Romanian Statistical Review
Revista Română de Statistică

THE JOURNAL OF NATIONAL INSTITUTE OF STATISTICS

EXAMINATION OF PENSION INVESTMENT FUNDS IN TURKEY WITH TIME SERIES ANALYSIS METHODS AND FORECASTING WITH ARIMA

TOPP-LEONE EXPONENTIAL DISTRIBUTION FOR ASYMMETRIC LOSS FUNCTIONS WITH IDENTICAL PRIORS

CLUSTERING COUNTRIES ACCORDING TO GLOBAL INNOVATION INDEX 2022 USING FUZZY CLUSTERING ANALYSIS

THE ROLE OF THE LABOR RESOURCES IN THE ECONOMIC GROWTH FOR SUSTAINABLE DEVELOPMENT
BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

“FORECASTING MAIZE PRODUCTION IN ROMANIA: A BSTS MODEL APPROACH”

3 /2024

www.revistadestatistica.ro

CONTENTS 3/2024

Examination of Pension Investment Funds in Turkey with Time Series Analysis Methods and Forecasting with ARIMA

Övgücan Karadağ Erdemir (ovgucan@hacettepe.edu.tr)
Hacettepe University, Faculty of Science, Department of Actuarial Sciences, Ankara, Türkiye,

Murat Kırkağaç (murat.kirkagac@dpu.edu.tr)
Kütahya Dumlupınar University, Faculty of Applied Sciences, Department of Insurance and Risk, Kütahya, Türkiye

ABSTRACT

In this study, the behavior of private pension investment funds in Turkey, one of the most important investment instruments, was examined using time series analysis methods over a six-year period. Daily price, daily number of shares in circulation, daily number of people, daily total fund value and daily logarithmic return data of selected low, medium and high risk pension investment funds were converted into weekly average data. The movements of the weekly average values of the funds over time were examined graphically using time series analysis methods. The stationarity of the weekly average logarithmic return values of ALZ, AZS and AMZ funds was examined with unit root tests, and the stationarity process was applied to non-stationary returns. Steady weekly average logarithmic return values were modeled with appropriate Autoregressive Integrated Moving Average (ARIMA) models, a one-year forecast was made and compared with the real values. It has been observed that in low risk ALZ funds, forecast values that are closer to reality and have lower errors are obtained with ARIMA methods.

Keywords: *Pension Investment Fund, Risk, Time Series Analysis, ACF, PACF, ADF Test, ARIMA*

1. INTRODUCTION

Most of the retirement plans in the world have a three-pillar structure. The first pillar is the national social security system created by the state, which grants retirement rights to the working individual. The second pillar consists of supplementary retirement funds and plans formed by the employers. The third pillar is private retirement funds established by the private sector, which are voluntary investment and savings systems (Dağlar, 2007).

In Turkey, to complement the public social security system, the Draft Law on the Individual Pension Savings and Investment System was submitted to the Presidency of the Turkish Grand National Assembly on May 16, 2000. The aim was to establish a system based on individual pension accounts to regulate the savings voluntarily made by individuals for retirement (Demirci, 2006).

The "Individual Pension Savings and Investment System Law" No. 4632 was accepted by the Turkish Grand National Assembly on March 28, 2001. The law was published in the Official Gazette No. 24366 on April 7, 2001, and put into practice on October 7, 2001.

The individual pension system is a private retirement system that directs the savings made by individuals during their active working years into long-term investments. It aims to provide additional income when individuals retire, thus enabling them to maintain their living standards (PMC, 2024).

Investment funds are one of the most important elements of the individual pension system. The reason is that the system is built on the basis of amounts accumulated through pension investment funds (Samancı, 2010).

According to the Individual Pension Savings and Investment System Law No. 4632, an individual pension investment fund is an asset designed within the framework of a pension contract by a pension company, where participants' contributions are monitored in individual pension accounts, and managed in accordance with the principles of fiduciary ownership and risk distribution. An individual pension investment fund does not have a legal personality. The individual pension investment fund cannot be used or established for purposes other than those stated in the law in force.

The two main parameters to be considered in individual pension investment funds, as in all investment funds, are the measurement of return and volatility. Return refers to the income obtained from an investment or movable value. Volatility refers to the risk of the change in value of a financial instrument over a certain period of time.

There are many studies on pension investment funds, which have become important investment instruments in financial markets. Studies are generally conducted using the returns of retirement investment funds. Value at Risk (VaR) values of the daily returns of retirement companies were calculated under both the constant variance assumption and conditional heteroscedasticity (Akduğan and Akın, 2013).

Apart from the returns related to pension investment funds, modeling studies are also carried out with price data. The price of five Turkish life, nonlife and pension insurance shares quoted on Istanbul Stock Market (BIST) was estimated via ARIMA models (Kurt and Senel, 2018).

It is realized that various time series analysis methods are used to model time series data observed in a certain period of time for various known investment instruments such as stock market, stocks, foreign currency, gold, dollar, euro and oil, apart from pension investment funds. Stock market, foreign exchange, gold and petroleum returns are predicted with ARIMA, ARCH, GARCH and EGARCH models using Turkish weekly data of BIST 100 (Altuntaş and Çolak, 2015; Değirmenci and Akay, 2017). Holzner et al. (2022) for instance analyzed the impact of public pension expenditures, the assets of pension investment funds, and the benefits paid on macroeconomic volatility.

Results of modelling with ARIMA and deep learning were compared in stock price prediction (Karadağ, 2022). The data of the EREGL share, which is traded in the main metal market on the Borsa Istanbul index, was modeled with ARIMA models and deep learning models using long-short-term memory (LSTM), gated recurrent unit (GRU) and recurrent neural networks (RNN) algorithms (Erden, 2023).

Apart from modeling financial investment instruments, ARIMA models are also preferred in the actuarial literature in modeling time series consisting of premium prices recorded at a certain time. Bortner et al. (2014) compared linear regression and ARIMA in the estimation of calculations related insurance policies. ARIMA models were used in the prediction of life insurance premium production (Çetinkaya, 2019) and in the prediction of fire and natural disaster insurance premiums (Dilmen et al., 2022). Insurance Penetration Rate was modelled via ARIMA models in other studies, as well (Hafiz et al., 2021). Eşsiz and Ordu (2024) employed ARIMA method for the S&P 500 Index basket fund, known for its high-risk, high-return profile among pension mutual funds.

Time series frequently play a crucial role in statistics and economics. A time series consists of a sequence of measurements taken at regular time intervals. This type of analysis is widespread in various scientific fields, but governments commonly use it to forecast economic trends for organizations based on economic data. In addition, time series methods can be used to estimate retirement funds, which are significant financial indicators. Studies have been conducted in which pension investment funds are modeled with artificial neural networks (Onocak and Koç, 2018; Çemrek and Demir, 2021) and ARIMA (Louisa et al., 2022).

In this study, the behaviors of low, medium and high risk private pension investment funds belonging to a private pension company operating in Turkey, which are ALZ, AZS and AMZ, was examined with the help of time series analysis methods over a six-year period. Daily price, daily number of shares in circulation, daily number of people, daily total fund value and daily logarithmic return data of selected low, medium and high risk pension investment funds were converted into weekly average data. The behaviors of the weekly average values of the funds over time were examined graphically using time series analysis methods. The stationarity of the weekly average logarithmic return values of ALZ, AZS and AMZ funds was examined with Augmented Dickey-Fuller (ADF) test, which is one of the unit root tests, and the stationarity process was applied to non-stationary returns. Steady weekly average logarithmic return values were modeled with appropriate ARIMA models, and a one year forecast was performed and compared with real values.

A review of the literature reveals that investment instruments like stocks and gold are typically modeled using ARIMA, while the returns of pension investment funds are often modeled with artificial neural networks. In this study, we analyzed three different pension investment funds with varying risk levels over a 5-year period. This approach is believed to contribute to the literature by providing a comparative analysis of different pension investment funds with distinct risk profiles.

The remainder of the paper is organized as follows. In the Second Section, unit root tests that test stationarity, and ARIMA models, which are linear time series models, will be briefly summarized. In the Third Section, an application will be carried out including time series graphs, ACF and PACF graphs, stationarity tests of some variables related retirement funds such as daily price, daily number of shares in circulation, daily number of people, daily total fund value and daily logarithmic return data and modeling of logarithmic returns of the pension investment fund at three different risk levels with ARIMA. In the last Section, the concluding remarks will be given.

2. METHODOLOGY: TIME SERIES ANALYSIS METHODS

Time series analysis is the examination of data measured over a certain period of time using the mathematical, statistical and econometric methods. Initial analysis can be conducted with time series chart, ACF and PACF charts. The main purpose of time series analysis is to make predictions about the future by using the behaviors of the past data. In most time series methods, stationarity is one of the prerequisites for modeling. After stability is achieved, the modeling and prediction phase can be started.

2.1. Stationarity and Unit Root Test

The concept of stationarity is divided into strictly and weakly stationarity. In strictly stationarity, the distribution function of the series does not change over time. It is very difficult to ensure strictly stationarity in real data applications. In case of weakly stationarity, the expected value ($E(Z_t) = \mu$) and the variance ($V(Z_t) = \sigma^2$) of the Z_t time series are fixed. In addition, its covariance is independent of time ($Cov(Z_t, Z_{t+k}) = \gamma_k$) (Kadılar and Çekim, 2020). What is mentioned with stationarity is generally the concept of weak stationarity, and there is a basic assumption that financial return data is weakly stationary (Tsay, 2005).

One of the most frequently used methods in testing stationarity and determining the degree of difference is unit root tests. The unit root expression is based on testing the hypothesis that the root is equal to the unit value ($\phi_1 = 1$) in the AR(1) model. In the case of unit root, the model is stationary (Eğrioğlu and Baş, 2020).

The most commonly used unit root test is the ADF test (Dickey and Fuller. 1981). The ADF test is based on the reparameterization $\gamma_0 = 1 - \phi_1$. Constant-free and trend-free model ($\Delta x_t = \gamma_0 x_{t-1} + \sum_{k=1}^{m} \gamma_k \Delta x_{t-k} + \varepsilon_t$), the model containing constant term ($\Delta x_t = \beta_0 + \gamma_0 x_{t-1} + \sum_{k=1}^{m} \gamma_k \Delta x_{t-k} + \varepsilon_t$), and the model including the terms constant and trend ($\Delta x_t = \beta_0 + \beta_1 t + \gamma_0 x_{t-1} + \sum_{k=1}^{m} \gamma_k \Delta x_{t-k} + \varepsilon_t$), are defined. In these models, the hypothesis $H_0: \gamma_0 = 0$ is tested (Eğrioğlu and Baş, 2020).

2.2. ARIMA Models

ARIMA models, also known as Box-Jenkins models, are the most basic linear time series models. ARIMA models are the generalization of exponential smoothing methods and are expressed in their most general form as *ARIMA(p,d,q)(P,D,Q)s*, where *p* and *q* are the degrees of auto regression (AR) and moving average (MA) models, respectively. *d* shows the number of differences required for a non-stationary process to become stationary. Box-Jenkins models can be expressed in two different ways as non-seasonal *((ARIMA(p,d,q))* and seasonal *(ARIMA(p,d,q)(P,D,Q)s)* models. In seasonal models, *P* and *Q* shows the degrees of seasonal auto regression (SAR) and seasonal moving average (SMA) models, respectively. *D* shows the number of seasonal differences and s is the period (Kadılar and Çekim, 2020).

The general representation of the $ARIMA(p,d,q)(P,D,Q)_s$ model is given below in Equation (1).

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-\Phi_1 B^s-\Phi_2 B^{2s}-\cdots-\Phi_P B^{Ps})(1-B)^d(1-B^s)^D z_t$$
$$=(1-\theta_1 B-\theta_1 B^2-\cdots-\theta_q B^q)(1-\Theta_1 B^s-\Theta_2 B^{2s}-\cdots-\Theta_Q B^{Qs})\varepsilon_t \quad (1)$$

In Equation 1, z_t is a stationary time series, ε_t is the error term which is white noise, and the term $(1-B)^d$ shows the closed form of the d-order difference operation. The terms ϕ and θ denote the coefficients of the auto regression and moving average models, respectively, while the terms Φ and Θ are the coefficients of the seasonal models.
In non-seasonal models, *P* and *Q,* which are the degrees of seasonal terms, take the value 0 and Equation 1 turns into Equation 2 as follows.

$$(1-\phi_1 B-\phi_2 B^2-\cdots-\phi_p B^p)(1-B)^d z_t=(1-\theta_1 B-\theta_1 B^2-\cdots-\theta_q B^q)\varepsilon_t \quad (2)$$

ARIMA models are modeled with the a four-stage modeling method called the Box-Jenkins method. These are listed as determining the appropriate model for the time series, estimating the model, diagnostic control and finally the prediction process. According to the ACF and PACF graphs, the models such as auto regression model, moving average model or autoregressive moving average model, which may be suitable for the stationary time series, and its degrees from the significant lag numbers are decided. In the second stage, coefficient significance analysis is performed among possible models. Models whose coefficients are statistically insignificant are eliminated. Then, the diagnostic detection phase is started, which includes joint graphs of the original data and prediction data, confidence interval control, and analysis of whether the errors comply with the white noise process. Finally, among available models, the most suitable model is decided according to the model selection criteria or error criteria. Later predictions are carried out (Kadılar and Çekim, 2020).

3. APPLICATION: EXAMINATION AND FORECASTING OF LOW, MEDIUM AND HIGH RISK PENSION INVESTMENT FUNDS WITH TIME SERIES ANALYSIS METHODS

3.1. Low, Medium and High Risk Pension Investment Funds Data

Pension investment funds in the individual pension system are one of the most crucial components, because the system is built on the principle of accumulated amounts through pension investment funds. An individual pension investment fund comprises investment tools created to manage the contributions of participants within the individual pension accounts under the company's pension contract framework (Elveren, 2002).

According to the Regulation on the Principles Regarding the Establishment and Operations of Pension Investment Funds, when determining the types of funds, attention must be paid to whether the fund's name includes an expression that gives the impression of investment in a particular sector, asset group, or sectors. If such an expression is used, at least 80% of the fund's assets must consist of assets belonging to the sector, asset group, or sectors indicated in the fund's name. Otherwise, this expression cannot be used in the fund's name. Taking these conditions into account, the fund types that guide the practices, which are not restrictive but have a guiding nature, are as follows:

- Income Funds
- Money Market Funds
- Growth Funds
- Fund of Funds
- Contribution Fund
- Precious Metals Funds
- Specialized Funds
- Other Funds

The latest data announced by the Pension Monitoring Center on 04.07.2024 is as follows:

- Participants' Fund Amount: 969.8 billion TL
- State Contribution Fund Amount: 118.5 billion TL
- Total Number of Participants: 9,147,935 people

Private Pension Fund Purchase and Sale Platform (BEFAS) is an electronic platform operated by Takasbank that allows the sale and repurchase of pension investment fund shares by the fund founder pension company to the participants of other pension companies. Participants attending in the private pension system can examine the daily, monthly and annual returns of their funds and obtain information about the risk levels of the funds via BEFAS.

The pension investment fund data were obtained from the websites of the Pension Monitoring Center (EGM) and the Individual Pension Fund Trading Platform (BEFAS). Three pension funds with low, medium, and high risk levels were selected (https://www.egm.org.tr/fonlar/bireysel-emeklilik-fon-alim-satim-platformu-befas/befas/). The codes and descriptions of the funds analyzed in the study are provided below in Table 1

Codes and Names of Low, Medium, and High Risk Pension Investment Funds

Table 1

Fund Code	Fund Name
ALZ	Startup Pension Investment Fund
AZS	Standard Pension Investment Fund
AMZ	Gold Pension Investment Fund

The study utilized daily data on the prices, number of shares in circulation, number of participants, total fund value, and logarithmic returns of the three selected low, medium, and high risk pension investment funds over a six-year period from 2018 to 2023. The daily data were recorded between January 2, 2018, and December 29, 2023, faced issues of missing observations due to weekends, public holidays, and official holidays. Since this issue of missing data arose, the solutions considered were either imputation methods such as replacing missing values with the mean or the previous value to complete the data and work with daily data, or calculating weekly averages to work with weekly data. The estimated missing data must be limited to 20% of the total data. Therefore, the idea of working with daily data was abandoned, and weekly average time series were obtained. Based on the generalization that a year consists of 52 weeks, the data for the 53rd week of 2020, which had 53 weeks, was excluded from the analysis. Missing (empty, 0) or erroneous (negative) observations were corrected by substituting the previous observation or the average values. As a result, a time series of 312 weeks was reached. The dataset includes the following columns for each fund: Date, Price, NumofShare (Number of Shares in Circulation), Numofpeople (Number of Participants), TotalFund (Total Fund Value), and Return (Log Return). Using the Price (P_t) data, logarithmic returns were calculated with the formula $r_t = ln\frac{P_t}{P_{t-1}}$.

The daily data on prices, number of shares in circulation, number of participants, total fund value, and logarithmic returns recorded over a six-year period for low, medium, and high risk pension investment funds were used to examine and interpret the general behaviors of the funds using time series analysis methods. The daily logarithmic return data for all three funds over the first 260 weeks of the five-year period from 2018 to 2022 were used to model and forecast with ARIMA. The last 52 weeks of 2023 were used to test the modeling results.

3.2. Examination of Low, Medium and High Risk Pension Investment Funds with Time Series Analysis Methods

Numerical calculations necessary to examine the behavior of pension investment fund variables over time and forecasting returns using time series methods were performed in the R Studio environment (http://www.rstudio.com/).

Firstly, descriptive statistics of the weekly average daily price, number of shares in circulation, number of participants, total fund value, and logarithmic return variables for low, medium, and high risk pension investment funds covering the period from 2018 to 2023, a 6-year period, were examined and presented in Table 2.

Descriptive Statistics of Price, Number of Share, Number of People, Total Fund and Logarithmic Return for ALZ, AZS and AMZ

Table 2

Statistics for the Price								
	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Variance**	**Std. Dev.**
ALZ	0.01112	0.01446	0.01728	0.01883	0.02231	0.03463	3.31E-05	0.005757
AZS	0.01120	0.01478	0.01865	0.02289	0.02583	0.05444	0.000129	0.011337
AMZ	0.01710	0.02704	0.04853	0.06754	0.10243	0.21212	0.002635	0.051327
Statistics for the Number of Share								
	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Variance**	**Std. Dev.**
ALZ	1.90E+09	2.16E+09	2.50E+09	3.26E+09	4.24E+09	7.29E+09	2.31E+18	1.52E+09
AZS	1.88E+09	2.03E+09	2.30E+09	5.09E+09	8.22E+09	1.13E+10	1.30E+19	3.60E+09
AMZ	1.98E+10	4.38E+10	7.81E+10	7.76E+10	1.08E+11	1.43E+11	1.31E+21	3.6199E+10
Statistics for the Number of People								
	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Variance**	**Std. Dev.**
ALZ	72103	90955	98367	103593	120443	140104	3.02E+08	17372.95
AZS	10460	12398	16180	21218	32435	33850	91399196	9560.293
AMZ	38382	77214	116634	169533	287977	437858	1.4E+10	118258.0
Statistics for the Total Fund								
	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Variance**	**Std. Dev.**
ALZ	26021256	33044634	39161913	67894026	85029172	2.5E+08	2.91E+15	53933074
AZS	24387671	31145908	37246222	1.48E+08	2.52E+08	4.39E+08	2.36E+16	1.54E+08
AMZ	3.38E+08	1.19E+09	3.75E+09	6.92E+09	1.11E+10	3.04E+10	5.73E+19	7.57E+09
Statistics for the Logarithmic Return								
	Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max**	**Variance**	**Std. Dev.**
ALZ	-0.00002	0.000497	0.00067	0.000771	0.000865	0.004314	2.01E-07	0.000449
AZS	-0.01262	-0.000735	0.000897	0.000935	0.002836	0.011805	1.19E-05	0.003456
AMZ	-0.05492	-0.001636	0.001647	0.001904	0.004675	0.042682	5.24E-05	0.007240

According to Table 2, it is observed that as the risk level increases, the average price, variance, and standard deviation values increase as expected. When the number of shares in circulation is examined, it is seen that high risk pension investment funds have the highest average value. It is noted that ALZ, AZS, and AMZ funds are the most preferred pension investment funds of the company. Following high risk pension investment funds, low risk pension investment funds are preferred. Total fund value is directly related to the number of shares and participants. It is observed that the average total fund value of high risk pension investment funds is higher than the others. Return value, which is one of the most important evaluation criteria for financial investment instruments, is examined as logarithmic return for these three funds. Since variance and standard deviation are the most basic risk measures, as expected, ALZ, being the low risk pension investment fund, has the lowest standard deviation value, while AMZ has the highest standard deviation value. As the risk level increases, the average logarithmic return values also increase. Time series plots ACF and PACF of the variables are also interpreted. ACF and PACF plots are provided in Appendix 1. Logarithmic return values for each fund are modeled using ARIMA models, and the forecasted values are compared with the actual data values. Error values such as MAE, RMSE, MAPE are computed.

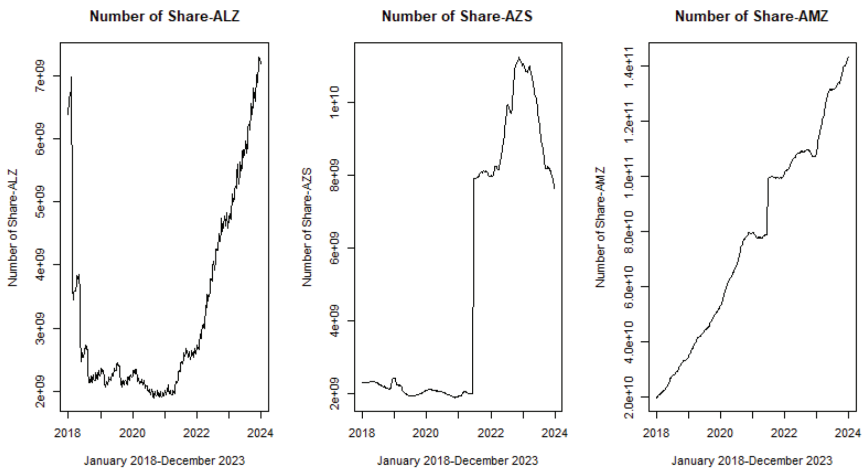
Graphs of Time Series for the Price of ALZ, AZS and AMZ Retirement Funds

Figure 1

The 312-week average daily price time series of low, medium, and high risk pension investment funds were analyzed using time series plots. The price plots of the three pension investment funds are provided in Figure 1. In all three funds, an increasing trend in prices from the first week of 2018 to the last week of 2023 is observed. In the low risk pension investment fund ALZ, almost no random fluctuations have been observed. There is a fairly smooth increasing trend. However, in the medium and high risk funds, random movements are observed along with the increasing trend.

Graphs of Time Series for the Number of Share of ALZ, AZS and AMZ Retirement Funds

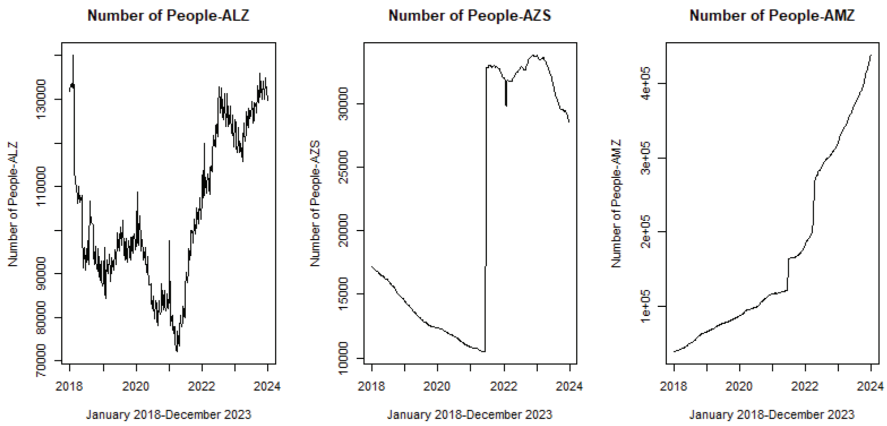
Figure 2



The time series plots of the weekly average number of shares in circulation for low, medium, and high risk pension investment funds over the 6-year period are provided in Figure 2. The number of shares in circulation, the number of participants, and the total fund value are closely related variables. Therefore, it would be beneficial to examine the time series plots of the number of participants provided in Figure 3 and the total fund values provided in Figure 4 together.

