **Berthine Nyunga Mpinda**, and **Olushina Olawale Awe** compare the predictive ability of Logistic Regression (LR) and a Multilayer Perceptron (MLP) using two types of data sets, with an advanced model explainability technique - Local Interpretable Model-Agnostic Explanations (LIME). The findings show that all models performed better after the data were balanced. MLP had higher scores than LR in terms of balanced accuracy, Matthews correlation coefficient, and F1 score. From the findings, this study recommends that lending companies with small amounts of data use a logistic regression model but for companies with vast amounts of data a multilayer perceptron will ease their credit offer processes. The study also highlights the importance of using explainable artificial intelligence. With the LIME explanation approach, authors were able to see how each feature influences the predicted class of a model for a given instance.

Anna Czapkiewicz's and **Katarzyna Brzozowska-Rup's** paper entitled *The Measurement of the Gross Domestic Product affected by the shadow economy* presents a method for balancing Gross Domestic Product (GDP) when the measurements of its components are distorted by the existence of the shadow economy using a multiple ultrastructural model (MUM), where the explanatory variables are subject to error. The expected value of GDP can be divided into two parts: the first part concerns data related to registered activities and the second part concerns unobserved data which may be partly related to the shadow economy. The empirical analysis utilizes the annual data of the Local Data Bank, for years 2000–2019. The results show that the unobservable part of the variables necessary to balance GDP on the production side does not exceed 1% of GDP, and on the expenditure side, it mostly reaches about 3% of GDP.

In the paper entitled *Extropy and entropy estimation based on progressive Type-I interval censoring* **Huda H. Qubbaj**, **Husam A. Bayoud**, and **Hisham M. Hilow** discuss the problem of estimation of the extropy and entropy measures based on

progressive Type-I interval censoring samples. Nonparametric-based methods involving moments and linear approximations have been proposed to this aim. The performance of the proposed estimates have been studied via simulation studies and real data sets considering various censoring schemes and three probability distributions: uniform, exponential and normal distributions. The proposed estimates of the extropy and entropy measures shown to be affected by the sample size, censoring schemes and the type of distribution. Yet, the estimates based on linear approximation outperform the other estimate in the majority of cases under study.

The paper by **R. R. Sinha, Bharti,** *Improved estimation of the mean through regressed exponential estimators based on sub-sampling non-respondents*, discusses the issue of estimating the population mean and presents and improved regressed exponential estimators using different parameters of an auxiliary character based on sub-sampling of non-respondents. The mean square error (MSE) of the proposed estimators for the most pragmatic simple random sampling without replacement (SRSWOR) scheme have been derived up to the first order of approximation (i.e. the expression containing errors up to the power of two so that the expectation comes only in terms of the mean, variance and covariance). The optimum value of the MSE of the estimators is found, along with the necessary conditions for optimizing the MSE. The effectiveness of the suggested estimators, outperforming the existing ones in terms of their MSE, has been studied theoretically, while the empirical illustration using the simulation studies have confirmed these findings.

Idowu Oluwasayo Ayodeji's paper, *Forecasts of the mortality risk of COVID-19 using the Markov-switching autoregressive model: a case study of Nigeria (2020–2022),* discusses some aspects of the global pandemic due to SARS-Cov-2 and attempts to predict future occurrences of such cases in order to prevent or combat effectively consequences of the virus. This study modeled daily fatality rate in Nigeria from March 23, 2020 to March 19, 2022 and forecasts future occurrences using Markov switching model (MSM). It revealed that as of 19th March, 2022, Nigeria remained at the low-risk regime in which 1 (95%CI: 0, 1) person, on the average, died of coronavirus daily; however, the most probable scenario in the nearest future was the medium-risk state in which an average of 4 persons would die daily with 48.7% probability. The study concluded that Nigerian COVID mortality risks followed a switching pattern which fluctuated within low-, medium- and high-risks; however, the medium-risk state was most likely in the future. The results indicated that the quarantine measures adopted by the governments yielded positive results. It also underscored the need for governments and individuals to intensify efforts to ensure that the country remained at the low-risk zone until the virus would be eventually eradicated.

In the paper *Nonparametric Bayesian optimal designs for Unit Exponential regression model with respect to prior processes (with the truncated normal as the base*

measure) **Anita Abdollahi Nanvapisheh**, **Soleiman Khazaei**, and **Habib Jafari** present a Bayesian optimal design by utilizing the Dirichlet process as *a prior*. The Dirichlet process serves as a fundamental tool in the exploration of Nonparametric Bayesian inference, offering multiple representations that are well-suited for application. Authors introduce a novel one-parameter model, referred to as the 'Unit-Exponential distribution', specifically designed for the unit interval. Additionally, a stick-breaking representation to approximate the D-optimality criterion considering the Dirichlet process as a functional tool was employed. This approach allows to explore and evaluate the performance of the nonparametric Bayesian optimal design under varying levels of concentration parameter α. The empirical results reveal interesting findings: for small parameter values, there are no two-point designs observed.

Jacek Białek's paper *The use of the Bennet indicators and their transitive versions for scanner data analysis* revises the price and quantity Bennet indicators and their multilateral versions for the analysis of scanner data. Specifically, instead of considering comparisons across firms, countries or regions, the transitive versions of the Bennet indicators are adapted to work on scanner data sets observed over a fixed time window. Since the scanner data sets have a high turnover of products, which can make it difficult to interpret the difference in sales values over the compared time periods, the paper also considers a matched sample approach. One of the objectives of the study is to compare bilateral and multilateral Bennet indicator results across all available products or strictly matched products over time. It also examines the impact of data filters used and the level of data aggregation on the price and quantity.

Other articles
XXXXI Multivariate Statistical Analysis 2023, Lodz, Poland. Conference Papers

The article by **Adam Idczak** and **Jerzy Korzeniewski,** *Language independent algorithm for clustering text documents with respect to their sentiment,* presents a novel unsupervised algorithm for documents written in any language using documents written in Polish as an example. The clustering of Polish language texts with respect to their sentiment is poorly developed in the literature on the subject. The novelty of the proposed algorithm involves the abandonment of stoplists and lemmatisation. Instead, the authors propose translating all documents into English and performing a two-stage document grouping. In the first step of the algorithm, selected documents are assigned to a class of positive or negative documents based on a set of lexical and grammatical rules as well as a set of key-terms. Key-terms do not have to be entered by the user, the algorithm finds them. In the second step, the remaining documents are attached to one of the classes according to the rules based on the vocabulary found in the documents grouped in the first step. The algorithm was tested on three corpora of documents and achieved very good results.

Research Communicates and Letters

In this section, the paper by **R. Sivasamy, *A finite state Markovian queue to let in impatient customers only during K-vacations*,** investigates a matrix analysis study for a single-server Markovian queue with finite capacity. During the vacation periods of the server, every customer becomes impatient and leaves the queues. If the server detects that the system is idle during service startup, the server rests. If the vacation server finds a customer after the vacation ends, the server immediately returns to serve the customer. Otherwise, the server takes consecutive vacations until the server takes a maximum number of vacation periods, e.g. K, after which the server is idle and waits to serve the next arrival. During vacation, customers often lose patience and opt for scheduled deadlines independently. If the customer's service is not terminated before the customer's timer expires, the customer is removed from the queue and will not return. Matrix analysis provides a computational form for a balanced queue length distribution and several other performance metrics.

Włodzimierz Okrasa
Editor