Graphs of Time Series for the Number of People of ALZ, AZS and AMZ Retirement Funds

Figure 3

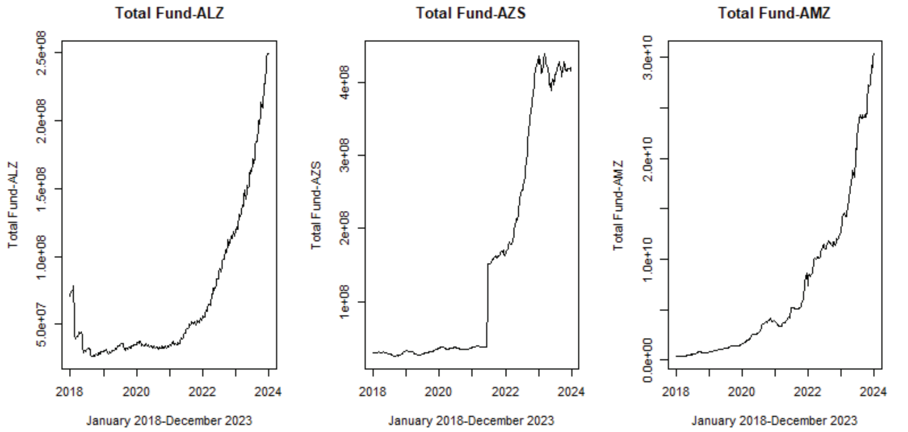


When examining the graphs of the number of shares in circulation and the number of participants for low risk pension investment funds, a sharp decreasing trend until 2019 and a rapidly increasing trend after 2021 are observed. However, the total fund value for low risk pension investment funds decreased between 2018 and 2019, remained stable between 2019 and 2021, and showed an increasing trend after 2021.

The number of shares in circulation for the medium risk pension investment fund AZS remained constant until the middle of 2021. After 2021 it started to increase. As directly related to the number of shares, the number of participants decreased from 2018 to the middle of 2021 and then showed a significant increase from the middle of 2021 onwards. It is observed that the total fund value is closely related to the number of shares in circulation, and their graphs are very similar.

Graphs of Time Series for the Total Fund of ALZ, AZS and AMZ Retirement Funds

Figure 4

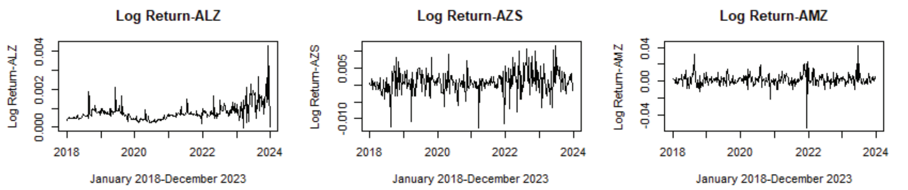


The number of shares in circulation, number of participants, and total fund size for the high risk pension investment fund AMZ shows an increasing trend from 2018 to the end of 2023.

The behaviors of the logarithmic returns for the three funds over the 6-year period are examined using the time series plots. The results are provided in Figure 5.

Graphs of Time Series for the Logarithmic Return of ALZ, AZS and AMZ Retirement Funds

Figure 5



When examining the graphs of the logarithmic return data for the funds, no trend or seasonal fluctuations are observed in any of the three funds. In the low risk pension investment fund ALZ, a slowly increasing trend is realized throughout 2023. However, the logarithmic return data for medium and high risk funds follow a stationary trend. The ACF and PACF plots provided in Appendix 1 for the logarithmic returns also support this stationary result.

3.3. Modeling Weekly Average Logarithmic Returns of Low, Medium, and High Risk Pension Investment Funds with *ARIMA(p,d,q)(P,D,Q)s* Models

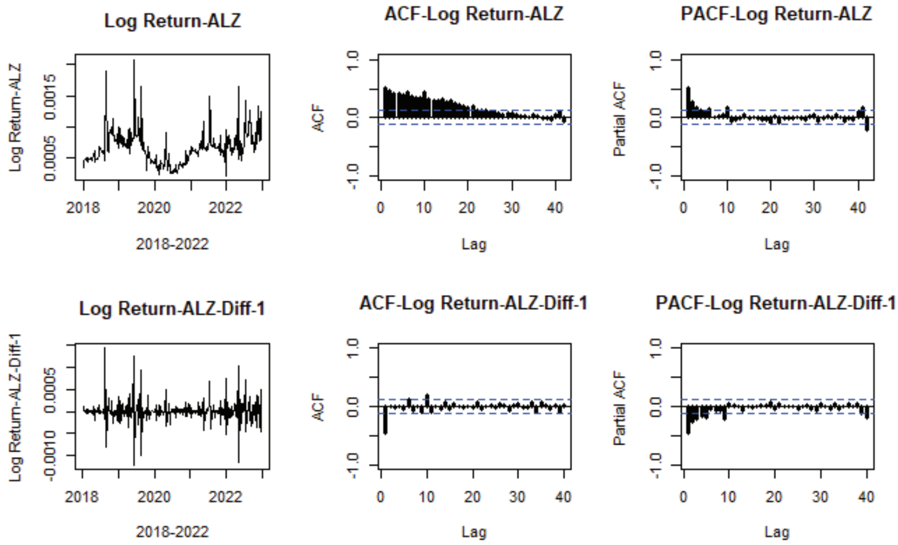
For ARIMA models, the arima() function under the "forecast" package in the R Studio, along with other packages such as "rJava," "XLConnect," "ggplot2," "zoo," "lmtest," and "readxl," are utilized. Weekly average logarithmic returns of low, medium, and high risk pension investment funds have been modeled using time series methods with a total of 260 weeks of data covering the first week of January 2018 to the last week of December 2022. ARIMA models have been used to make 52-week, or one-year, forecasts. The 52 weeks of data observed from the first week of January to the last week of December 2023 have been selected as out-of-sample test data for comparing forecast values to actual data. It should be noted that while 6 years of data were examined in the previous section to analyze the general behaviors of pension investment funds, in this section, only 5 years of data are used for forecasting purposes.

3.3.1. Modeling the Average Logarithmic Return Time Series of the Low Risk ALZ Pension Investment Fund with ARIMA Models

The 260-week time series plot of the logarithmic returns and the differentiated time series plot for the low risk ALZ pension investment fund from 2018 to 2022, along with the ACF and PACF plots, are provided in Figure 6 below. When examining the ACF-PACF plots of the original time series, it is observed that the time series is non-stationary. However, when a difference is taken, the series becomes stationary.

Time Series, ACF, and PACF Plots for the Weekly Average Logarithmic Returns of Original and Differenced Low Risk ALZ Pension Investment Fund for the Period 2018-2022

Figure 6



After graphical examination, the stationarity of the series is examined using the ADF test, one of the most commonly used unit root tests, and the results are provided in Table 3 below.

Stationarity Test Results for the Weekly Average Logarithmic Time Series of ALZ Pension Investment Fund

Table 3

	ADF Unit Root Test - Original Data		
	Draft	**Draft and Trend**	**None**
p-value	< 2.2e-16	< 2.2e-16	1.435e-14
ADF Test Statistics	-5.8643	-5.9322	-1.6269
Adjusted R^2	0.3006	0.3001	0.2143
	ADF Unit Root Test - Differenced Data		
	Draft	**Draft and Trend**	**None**
p-value	< 2.2e-16	< 2.2e-16	< 2.2e-16
ADF Test Statistics	-18.0167	-17.9814	-18.0484
Adjusted R^2	0.7493	0.7463	0.7473
MacKinnon Critical Value (%1)	-3.46	-3.99	-2.58
MacKinnon Critical Value (%5)	-2.88	-3.43	-1.95
MacKinnon Critical Value (%10)	-2.57	-3.13	-1.62

Although the graphical examination of the original series suggests non-stationarity according to the unit root test results provided in Table 3, the time series of the ALZ fund is stationary even without differencing. According to the unit root test results in Table 3, the ADF test statistic values for all models are smaller than the critical values and $p<0.05$, rejecting the hypothesis that there is a unit root in the series. Since the adjusted R^2 values, which indicate goodness of fit, increase after differencing, it is believed that differencing improves the modeling of the series.

Since the ACF plot in Figure 6 appears to decay faster than the PACF plot, it is considered that moving average models might be more appropriate. As the first value in the ACF plot exceeds the confidence limit, models with $q=1$ degrees are established. Within the scope of first-order moving average (MA(1)) models, ARIMA(0,1,1)(0,0,0), ARIMA(0,1,1)(0,0,1)[52], ARIMA(0,1,1)(1,0,0)[52], and ARIMA(0,1,1)(1,0,1)[52] models are established. Although visually, the moving average model seems more suitable for the logarithmic return time series, some first-order autoregressive (AR(1)) models and some autoregressive moving average (ARMA(1,1)) models are also tested. Some of the tested models include ARIMA(1,1,0)(0,0,0)[52], ARIMA(1,1,0)(0,0,1)[52], ARIMA(1,1,0)(1,0,0)[52], ARIMA(1,1,1)(0,0,0)[52], ARIMA(1,1,1)(1,0,0)[52], and ARIMA(1,1,1)(0,0,1)[52]. In some of these models, the coefficients were not statistically significant ($p>0.05$). Among the models with statistically significant coefficients, the ARIMA(0,1,1)(0,0,0) and ARIMA(0,1,1)(10,0)[52] models, which have mismatched real data and forecast data graphs, are eliminated. The adequacy of errors for the last two suitable models is also examined. When the ACF and PACF plots of the errors are compared, it is observed that the lag values of the errors for the ARIMA(1,1,1)(1,0,1)[52] model fall within the confidence limits. The adequacy of errors to the white noise process is tested with the Box-Ljung test, and although the test result does not accept the adequacy to the white noise process, according to the graphical examination, the errors can be considered adequate for the white noise process ($p<0.05$). Additionally, when a comparison is made based on information criteria such as AIC, AICC, and BIC, and error values such as RMSE, MAE, MPE, MAPE, and MASE, it is observed that the ARIMA(1,1,1)(1,0,1)[52] model has the smallest information criteria and error values.

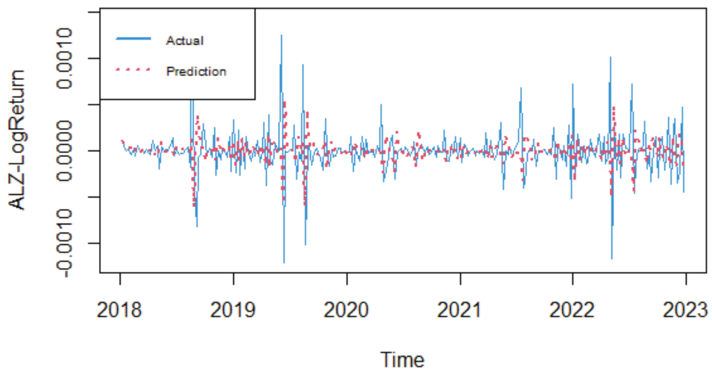
Results of ARIMA Models for Low Risk ALZ Retirement Fund

Table 4

Models	AIC	AICC	BIC	RMSE	MAE	MASE
ARIMA (0,1,1)(0,0,0)	-3509.81	-3509.71	-3499.15	0.0002625092	0.0001473876	0.6746636
ARIMA (0,1,1)(1,0,0)[52]	-3510.66	-3510.50	-3496.45	0.0002605451	0.0001491751	0.6828458
ARIMA (1,1,0)(0,0,1)[52]	-3388.99	-3388.84	-3374.78	0.0003324907	0.0002030711	0.9295536
ARIMA (1,1,1)(1,0,1)[52]	-3567.96	-3567.63	-3546.65	0.0002305417	0.0001345736	0.6160080

The joint plot of ALZ log-return data and the log-return data predicted by the ARIMA(1,1,1)(1,0,1)[52] model

Figure 7



According to Figure 7 and Table 4, among the models tested for the low risk ALZ pension investment fund, the most suitable model is considered to be the ARIMA(1,1,1),(1,0,1)[52] model, and the model results are provided below in Table 5.

Parameter Estimation of ARIMA(1,1,1)(1,0,1)[52] Model for ALZ Retirement Fund

Table 5

Variables	Estimate	Std. Error	z value	Pr(>|z|)
AR (1)	-0.4653	3.0792e-02	-15.1120	< 2.2e-16*
MA (1)	-1.000	9.9765e-03	-100.2334	< 2.2e-16*
SAR (1)	-0.6798	1.4785e-01	-4.5976	4.274e-06*
SMA(1)	0.5686	1.6001e-01	3.5534	0.0003802*

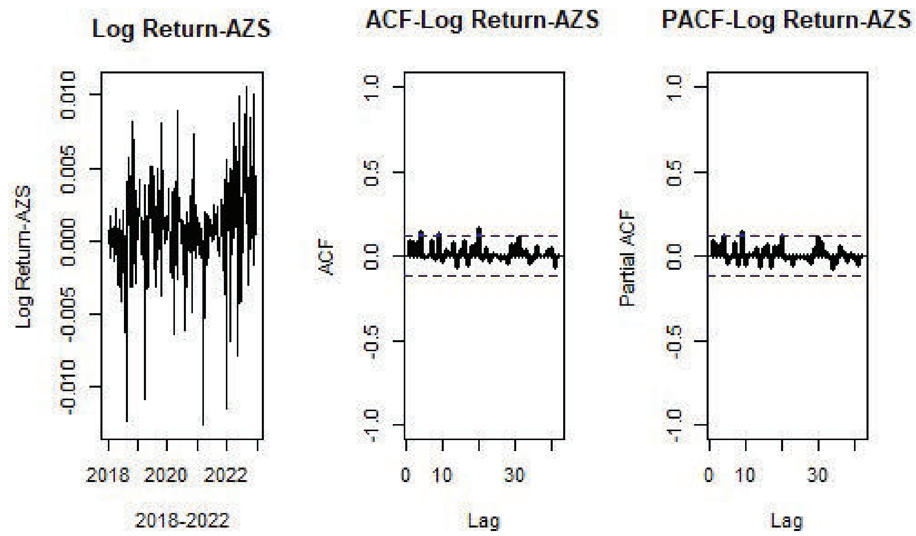
* Statistically significant coefficient (α=0,05)

3.3.2. Modeling the Average Logarithmic Return Time Series of the Medium Risk AZS Pension Investment Fund with ARIMA Models

The graph of the 260-week logarithmic return time series of the medium risk AZS pension investment fund, along with the ACF and PACF plots, is provided in Figure 8 below. Upon examining the ACF and PACF plots of the AZS fund time series, it is observed that this time series is stationary.

Time Series, ACF, and PACF Plots for the Weekly Average Logarithmic Returns of Medium Risk AZS Pension Investment Fund for the Period 2018-2022

Figure 8



The stationarity of the AZS pension investment fund is examined with the ADF test, and the results are presented in Table 5 below. The ADF test results given in Table 6 also support the results obtained in Figure 8.

Stationarity Test Results for the Weekly Average Logarithmic Time Series of AZS Pension Investment Fund

Table 6

	ADF Unit Root Test		
	Draft	**Draft and Trend**	**None**
p-value	< 2.2e-16*	< 2.2e-16*	< 2.2e-16*
ADF Test Statistics	-10.0674	-10.6069	-9.3688
Adjusted R^2	0.4547	0.4698	0.4325
MacKinnon Critical Value (%1)	-3.44	-3.98	-2.58
MacKinnon Critical Value (%5)	-2.87	-3.42	-1.95
MacKinnon Critical Value (%10)	-2.57	-3.13	-1.62

* Statistically significant (α=0,05)

According to the unit root test results shown in Table 6, the AZS pension investment fund is a stationary time series. The ADF test statistics values for all models in Table 6 are smaller than the critical values, and since $p<0.05$, the hypothesis suggesting a unit root in the series is rejected. Although the adjusted R^2 values are not very high, they are acceptable, and modeling can proceed without any adjustments to the data.

AR and MA effects could not be fully determined from the ACF and PACF plots; hence, autoregressive moving average (ARMA) models were primarily tested. In addition to ARMA models, several AR and MA models were also tried. The tested models include ARIMA(1,0,1)(0,0,0), ARIMA(1,0,1)(1,0,1)[52], ARIMA(1,0,1)(1,0,0)[52], ARIMA(1,0,1)(0,0,1)[52], ARIMA(0,1,1)(1,0,0)[52], ARIMA(0,0,1)(0,0,0), ARIMA(0,0,1)(0,0,1)[52], ARIMA(1,0,0)(0,0,1)[52], ARIMA(1,0,0)(0,0,0), ARIMA(1,0,0)(1,0,0)[52], and ARIMA(1,0,0)(0,0,1)[52]. In some of these models, the coefficients were not statistically significant (p > 0.05). Among the models with statistically significant coefficients, ARIMA(1,0,1)(0,0,0) and ARIMA(0,1,1)(1,0,0)[52] showed consistent alignment between real data and forecast data, and their errors seemed to adhere to the ACF-PACF graphs, indicating adherence to a white noise process. The adherence of errors to the white noise process was tested using the Box-Ljung test (p > 0.05). The information criteria such as AIC, AICC, and BIC, as well as error metrics including RMSE, MAE, MPE, MAPE, and MASE, for these two models are presented in Table 7 below. According to Table 7, the ARIMA(1,0,1)(0,0,0) model has lower information criteria values.

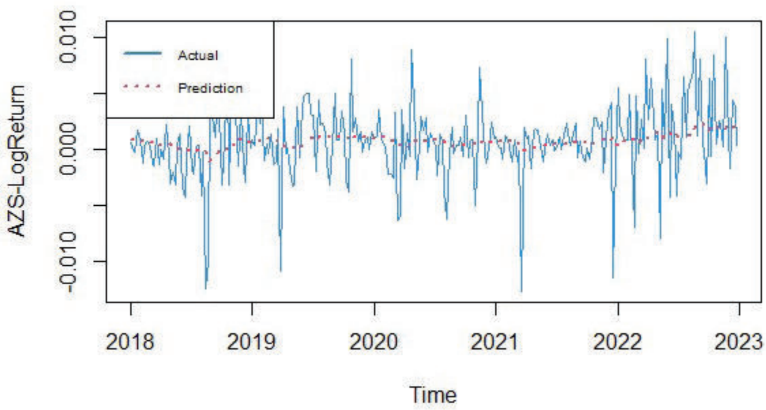
Results of ARIMA Models for Medium Risk AZS Retirement Fund

Table 7

Models	AIC	AICC	BIC	RMSE	MAE	MASE
ARIMA (1,0,1)(0,0,0)	-2219.01	-2218.86	-2204.77	0.003239225	0.002316880	0.7049076
ARIMA (0,1,1)(1,0,0)[52]	-2208.58	-2208.43	-2194.36	0.003327992	0.002318368	0.7053177

The joint plot of AZS log-return data and the log-return data predicted by the ARIMA(1,0,1)(0,0,0) model

Figure 9



According to Table 6 and Figure 9, the most suitable model among the ones tested for the medium risk AZS pension investment fund is accepted to be the ARIMA(1,0,1),(0,0,0) model. The model results are exhibited below in Table 8.

Parameter Estimation of ARIMA(1,0,1)(0,0,0) Model for AZS Retirement Fund

Table 8

Variables	Estimate	Std. Error	*z* value	Pr(>|*z*|)
AR (1)	0.9656	0.05782947	16.6972	< 2.2e-16*
MA (1)	-0.9190	0.08196854	-11.2116	< 2.2e-16*

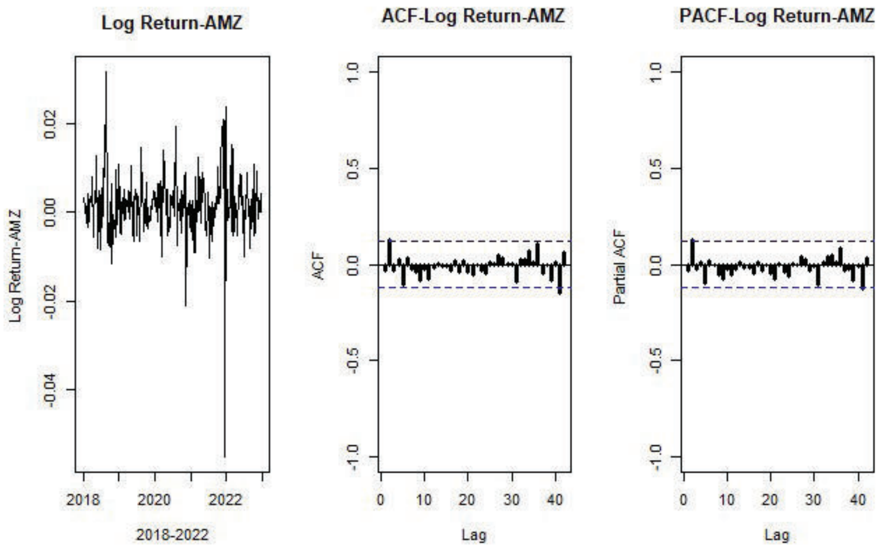
* Statistically significant coefficient (α=0,05)

3.3.3. Modeling the Average Logarithmic Return Time Series of the High Risk AMZ Pension Investment Fund with ARIMA Models

The 260-week logarithmic return time series plot and ACF and PACF plots for the high risk AMZ pension investment fund from 2018 to 2022 are displayed in Figure 10 below. Upon examination of the ACF and PACF plots of the AMZ fund, it is found out that this time series is stationary.

Time Series, ACF, and PACF Plots for the Weekly Average Logarithmic Returns of High Risk AMZ Pension Investment Fund for the Period 2018-2022

Figure 9



The stationarity of the AMZ pension investment fund is examined using the ADF test, and the results are provided in Table 9 below. The ADF test results given in Table 8 also support the results obtained in Figure 9.

Stationarity Test Results for the Weekly Average Logarithmic Time Series of AMZ Pension Investment Fund

Table 9

	ADF Unit Root Test		
	Draft	**Draft and Trend**	**None**
p-value	< 2.2e-16*	< 2.2e-16*	< 2.2e-16*
ADF Test Statistics	-10.0409	-10.0219	-9.3257
Adjusted R^2	0.5235	0.5235	0.5037
MacKinnon Critical Value (%1)	-3.44	-3.98	-2.58
MacKinnon Critical Value (%5)	-2.87	-3.42	-1.95
MacKinnon Critical Value (%10)	-2.57	-3.13	-1.62

* Statistically significant coefficient (α=0,05)

According to the unit root test results given in Table 9, the AMZ pension investment fund exhibits a stationary time series. For all models, the ADF test statistic values are smaller than the critical values, and $p < 0.05$, rejecting the hypothesis of a unit root in the series. The adjusted R^2 values are around 50%, which is an acceptable limit for modeling the data, and modeling can proceed without any adjustments to the data.

Similar to medium risk pension investment funds, ARMA models were primarily attempted due to the inability to accurately determine the AR and MA effects from the ACF and PACF plots. In addition to ARMA models, a few AR and MA models were also tested. Since the 2nd lags are significant in the ACF and PACF plot, 2nd order models will also be attempted. The models tested include ARIMA(1,0,1)(0,0,0), ARIMA(1,0,1)(1,0,1)[52], ARIMA(1,0,1)(1,0,0)[52], ARIMA(1,0,1)(0,0,1)[52], ARIMA(2,0,2)(0,0,0), ARIMA(2,0,2)(1,0,1), ARIMA(2,0,1)(0,0,0), and ARIMA(1,0,2)(0,0,0). In some of these models, all coefficients were not statistically significant, while in others, seasonal coefficients were not statistically significant (p>0.05). Among the models with statistically significant coefficients, the ARIMA(1,0,1)(0,0,0), ARIMA(2,0,2)(1,0,1), and ARIMA(2,0,1)(0,0,0) models exhibited consistent real data and prediction data graphs, and their errors were found to be suitable according to the ACF-PACF plots. The adequacy of errors to the white noise process was tested using the Box-Ljung test (p>0.05). Information criteria such as AIC, AICC, and BIC, as well as error values such as RMSE, MAE, MPE, MAPE, and MASE for these three models, are provided below in Table 10.

Results of ARIMA Models for High Risk AZS Retirement Fund

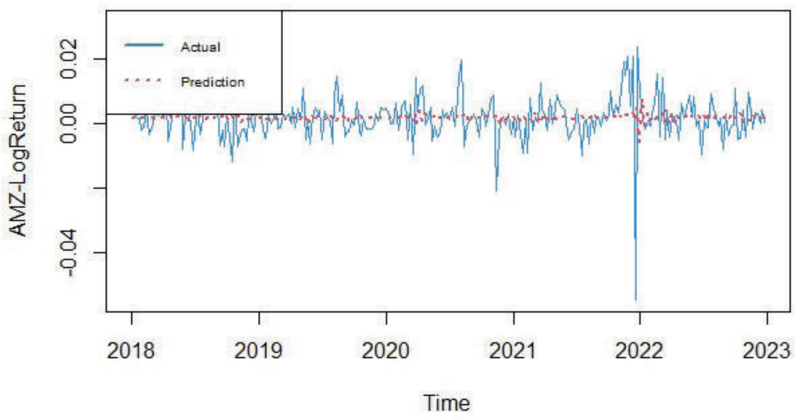
Table 10

Models	AIC	AICC	BIC	RMSE	MAE	MASE
ARIMA (1,0,1)(0,0,0)	-1824.24	-1824.08	-1810.00	0.007106399	0.004668171	0.6700102
ARIMA (2,0,1)(0,0,0)	-1823.66	-1823.43	-1805.86	0.007116759	0.004684180	0.6779852
ARIMA (2,0,2)(1,0,1)[52]	-1818.12	-1817.54	-1789.63	0.007110260	0.004676235	0.6768353

According to Table 10, it is observed that the ARIMA(1,0,1)(0,0,0) model has lower information criteria.

The joint plot of AMZ log-return data and the log-return data predicted by the ARIMA(1,0,1)(0,0,0) model

Figure 10



According to Table 10 and Figure 10, among the models tested for the high risk AMZ pension investment fund, the most suitable model, similar to the medium risk pension fund AZS, is considered to be the first-order autoregressive moving average model, ARIMA(1,0,1),(0,0,0), and the model results are presented below in Table 11.

Parameter Estimation of ARIMA(1,0,1)(0,0,0) Model for AMZ Retirement Fund

Table 11

Variables	Estimate	Std. Error	z value	Pr(>|z|)
AR (1)	-0.8484	0.13372459	-6.3441	2.238e-10 *
MA (1)	0.7856	0.15472554	5.0775	3.824e-07 *

* Statistically significant coefficient (α=0,05)

3.4 Comparison of Forecast Results Obtained with ARIMA Models for Low, Medium, and High Risk Pension Investment Funds

The ARIMA models, as stated in the previous section and detailed in Tables 5, 8, and 11, were used to forecast 52-week (one-year) log returns for low, medium, and high risk pension investment funds. To compare the performance of forecast values, the data from the first 260 weeks were used to determine the models, while the 52-week data observed from the first week of January 2023 to the last week of December was designated as out-of-sample test data to compare forecast values with actual data. Various comparison criteria such as Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) were calculated using the forecast values from the models and the actual values for the three funds. The results are presented below in Table 12.

Comparison Criteria of ARIMA Models for Low, Medium and High Risk Pension Investment Funds

Table 12

	MSE	RMSE	MAE	MAPE
ALZ	2.191588e-06	0.001480401	0.001292598	1.792119
AZS	1.438117e-05	0.003792251	0.003034463	1.266943
AMZ	5.532536e-05	0.007438102	0.004352099	1.842569

When examining the error values, it is observed that forecasts obtained with the ARIMA models determined for all three funds have low MSE values. If we compare across risk levels, it is evident that for low risk pension investment funds, ARIMA models have lower MSE, RMSE, MAE, and MAPE values, indicating that they compute forecast values closer to real data.

4. CONCLUSION

In this study, the behavior of private pension investment funds in Turkey were examined with ARIMA models. For this purpose three private pension investment funds, one low risk (ALZ), one medium risk (AZS) and the other high risk (AMZ), belonging to a private pension company operating in Turkey were analyzed. Weekly return of the data for the time period of 02.01.2018-29.12.2023, covering a 6-year period, was used. The stationarity of the weekly average logarithmic return values of ALZ, AZS, and AMZ funds was assessed using unit root tests, and non-stationary returns were adjusted to achieve stationarity. The stabilized weekly average logarithmic return values were then modeled using suitable ARIMA models. A one-year forecast was generated and compared with actual values.

When the descriptive statistics are examined, it was seen that as the risk level increases, the average logarithmic return values increase as well. It was observed that the average total value of high risk pension investment funds is higher than that of others. The return value, a crucial evaluation criterion for financial investment instruments, was analyzed as the logarithmic return for these three funds.

Upon examining the number of shares in circulation and the number of participants for low risk pension investment funds, a sharp declining trend until 2019 followed by a rapid increase after 2021 was observed. However, the total fund value for low risk pension investment funds decreased between 2018 and 2019, remained stable from 2019 to 2021, and then exhibited an upward trend after 2021.

While analyzing the average logarithmic return time series of the low risk (ALZ) pension investment fund with ARIMA Models, the ACF-PACF plots of the original time series have been examined. It is evident that the series is non-stationary. However, applying differencing to the series renders it stationary. According to modeling results, it is observed that the ARIMA(1,1,1)(1,0,1) model has the smallest information criteria and error values.

When analyzing the average logarithmic return time series of the medium risk (AZS) pension investment fund, it was observed that this time series is stationary. Among the models with statistically significant coefficients, ARIMA(1,0,1)(0,0,0) and ARIMA(0,1,1)(1,0,0)[52] demonstrated consistent alignment between actual data and forecast data. Additionally, their errors appeared to follow the ACF-PACF graphs, suggesting they conformed to a white noise process. Also the ARIMA(1,0,1)(0,0,0) model had lower information criteria values.

When analyzing the average logarithmic return time series of the high risk (AMZ) pension investment fund with ARIMA Models, it was noted that this time series is stationary. The ARIMA(1,0,1)(0,0,0), ARIMA(2,0,2)(1,0,1), and ARIMA(2,0,1)(0,0,0) models exhibited consistent real data and prediction data graphs, and their errors were found to be suitable according to the ACF-PACF plots. Furthermore, ARIMA(1,0,1)(0,0,0) model has lower information criteria.

When examining the error values, it was observed that the forecasts obtained using the ARIMA models for all three funds have low MSE values. Comparing across risk levels, it is evident that for low risk pension investment funds, ARIMA models exhibit lower MSE, RMSE, MAE, and MAPE values, indicating that they generate forecast values closer to the actual data.

When the literature is examined, it is seen that investment instruments such as stocks and gold are generally modeled with ARIMA. The returns of pension investment funds are generally modeled with artificial neural networks. From this perspective, it is thought that this study contributes to the literature in this respect. Eşsiz and Ordu (2024), which is a similar study to this study, worked for 1 fund and a 3-year period. In this study, we studied three different funds with different risk levels and a period of 5 years.

This study is open to improvement in various aspects. This study can be improved by modeling logarithmic returns and by predicting with artificial neural networks and different time series analysis methods such as VAR, ARCH, GARCH models and the results can be compared with ARIMA results.

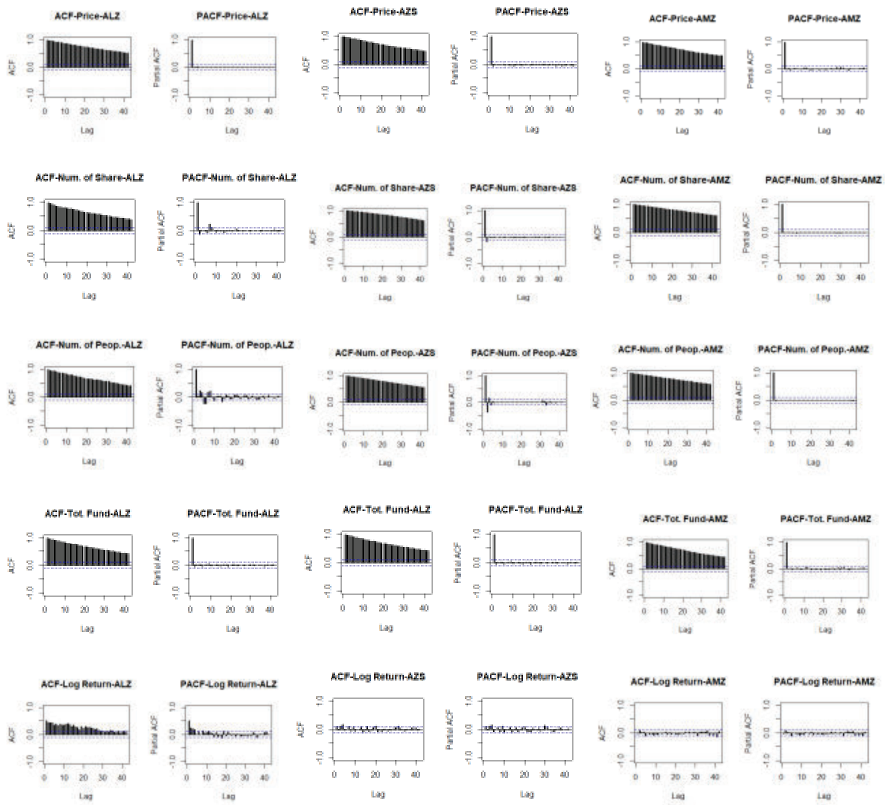
REFERENCES

1. Altuntaş, S. T. and Çolak, F. D. (2015). BİST-100 endeksinde volatilitenin modellenmesi ve öngörülmesinde ARCH modelleri. İstanbul Üniversitesi İşletme Fakültesi İşletme İktisadı Enstitüsü Yönetim Dergisi, 26(79), 208-223.
2. Akduğan, U. and Akın, Y. K. (2013). Parametrik Riske Maruz Değer Hesaplamasında Volatilitenin Modellenmesi: Türkiye'de Emeklilik Yatırım Fonları Üzerine Bir Uygulama Volatility Modelling in Parametric Value at Risk Calculation: An Application on Pension Funds in Turkey. In İnternational Conference on Eurasıan Economies (pp. 608-616).
3. BEFAS, Pension Fund Trading Platform., https://www.tefas.gov.tr/ (June, 2024).
4. Bortner, D., Pulliam, C. and Yam, W. (2014). Trend analysis for quarterly insurance time series. University of California Santa Balbara. Department of Statistics and Applied Probability, 1, 1-2.
5. Çemrek, F. and Demir, Ö. (2021). Forecasting the returns of pension investment funds in Turkey with artificial neural network. Cumhuriyet Science Journal, 42(4), 942-950.
6. Çetinkaya, T. (2019). Hayat sigortası prim üretimlerini tahminleme yöntemlerini karşılaştırarak gelecek yıllar prim üretimini tahminleme (Doctoral dissertation, Marmara Universitesi (Turkey)).
7. Dağlar, H. (2007). Kurumsal Yatırımcılar Olarak Emeklilik Yatırım Fonları ve Performanslarının Değerlendirilmesi, Türkiye Bankalar Birliği, Yayın No:249, Burdur.
8. Değirmenci, N. and Akay, A. (2017). Finansal verilerin ARIMA ve ARCH modelleriyle öngörüsü: Türkiye örneği. Eskişehir Osmangazi Üniversitesi İktisadi ve İdari Bilimler Dergisi, 12(3), 15-36.
9. Demirci, A. and Şen, A. (2006). Bireysel Emeklilik Sistemi, Maliye Hesap Uzmanları Derneği, Ankara.
10. Dickey, D.A. and Fuller W.A. (1979), Distributions of the Estimators For Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, 75, 427–431.
11. Dilmen, B., Gencer, Ş., Arıkel, F., Kayır, Ş. and Erdemir, Ö. K. (2022). Yangın ve doğal afet sigortası priminin Box-Jenkins modelleri ve yapay sinir ağları ile tahmin edilmesi. İstatistikçiler Dergisi: İstatistik ve Aktüerya, 15(2), 60-71.
12. Eğrioğlu E. and Baş E. (2020). Zaman serileri ve öngörü yöntemleri (R uygulamalı), Atlas Akademik Basım Yayın Dağıtım Tic. Ltd. Şti., 1-270.

13. Elveren, A.H. (2002). Bireysel Emeklilik Sistemi ve Emeklilik Şirketleri, Bireysel Emeklilik Sistemi ve Son Gelişmeler Semineri, 19, İstanbul.
14. Erden, C. (2023). Derin Öğrenme ve ARIMA Yöntemlerinin Tahmin Performanslarının Kıyaslanması: Bir Borsa İstanbul Hissesi Örneği. Yönetim Ve Ekonomi Dergisi, 30(3), 419-438. https://doi.org/10.18657/yonveek.1208807
15. Eşsiz, S. and Ordu, M. (2024). A Univariate Time-Series Analysis For Estimating The Daily Price Of The S&P 500 Fund Basket Pension Investment Fund, Hodja Akhmet Yassawi 8th International Congress On Scientific Research. Konya, Türkiye.
16. Hafiz, U. A., Salleh, F., Garba, M. and Rashid, N. (2021). Projecting insurance penetration rate in Nigeria: An ARIMA approach. Revista Geintec-Gestao Inovacao E Tecnologias, 11(3), 63-75.
17. Holzner, M., Jestl, S. and Pichler, D. (2022). Public and private pension systems and macroeconomic volatility in OECD countries. Scottish Journal of Political Economy, 69(2), 131-168.
18. Kadılar, C. and Öncel Çekim H. (2020), SPSS ve R uygulamalı zaman serileri analizine giriş, Seçkin Yayıncılık, Ankara, 1-379.
19. Karadağ, K. (2022). Hibrit derin öğrenme modelleri ile hisse senedi fiyat tahmini (Master's thesis, Trakya Üniversitesi Sosyal Bilimler Enstitüsü.
20. Kurt, F. E. and Senal, S. (2018). Borsa İstanbul'da Kote Bireysel Emeklilik, Hayat Ve Hayat-Dışı Sigorta Şirketlerinin Hisse Senedi Fiyat Tahmininde Box-Jenkıns Yöntemi. Muhasebe ve Finansman Dergisi, (80), 233-252.
21. Louisa, L., Fauzi, R. and Nugraha, E. S. (2022). Forecasting of Retirement Insurance Filled via Internet by ARIMA Models. Journal of Actuarial, Finance, and Risk Management, 1(1), 1-8.
22. Onocak, D. and Koç, S. (2018). Yapay sinir ağlari ile emeklilik yatirim fonu hisse senedi fiyatlarinin tahmini. Finans Ekonomi ve Sosyal Araştırmalar Dergisi, 3(3), 590-600.
23. PMC, Penison Monitoring Center A.Ş., http://www.egm.org.tr (June, 2024).
24. Private Pension Savings and Investment System Law No. 4632, Official Gazette, 24366, 28.03.2001.
25. R. RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.
26. Samancı, S. (2010). Türkiye'de Bireysel Emeklilik Sisteminin Analizi Ve Tüketicilerin Yaklaşımları Üzerine Bir Araştırma, Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
27. Tsay, R. S. (2005). Analysis of financial time series. John wiley & sons.

Appendix.

A1. ACF and PACF Graphs for the Price, the Number of Share, the Number of People, the Total Fund and the Log-Return of ALZ, AZS and AMZ Retirement Funds



Topp-Leone Exponential Distribution For Asymmetric Loss Functions With Identical Priors

Saridha D (saridhareddy_07@yahoo.com)
Research Scholar, Department of Statistics, Presidency College, Chennai-5.India

Radha R.K (radhasai66@gmail.com)
Associate Professor, Department of Statistics, Presidency College, Chennai-5.India

ABSTRACT

The present paper considers estimating the shape and scale parameters of the Topp-Leone Exponential distribution. Bayes estimators are obtained using exponential, gamma, log-normal, and Weibull distributions as the identical priors under asymmetric loss functions and integrated with the Lindley approximation method. These priors are compared using Bayes risk through simulation study with varying sample sizes and real data sets. Specifically, for the shape parameter of the Topp-Leone exponential distribution, the study identifies that Gamma prior under the Entropy loss function is most preferred.

Keywords: *Prior, Lindley's approximation, Asymmetric loss functions, Bayes Estimator, Bayes Risk*

JEL Classification: *C11*

1. INTRODUCTION

Bayesian estimation, a non-classical approach to statistical inference, is widely applied globally. The Topp-Leone distribution, a bounded J-shaped distribution, is an alternative to the Beta distribution. Various authors have studied this distribution. The Exponential distribution proposed by Epstein (1954) plays a significant role in real-life scenarios. Topp and Leone (1955) proposed that the Topp-Leone distribution includes discussions on its bounded variant, for analyzing empirical data characterized by J-shaped histograms. Nadarajah and Kotz (2003) determined some J-shaped distribution moments of Topp-Leone distribution. Kotz and Seier (2007) compared the kurtosis of the Topp-Leone and left triangular distributions. Genc (2012) presented recurrence relations for the moments of order statistics from the Topp–Leone distribution. Al-Shomrani et al (2016) introduced the Topp-Leone family of distributions, providing a comprehensive overview of its characteristics and practical applications. Mohammed et al (2018) studied the comparison of the Topp-Leone Exponential, Topp-Leone Exponentiated exponential and Topp-Leone Exponentiated expansion models for ovarian cancer patient data. Fatoki Olayode (2019) discussed the moment generation function, survival function and ordinal statistics of the Topp-Leone Rayleigh distribution. Kawsar et al (2017) estimated the shape parameter of Exponentiated moment Exponential distributions using informative and non-informative priors under the SELF, PLF and Entropy loss functions. Hind Jawad Kadhim Albderi (2021) discussed the survival function of the Topp-Leone exponential distribution and its application. Noman Rashed (2019) studied the properties and applications of the Topp-Leone compound Rayleigh distribution. Radha et al (2017) discussed the classical and Bayesian estimation of Power function distribution. Randhir Singh (2021) investigated Bayesian parameter estimation of the Exponential distribution using type II censored samples, employing various loss functions such as Squared Error, DeGroot, Minimum expected loss, and Exponentially weighted minimum expected loss. Fithriani et al (2019) utilized SELF and PLF to estimate the shape parameter k of the Burr distribution, by comparing both symmetric and asymmetric loss functions. Sanjay Kumar Singh et al (2011) discussed parameter estimation of the Exponentiated Exponential distribution and its reliability function for type II censored data using the entropy loss function. Farouk et al (2019) focused on parameter estimation of the Lindley distribution using informative and non-informative priors under the Linex loss function. Saridha et al (2024) discussed the Topp-Leone Exponential distribution, emphasizing the role of symmetric loss functions with identical priors. This paper adopts the Bayesian approach for Topp-Leone Exponential Distribution to estimate the parameters with identical priors using asymmetric loss functions. The unknown shape and scale parameter are assumed to follow identical priors presented in Table:1 for Topp -Leone Exponential distribution.

Priors Selection

Table 1

Priors	***Identical priors***	
	Shape Parameter $\boldsymbol{\eta}$	*Scale Parameter* $\boldsymbol{\delta}$
Exponential	*Exponential*	*Exponential*
Gamma	*Gamma*	*Gamma*
Log-Normal	*Log-Normal*	*Log-Normal*
Weibull	*Weibull*	*Weibull*

2. MAXIMUM LIKELIHOOD ESTIMATION

The p.d.f of Topp-Leone Exponential distribution (Al-Shomrani et al,2016) is given by

$$f(x;\eta,\delta) = 2\eta\delta\ e^{-2\delta x}\left(1-e^{-2\delta x}\right)^{\eta-1}\ ; x,\eta,\delta \geq 0 \qquad [1]$$

with η as shape parameter and δ the scale parameter.

Then the likelihood function:

$$L(x;\eta,\delta) = (2\eta\delta)^n\ e^{-2\delta\sum_{i=1}^{n} x_i}\prod_{i=1}^{n}\left(1-e^{-2\delta x_i}\right)^{\eta-1} \qquad [2]$$

Taking the log of likelihood equation [2] and differentiating $w.r.to.\eta$ and δ gives

$$\frac{\partial LogL}{\partial\eta} = \frac{n}{\eta} + \sum_{i=1}^{n} log\left(1-e^{-2\delta x_i}\right) = 0 \qquad [3]$$

$$\frac{\partial LogL}{\partial\delta} = \frac{n}{\delta} - 2\sum_{i=1}^{n} x_i + (\eta-1)\sum_{i=1}^{n}\frac{(-2x_i)e^{-2\delta x_i}}{(1-e^{-2\delta x_i})} = 0 \qquad [4]$$

The maximum likelihood estimates (MLEs) of η and δ, say $\hat{\eta}$ and $\hat{\delta}$, respectively, are the solution of the equations [3] and [4]. Unfortunately, analytic solutions for η and δ are not in the closed form. To estimate these parameters η and δ Newton Raphson's method is used..

3. PRIORS AND POSTERIOR DISTRIBUTIONS

In Topp-Leone Exponential distribution, it is assumed that the shape parameter η and scale parameter δ both have identical priors namely, Exponential(E) - Exponential(E), Gamma (G)- Gamma (G), Log Normal (LN)-Log Normal (LN) and Weibull(W)-Weibull (W). The posterior distribution with identical priors for the shape and scale parameters are discussed as follows:

3.1. Posterior Distribution for Topp-Leone Exponential distribution using Identical Priors

3.1.1 Exponential Prior

The joint prior distribution using Exponential priors for both η and δ $i.e., \eta \sim exp(a_1)$ and $\delta \sim exp(a_2)$ is :

$$p_1(\eta,\delta) = a_1 a_2 e^{-(a_1\eta + a_2\delta)} \; ; a_1, a_2 > 0 \qquad [5]$$

$$\pi_1(\eta,\delta|x) = \frac{1}{C_1} e^{-(a_1\eta+a_2\delta)} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^{n} \left(1 - e^{-2\delta x_i}\right)^{\eta-1} \qquad [6]$$

where

$$C_1 = \int_0^\infty \int_0^\infty e^{-(a_1\eta+a_2\delta)} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta$$

3.1.2. Gamma Prior

The joint prior distribution using Gamma priors for both η and δ $i.e., \eta \sim G(a_3, b_1)$ and $\delta \sim G(a_4, b_2)$ is:

$$p_2(\eta,\delta) = \frac{b_1 b_2}{\Gamma a_3 \Gamma a_4} \eta^{a_3-1} \delta^{a_4-1} e^{-(b_1\eta+b_2\delta)}; \; \eta, \delta, a_3, a_4, b_1, b_2 > 0 \qquad [7]$$

The joint posterior distribution of η and δ is given by:

$$\pi_2(\eta,\delta|x) = \frac{1}{C_2} \eta^{a_3-1} \delta^{a_4-1} e^{-(b_1\eta+b_2\delta)} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^{n} \left(1 - e^{-2\delta x_i}\right)^{\eta-1} \qquad [8]$$

where

$$C_2 = \int_0^\infty \int_0^\infty \eta^{a_3-1} \delta^{a_4-1} e^{-(b_1\eta+b_2\delta)} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta$$

3.1.3. Log-Normal Prior

The joint prior distribution using Log-Normal prior for both η and δ $i.e., \eta \sim LN(a_5, b_3)$ and $\delta \sim LN(a_6, b_4)$

$$p_3(\eta,\delta) = \frac{1}{\eta b_3 \sqrt{2\pi_1}} e^{-\frac{(log\eta - a_5)^2}{2b_3^2}} \frac{1}{\delta b_4 \sqrt{2\pi_2}} e^{-\frac{(log\delta - a_6)^2}{2b_4^2}} , a_5, a_6, b_3, b_4 > 0 \qquad [9]$$

The joint posterior distribution of η and δ is given by:

$$\pi_3(\eta,\delta|x) = \frac{1}{C_3} \frac{1}{\eta\delta} e^{-\frac{(log\eta - a_5)^2}{2b_3^2}} e^{-\frac{(log\delta - a_6)^2}{2b_4^2}} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta \qquad [10]$$

where

$$C_3 = \int_0^\infty \int_0^\infty \frac{1}{\eta\delta} e^{-\frac{(log\eta - a_5)^2}{2b_3^2}} e^{-\frac{(log\delta - a_6)^2}{2b_4^2}} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta$$

3.1.4. Weibull Prior

The joint prior distribution using Weibull prior for both η and δ $i.e., \eta \sim W(a_7, b_5)$ and $\delta \sim W(a_8, b_6)$ is:

$$p_4(\eta,\delta) = \frac{a_7}{b_5^{a_7}} \eta^{a_7-1} e^{-\left(\frac{\eta}{b_5}\right)^{a_7}} \frac{a_8}{b_6^{a_8}} \delta^{a_8-1} e^{-\left(\frac{\delta}{b_6}\right)^{a_8}} \quad [11]$$

The joint posterior distribution of η and δ is given by:

$$\pi_4(\eta,\delta|x) = \frac{1}{c_4} \eta^{a_7-1} e^{-\left(\frac{\eta}{b_5}\right)^{a_7}} \delta^{a_8-1} e^{-\left(\frac{\delta}{b_6}\right)^{a_8}} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta \quad [12]$$

where

$$C_4 = \int_0^\infty \int_0^\infty \eta^{a_7-1} e^{-\left(\frac{\eta}{b_5}\right)^{a_7}} \delta^{a_8-1} e^{-\left(\frac{\delta}{b_6}\right)^{a_8}} (2\eta\delta)^n e^{-2\delta\sum_{i=1}^n x_i} \prod_{i=1}^n \left(1 - e^{-2\delta x_i}\right)^{\eta-1} d\eta d\delta$$

4. BAYES ESTIMATES UNDER DIFFERENT LOSS FUNCTIONS

To estimate the parameters of Topp-Leone Exponential distribution for asymmetric loss functions namely, DeGroot, Linear Exponential loss function (LINEX) and General Entropy loss function presented in TABLE: 2 are considered.

Bayes estimators and Bayes risk for various loss functions

Table 2

Loss Function Expression	***Bayes Estimator***		***Bayes risk***	
	Parameter η	Parameter δ	Parameter η	Parameter δ
DEGROOT $= L((\hat{\eta} - \eta) \propto \left(\frac{\hat{\eta}-\eta}{\hat{\eta}}\right)^2$	$\hat{\eta}_{DEGROOT} = \frac{E(\eta^2\|x)}{E(\eta\|x)}$	$\hat{\delta}_{DEGROOT} = \frac{E(\delta^2\|x)}{E(\delta\|x)}$	$\frac{Var(\eta\|x)}{E(\eta^2\|x)}$	$\frac{Var(\delta\|x)}{E(\delta^2\|x)}$
$LINEX = L((\hat{\eta} - \eta) \propto \exp(a_1(\hat{\eta} - \eta)) - a_1(\hat{\eta} - \eta) - 1$	$\hat{\eta}_{LINEX} = -\frac{1}{m}\log[E(e^{-m\eta x}\|x)]$	$\hat{\delta}_{LINEX} = -\frac{1}{m}\log[E(e^{-m\delta x}\|x)]$	$\log[E(e^{-m\eta x}\|x)] + mE(\eta\|x)$	$\log[E(e^{-m\delta x}\|x)] + mE(\delta\|x)$
$ENTROPY = L((\hat{\eta} - \eta) \propto \left(\frac{\hat{\eta}}{\eta}\right) - a\log\left(\frac{\hat{\eta}}{\eta}\right) - 1$	$\hat{\eta}_{ENTROPY} = [E(\eta^{-c}\|x]^{-\frac{1}{c}}$	$\hat{\delta}_{ENTROPY} = [E(\delta^{-c}\|x]^{-\frac{1}{c}}$	$cE(log\eta\|x) - clog[E(\eta^{-c}\|x]^{-\frac{1}{c}}$	$cE(log\delta\|x - clog[E(\delta^{-c}\|x]^{-\frac{1}{c}}$

The joint posterior distribution given in the equations [6], [8], [10] and [12] cannot be solved analytically to estimate the parameters η and δ. Hence, the Lindley approximation method is adopted. The posterior expectation can be expressed as (Anitta et al,2020).

$$E[u(\eta,\delta)|x] = \frac{\int u(\eta,\delta)\exp[L(\eta,\delta) + \rho(\eta,\delta)]\, d(\eta,\delta)}{\int \exp[L(\eta,\delta) + \rho(\eta,\delta)]\, d(\eta,\delta)} \quad [13]$$

where $u(\eta,\delta)$ is a function of η and δ only, $L(\eta,\delta)$ is the log-likelihood and $\rho(\eta,\delta)$ is the log of the joint prior of η and δ.

According to Lindley (1980), if the sample size nn is sufficiently large, the above equation can be approximately evaluated through:

$$I(x) = u(\hat{\eta},\hat{\delta}) + \frac{1}{2}[(u_{\eta\eta} + 2u_\eta\rho_\eta)\sigma_{\eta\eta} + (u_{\delta\eta} + 2u_\delta\rho_\eta)\sigma_{\delta\eta} + (u_{\eta\delta} + 2u_\eta\rho_\delta)\sigma_{\eta\delta}$$
$$+(u_{\delta\delta} + 2u_\delta\rho_\delta)\sigma_{\delta\delta} + \frac{1}{2}[(u_\eta\sigma_{\eta\eta} + u_\delta\sigma_{\eta\delta})(L_{\eta\eta\eta}\sigma_{\eta\eta} + L_{\eta\delta\eta}\sigma_{\eta\delta} + L_{\delta\eta\eta}\sigma_{\delta\eta} + L_{\delta\delta\eta}\sigma_{\delta\delta})]$$
$$+\frac{1}{2}[(u_\eta\sigma_{\delta\eta} + u_\delta\sigma_{\delta\delta})(L_{\eta\eta\delta}\sigma_{\eta\eta} + L_{\eta\delta\delta}\sigma_{\eta\delta} + L_{\delta\eta\delta}\sigma_{\delta\eta} + L_{\delta\delta\delta}\sigma_{\delta\delta})] \quad [14]$$

$u_\eta = \frac{\partial u(\eta,\delta)}{\partial \eta}$; $u_\delta = \frac{\partial u(\eta,\delta)}{\partial \delta}$; $u_{\eta\eta} = \frac{\partial^2 u(\eta,\delta)}{\partial \eta^2}$; $u_{\delta\delta} = \frac{\partial^2 u(\eta,\delta)}{\partial \delta^2}$; $u_{\eta\delta} = \frac{\partial^2 u(\eta,\delta)}{\partial\eta\partial\delta}$; $\frac{\partial^2 LogL}{\partial\eta^2} = L_{\eta\eta}$ and so on.

with the above-defined expressions, the values of the estimates for the Topp-Leone Exponential distribution are as follows.

$$E[u(\hat{\eta},\hat{\delta})|x] = u(\hat{\eta},\hat{\delta}) + \frac{1}{2}[(u_{\eta\eta} + 2u_\eta\rho_\eta)\sigma_{\eta\eta} + (u_{\delta\eta} + 2u_\delta\rho_\eta)\sigma_{\delta\eta} + (u_{\eta\delta} + 2u_\eta\rho_\delta)\sigma_{\eta\delta}$$
$$+(u_{\delta\delta} + 2u_\delta\rho_\delta)\sigma_{\delta\delta} + \frac{1}{2}[(u_\eta\sigma_{\eta\eta} + u_\delta\sigma_{\eta\delta})(S_1)] + \frac{1}{2}[(u_\eta\sigma_{\delta\eta} + u_\delta\sigma_{\delta\delta})(S_2)] \quad [15]$$

Where $S_1 = L_{\eta\eta\eta}\sigma_{\eta\eta} + L_{\delta\delta\eta}\sigma_{\delta\delta} S_2 = L_{\eta\delta\delta}\sigma_{\eta\delta} + L_{\delta\eta\delta}\sigma_{\delta\eta} + L_{\delta\delta\delta}\sigma_{\delta\delta}$

Then the logarithmic joint prior density of:

(i) Exponential prior :

$$\rho(\eta,\delta) = loga_1 + loga_2 - a_1\eta - a_2\delta$$

$$\rho(\eta) = \frac{\partial\rho(\eta,\delta)}{\partial\eta} = -a_1 \quad [16]$$

$$\rho(\delta) = \frac{\partial\rho(\eta,\delta)}{\partial\delta} = -a_2 \quad [17]$$

(ii) Gamma prior :

$$\rho(\eta,\delta) = (a_3 - 1)log\eta - b_1\eta + (a_4 - 1)log\delta - b_2\delta$$

$$\rho(\eta) = \frac{\partial\rho(\eta,\delta)}{\partial\eta} = \frac{a_3 - 1}{\eta} - b_1 \quad [18]$$

$$\rho(\delta) = \frac{\partial\rho(\eta,\delta)}{\partial\delta} = \frac{a_4 - 1}{\delta} - b_2 \quad [19]$$

(iii) Log-Normal prior :

$$\rho(\eta,\delta) = \log\left(\frac{1}{\eta\delta}\right) - \frac{(log\eta - a_5)^2}{2b_3^2} - \frac{(log\delta - a_6)^2}{2b_4^2}$$

$$\rho(\eta) = \frac{\partial \rho(\eta,\delta)}{\partial \eta} = -\frac{1}{\eta} - \frac{log\eta - a_5}{\eta b_3^2} \quad [20]$$

$$\rho(\delta) = \frac{\partial \rho(\eta,\delta)}{\partial \delta} = -\frac{1}{\delta} - \frac{log\delta - a_6}{\delta b_4^2} \quad [21]$$

(iv) Weibull prior :

$$\rho(\eta,\delta) = (a_7 - 1)log\eta + (a_8 - 1)log\delta - \left(\frac{\eta}{b_5}\right)^{a_7} - \left(\frac{\delta}{b_6}\right)^{a_8}$$

$$\rho(\eta) = \frac{\partial \rho(\eta,\delta)}{\partial \eta} = \frac{a_7 - 1}{\eta} - \frac{a_7}{b_5}\left(\frac{\eta}{b_5}\right)^{a_7 - 1} \quad [22]$$

$$\rho(\delta) = \frac{\partial \rho(\eta,\delta)}{\partial \delta} = \frac{a_8 - 1}{\delta} - \frac{a_8}{b_6}\left(\frac{\delta}{b_6}\right)^{a_8 - 1} \quad [23]$$

4.1 Lindley's Approximation of η and δ using DEGROOT:

The Bayes estimate for the parameter η of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\eta}_E = \left[\frac{\hat{\eta}^2 + \sigma_{\eta\eta} - 2\hat{\eta}(a_1)\sigma_{\eta\eta} - 2\hat{\eta}(a_2)\sigma_{\eta\delta} + \hat{\eta}\sigma_{\eta\eta}S_1 + \hat{\eta}\sigma_{\delta\eta}S_2}{\hat{\eta} - (a_1)\sigma_{\eta\eta} - (a_2)\sigma_{\eta\delta} + \frac{1}{2}\sigma_{\eta\eta}S_1 + \frac{1}{2}\sigma_{\delta\eta}S_2}\right] \quad [24]$$

$$\hat{\eta}_G = \left[\frac{\hat{\eta}^2 + \sigma_{\eta\eta} + 2\hat{\eta}\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\eta\eta} + 2\hat{\eta}\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\eta\delta} + \hat{\eta}\sigma_{\eta\eta}S_1 + \hat{\eta}\sigma_{\delta\eta}S_2}{\hat{\eta} + \left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\eta\eta} + \left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\eta\delta} + \frac{1}{2}\sigma_{\eta\eta}S_1 + \frac{1}{2}\sigma_{\delta\eta}S_2}\right] \quad [25]$$

$$\hat{\eta}_{LN} = \left[\frac{\hat{\eta}^2 + \sigma_{\eta\eta} + 2\hat{\eta}\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\eta\eta} + 2\hat{\eta}\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\eta\delta} + \hat{\eta}\sigma_{\eta\eta}S_1 + \hat{\eta}\sigma_{\delta\eta}S_2}{\hat{\eta} + \left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\eta\eta} + \left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\eta\delta} + \frac{1}{2}\sigma_{\eta\eta}S_1 + \frac{1}{2}\sigma_{\delta\eta}S_2}\right] \quad [26]$$

$$\hat{\eta}_W = \left[\frac{\hat{\eta}^2 + \sigma_{\eta\eta} + 2\hat{\eta}\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\eta\eta} + 2\hat{\eta}\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\eta\delta} + \hat{\eta}\sigma_{\eta\eta}S_1 + \hat{\eta}\sigma_{\delta\eta}S_2}{\hat{\eta} + \left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\eta\eta} + \left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\eta\delta} + \frac{1}{2}\sigma_{\eta\eta}S_1 + \frac{1}{2}\sigma_{\delta\eta}S_2}\right] \quad [27]$$

The Bayes estimate for the parameter δ of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\delta}_E = \left[\frac{\hat{\delta}^2 + \sigma_{\delta\delta} - 2\hat{\delta}(a_1)\sigma_{\delta\eta} - 2\hat{\delta}(a_2)\sigma_{\delta\delta} + \hat{\delta}\sigma_{\eta\delta}S_1 + \hat{\delta}\sigma_{\delta\delta}S_2}{\hat{\delta} - (a_1)\sigma_{\delta\eta} - (a_2)\sigma_{\delta\delta} + \frac{1}{2}\sigma_{\eta\delta}S_1 + \frac{1}{2}\sigma_{\delta\delta}S_2}\right] \qquad [28]$$

$$\hat{\delta}_G = \left[\frac{\hat{\delta}^2 + \sigma_{\delta\delta} + 2\hat{\delta}\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\delta\eta} + 2\hat{\delta}\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\delta\delta} + \hat{\delta}\sigma_{\eta\delta}S_1 + \hat{\delta}\sigma_{\delta\delta}S_2}{\hat{\delta} + \left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\delta\eta} + \left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\delta\delta} + \frac{1}{2}\sigma_{\eta\delta}S_1 + \frac{1}{2}\sigma_{\delta\delta}S_2}\right] \qquad [29]$$

$$\hat{\delta}_{LN} = \left[\frac{\hat{\delta}^2 + \sigma_{\delta\delta} + 2\hat{\delta}\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\delta\eta} + 2\hat{\delta}\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\delta\delta} + \hat{\delta}\sigma_{\eta\delta}S_1 + \hat{\delta}\sigma_{\delta\delta}S_2}{\hat{\delta} + \left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\delta\eta} + \left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\delta\delta} + \frac{1}{2}\sigma_{\eta\delta}S_1 + \frac{1}{2}\sigma_{\delta\delta}S_2}\right] \qquad [30]$$

$$\hat{\delta}_W = \left[\frac{\hat{\delta}^2 + \sigma_{\delta\delta} + 2\hat{\delta}\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\delta\eta} + 2\hat{\delta}\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\delta\delta} + \hat{\delta}\sigma_{\eta\delta}S_1 + \hat{\delta}\sigma_{\delta\delta}S_2}{\hat{\delta} + \left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\delta\eta} + \left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\delta\delta} + \frac{1}{2}\sigma_{\eta\delta}S_1 + \frac{1}{2}\sigma_{\delta\delta}S_2}\right] \qquad [31]$$

4.2 Lindley's Approximation of η and δ using LINEX:

The Bayes estimate for the parameter η of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\eta}_E = -\frac{1}{m}\log\left[\begin{array}{c} e^{-m\hat{\eta}} + \frac{1}{2}m^2e^{-m\hat{\eta}}\sigma_{\eta\eta} + me^{-m\hat{\eta}}(a_1)\sigma_{\eta\eta} + me^{-m\hat{\eta}}(a_2)\sigma_{\eta\delta} \\ -\frac{me^{-m\hat{\eta}}}{2}\sigma_{\eta\eta}S_1 - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\delta\eta}S_2 \end{array}\right] \qquad [32]$$

$$\hat{\eta}_G = -\frac{1}{m}\log\left[\begin{array}{c} e^{-m\hat{\eta}} + \frac{1}{2}m^2e^{-m\hat{\eta}}\sigma_{\eta\eta} - me^{-m\hat{\eta}}\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\eta\eta} \\ -me^{-m\hat{\eta}}\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\eta\delta} - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\eta\eta}S_1 - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\delta\eta}S_2 \end{array}\right] \qquad [33]$$

$$\hat{\eta}_{LN} = -\frac{1}{m}\log\left[\begin{array}{c} e^{-m\hat{\eta}} + \frac{1}{2}m^2e^{-m\hat{\eta}}\sigma_{\eta\eta} - me^{-m\hat{\eta}}\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\eta\eta} \\ -me^{-m\hat{\eta}}\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\eta\delta} - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\eta\eta}S_1 - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\delta\eta}S_2 \end{array}\right] \qquad [34]$$

$$\hat{\eta}_W = -\frac{1}{m}\log\left[\begin{array}{c} e^{-m\hat{\eta}} + \frac{1}{2}m^2e^{-m\hat{\eta}}\sigma_{\eta\eta} - me^{-m\hat{\eta}}\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\eta\eta} \\ -me^{-m\hat{\eta}}\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\eta\delta} - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\eta\eta}S_1 - \frac{me^{-m\hat{\eta}}}{2}\sigma_{\delta\eta}S_2 \end{array}\right] \qquad [35]$$

The Bayes estimate for the parameter δ of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\delta}_E = -\frac{1}{m}\log\begin{bmatrix} e^{-m\hat{\delta}} + \frac{1}{2}m^2 e^{-m\hat{\delta}}\sigma_{\delta\delta} + me^{-m\hat{\delta}}(a_1)\sigma_{\delta\eta} + me^{-m\hat{\delta}}(a_2)\sigma_{\delta\delta} \\ -\frac{me^{-m\hat{\delta}}}{2}\sigma_{\eta\delta}S_1 - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\delta\delta}S_2 \end{bmatrix} \quad [36]$$

$$\hat{\delta}_G = -\frac{1}{m}\log\begin{bmatrix} e^{-m\hat{\delta}} + \frac{1}{2}m^2 e^{-m\hat{\delta}}\sigma_{\delta\delta} - me^{-m\hat{\delta}}\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\delta\eta} \\ -me^{-m\hat{\delta}}\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\delta\delta} - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\eta\delta}S_1 - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\delta\delta}S_2 \end{bmatrix} \quad [37]$$

$$\hat{\delta}_{LN} = -\frac{1}{m}\log\begin{bmatrix} e^{-m\hat{\delta}} + \frac{1}{2}m^2 e^{-m\hat{\delta}}\sigma_{\delta\delta} - me^{-m\hat{\delta}}\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\delta\eta} \\ -me^{-m\hat{\delta}}\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\delta\delta} - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\eta\delta}S_1 - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\delta\delta}S_2 \end{bmatrix} \quad [38]$$

$$\hat{\delta}_W = -\frac{1}{m}\log\begin{bmatrix} e^{-m\hat{\delta}} + \frac{1}{2}m^2 e^{-m\hat{\delta}}\sigma_{\delta\delta} - me^{-m\hat{\delta}}\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\delta\eta} \\ -me^{-m\hat{\delta}}\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\delta\delta} - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\eta\delta}S_1 - \frac{me^{-m\hat{\delta}}}{2}\sigma_{\delta\delta}S_2 \end{bmatrix} \quad [39]$$

4.3 Lindley's Approximation of η and δ using ENTROPY:

The Bayes estimate for the parameter $\eta\eta$ of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\eta}_E = \Big[\hat{\eta}^{-c} + (\frac{\sigma_{\eta\eta}}{2}c(c+1)\hat{\eta}^{-c-2}\Big) + (c\hat{\eta}^{-c-1})(a_1)\sigma_{\eta\eta} + (c\hat{\eta}^{-c-1})(a_2)\sigma_{\eta\delta}$$
$$-\frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\eta\eta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\eta}S_2]^{-\frac{1}{c}} \quad [40]$$

$$\hat{\eta}_G = \Big[\hat{\eta}^{-c} + (\frac{\sigma_{\eta\eta}}{2}c(c+1)\hat{\eta}^{-c-2}\Big) - (c\hat{\eta}^{-c-1})\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\eta\eta}$$
$$-(c\hat{\eta}^{-c-1})\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\eta\delta}$$
$$-\frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\eta\eta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\eta}S_2]^{-\frac{1}{c}} \quad [41]$$

$$\hat{\eta}_{LN} = \Big[\hat{\eta}^{-c} + (\frac{\sigma_{\eta\eta}}{2}c(c+1)\hat{\eta}^{-c-2}\Big) - (c\hat{\eta}^{-c-1})\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\eta\eta}$$
$$-(c\hat{\eta}^{-c-1})\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\eta\delta} - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\eta\eta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\eta}S_2]^{-\frac{1}{c}} \quad [42]$$

$$\hat{\eta}_W = \Big[\hat{\eta}^{-c} + (\frac{\sigma_{\eta\eta}}{2}c(c+1)\hat{\eta}^{-c-2}\Big) - (c\hat{\eta}^{-c-1})\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\eta\eta}$$
$$-(c\hat{\eta}^{-c-1})\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\eta\delta} - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\eta\eta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\eta}S_2]^{-\frac{1}{c}} \quad [43]$$

The Bayes estimate for the parameter δ of Exponential, Gamma, Log-Normal, and Weibull priors using equation [14] are given by:

$$\hat{\delta}_E = [\hat{\delta}^{-c} + (c\hat{\delta}^{-c-1})(a_1)\sigma_{\delta\eta} + \frac{1}{2}(c(c+1)\hat{\delta}^{-c-2}\sigma_{\delta\delta}) + (c\hat{\delta}^{-c-1})(a_2)\sigma_{\delta\delta} - \frac{(c\hat{\delta}^{-c-1})}{2}\sigma_{\eta\delta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\delta}S_2]^{-\frac{1}{c}} \quad [44]$$

$$\hat{\delta}_G = [\hat{\delta}^{-c} - (c\hat{\delta}^{-c-1})\left(\frac{a_3-1}{\hat{\eta}} - b_1\right)\sigma_{\delta\eta} + \frac{1}{2}(c(c+1)\hat{\delta}^{-c-2}\sigma_{\delta\delta} - (c\hat{\delta}^{-c-1})\left(\frac{a_4-1}{\hat{\delta}} - b_2\right)\sigma_{\delta\delta} - \frac{(c\hat{\delta}^{-c-1})}{2}\sigma_{\eta\delta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\delta}S_2]^{-\frac{1}{c}} \quad [45]$$

$$\hat{\delta}_{LN} = [\hat{\delta}^{-c} - (c\hat{\delta}^{-c-1})\left(-\frac{1}{\hat{\eta}} - \frac{log\hat{\eta} - a_5}{\hat{\eta}b_3^2}\right)\sigma_{\delta\eta} + \frac{1}{2}(c(c+1)\hat{\delta}^{-c-2}\sigma_{\delta\delta}) - (c\hat{\delta}^{-c-1})\left(-\frac{1}{\hat{\delta}} - \frac{log\hat{\delta} - a_6}{\hat{\delta}b_4^2}\right)\sigma_{\delta\delta} - \frac{(c\hat{\delta}^{-c-1})}{2}\sigma_{\eta\delta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\delta}S_2]^{-\frac{1}{c}} \quad [46]$$

$$\hat{\delta}_W = [\hat{\delta}^{-c} - (c\hat{\delta}^{-c-1})\left(\frac{a_7-1}{\hat{\eta}} - \frac{a_7}{b_5}\left(\frac{\hat{\eta}}{b_5}\right)^{a_7-1}\right)\sigma_{\delta\eta} + \frac{1}{2}(c(c+1)\hat{\delta}^{-c-2}\sigma_{\delta\delta}) - (c\hat{\delta}^{-c-1})\left(\frac{a_8-1}{\hat{\delta}} - \frac{a_8}{b_6}\left(\frac{\hat{\delta}}{b_6}\right)^{a_8-1}\right)\sigma_{\delta\delta} - \frac{(c\hat{\delta}^{-c-1})}{2}\sigma_{\eta\delta}S_1 - \frac{(c\hat{\eta}^{-c-1})}{2}\sigma_{\delta\delta}S_2]^{-\frac{1}{c}} \quad [47]$$

5. SIMULATION STUDY

This study was conducted to compare the performance of Bayes estimates under different loss functions for the Topp-Leone Exponential distribution. Data sets of sizes n=20,50 and 100 representing small, moderate and large samples respectively, were generated with hyperparameters $a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = a_7 = a_8 = m = 1$, $b_1 = b_2 = 1.5$, $c = 0.5$, $b_3 = b_4 = 1$, $b_5 = b_6 = 2$ and N=5000 replications. The simulation results for estimating the shape and scale parameters with different loss functions using identical priors are presented in Tables 3-4, utilizing the R package.

5.1 Results and Discussion

A comparative study based on Bayes risk for different loss functions to estimate the parameters of the Topp-Leone Exponential distribution is summarized in Table:3-4.

Bayes estimate of the parameters along with their Bayes Risk * for identical prior with different loss functions when $\eta = 0.5\ and\ \delta = 1$
$\eta = 0.5\ and\ \delta = 1$.

Table 3

SAMPLE SIZES	LOSS FUNCTIONS	PARAMETERS	PRIORS			
			EXPONENTIAL	GAMMA	LOG-NORMAL	WEIBULL
20	DEGROOT	η	0.60293 (0.066475)	0.568878 (0.06838)	0.669454 (0.039294)	0.632375 (0.059097)
		δ	1.28250 (0.1125307)	1.1368 (0.113649)	1.475939 (0.064673)	1.391362 (0.09713)
	LINEX	η	0.549001 (0.012573)	0.518817 (0.010519)	0.629877 (0.011411)	0.580594 (0.013218)
		δ	1.064835 (0.074388)	0.976691 (0.045472)	1.283620 (0.090761)	1.164008 (0.092274)
	ENTROPY	η	0.532706 (0.008395)	0.505298 (0.007186)	0.637953 (0.007949)	0.562756 (0.008395)
		δ	1.048895 (0.013683)	0.961867 (0.010017)	1.270560 (0.014653)	1.149341 (0.015549)
50	DEGROOT	η	0.539982 (0.027139)	0.530182 (0.027842)	0.565995 (0.021759)	0.5494 (0.027139)
		δ	1.114700 (0.049272)	1.078422 (0.050909)	1.187738 (0.038550)	1.148499 (0.045834)
	LINEX `	η	0.521148 (0.004037)	0.51136 (0.003932)	0.54978 (0.003728)	0.53105 (0.00403)
		δ	1.031183 (0.029049)	0.997515 (0.026656)	1.112649 (0.029183)	1.0662 (0.030094)
	ENTROPY	η	0.514218 (0.003463)	0.504766 (0.003331)	0.553462 (0.003343)	0.523951 (0.003463)
		δ	1.021257 (0.006197)	0.988519 (0.005705)	1.101983 (0.006283)	1.055699 (0.006429)
100	DEGROOT	η	0.518808 (0.013712)	0.51428 (0.013903)	0.531918 (0.012256)	0.523255 (0.013712)
		δ	1.056347 (0.025385)	1.039676 (0.025865)	1.093482 (0.022376)	1.072476 (0.024445)
	LINEX	η	0.509804 (0.001861)	0.505258 (0.001844)	0.523584 (0.001781)	0.514373 (0.001856)
		δ	1.015836 (0.013843)	0.99957 (0.013384)	1.0553295 (0.0137501)	1.032392 (0.014013)
	ENTROPY	η	0.506328 (0.001736)	0.501856 (0.001706)	0.5252350 (0.0017025)	0.510861 (0.001736)
		δ	1.010079 (0.003191)	0.994096 (0.003075)	1.0493324 (0.0032090)	1.026456 (0.003245)

* Bayes Risk are given in the parenthesis.

Bayes estimate of the parameters along with their Bayes Risk * for identical prior with different loss functions when $\eta = 1\ and\ \delta = 1$
$\eta = 1\ and\ \delta = 1$.

Table 4

SAMPLE SIZES	LOSS FUNCTIONS	PARAMETERS	PRIORS			
			EXPONENTIAL	GAMMA	LOG-NORMAL	WEIBULL
20	DEGROOT	η	1.19557 (0.087745)	0.587241 (0.064522)	1.380657 (0.059055)	1.318152 (0.087745)
		δ	1.16295 (0.077983)	1.044462 (0.067412)	1.316177 (0.054472)	1.258488 (0.069892)
	LINEX	η	1.04219 (0.050776)	0.954938 (0.01798)	1.218846 (0.072767)	1.14139 (0.071614)
		δ	1.03005 (0.044047)	0.950922 (0.02583)	1.190186 (0.052897)	1.117588 (0.053853)
	ENTROPY	η	1.030779 (0.009151)	0.944224 (0.005293)	1.001292 (0.010817)	1.132481 (0.009151)
		δ	1.017738 (0.008829)	0.941671 (0.005936)	1.176555 (0.009957)	1.104033 (0.010233)
50	DEGROOT	η	1.083547 (0.034225)	1.048591 (0.034513)	1.142298 (0.028866)	1.116139 (0.034225)
		δ	1.068899 (0.033231)	1.036301 (0.033439)	1.125085 (0.028216)	1.09951 (0.031374)
	LINEX	η	1.025803 (0.020069)	0.994102 (0.017927)	1.088047 (0.020593)	1.058753 (0.020962)
		δ	1.01566 (0.018054)	0.985772 (0.016337)	1.075033 (0.018413)	1.046562 (0.018755)
	ENTROPY	η	1.019124 (0.004167)	0.988146 (0.00375)	1.053184 (0.004291)	1.051676 (0.004167)
		δ	1.008683 (0.004043)	0.979589 (0.003656)	1.06748 (0.004172)	1.03912 (0.004218)
100	DEGROOT	η	1.039334 (0.01709)	1.023898 (0.017228)	1.067826 (0.015651)	1.054301 (0.01709)
		δ	1.033935 (0.016888)	1.018909 (0.016992)	1.061929 (0.015524)	1.048523 (0.016385)
	LINEX	η	1.012196 (0.009255)	0.99727 (0.00888)	1.04171 (0.00927)	1.02736 (0.009382)
		δ	1.007745 (0.008823)	0.993234 (0.008469)	1.036651 (0.008855)	1.022482 (0.008951)
	ENTROPY	η	1.008337 (0.002117)	0.993628 (0.002025)	1.03264 (0.002141)	1.023383 (0.002117)
		δ	1.003854 (0.002086)	0.989561 (0.001995)	1.032598 (0.002116)	1.018462 (0.002127)

* Bayes Risks are given in the parenthesis.

The following results concerning different loss functions with identical priors are observed as follows:

(i)for the scale parameter $\delta = 1$ and shape parameter $\eta = 0.5$

- **DEGROOT Loss Function**: The Log-Normal prior exhibits a lower Bayes risk for shape and scale parameters as sample size increases.
- **LINEX Loss Function**: Gamma prior shows lower Bayes risk for both parameters at a sample size of 20. At sample sizes 50 and 100, Log-Normal prior is preferred for the shape parameter, while Gamma prior is preferred for the scale parameter.
- **ENTROPY Loss Function**: Gamma prior shows lower Bayes risk for both parameters at sample sizes 20 and 50. At sample size 100, Log-Normal prior is preferred for the shape parameter, while Gamma prior is preferred for the scale parameter.

(ii)for the scale parameter $\delta = 1$ and shape parameter $\eta = 1$

- **DEGROOT Loss Function**: Log-Normal prior shows lower Bayes risk for shape and scale parameters.
- **LINEX and ENTROPYLoss Functions:** Gamma prior shows lower Bayes risk for shape and scale parameters

Comparing DEGROOT, LINEX, and ENTROPY loss functions across different priors, the combination of Log-Normal prior for the shape parameter and Gamma prior for the scale parameter shows consistently lower Bayes risk with the ENTROPY loss function.

5.2 Real Data Set:

This data set is taken from M.D. Nicholas. et al (2006) and the data represent the tensile strength of 100 observations of carbon fibres. The results are presented in Table 5.

Bayes estimate of the parameters along with their Bayes Risk *
for identical prior with different loss functions

Table 5

LOSS FUNCTIONS	PARAMETERS	PRIORS			
		EXPONENTIAL	GAMMA	LOG-NORMAL	WEIBULL
DEGROOT	η	2.54771 (0.019914)	2.472686 (0.018527)	2.646805 (0.019188)	2.618629 (0.019914)
	δ	0.31861 (0.008936)	0.314108 (0.008603)	0.323512 (0.008854)	0.322994 (0.008885)
LINEX	η	2.439961 (0.05701)	2.380377 (0.046496)	2.530708 (0.06531)	2.50332 (0.063746)
	δ	0.315318 (0.000447)	0.310989 (0.000417)	0.320188 (0.000459)	0.319665 (0.000459)
ENTROPY	η	2.462061 (0.002258)	2.39772 (0.001874)	1.467531 (0.002496)	2.529028 (0.002258)
	δ	0.313713 (0.001068)	0.309528 (0.000976)	0.318503 (0.001117)	0.317984 (0.001114)

* Bayes Risks are given in the parenthesis.

From the above Table 5, we observe that Gamma prior has a lower Bayes risk. In the case of shape and scale parameters, the ENTROPY and LINEX loss functions perform better.

6. CONCLUSION

In this study, we have discussed the problem of Bayesian estimation for the Topp-Leone Exponential distribution with identical priors under an asymmetric loss function by applying Lindley's approximation method and illustrated the methodology through simulation technique and real data set. On comparing the estimated Bayes risk values of the Topp-Leone Exponential distribution using asymmetric loss functions, it is found that the risk under the Entropy loss function is the minimum among the other loss functions. Finally, it is observed that two parameters Topp-Leone Exponential distribution with gamma prior under Entropy loss function performed well in this study. Further studies are necessary to confirm these findings.

REFERENCES

1. Al-Shomrani A, Arif O, Ibrahim S, Hanif S and Shahbaz M (2016), "Topp–Leone Family of Distributions: Some Properties and Application", Pakistan Journal of Statistics and Operation Research, 12(3), 443-451.
2. Anitta SA, and Dais George (2020), "Bayesian Analysis of Two Parameter Weibull Distribution using Different Loss Functions", Stochastic Modeling and Applications,24(2).

3. Epstein B, and Sobel M (1954), “Some theorems to life resting form an exponential distribution”, Annals of Mathematical Statistics, 25(2):373-381.
4. Farouk Metiri, Halim Zeghdoudi, and Mohamed Riad Remiata, On Bayes estimates of Lindley distribution under Linex loss function: Informative and Non-informative priors, Global Journal of Pure and Applied Mathematics 12 (2016), 391- 400.
5. Fatoki Olayode (2019), “The Topp-Leone Rayleigh Distribution with Application”,American Journal of Mathematics and Statistics, 9 (6): 215-220.
6. Fithriani I, Arief Hakim, and Mila Novita (2019), “A comparison of the Bayesian method under symmetric and asymmetric loss functions to estimate the shape parameter K of Burr distribution”, Journal of Physics Conference Series (2019),1218.
7. Genc A (2012), “ Moments of order statistics of Topp Leone distribution”, Statistical Papers,53(1):117-31.
8. Hind Jawad Kadhim Albderia (2021), “Estimate survival function of the Topp-Leone exponential distribution with application”, International journal Nonlinear Analysis and Applications,12(2): 53-60.
9. Kawsar F, Ahmed SP (2017), “Bayesian Approach in Estimation of shape parameter of Exponentiated moment Exponential distribution” Journal of Statistical Theory and Applications,17(2):359-74.
10. Kotz S, Seier E (2007), “ Kurtosis of the Topp Leone distributions”, Interstat, 1-15.
11. Lindley DV, (1980),Approximate Bayesian methods, Journal of Statistical Computation and Simulation. Trabajos de Estadistica y de Investigacion Operativa 31,223-45.
12. Mohammed H, AbuJarad, Athar Ali Khan (2018), “Bayesian Survival Analysis of Topp-Leone Generalized Family with Stan”, International Journal of Statistics and Applications,8(5):274-90.
13. Noman Rasheed (2019), “Topp–Leone compound Rayleigh distribution: properties and applications”, Research Journal of Mathematical and Statistical Science,7(3): 51–58.
14. Nadarajah S, Kotz S (2003), “Moments of some J-shaped distributions”, Journal of Applied Statistics, 30(3):311-17.
15. Nicholas MD and Padgett WJ (2006), “A Bootstrap control chart for Weibull percentiles,”,Quality and Reliability engineering international ,22 : 141-151.
16. Radha RK, Venkatesan P (2013), “Bayes Estimator as a function of some classical estimators for Power Function Distribution”, International Journal of Statistics and Analysis,3(2):105-09.
17. Randhir Singh (2021), “Bayesian estimation of the unknown parameter and reliability of the Exponential distribution with a non-natural conjugate prior”,Journal of Emerging Technologies and Innovative Research, 8: 228-236.
18. Singh SK, Umesh Singh, and Dinesh Kumar (2011), “Estimation of parameters and reliability function of Exponentiated Exponential distribution: Bayesian approach under General Entropy loss function”, Pakistan Journal of Statistics and Operation Research, 7:199-216.
19. Saridha D, Radha RK and Venkatesan P(2024), “Bayesian estimation of Topp-Leone Exponential distribution using symmetric loss functions for identical priors”, Sirjana Journal, 54(3), 19-26.
20. Topp CW and Leone FC.(1955), “A family of J-shaped frequency functions”, Journal of the American Statistical Association, 50: 209-19.

Clustering Countries According to Global Innovation Index 2022 Using Fuzzy Clustering Analysis

Esra POLAT (espolat@hacettepe.edu.tr)
Department of Statistics, Faculty of Science, Hacettepe University, Ankara, Turkey

ABSTRACT

In the dynamic landscape of global competition, characterized by the escalating significance of technology, innovation emerges as a pivotal determinant for nations seeking to enhance and sustain their competitiveness. In this research, the dataset encompassing seven subcategories within each primary indicator of both innovation input and output subscales, as delineated in the 2022 Global Innovation Index (GII) report, was employed for clustering 132 countries with a fuzzy c-means clustering algorithm. Cluster 1 encompasses a total of 97 countries, while Cluster 2 comprises 35 countries. Following the analysis, the countries with high-income levels in Cluster 2 ranked first. These countries are also positioned among the foremost countries in the GII rankings, which means the ones exhibiting high-income levels attain leading positions similarly across innovation indicators. However, all of the low-income countries and all low-middle-income countries except India clustered in Cluster 1. The cluster analysis results and index rankings are parallel for the countries with high and low GII values. The top countries in GII rankings clustered in Cluster 2. The countries at the bottom of GII rankings clustered in Cluster 1. The fuzzy c-means clustering algorithm revealed the power of the GII to reflect the data.

Keyword: *Countries, Global innovation index, innovation, Fuzzy Clustering Analysis*

JEL Classification: *B41, C13, C22, R21, R29*

1. INTRODUCTION

Innovation indices aspire to methodically assess the impact of innovation on diverse variables across technological, macro, micro, and other dimensions. Global Innovation Index (GII) is one of the evaluators of innovation performance and efficiency. The GII may be considered as a paramount metric of a country's ability for innovation. The GII is derived by computing the mean of the sub-index values about innovation output and innovation input. Hancioglu (2016) observed that the GII can facilitate the computation of innovation efficiency performance values for countries. The determination of countries' innovation efficiency performance involves the computation of the ratio between innovation output sub-index values and innovation input sub-index values. This ratio elucidates the extent to which outputs can generated per unit of input (Aytekin et al., 2022).

The overarching objective of the GII is to enhance the precision of innovation measurement, thereby contributing to a more comprehensive depiction of global innovation ecosystems (WIPO, 2022).

The comprehensive ranking within the GII hinges upon two pivotal sub-components, the Innovation Output Sub-Component and the Innovation Input Sub-Component, both of equal significance in delineating a comprehensive overview of innovation. As a result, the computation necessitates the derivation of three distinct indices (WIPO, 2022).

- Innovation Input Sub-Component: 5 input components encapsulate facets of the economic framework that foster and facilitate innovative endeavors.
- Innovation Output Sub-Component: Innovation outputs manifest as outcomes of inventive activities within the economic sphere. Despite the Output Sub-Component incorporating solely two components, its significance equals that of the Input Sub-Component in the computation of the overarching GII scores.
- The aggregate GII score is calculated as the mean of the Output and Sub-Components, serving as the basis for the generation of GII economy rankings.

The 2022 GII Report, authored by the World Intellectual Property Organization (WIPO), a Switzerland-based agency operating under the auspices of the United Nations (UN), has been released.

There are studies on analyzing GII report data sets using cluster analysis. Some of them are listed below.

Jankowska et al.(2017) utilized k-means cluster analysis on GII (2015) data to discern countries exhibiting varying levels of innovation inputs, delineated as high, medium, or low, thereby reflecting their capacity to generate innovation output. Then, they wanted to identify countries deviating from expected patterns, i.e., even though they were well (poorly, moderately) equipped, performed better (or worse) than foreseen. Furthermore, they conducted a focused examination of Poland and Bulgaria to ascertain the underlying reasons for their challenges in sustaining innovations.

In their study, Unlu (2019) empirically examined variations in innovation performance efficiency across middle-income countries. They used Ward's agglomerative hierarchical technique for cluster analysis. Subsequently, cluster analyses were performed individually for both input and output indicators. Additionally, discriminant analysis was employed to ascertain the determinants of efficiency discrepancies. The study encompassed 54 countries classified based on the World Bank's income categorization, comprising 31 upper-middle-income and 23 lower-middle-income nations. The dataset is taken from the 2018 GII. The results substantiate the presence of inefficiency issues regarding innovation performance within middle-income economies.

Gurtuna and Polat (2020) examined the three subcategories associated with each primary indicator within the innovation output and input subscales of the 2018 GII report. The dataset comprised 126 countries, which underwent analysis using the clustering method. This investigation employs Ward's Technique and the k-means method. Their first aim was to assign the countries into 3 and 5 clusters using their GII values. They mentioned that sorting countries by GII values was possible, however, the issue of determining clusters was uncertain. Cluster analysis of countries made it possible to cluster countries such as Low - Medium - High or Low - Low Medium - Medium - Medium High – High. The second purpose of the analysis was to use the 21 variables when creating the index to determine similar countries in terms of innovation. This target was accomplished by using various cluster numbers, such as 3, 4, and 5, and different methods, such as Ward's Technique and the k-means method. Although ranking countries according to cluster analysis results or according to GII values were consistent with each other in some situations, it also observed that they behaved differently at some points.

Famalika and Sihombing (2021) employed the k-medians and k-means techniques to cluster countries using the GII 2018 dataset in their research. The sub-component within the GII comprises 7 components: Human Capital and Research, Institutions, Market sophistication, Infrastructure, Creative Outputs, Knowledge and Technology Outputs, and Business Sophistication. The clustering analyses applied to these seven variables. Upon conducting the research, the derived clustering outcomes employing both the k-medians and k-means methods revealed that k-medians outperformed the k-means technique, evidenced by the smaller variance value associated with k-medians. In each method, 3 clusters were created. In the k-means method, Cluster 1 comprises 48 countries, Cluster 2 includes 45 countries, and Cluster 3 encompasses 33 countries. Notably, Cluster 1 exhibits a relatively high average value across seven variables. However, Cluster 2 demonstrates a low average value for these variables, while Cluster 3 manifests the highest average value among the 3 clusters. Transitioning to the k-medians method, Cluster 1 encompasses 33 countries, Cluster 2 involves 53 countries, and Cluster 3 includes 40 countries. Cluster 1, in this context, displays the highest average value across the seven variables. Cluster 2 demonstrates a relatively high average value, whereas Cluster 3 exhibits a low average value for the mentioned variables.

In their study, Eren and Gelmez (2022) ranked 132 countries based on the GII (2021) report dataset, employing ARAS and COPRAS methods across seven criteria. The ENTROPY weighting method was applied as the primary approach for ranking countries based on their innovation performance. After ranking innovation performances of 132 countries within the index, they categorized into clusters based on their innovation indicators. The cluster analysis was applied utilizing the WEKA program. Switzerland, Sweden, and the USA emerged as the nations with the most favorable rankings concerning innovation indicators, as determined through the ARAS and COPRAS techniques. Conversely, Benin, Angola, and Guinea were identified as the countries with the least favorable rankings. The outcome of the clustering analysis conducted using the WEKA program revealed the subdivision of these countries into eight distinct clusters.

In their investigation, Alqararah and Alnafrah (2023) utilized a multi-dimensional innovation-driven clustering methodological analysis for the data set of GII for the year 2019. k-means and hierarchical cluster analysis approaches were employed, utilizing diverse sets of distance matrices to unveil and scrutinize discrete innovation patterns. They categorized 129 countries into 4 clusters: Advanced, Specials, Primitives, and Intermediates. Each cluster demonstrates distinct weaknesses and strengths concerning innovation performance. The Specials cluster demonstrates notable proficiency in knowledge commercialization and institutions, whereas, the Advanced cluster exhibits strengths in education and ICT-related services, albeit with a weakness apparent in patent commercialization. The Intermediates cluster exhibits strengths in venture capital and labor productivity, while simultaneously manifesting weaknesses in R&D expenditure and the quality of higher education. The Primitives cluster demonstrates proficiency in creative actions but it presents deficiencies in training, education, and digital skills. Moreover, they specified 35 indicators characterized by minimal variance parts across nations (Alqararah and Alnafrah, 2023).

The countries can be ranked using innovation indices. However, countries could be similar or different regarding innovation indicators, and this may not have reflected in the indexes. One of the goals of this study is to examine the similarities and differences between countries with each other within the scope of innovation performances and evaluate how much GII index values reflect these similarities and differences. By applying fuzzy cluster analysis, it is aimed to bring together countries with similar characteristics.

The comparison of the results of fuzzy cluster analysis with the results of an innovation measurement index could be used to measure the consistency of the index for future studies.

The following sections of this study are organized as in below. In Section 2, a brief theory for Fuzzy Clustering based on the fuzzy c-means clustering algorithm is presented. In Section 3, presentation of the methodology and variables used for clustering the 132 countries by fuzzy clustering analysis is presented. Finally, general comments and a summary of the results are presented in the last section.

2.MATERIAL AND METHOD

2.1.Fuzzy Clustering

The Fuzzy Clustering technique is recognized as a generalized variant that incorporates elements from both the medoids and k-means clustering techniques, both of which exemplify non-hierarchical clustering approaches. The Fuzzy Clustering technique involves the separation of n units into k clusters, allowing for the non-compulsory inclusion of units in clusters and permitting their divergence. In traditional clustering methodologies, units are unequivocally allocated to a specific cluster. Nevertheless, within the fuzzy clustering technique, it is necessary to compute the membership coefficient and membership probability of units across various clusters. In clustering analysis, the allocation of units to a cluster is examined within three distinctive scenarios: probabilistic, fuzzy, and absolute. In the paradigm of absolute clustering, units exhibit an exclusive affiliation wherein they are either a member or not a member of a single cluster. Conversely, in fuzzy clustering, elements can concurrently belong to multiple clusters. In probabilistic clustering, a unit is assigned to a cluster or not. Nevertheless, the allotment of a unit to a cluster is contingent upon the underlying probability distribution (Alptekin and Yesilaydin, 2015). The definitiveness inherent in traditional clustering methodologies occasionally gives rise to inaccuracies in results. In instances where observational units are equidistant from each homogeneous cluster, ambiguity arises concerning the assignment of these units to specific clusters. This scenario underscores the significance of the conception of the probability of membership to clusters (Bulbul and Camkiran, 2018). Given that the fuzzy clustering technique facilitates membership determination based on the degree of affiliation with clusters, it often yields more robust and natural outcomes compared to conventional methods (Cebeci and Yildiz, 2015).

Fuzzy clustering affords a nuanced exploration of data, offering more detailed insights. However, challenges arise when summarizing and classifying information when dealing with an abundance of units and clusters, leading to an excess of generated outputs (Zorlutuna and Erilli, 2018).

The predominant technique in fuzzy clustering is the fuzzy c-means clustering algorithm, initially introduced by Bezdek and Hathaway (1987) and subsequently refined by Kaufman and Rousseeuw (1990). As an alternative approach to the conventional k-means method, where each unit is exclusively assigned to a single cluster, fuzzy clustering assigns each unit a probability of belonging to every cluster individually and distinctively from other clusters. The Fuzzy c-means algorithm addresses situations where units are positioned in a manner that makes it difficult to determine the optimal center to which they should belong. This challenge arises when the distances between a unit and neighboring centers are almost identical to each other. Fuzzy c-means determines centroids according to these probabilities. The applied procedures for iteration, termination, and initialization are identical to the ones used in the k-means algorithm. It is discerned that fuzzy c-means and k-means diverge in their treatment of assigning probabilities to individual data points, with k-means assigning a probability of 1 if the unit is closest to a centroid and 0 otherwise. Challenges arise in case of the distances between a unit and neighboring centers are almost identical to each other (Al Rahhal and Rencber, 2022).

This algorithm was designed to minimize the cost function, computed based on cluster memberships and distances. The cost function is presented in Eqn [1] (Bagdatli Kalkan, 2019).

$$J = \left(U, c_1, \cdots, c_C\right) = \sum_{i=1}^{c} J_i = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m} d_{ij}^{2} \qquad [1]$$

In Eqn [1], U represents the Membership Matrix consisting of uij, denoting membership probabilities. These probabilities range between zero and one, with the sum of membership probabilities for each point equating to one. ci is the cluster center of fuzzy group i; and dij signifies the Euclidean distance between the ith cluster center and the jth unit. The parameter m serves as a weighting exponent. The necessary circumstances for Eqn [1] to achieve its minimum are as follows:

$$c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \qquad [2]$$

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{d_{ij}}{d_{kj}} \right)^{2/(n-1)}} \qquad [3]$$

The Fuzzy c-means algorithm encompasses four distinct steps. The initial step involves the initialization of the membership matrix U with subjective values ranging between zero and one. In the second step, cluster centers are calculated using Eqn [2]. In the third step, Eqn [1] is used for the calculation of the cost function. Stop if either cost function falls below a specified acceptance value or its betterment over former iterations below a specific threshold. The final step involves the formulation of the U matrix utilizing Eqn [3], subsequently iterating back to the second step (Saravananathan and Velmurugan, 2018). Given that the outcome of this algorithm is contingent upon the initially created random values, various algorithms have been and continue to be developed to address challenges arising from inherent randomness (Zorlutuna and Erilli, 2018).

3. RESULTS AND DISCUSSION

GII measures the innovation of countries by employing a multitude of indicators that have an impact on innovation. In this study, using data obtained from the GII (2022) Report, seven criteria (Institutions, Infrastructure, Market sophistication, Knowledge and technology outputs, Human capital and research, Business sophistication, and Creative outputs), 132 countries were analyzed by fuzzy clustering analysis. The data was analyzed using a fuzzy c-means clustering algorithm.

This study utilized secondary data obtained from collaborative efforts involving the World Intellectual Property Organization in conjunction with INSEAD and Cornell University. These three institutions assessed a nation's global innovation standing based on seven components, as given in Table 1 (Famalika and Sihombing, 2021; Aytekin et al., 2022).

Global Innovation Index components

Table 1

Input innovation	**Output innovation**
Institutions	Knowledge and technology outputs
Human capital and research	Creative outputs
Infrastructure	
Market sophistication	
Business sophistication	

Initially, validity indices are employed for the determination of the suitable number of clusters. The validity values are presented for various numbers of clusters in Table 2.

Fuzzy C-Means Clustering Validity values

Table 2

	k=2	k=3	k=4	k=5	k=6	k=7	k=8	k=9	k=10
Partition Entropy Index	0.350	0.606	0.873	1.055	1.228	1.371	1.497	1.607	1.703
Partition Coefficient	0.790	0.659	0.529	0.453	0.386	0.343	0.307	0.279	0.252
Modified Partition Coefficient	0.579	0.489	0.372	0.316	0.263	0.234	0.208	0.189	0.169
Fuzzy Silhouette Index	0.780	0.676	0.568	0.504	0.403	0.427	0.416	0.402	0.381

Validity indices are commonly employed for determination the optimal number of clusters; however, they cannot inherently furnish definitive insights into the quality of clustering outcomes. The computation of the Partition Coefficient Index involves the utilization of the clustering degrees matrix (U), to achieve a maximum value. The Modified Partition Coefficient Index is characterized as a linear transformation of the Partition Coefficient, with its values constrained within the range of 0 to 1. The Modified Partition Coefficient Index is characterized as a linear transformation of the Partition Coefficient, with its values constrained within the range of 0 to 1. The Fuzzy Silhouette Index is a more sophisticated metric in comparison to other indices, leveraging a broader spectrum of information. The objective is to maximize this value (Ferraro and Giordani, 2015; Bagdatli Kalkan, 2019). Consequently, several indices listed in Table 2 do not serve as conclusive evidence for the quality of clustering. Nevertheless, the current quantity is computed to be the most optimal among alternative cluster numbers. It is important to note that no validity index produces definitive outcomes, thereby necessitating ongoing developments in the refinement of these indices. According to these indexes, the number of clusters was determined as 2. After fuzzy c-means clustering, obtained 2 clusters of countries were based on the GII 2022. Membership values of countries to clusters are shown in Table 3.

Membership values of countries to clusters

Table 3

	Country	Cluster 1 Membership Degree	Cluster 2 Membership Degree		Country	Cluster 1 Membership Degree	Cluster 2 Membership Degree
C-1	Albania	0.956	0.044	C-67	Lithuania	0.324	0.676
C-2	Algeria	0.947	0.053	C-68	Luxembourg	0.090	0.910
C-3	Angola	0.920	0.080	C-69	Madagascar	0.925	0.075
C-4	Argentina	0.896	0.104	C-70	Malaysia	0.261	0.739
C-5	Armenia	0.957	0.043	C-71	Mali	0.925	0.075
C-6	Australia	0.048	0.952	C-72	Malta	0.105	0.895
C-7	Austria	0.042	0.958	C-73	Mauritania	0.909	0.091
C-8	Azerbaijan	0.949	0.051	C-74	Mauritius	0.546	0.454
C-9	Bahrain	0.765	0.235	C-75	Mexico	0.822	0.178
C-10	Bangladesh	0.962	0.038	C-76	Mongolia	0.902	0.098
C-11	Belarus	0.796	0.204	C-77	Montenegro	0.783	0.217
C-12	Belgium	0.056	0.944	C-78	Morocco	0.921	0.079
C-13	Benin	0.931	0.069	C-79	Mozambique	0.927	0.073
C-14	Bosnia and Herzegovina	0.825	0.175	C-80	Myanmar	0.936	0.064
C-15	Botswana	0.899	0.101	C-81	Namibia	0.942	0.058
C-16	Brazil	0.701	0.299	C-82	Nepal	0.940	0.060
C-17	Brunei Darussalam	0.802	0.198	C-83	Netherlands	0.072	0.928
C-18	Bulgaria	0.417	0.583	C-84	New Zealand	0.042	0.958
C-19	Burkina Faso	0.950	0.050	C-85	Nicaragua	0.931	0.069
C-20	Burundi	0.907	0.093	C-86	Niger	0.922	0.078
C-21	Cote d’Ivoire	0.950	0.050	C-87	Nigeria	0.940	0.060
C-22	Cambodia	0.954	0.046	C-88	North Macedonia	0.817	0.183
C-23	Cameroon	0.916	0.084	C-89	Norway	0.056	0.944
C-24	Canada	0.058	0.942	C-90	Oman	0.861	0.139
C-25	Chile	0.595	0.405	C-91	Pakistan	0.952	0.048
C-26	China	0.068	0.932	C-92	Panama	0.937	0.063
C-27	Colombia	0.862	0.138	C-93	Paraguay	0.969	0.031
C-28	Costa Rica	0.877	0.123	C-94	Peru	0.796	0.204
C-29	Croatia	0.541	0.459	C-95	Philippines	0.838	0.162

C-30	Cyprus	0.054	0.946	C-96	Poland	0.400	0.600
C-31	Czech Republic	0.178	0.822	C-97	Portugal	0.151	0.849
C-32	Denmark	0.049	0.951	C-98	Qatar	0.575	0.425
C-33	Dominican Republic	0.979	0.021	C-99	Republic of Korea	0.084	0.916
C-34	Ecuador	0.965	0.035	C-100	Republic of Moldova	0.841	0.159
C-35	Egypt	0.986	0.014	C-101	Romania	0.654	0.346
C-36	El Salvador	0.984	0.016	C-102	Russian Federation	0.568	0.432
C-37	Estonia	0.088	0.912	C-103	Rwanda	0.909	0.091
C-38	Ethiopia	0.935	0.065	C-104	Saudi Arabia	0.527	0.473
C-39	Finland	0.078	0.922	C-105	Senegal	0.957	0.043
C-40	France	0.047	0.953	C-106	Serbia	0.706	0.294
C-41	Georgia	0.843	0.157	C-107	Singapore	0.120	0.880
C-42	Germany	0.066	0.934	C-108	Slovakia	0.628	0.372
C-43	Ghana	0.972	0.028	C-109	Slovenia	0.190	0.810
C-44	Greece	0.562	0.438	C-110	South Africa	0.865	0.135
C-45	Guatemala	0.949	0.051	C-111	Spain	0.054	0.946
C-46	Guinea	0.891	0.109	C-112	Sri Lanka	0.938	0.062
C-47	Honduras	0.956	0.044	C-113	Sweden	0.110	0.890
C-48	Hong Kong	0.161	0.839	C-114	Switzerland	0.125	0.875
C-49	Hungary	0.258	0.742	C-115	Tajikistan	0.960	0.040
C-50	Iceland	0.044	0.956	C-116	Thailand	0.587	0.413
C-51	India	0.464	0.536	C-117	Togo	0.947	0.053
C-52	Indonesia	0.893	0.107	C-118	Trinidad and Tobago	0.957	0.043
C-53	Iran (Islamic Republic of)	0.646	0.354	C-119	Türkiye	0.463	0.537
C-54	Iraq	0.905	0.095	C-120	Tunisia	0.898	0.102
C-55	Ireland	0.054	0.946	C-121	Uganda	0.920	0.080
C-56	Israel	0.099	0.901	C-122	Ukraine	0.803	0.197
C-57	Italy	0.106	0.894	C-123	United Kingdom	0.097	0.903
C-58	Jamaica	0.883	0.117	C-124	United Arab Emirates	0.142	0.858
C-59	Japan	0.044	0.956	C-125	United Republic of Tanzania	0.959	0.041
C-60	Jordan	0.895	0.105	C-126	United States of America	0.143	0.857

C-61	Kazakhstan	0.900	0.100	C-127	Uruguay	0.809	0.191
C-62	Kenya	0.966	0.034	C-128	Uzbekistan	0.933	0.067
C-63	Kuwait	0.828	0.172	C-129	Viet Nam	0.674	0.326
C-64	Kyrgyzstan	0.934	0.066	C-130	Yemen	0.866	0.134
C-65	Lao People's Democratic Republic	0.946	0.054	C-131	Zambia	0.945	0.055
C-66	Latvia	0.437	0.563	C-132	Zimbabwe	0.932	0.068

It is clear that from Table 4, 97 countries are assigned to Cluster 1, 35 countries are assigned to Cluster 2. The ranks of countries according to their GII values are given in Table 4 in parenthesis.

Clustering results of countries and GII ranks

Table 4

Cluster	Countries
1	Albania (84), Algeria (115), Angola (127), Argentina (69), Armenia (80), Azerbaijan (93), Bahrain (72), Bangladesh (102), Belarus (77), Benin (124), Bosnia and Herzegovina (70), Botswana (86), Brazil (54), Brunei Darussalam (92), Burkina Faso (120), Burundi (130), Cote d'Ivoire (109), Cambodia (97), Cameroon (121), Chile (50), Colombia (63), Costa Rica (68), Croatia (42), Dominican Republic (90), Ecuador (98), Egypt (89), El Salvador (100), Ethiopia (117), Georgia (74), Ghana (95), Greece (44), Guatemala (110), Guinea (132), Honduras (113), Indonesia (75), Iran (Islamic Republic of) (53), Iraq (131), Jamaica (76), Jordan (78), Kazakhstan (83), Kenya (88), Kuwait (62), Kyrgyzstan (94), Lao People's Democratic Republic (112), Madagascar (106), Mali (126), Mauritania (129), Mauritius (45), Mexico (58), Mongolia (71), Montenegro (60), Morocco (67), Mozambique (123), Myanmar (116), Namibia (96), Nepal (111), Nicaragua (108), Niger (125), Nigeria (114), North Macedonia (66), Oman (79), Pakistan (87), Panama (81), Paraguay (91), Peru (65), Philippines (59), Qatar (52), Republic of Moldova (56), Romania (49), Russian Federation (47), Rwanda (105), Saudi Arabia (51), Senegal (99), Serbia (55), Slovakia (46), South Africa (61), Sri Lanka (85), Tajikistan (104), Thailand (43), Togo (122), Trinidad and Tobago (101), Tunisia (73), Uganda (119), Ukraine (57), United Republic of Tanzania (103), Uruguay (64), Uzbekistan (82), Viet Nam (48), Yemen (128), Zambia (118), Zimbabwe (107)
2	Australia (25), Austria (17), Belgium (26), Bulgaria (35), Canada (15), China (11), Cyprus (27), Czech Republic (30), Denmark (10), Estonia (18), Finland (9), France (12), Germany (8), Hong Kong (14), Hungary (34), Iceland (20), India (40), Ireland (23), Israel (16), Italy (28), Japan (13), Latvia (41), Lithuania (39), Luxembourg (19), Malaysia (36), Malta (21), Netherlands (5), New Zealand (24), Norway (22), Poland (38), Portugal (32), Republic of Korea (6), Singapore (7), Slovenia (33), Spain (29), Sweden (3), Switzerland (1), Türkiye (37), United Kingdom (4), United Arab Emirates (31), United States of America (2)

The countries with the highest GII values are in Cluster 2. These countries in Cluster 2 are mostly upper-income or upper-middle-income countries. In Cluster 2, only India is a low-middle-income country, and Bulgaria, China, Malaysia, and Türkiye are upper-middle-income countries. Therefore, the countries with high-income levels, as well as rank high in terms of innovation indicators. This result is consistent with the study of Eren and Gelmez (2022) that clustered countries by using the GII (2021) data set. The countries in Cluster 1 consist of mostly lower-middle and low-income countries. Therefore, countries with low-income levels are at the bottom regarding innovation indicators. However, countries in Cluster 1 such as Albania, Armenia, Azerbaijan, Argentina, Belarus, Bahrain, Bosnia and Herzegovina, Brazil, Botswana, Brunei Darussalam, Chile, Colombia, Croatia, Costa Rica, Dominican Republic, Ecuador, El Salvador, Georgia, Greece, Guatemala, Indonesia, Iraq, Jamaica, Kazakhstan, Kuwait, Mauritius, Mexico, Montenegro, Namibia, North Macedonia, Oman, Panama, Paraguay, Peru, Republic of Moldova, Romania, Qatar, Saudi Arabia, Russian Federation, Serbia, South Africa, Slovakia, Thailand, Trinidad and Tobago are upper income or upper-middle-income countries.

Türkiye is located in Cluster 2, and Türkiye ranked 37th according to the 2022 GII rankings. Türkiye rose four places compared to the previous year. Per the findings in the report, while Türkiye had the highest performance in the human capital and research index, it showed the lowest performance in the institutions sub-component.

Final cluster prototype

Table 5

Sub-indexes	**Cluster 1**	**Cluster 2**
Institutions	50.38571	73.68434
Human capital and research	23.14634	50.74311
Infrastructure	35.81554	57.66978
Market sophistication	25.98683	48.68099
Business sophistication	23.16648	49.61126
Knowledge and technology outputs	15.23638	42.46534
Creative outputs	12.64314	39.36598

Upon examining Table 5, it becomes evident that all variables exhibit their highest values within Cluster 2. Consequently, countries that are members of the second cluster demonstrate superior performance regarding the GII. This result verifies the results presented in Table 4.

Summary statistics of the variables on the clusters is presented in Table 6.

Summary statistics of the variables on the clusters

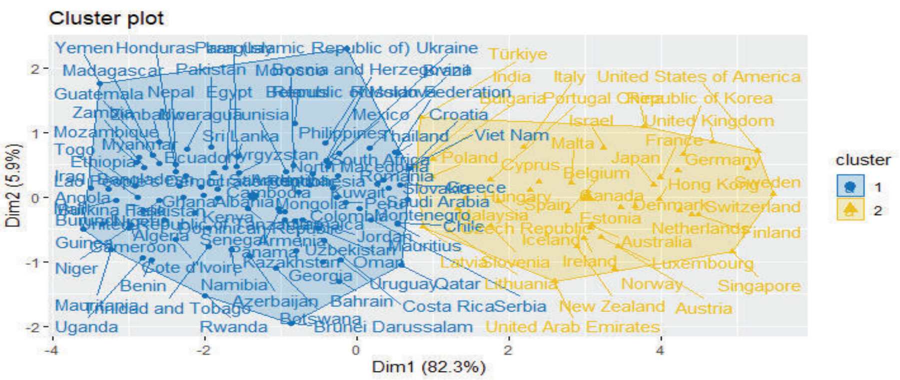
Table 6

Variable	Mean±SD	Median	Min-Max
Institutions	58.07273±14.90591	56.05	17.5 - 95.9
Human capital and research	32.62424± 15.47464	30.7	6-66.4
Infrastructure	43.50455± 12.9991	43.4	17.5- 95.9
Market sophistication	33.8697 ± 14.89748	32.45	4.4-80.8
Business sophistication	31.85379 ± 14.22105	27.15	10.2-69.8
Knowledge and technology outputs	24.46742 ± 15.51748	20.75	1.6-67.1
Creative outputs	21.70379 ± 15.26518	19.4	0.3-56.3

Figure 1 illustrate the distribution of countries in the two clusters.

Distribution of countries into two clusters

Figure 1



4.CONCLUSION AND DISCUSSION

Upon reviewing the clustering outcomes for the countries, it is clear that the two clusters consistently align with the rankings of the GII for the year 2022. Consequently, the reliability of the analytical findings coincides with our analysis. Furthermore, upon scrutinizing the clusters in conjunction with country profiles, it was evident that the employed analyses complemented each other. Countries characterized by high-income levels in Cluster 2 attained the top ranking. These countries also feature prominently among the leading countries in the GII. This observation underscores the correlation between high-income countries and their prominent positions regarding innovation indicators. Cluster 1 comprises primarily low-income and lower-middle countries, which illustrates countries with lower income levels are ordered similarly in the lower echelons of innovation indicators. Türkiye, our country, is in Cluster 2, characterized by high-income and upper-middle-income countries. Türkiye was positioned 41st according to the GII for the year 2021. Türkiye rose four places in 2022 to 37th place. Türkiye entered the top 40 for the first time, climbing 14 places in the Index in the last two years. Türkiye also maintained its 4th place among 36 upper-middle-income countries.

Since fuzzy cluster analysis evaluates the whole data set, it has the chance to reveal some similarities that indices expressing a single numerical value cannot reveal. The cluster analysis results and index rankings are parallel for the countries with high and low GII values. The top countries in GII rankings clustered in Cluster 2. The countries at the bottom of GII rankings clustered in Cluster 1. Cluster analysis is a method based on whole data having the chance to reveal some similarities that indices based on a single numerical value could be incapable. This type of clustering analysis shows the power of the index to reflect the data. Our study reveals the consistency of the rankings according to the GII index.

Future studies may consider comparing the results of these analyses through the application of additional or alternative quantitative methodologies for assessing the innovation performances of countries. Moreover, the measurement of innovation performances could be examined for the diverse categories of countries, such as according to their income levels or other categories that could logically have a relationship or connection with their innovation levels. Furthermore, the various clustering techniques can be compared to each other for different numbers of clusters.

REFERENCES

1. Alqararah, K. and Alnafrah, I. (2023), "Innovation-Driven Clustering for Better National Innovation Benchmarking", *Journal of Entrepreneurship and Public Policy*, 2045-2101. DOI 10.1108/JEPP-01-2023-0007.
2. Al Rahhal, T. and Rencber, O. F. (2022), *Comparisons of K-Means and Fuzzy C-Means Algorithm by Clustering Countries*, Ankara, Turkey: Iksad Publications.
3. Alptekin, N. and Yesilaydin, G. (2015), "OECD Ülkelerinin Sağlık Göstergelerine Göre Bulanık Kümeleme Analizi İle Sınıflandırılması", *Journal of Business Research Turk*, 7(4), pp. 137-155.
4. Aytekin, A., Ecer, F., Korucuk, S. and Karamasa, C. (2022), "Global innovation efficiency assessment of EU member and candidate countries via DEA-EATWIOS multi-criteria methodology", *Technology in Society*, 68, 101896. https://doi.org/10.1016/j.techsoc.2022.101896.
5. Bagdatli Kalkan, S. (2019), "Classification of European Union Member and Candidate Countries According to Ease of Doing Business Index Using Fuzzy Clustering", *Eurasian Econometrics*, *Statistics & Emprical Economics Journal*, 12, pp.25-35. http://dx.doi.org/10.17740/eas.stat.2019-V12-03.
6. Bezdek, J. C. and Hathaway, R. J. (1987), "Clustering with Relational c-Means Partitions From Pairwise Distance Data", *Mathematical Modelling*, 9(6), pp.435-439. https://doi.org/10.1016/0270-0255(87)90509-4.

7. Bulbul, S. and Camkiran, C. (2018), “Bankalarin Klasik ve Bulanik Yaklaşimlarla Siniflandirilmasi”, *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 20(2), pp.367-385. https://doi.org/10.26468/trakyasobed.464442.
8. Cebeci, Z. and Yildiz, F. (2015)*, “Bulanık C-Ortalamalar Algoritmasının Farklı Küme Büyüklükleri için Hesaplama Performansı ve Kümeleme Geçerliliğinin Karşılaştırılması”,* paper presented at the 9. Ulusal Zootekni Bilim Kongresi, Konya, pp.227-239.
9. Eren, H. and Gelmez, E. (2022), “Ülkelerin İnovasyon Performansına Göre Kümelenmesi; Entropi, Copras ve Aras Yöntemleriyle Değerlendirilmesi”, *Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi*, 12(3), pp.1546-1565. https://doi.org/10.30783/nevsosbilen.1153211.
10. Famalika, A. and Sihombing, P. R. (2021), “Implementation of k-means and K-Medians Clustering in Several Countries Based on Global Innovation Index (GII) 2018”, *Advance Sustainable Science, Engineering and Technology (ASSET)*, 3(1), pp.0210107-1 ~ 0210107-10. https://doi.org/10.26877/asset.v3i1.8461.
11. Ferraro, M. B. and Giordani, P. (2015), “A toolbox for fuzzy clustering using the Rprogramming language”, *Fuzzy Sets and Systems*, 279, pp.1-16. http://dx.doi.org/10.1016/j.fss.2015.05.001.
12. Gurtuna, F. and Polat, U. (2020), “Küresel İnovasyon Endeksi Verilerinin Kümeleme Analizi ile Değerlendirilmesi”, *Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, 35(2), pp.551-565. https://doi.org/10.21605/cukurovaummfd.792484.
13. Hancioglu, Y. (2016), “Küresel İnovasyon Endeksini Oluşturan İnovasyon Girdi ve Çıktı Göstergeleri Arasındaki İlişkinin Kanonik Korelasyon Analizi ile İncelenmesi: OECD Örneği”, *AİBÜ Sosyal Bilimler Enstitüsü Dergisi*, 16(4), pp.131-157. https://doi.org/10.11616/basbed.vi.455411.
14. Jankowska, B., Matysek-Jędrych, A. and Mroczek-Dąbrowska, K. (2017), “Efficiency of National Innovation Systems – Poland and Bulgaria in The Context of the Global Innovation Index”, *Comparative Economic Research*, 20(3), pp.77-94. https://doi.org/10.1515/cer-2017-0021.
15. Kaufman, L. and Rousseeuw, P.J. (1990). *Findings Groups in Data: An Introduction to Cluster Analysis*. John Wiley: USA.
16. Saravananathan, K. and Velmurugan, T. (2018), “Cluster Based Performance Analysis for Diabetic Data”, *International Journal of Pure and Applied Mathematics*, 119(16), pp.399-410.
17. Unlu, F. (2019), “Differences in Efficiency of Innovation Performance among Middle-income Countries: An Empirical Approach”, *International Journal of Economics and Innovation***,** 5(2), pp.213-229. https://doi.org/10.20979/ueyd.567962.
18. Zorlutuna, S. and Erilli, N. A. (2018), “Sosyo-Ekonomik Verilere Göre İllerin Bulanık C-Ortalamalar Yöntemi ile Sınıflandırılması: 2002-2008-2013 Dönemleri Karşılaştırması”, *İktisadi Yenilik Dergisi*, 5(2), pp.13-31.
19. WIPO (2022). *Global Innovation Index 2022*. [online] Avaliable at: https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2022-en-main-report-global-innovation-index-2022-15th-edition.pdf [Accesed 22 October 2023].

The role of the labor resources in the economic growth for sustainable development

Prof. Constantin ANGHELACHE PhD (actincon@yahoo.com)
Bucharest University of Economic Studies / Artifex University of Bucharest

Assoc. prof. Mădălina-Gabriela ANGHEL PhD (madalinagabriela_anghel@yahoo.com)
Artifex University of Bucharest

PhD Student Marius-Cristian RADUŢ (radutmc@gmail.com)
Bucharest University of Economic Studies

PhD Student Denis-Arthur STRIJEK (denis.strijek@gmail.com)
Bucharest University of Economic Studies

ABSTRACT

The article presents in its content a series of theoretical, statistical and econometric concepts regarding the situation on the labor market in Romania. This article starts from the premise that in order to achieve the sustainable development defined by Agenda 2023 and the document A sustainable future of Europe, it is necessary for the management of human resources to be performed in order to ensure economic growth. After the presentation of the theoretical concepts, we expanded the analysis on the situation of the labor market in Romania, with the help of statistical tools, presenting evolutions, graphically and in table form, of the labor resources, of the employed population by gender, residential areas and activities of the national economy, developments in the unemployment rate and others. Also, the distribution of the labor force on the territory of Romania was analyzed and the regions with the highest activation, employment and unemployment rates were highlighted.

Keywords: *active population, unemployment rate, employed population, indicator, jobs*

JEL Classification: *E24, J21, J64*

INTRODUCTION

In this article, the authors focused on the analysis of the human resources, which are the users or manipulators of the capital and resources necessary for economic activity, economic growth and sustainable development.

The way human resources are managed, which represents one of the sources of the labor force, largely depends on employment, materialized in the increase of labor productivity. It is true that until the concept of productivity and its increase is reached, the volume of human resources required, which are the conditions in which the employees carry out their work, the level of training, education, respectively the structure in which they will operate within the national economy, is addressed.

Employee performance also plays an important role, being the variable through which human resources contribute to economic growth.

The article is mainly based on the management of human resources at the macroeconomic level and how the transition from economic growth to sustainable development is achieved.

The data used are extracted from the databases of the National Institute of Statistics of Romania, Eurostat and other institutions, which ensure the comparability and harmonization of results in the process of obtaining management activity in sustainable development and human resources.

What is the role of the workforce in sustainable development and economic growth is a priority at the macroeconomic level.

In addition to the existence of the problem of performance dependence, which is subjectively perceived by employees, with the main aspects related to the requirements of the labor market, there are also a number of indicators that can refer to the workload, the size of lost time and the productivity of work. They have the role of highlighting the quantitative and qualitative results of human resource management in sustainable development.

One cannot ignore the latest events that have marked the world economy. Thus, the crisis caused by the COVID-19 pandemic, which later turned into an economic-financial crisis, puts in difficulty the states wishing for harmonious development and sustainable developments. More recently, with the outbreak of the Russian-Ukrainian conflict, the difficulties that have stood in front of sustainable developments are the energy and food crises.

Sustainable economic growth is that which has the main purpose of ensuring a decent living for all citizens and that which encourages and delimits itself from the negative impact on the environment or society.

One of the essential conditions in the effective management of labor resources is that it ensures some salary values high enough to ensure a decent living. Labor users need to break away from the idea of low income to maximize short-term profits. There are not only two pure categories in society (producers and consumers), each of the two categories is constantly in the role of the other (producers are also consumers in their turn).

LITERATURE REVIEW

Biea, D'Adamo, Hartley and Hesse (2019) analyzed what the salary dynamics were in Romania, while Chéron, Hairault and Langot (2013) carried out a life cycle analysis, analyzing the places of job vacancies, unemployment and focused on identifying a balance point between the indicators. Pulignano (2009) addressed the theme of international cooperation, transnational restructuring and virtual networks in Europe, while Schneider and Häge (2007) had previously addressed a somewhat bolder theme of withdrawing the authority of nation states and making way for Europeanization. Maestas, Mullen and Powell (2016) analyzed the effects of population aging on economic growth and the correlation between labor resources and productivity. Südekum (2003) talks about the macroeconomic theories and models used by the European Union, making a brief review of the economic doctrines that address the problem of unemployment. Kroft, Lange and Notowidigdo (2013) analyzed what employer behavior is and how it influences the labor market. Klein and Ventura (2009) analyzed productivity differences and how they influence labor relocation. Anghelache, Avram, Burea and Mirea (2019) emphasize the importance of access to financing from European capitals, there being a dependency between them and Romania's economic development. Crouch (2014) addresses the theme of labor market insecurity in times of crisis, what is the role of the state in these times and makes a grouping of states into areas, taking into account how labor market governance is achieved. Adda, Monti, Pellizzari, Schivardi and Trigari (2017) analyzed unemployment developments in Italy through the lens of the lack of correlation between employees' professional skills and the labor market. Moxon, Bacalso and Şerban (2021) made an analysis on how the life of young people is influenced by the Covid-19 pandemic. Hili, Lahmandi-Ayed and Lasram (2016) write a paper on how the labor market differentiates itself in the context of globalization. Lengyel, Borbála and Lilla (2017) study the labor market at the level of the European Union, after the economic crisis of 2008, focusing on long-term unemployment. Dorsett and Luccino (2018) describe the labor market as being in transition and talk about the role played by early experience in the employment decision of young people. Radu (2022) makes a synthesis of the existing situation on the labor market at the level of the European Union, emphasizing how young people have been affected by the effects of the Covid-19 crisis.

METHODOLOGY

The analysis is based on the use of the main indicators established at the time of the realization of the national strategy for the sustainable development of Romania. Strategy with a time horizon of 2030.

The statistical indicators were synthesized in order to highlight the basic elements in establishing the scientific framework for data processing. We can list some of these indicators that we have considered (employed population, labor force, unemployment, labor productivity, population waiting for a job and others).

Romania, being a member state of the UN and the European Union, has chosen to align itself with the sustainable development embodied in the 2023 Agenda, adopted by the UN during the September 2015 Summit.

Later, in 2017, the Council of the European Union adopts the document A sustainable future of Europe, which represents the European Union's response to the 2030 Agenda for sustainable development.

Within the states of the European Union and the states in the accession process, sustainable development represents the means by which national strategies are adjusted so as to achieve the extension and consolidation of the sustainability of a state.

Romania's strategy is based on three important pillars, economic, social and the state of the environment, and is based on the interest of the citizen, it focuses on innovation, optimism, resilience and confidence that through the set objectives, citizens will be ensured living conditions in a clean environment and a adequate standard of living, the way to achieve these desired (in a balanced, fair and efficient way).

The main objective of the Romanian sustainable development agenda is based on sustained economic growth, sustainable and open to all, on the full and productive employment of the labor force and the provision of decent jobs for employees.

The motto of the European Union, which wants to be implemented, is ”No one is left behind!”. The problem that arises is that no one is left behind in the European Union if they have an adequate management of human resources, if they increase their efficiency, embodied in productivity and if they have the possibility to use the available labor force.

Until 2030, the sustainable development strategy has as its first target the maintenance of the growth rate of the gross domestic product, if possible even above the average achieved in the European Union. Cooperation plays a role in the efficient use of capital resources, but also of labor, and the application of the principles of sustainable development and the constant improvement of the population's standard of living must represent a permanent objective, at least in the framework of human resources management throughout the next period.

DATA, RESULTS AND DISCUSSION

After analyzing the data obtained from Eurostat and the National Institute of Statistics, the following situation can be identified on the labor market in Romania: At the beginning of 2023, the population over the age of 15 numbered 15,957.4 thousand people, 51.7% being the share majority owned by women. Employed population numbered 7,806.4 thousand people, representing approximately 48.9% of the population with the right to work and the majority share was owned by men (57.6%).

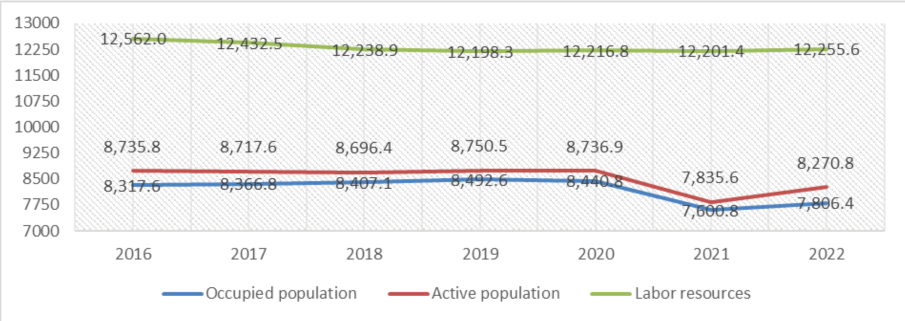
The inactive population over the age of 15 exceeded the number of the active population, at the end of 2022, being 7,686.5 thousand people, 62% of the majority being owned by women.

The total number of unemployed was 464.4 thousand people, 37.9% representing the share of women, 33% were found in the urban environment, and the share among young people between the ages of 15-25 was 25.5%.

Labor resources represent 63% of the total population, representing 12,255.6 thousand people, 3.8% are unemployed and 32.2% inactive people.

The evolution of labor resources, the active population and the employed population at the end of 2022

- thousands of people -



Data source: National Institute of Statistics. Data processed by the authors

Interpreting the data presented in figure number 1, it can be seen that on January 1, 2023, the employed population was 7,806.4 thousand people. Making the proportion of employed people in labor resources we observe that we obtain an employment rate of approximately 63.7%.

We notice that after the crisis generated by the Covid-19 pandemic, the population is starting to become active again on the labor market, the number increasing from 7,835.6 thousand people to 8,270.8 thousand people. This growth is not maintained when we look at the employed population, although we have an increase from 7,600.8 thousand people to 7,806.4 thousand people, the difference between the employed and the active population is increasing, which leads us to think that the labor market still does not have a sufficient offer for all job seekers.

The employed population by gender, residence and activities of the national economy

Table no. 1

Economic activities	Occupied population					in % of the total		
	Total	*Men*	*Women*	*Urban*	*Rural*	*Total*	*Men*	*Women*
TOTAL	***7.806.452***	***4.492.654***	***3.313.799***	***4.623.522***	***3.182.931***	***100%***	***100%***	***100%***
Agriculture, forestry and fishing	*878.389*	*591.171*	*287.218*	*83.687*	*794.703*	*11,25%*	*13,16%*	*8,67%*
Total industry	*1.797.700*	*1.090.704*	*706.995*	*1.091.829*	*705.870*	*23,03%*	*24,28%*	*21,33%*
Extractive industry	*56.192*	*48.870*	*7.322*	*29.200*	*26.992*	*0,72%*	*1,09%*	*0,22%*
Manufacturing industry	*1.541.083*	*888.758*	*652.325*	*933.971*	*607.112*	*19,74%*	*19,78%*	*19,69%*
Production and supply of electricity and thermal energy, gas, hot water and air conditioning	*82.542*	*66.101*	*16.441*	*64.176*	*18.365*	*1,06%*	*1,47%*	*0,50%*
Water distribution; sanitation, waste management, decontamination activities	*117.883*	*86.975*	*30.907*	*64.482*	*53.401*	*1,51%*	*1,94%*	*0,93%*
construction	*765.179*	*715.583*	*49.596*	*369.565*	*395.614*	*9,80%*	*15,93%*	*1,50%*
Wholesale and retail trade; repair of motor vehicles and motorcycles	*1.382.299*	*612.023*	*770.276*	*941.770*	*440.529*	*17,71%*	*13,62%*	*23,24%*
Transport and storage	*555.470*	*480.821*	*74.649*	*337.747*	*217.723*	*7,12%*	*10,70%*	*2,25%*
Hotels and restaurants	*191.440*	*73.554*	*117.886*	*135.501*	*55.939*	*2,45%*	*1,64%*	*3,56%*
Information and communications	*201.830*	*134.264*	*67.566*	*178.190*	*23.640*	*2,59%*	*2,99%*	*2,04%*
Financial intermediation and insurance	*115.102*	*39.651*	*75.450*	*95.013*	*20.089*	*1,47%*	*0,88%*	*2,28%*
Real estate transactions	*25.232*	*12.645*	*12.588*	*17.726*	*7.506*	*0,32%*	*0,28%*	*0,38%*
Professional, scientific and technical activities	*220.824*	*92.541*	*128.283*	*191.492*	*29.331*	*2,83%*	*2,06%*	*3,87%*
Administrative service activities and support service activities	*214.125*	*153.403*	*60.722*	*129.185*	*84.941*	*2,74%*	*3,41%*	*1,83%*

Economic activities	Occupied population					in % of the total		
Public administration and defense; social insurance from the public system	*422.081*	*253.480*	*168.600*	*293.208*	*128.872*	*5,41%*	*5,64%*	*5,09%*
Education	*370.401*	*84.846*	*285.555*	*274.525*	*95.876*	*4,74%*	*1,89%*	*8,62%*
Health and social assistance	*449.097*	*75.584*	*373.513*	*334.277*	*114.820*	*5,75%*	*1,68%*	*11,27%*
Performing, cultural and recreational activities	*67.180*	*31.434*	*35.746*	*54.597*	*12.583*	*0,86%*	*0,70%*	*1,08%*
Other activities of the national economy	*150.104*	*50.948*	*99.156*	*95.211*	*54.893*	*1,92%*	*1,13%*	*2,99%*

Data source: National Institute of Statistics. Data processed by the authors

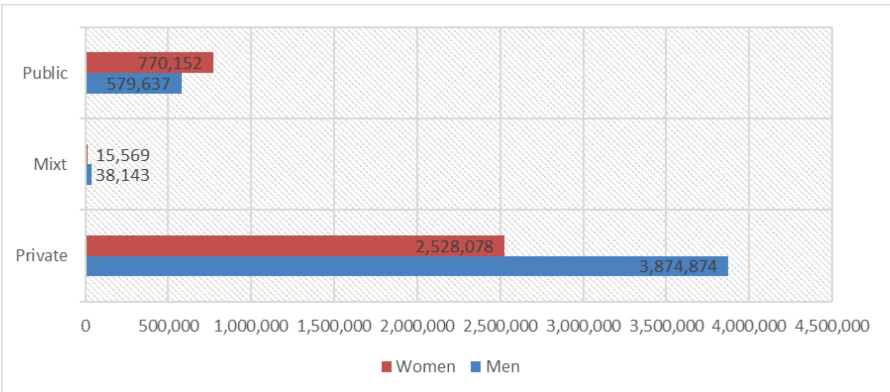
From table no. 1 we notice that the distribution of the employed population by fields suggests that the weight of non-agricultural activities has a weight of 88.75%. Industry has a weight of 23.03% and trade has a weight of 17.71% of the total employed population.

At the same time, we observe a pronounced degree of feminization of certain economic activities, Health and social assistance (83.17%), Education (77.09%), Financial intermediation and insurance (65.55%), respectively Hotels and restaurants (61.58 %).

On the other hand, men have higher employment rates in the following economic activities, Construction (93.52%), Extractive Industry (86.97%), Transport and storage (86.56%) but also in Public Administration and Defense; social insurance from the public system (60.05%).

The employed population according to the form of owners by gender on December 31, 2022

Figure no. 2



Data source: National Institute of Statistics. Data processed by the authors

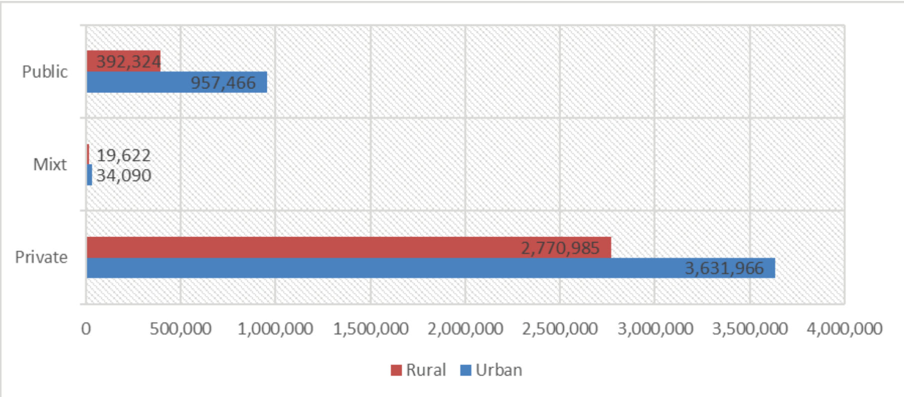
From the previous figure we see that the form of ownership of employers plays a role in attracting employees of a certain gender. Employers with public capital tend to employ female employees, a proportion of 57.06% female employees can be observed, compared to 42.94% male employees.

Things change in the opposite direction when we observe the private environment and employers with mixed capital, in the private environment, 60.52% represent male employees and only 39.48% represent female employees. Where the forms of ownership intersect, we observe an employment rate of male employees of (71.01%) compared to the employment rate of female employees (28.99%).

The total employment rate among men, regardless of the form of ownership, is 57.55% and among women is 42.45%. We can say that companies with public capital ensure a balance in the labor market, ensuring jobs to a greater extent for women, compared to companies with private or mixed capital, which predominantly employ men.

Employed population by type of owners, by residence environments on December 31, 2022

Figure no. 3



Data source: National Institute of Statistics. Data processed by the authors

After analyzing the situation of the employed population structured by gender, we studied the role of the residence environment on the labor market. Thus, we find that 59.23% of the employed population is found in urban areas, while 40.77% is found in rural areas.

We practically observe that discrimination by residence environment is approximately 1.68 percentage points higher than by gender.

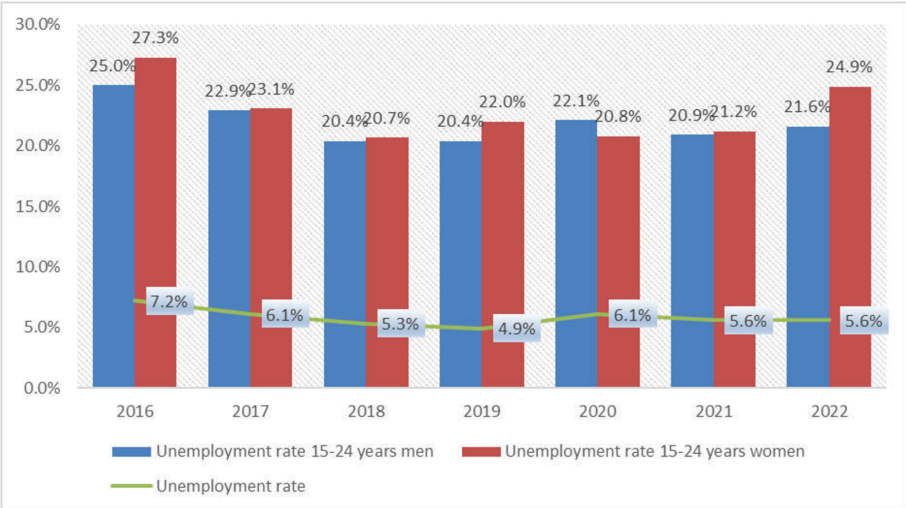
The employed population among employers with private capital comes predominantly from the urban environment (70.93%) compared to the rural environment (29.07%). The form of mixed ownership presents an imbalance of 63.47% compared to 43.28%, between the urban and rural environment.

This time, we see from figure no. 3, that the form of ownership of the private workplace comes closest to equilibrium, offering jobs in proportion to 56.72% in the urban environment and 43.28% in the rural environment.

The unemployment rate in 2022 remained at the level of the previous year, being 5.6%. By gender, a difference of 1 percentage point can be observed (6% unemployment rate for men, respectively 5% for women). The area of residence shows a difference of 5.7 percentage points (unemployment registered in the urban area is 3.2% compared to 8.9% in the rural area).

The evolution of the unemployment rate and unemployment among young people, by sex

Figure no. 4



Data source: National Institute of Statistics. Data processed by the authors

From figure no. 4 it can be observed that the highest level of unemployment among young people between the ages of 15-24 was recorded, 22.8% being the highest in the last 5 years.

The structure of active and inactive people, the activity rate (R.A.), the employment rate (R.O.) and the unemployment rate (R.Ș.) by macro-regions, regions and age groups, as of December 31, 2022

Table no. 2

Macroregions Regions Age groups	Total popuation	Active people - *persons* -			Inactive people	R.A.	R.O.	R.Ș.
		Total	occupied	unemployed		*percentages*		
TOTAL								
working age population (15-64 years)	12.255.585	8.191.120	7.728.335	462.785	4.064.466	66,8%	63,1%	5,6%
1. MACROREGION 1								
working age population (15-64 years)	3.108.705	2.090.249	2.003.105	87.144	1.018.456	67,2%	64,4%	4,2%
1.1. NORTHWEST								
working age population (15-64 years)	1.647.767	1.133.846	1.098.492	35.354	513.921	68,8%	66,7%	3,1%
1.2. CENTER								
working age population (15-64 years)	1.460.938	956.403	904.612	51.790	504.535	65,5%	61,9%	5,4%
2. MACROREGION 2								
working age population (15-64 years)	3.439.856	2.252.851	2.097.169	155.681	1.187.005	65,5%	61,0%	6,9%
2.1. NORTH EAST								
working age population (15-64 years)	1.974.831	1.333.388	1.245.029	88.359	641.443	67,5%	63,0%	6,6%
2.2. SOUTH EAST								
working age population (15-64 years)	1.465.024	919.462	852.140	67.323	545.562	62,8%	58,2%	7,3%
3. MACROREGION 3								
working age population (15-64 years)	3.368.326	2.381.498	2.255.224	126.274	986.828	70,7%	67,0%	5,3%
3.1. SOUTH MUNTENIA								
working age population (15-64 years)	1.789.655	1.176.635	1.083.260	93.375	613.020	65,7%	60,5%	7,9%
3.2. BUCHAREST-ILFOV								
working age population (15-64 years)	1.578.670	1.204.862	1.171.964	32.899	373.808	76,3%	74,2%	2,7%
4. MACROREGION 4								
working age population (15-64 years)	2.338.699	1.466.522	1.372.837	93.685	872.177	62,7%	58,7%	6,4%
4.1. SOUTH-WEST OLTENIA								
working age population (15-64 years)	1.196.915	750.130	683.102	67.028	446.785	62,7%	57,1%	8,9%
4.2. WEST								
working age population (15-64 years)	1.141.784	716.392	689.735	26.657	425.392	62,7%	60,4%	3,7%

Data source: National Institute of Statistics. Data processed by the authors

From table no. 2 we note that the third macroregion, which includes the Bucharest-Ilfov Region, has the largest number of people of working age, has the largest number of active persons, the largest number of employed persons but loses the largest number of unemployed, in the detrimeter of Macroregion two.

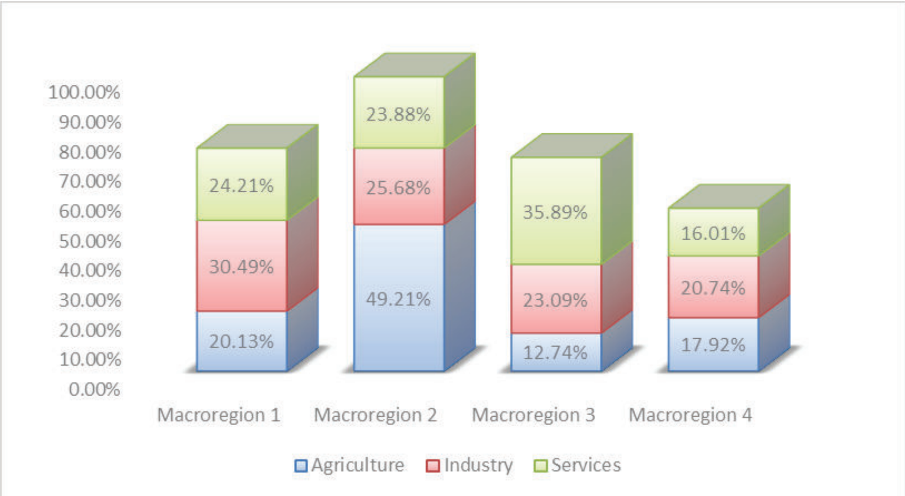
We note that the lowest unemployment rates are recorded in the Bucharest-Ilfov region (2.7%), the North-West region (3.1%) and the West region (3.7%), below the national unemployment rate of 5 .6%.

The highest activity rate is found in the Bucharest-Ilfov region (76.3%), the North-West region (68.8%) and the North-East region (67.5%). The highest occupancy rate is also recorded in the Bucharest-Ilfov region (74.2%), followed by the North-West region (66.7%) and the North-East region (63%).

We also observe a configuration of four macro-regions that group two regions of the country each, in order to analyze which sectors of the national economy they emphasize for attracting human resources.

The structure of the employed population in the four macro-regions (grouped by two regions each), by sector of the national economy

Figure no. 5



Data source: National Institute of Statistics. Data processed by the authors

From figure no. 5 we note that Macroregion 2 is mainly occupied in agriculture (49.21%) while the other three macroregions register shares of employment in agriculture of 20% or less. The top of the most industrialized macroregions is the following Macroregion 1 (30.49%), Macroregion 2 (25.68%) and Macroregion 3 (23.09%).

Services are highly developed in Macroregion 3 (35.89%), Macroregion 1 (24.21%) and Macroregion 2 (23.88%).

It can be observed that Macroregion 4 has the lowest weights among the four macroregions, the only category that does not occupy the last position is agriculture, with 17.92%. Of course, one explanation for these results is the share of the employed population in this macro-region. Macroregion 3 holds 29.08% of the total employed population, while Macroregion 2 holds 27.32% of the employed population, Macroregion 1 holds 25.81% of the total employed population, while Macroregion 4 holds only 17.78% of employed population.

Under these conditions, the result from the industry activity sector is a significant one, with 20.74% of the population employed in industry, being only 2.35 percentage points less than Macroregion 3.

CONCLUSIONS

The research and development is one of the important pillars of sustainable development, it is necessary to pay more attention, it also requires optimal funding, so as to ensure the contribution of experts to the fulfillment of the sustainable development strategy. If until the emergence of the covid-19 pandemic crisis, Romania had managed to take steps towards reducing the gaps with other countries, we notice that, post-crisis, it fails to fully use the labor resources at its disposal, employment requiring investments in industry, in developing a strategy for digitization, robotization and increasing the professional capacity of employees.

The financial-material resource also plays an important role in sustainable development, through the macroeconomic correlations that are established. Although the capitalization is sufficient, a restructuring and a consolidation of the banking sector is necessary. Because by maintaining only two banks with majority Romanian capital, the rest being territorial branches of European banks, registered in Romania, cooperation in the development of large-scale Romanian projects is not guaranteed or achieved.

Romania must make better use of the tourism potential, by supporting commercial companies with a tourism profile (HORECA), in such a way that the increase in tourism capacities can ensure the absorption of unoccupied human resources or of the population reconverted after giving up a series of economic sectors unfriendly to the environment.

Also, an important problem is aging in the field of agriculture, by practicing agritourism, human resources could be identified that could choose to move, voluntarily, to the countryside and start practicing agro-industrial activities, the infusion with the young generation could ensure the development of rural personnel. In this way, the possibility of implementing advanced production methods (irrigation, fertilization or advanced agrotechnical methods) is created, which can supply the markets with products and services for the Romanian consumer.

The sustainable development requires ensuring a harmonious framework for the employment of human resources in order to be able to use financial and capital resources efficiently.

References

1. Adda, J., Monti, P., Pellizzari, M., Schivardi, F., Trigari, A., *Unemployment and Skill Mismatch in the Italian Labor Market*. A project coordinated by IGIER-Bocconi, supported by J.P. Morgan, 2017
2. Anghelache C., Avram D., Burea D., Mirea M., *The accession of european funds -priority for Romania*. Romanian Statistical Review Supplement, 1, 124-137, 2019
3. Biea N., D'Adamo G., Hartley J., Hesse N., *Wage dynamics in România*, European Economy – Economic briefs, Comisia Europeană, 2019
4. Chéron, A., Hairault, J.O., Langot, F., *Life-Cycle Equilibrium Unemployment*. Journal of Labor Economics*, 31* (4), 843-882, 2013
5. Crouch, C., *The Governance of Labour Market Insecurity during the Crisis*, Stato e Mercato, no. 100 (1) (2014): 69–86
6. Dorsett, R., Luccino, P., *Young people's labour market transitions: The role of early experiences*. Labour Economics, Vol 54, pp. 29-46, 2018
7. Hili, A., Lahmandi-Ayed, R., Lasram, H., *Differentiation, labor market and globalization.* The Journal of International Trade & Economic Development, 25 (6), 809-833, 2016.
8. Klein, P., Ventura, G., *Productivity differences and the dynamic effects of labor movements*. Journal of Monetary Economics, 56 (8), 1059-1073, 2009.
9. Kroft, K., Lange, F. and Notowidigdo, M., *Duration Dependence and Labor Market Conditions: Evidence from a Field Experiment*. Quarterly Journal of Economics, 128 (3), 1123-1167, 2013.
10. Lengyel, G., Borbála, G., Lilla, T., *Expected long-term labour market resilience and the European Uunion's perceived role results of an online policy delphi*. Society and Economy 39, no. 2 (2017): 165–82
11. Maestas, N., Mullen, K., Powell, D., *The effect of population ageing on economic growth, the labor force and productivity*. National Bureau Of Economic Research, Cambridge, Working Paper no. 22452, 2016.
12. Moxon, D., Bacalso, C., Șerban, A., *Beyond the pandemic: The impact of COVID-19 on young people in Europe*. Brussels. EuropeanYouth Forum, 2021
13. Pulignano V., *International Cooperation, Transnational Restructuring and Virtual Networking in Europe.* European Journal of Industrial Relations, 15(2), 187-205, 2009
14. Radu, I., T*he socio-economic situation of youth in Europe under the effects of the Covid-19 pandemic*, Romanian Statistical Review, Supplement, 7, 23-38, 2022
15. Schneider V., Häge F., *Europeanization and the retreat of the state*. Journal of European Public Policy, 15 (1), 1-19, 2007
16. Südekum, J., *Macroeconomic Theories of Unemloyment and the 'European Labour Market Model.' In Agglomeration and Regional Unemployment Disparities: A Theoretical Analysis with Reference to the European Union*, NED-New edition., 47–70. Peter Lang AG, 2003
17. www.insse.ro – official site of the National Institute of Statistics of Romania
18. https://ec.europa.eu/eurostat

"Forecasting Maize Production In Romania: A BSTS Model Approach"

AMIR KHALEEL HASSOO (amir.hasoo@soran.edu.iq)
Soran University Department of Statistics

ABSTRACT

Forecasting agricultural production is crucial for strategic planning and policy-making. This study employs the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania for the period 2023-2027. The BSTS model, known for its flexibility and ability to incorporate multiple components like trends, seasonality, and regression effects, is particularly suitable for capturing the complex dynamics of agricultural time series data. Historical data on maize production from 1961 to 2022 in FAOSTAT website was used to train the model, ensuring robust and accurate forecasts. The results indicate a steady increase in maize production over the forecast period, with projected figures of 11,341,460 metric tons in 2023, rising to 11,437,732 metric tons in 2024, 11,558,277 metric tons in 2025, 11,594,832 metric tons in 2026, and 11,578,402 metric tons in 2027. These forecasts provide valuable insights for policymakers, farmers, and stakeholders in the agricultural sector, enabling them to make informed decisions regarding resource allocation, market strategies, and food security planning. The study highlights the efficacy of the BSTS model in agricultural forecasting and underscores its potential application in other areas of economic and environmental planning. Future research could enhance the model by incorporating additional variables such as climate data and economic indicators, further improving the accuracy and reliability of agricultural forecasts.

Keywords: *BSTS Model, Maize Production Forecasting, Agricultural Planning, Romania.*

1. INTRODUCTION

Agricultural production plays a pivotal role in the global economy, providing essential food security and raw materials for various industries. In Romania, maize is a significant crop, contributing substantially to the country's agricultural output and economic stability. Accurate forecasting of maize production is crucial for effective planning and decision-making, enabling stakeholders to optimize resource allocation, market strategies, and policy development. Despite its importance, maize production forecasting in Romania often relies on traditional models that fail to capture the complex dynamics and uncertainties inherent in agricultural time series data. This shortfall results in suboptimal planning and increased vulnerability to market and climatic fluctuations (Dragomir et al., 2022). This study aims to apply the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania from 2023 to 2027. The BSTS model is chosen for its flexibility and capability to incorporate multiple components, such as trends, seasonality, and regression effects, making it well-suited for agricultural forecasting. By leveraging historical production data, this research seeks to generate precise forecasts that can enhance strategic planning and policy formulation. Accurate forecasts are critical for stakeholders including policymakers, farmers, and market analysts, as they enable better resource management, enhance market efficiency, and contribute to food security (Popescu et al., 2018).

Despite numerous studies on agricultural forecasting, there is a noticeable gap in applying advanced statistical models like BSTS in this domain. Existing research predominantly relies on traditional methods that often fail to address complex and non-linear patterns in agricultural data. This study addresses this gap by demonstrating the efficacy of the BSTS model for maize production forecasting in Romania, and suggests potential enhancements by integrating additional variables such as climate data and economic indicators (Petre, 2017). Further research could also explore comparative studies between BSTS and other advanced forecasting models to evaluate their relative strengths and limitations (Jun, 2019).

The provided data highlights the dominance of the United States and China in global maize production, with Romania ranking 16th among the top producers. This underscores the need for advanced forecasting techniques to better manage and predict maize production in both major and smaller-scale producing countries (FAOSTAT website).

Top Ten Countries Maize Production in the World

Table 1

Number	Country	Million metric tonnes
1	United States	348.8
2	China	277.2
3	Brazil	109.4
4	Argentina	59.0
5	European Union	53.0
6	India	33.7
7	Mexico	26.6
8	Ukraine	26.2
9	Indonesia	23.6
10	South Africa	16.1
16	**Romania**	**8.0**

“Faostat”. Retrieved 27 February 2024.

2. METHODOLOGY

2.1. Materials

The primary materials used in this study include historical maize production data from Romania, as well as software tools for data analysis and modeling. The historical data, which spans from 1961 to 2022, was obtained from the National Institute of Statistics of Romania and the Food and Agriculture Organization (FAO) database. For the data analysis and modeling, the R programming language was employed, leveraging packages such as `bsts` for the (BSTS) model time series analysis techniques.

2.2. Data Collected

The data collected for this study encompasses annual maize production figures in metric tons from 1961 to 2022. This data includes information on total production per hectare per hectare the FAOSTAT database.

2.3. Bayesian Structural Time Series (BSTS)

The BSTS model was specified to include components such as local linear trends, seasonal effects, and regression terms for the selected features. The model was initialized using historical data, and hyperparameters were tuned to optimize the model's performance. The BSTS model can be expressed as a combination of different components (Scott and Varian , 2013).

$$Yt = \mu_t + S_t + \beta' X_t + \epsilon_t \dots\dots\dots (1)$$

Where:

• Y_t is the observed value at time t.

• μ_t represents the local linear trend (level and slope).

• S_t represents the seasonal effect.

• $\beta' X_t$ represents the regression terms.

• ϵ_t is the observation noise, typically assumed to be Gaussian with variance $\sigma\epsilon_2$.

The study began with data preprocessing to clean the raw data, address missing values, and identify underlying patterns. A Bayesian Structural Time Series (BSTS) model was then specified to include local linear trends, seasonal effects, and regression terms, initialized using historical data with optimized hyperparameters. The BSTS model, grounded in state space modeling, utilizes a mathematical framework where observed data is influenced by unobserved variables, incorporating both a state equation (for hidden state evolution) and an observation equation (linking states to observed data). These models are crucial for techniques like the Kalman filter and are widely used for smoothing, filtering, and forecasting in time series analysis. Bayesian methods, including Markov Chain Monte Carlo (MCMC), were employed for parameter estimation, ensuring convergence by sampling from the posterior distribution. The model was trained on data from 1961 to 2022, with cross-validation to avoid overfitting, and was then used to forecast maize production from 2023 to 2027, providing predictions with confidence intervals (Wang and Zivot, 2000).

2.4. Approach Bayesian Structural Time Series (BSTS) modeling in R software

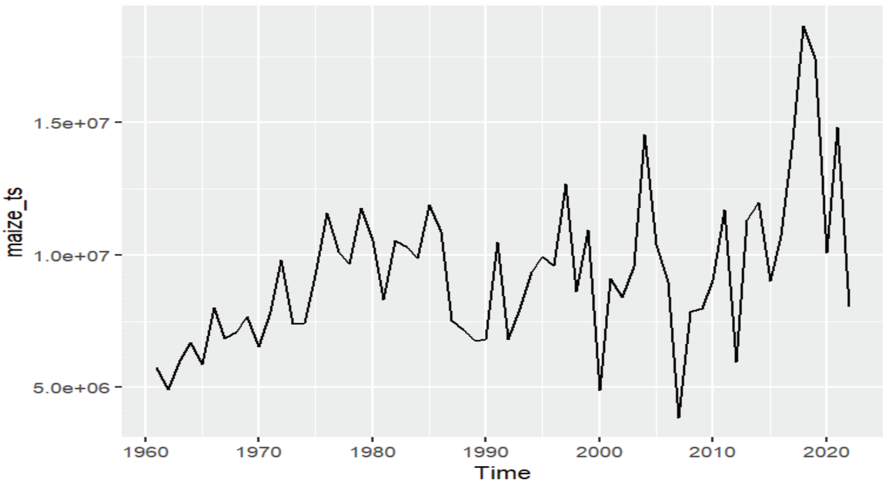
To approach Bayesian Structural Time Series (BSTS) modeling in R, start by installing and loading the necessary packages, such as (bsts, ggplot2, tseries and dplyr). Prepare your data, ensuring it's formatted as a time series object or a data frame with date and value columns. Specify the BSTS model by defining its components—trend, seasonality, and any regression components if needed—using the bsts function. Fit the model to data, then evaluate its performance through summaries and diagnostics. Generate forecasts with the fitted model and visualize the results using plotting functions like ggplot2. Refine the model as needed by adjusting components, priors, or hyperparameters based on initial results. For additional guidance, consult the bsts package documentation and seek out online tutorials (Pol et,. al 2018).

3. RESULTS AND DISSCUSION

Romania's maize production is influenced by various factors, including climatic conditions, agricultural practices, and economic variables. The substantial range and variability in production highlight the sensitivity of maize yields to these factors. Accurate forecasting and understanding of these patterns are essential for effective agricultural planning and policy-making.

Maize production in Romania

Figure 1



The descriptive statistics for maize production in Romania reveal a wide range in values, with a minimum production of 3,853,918 and a maximum of 18,663,940. The mean production is 9,293,781, with a standard deviation of 2,833,017, indicating substantial variability in maize production over the observed period. Understanding these descriptive statistics is crucial for forecasting as they provide context for the data, helping to identify trends and patterns that can inform model specifications and improve the accuracy of predictions.

Descriptive statistics

Table 2

Variable	**Min**	**Max**	**Mean**	**S.D**
Maize	3853918	18663940	9293781	2833017

The p-value is 0.04294, which is below the commonly used significance level of 0.05. This p-value indicates that there is significant evidence to reject the null hypothesis of non-stationarity, suggesting that the time series is likely stationary.

Augmented Dickey-Fuller Test

Table 3

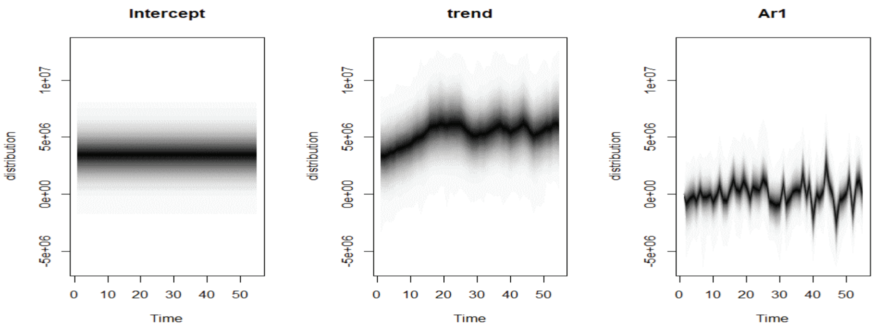
Statistic	**Value**
Dickey-Fuller Statistic	-3.5725
Lag Order	3
P-value	**0.04294**

3.1. Identification

Initialized with three components: a static intercept to adjust for a constant baseline, a local level to account for changes in the average level over time, and an autoregressive component with one lag to handle autocorrelation. These components collectively help the model capture different aspects of the time series data and improve its forecasting ability.

Components maize of BSTS Model

Figure 2

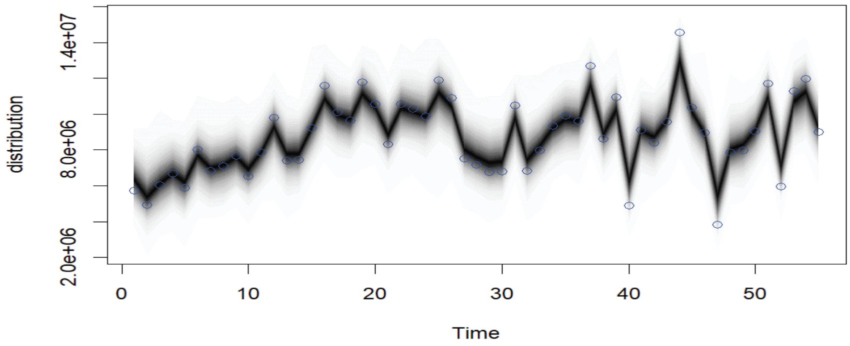


3.2. SELECT FIT MODEL

Markov Chain Monte Carlo (MCMC) is a method used in Bayesian statistics to estimate model parameters when direct computation is complex. It involves generating a sequence of samples from the posterior distribution of the parameters by iteratively updating values based on the likelihood of the observed data. To fit the model, MCMC samples are used to perform posterior predictive checks and evaluate model fit using criteria such as Deviance Information Criterion (DIC), Widely Applicable Information Criterion (WAIC), or leave-one-out cross-validation (LOO). A “blue point” typically represents a data point or criterion score that indicates the model’s fit. Effective use of MCMC allows for assessing how well the Bayesian model captures the data, with favorable fit criteria suggesting a better model fit.

Training data for predicted value and actual values of maize production time series by using BSTS

Figure 3



3.3. Forecast from 2023 to 2027

Table 3 provides forecasted values for the time series from 2023 to 2027. The projections show an upward trend over the initial years, with values increasing from 11,341,460 in 2023 to 11,558,277 in 2025. However, the growth rate slows down in 2026, with the forecasted value reaching 11,594,832, and slightly decreases to 11,578,402 in 2027. This indicates a general upward trend with some stabilization or minor decline towards the end of the forecast period.

The predicted values of maize production in Romania from 2023 to 2027

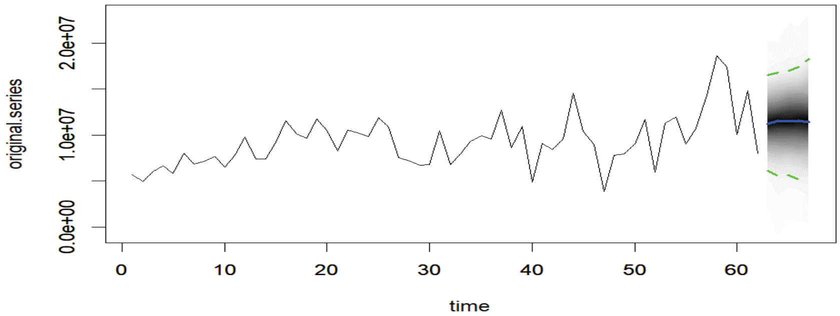
Table 4

Date	Forecast
2023	11341460
2024	11437732
2025	11558277
2026	11594832
2027	11578402

Furthermore, the BSTS model was used to forecast Romania annual maize output for the year 2022. The figure, as shown in Figure 5, demonstrates that the anticipated values for 2022 roughly coincide with the actual values, suggesting convergence between the expected and observed series.

Predicted values of maize production in 2027

Figure 5

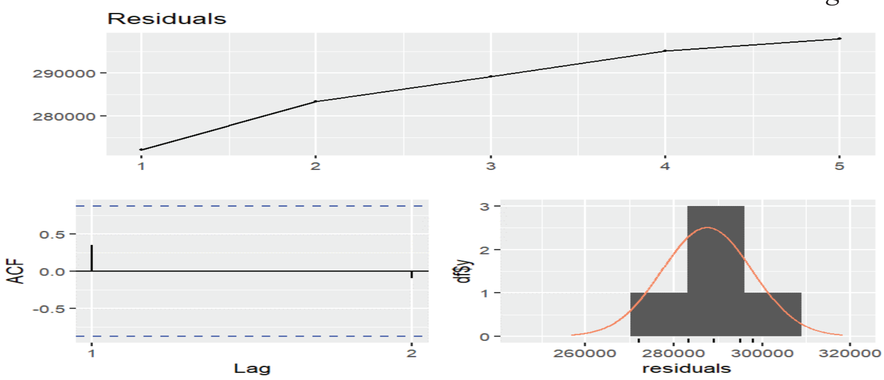


Model Checking

The Box-Ljung test, a statistical test designed to assess the presence of autocorrelation in time series residuals, was performed on the BSTS model residuals using the given output. The obtained p-value of 0.9585 was more than the 0.05 criterion of significance. This means that there isn't enough data to justify the presence of autocorrelation in the model's residuals. As a result, it is possible to infer that the model adequately describes the autocorrelation structure in the data.

Residuals from BSTS

Figure 6



DISCUSSION

The study demonstrates the effectiveness of the Bayesian Structural Time Series (BSTS) model in forecasting maize production in Romania, revealing a steady increase in projected yields from 2023 to 2027. The model's capability to incorporate trend, seasonality, and regression components makes it highly suitable for capturing the complex dynamics of agricultural data, which is crucial for strategic planning and policy-making. However, the study also identifies several limitations. The reliance on historical data may lead to overfitting, potentially reducing the model's accuracy when applied to new or unseen data (Osiewalski et al., 2020). Overfitting can occur if the model becomes too attuned to historical patterns that may not persist in the future. To address these limitations and enhance the model's predictive power, future research should consider incorporating additional variables, such as climate data (e.g., temperature, precipitation) and economic indicators (e.g., market prices, trade policies). These variables could provide a more comprehensive understanding of the factors influencing maize production. Additionally, validating the model's generalizability across different contexts and regions could help assess its robustness and adaptability. The findings hold significant implications for stakeholders, including policymakers, farmers, and market analysts. Improved forecasting accuracy enables better resource management, more effective policy decisions, and enhanced market strategies. By addressing the identified limitations and expanding the model's scope, future research can contribute to more reliable and actionable agricultural forecasts (Steel , 2010).

CONCLUSION AND RECOMMENDATION

The study applied the Bayesian Structural Time Series (BSTS) model to forecast maize production in Romania, revealing a positive trend from 2023 to 2027. While the model effectively captures the dynamics of agricultural data, its reliance on historical data and the risk of overfitting may limit its accuracy for future predictions. To enhance the model's robustness, future research should incorporate external variables, such as climate and economic factors, and apply cross-validation techniques to prevent overfitting. Additionally, exploring the model's generalizability to other crops and regions will provide valuable insights into its broader applicability. Regular updates and integration with complementary forecasting methods are also recommended to maintain accuracy and relevance.

To improve maize production in Romania, it is crucial to implement climate-resilient practices. This includes adopting drought-resistant varieties and employing efficient water management techniques to ensure stable yields despite weather variability. Precision agriculture technologies, such as satellite imagery and soil sensors, can further optimize input use, boost productivity, and minimize environmental impact. Enhancing farmer education and training on modern techniques and technologies will empower farmers to implement best practices, leading to higher and more sustainable maize yields.

Future research should continue to explore innovative forecasting methods and integrate findings from diverse agricultural studies to further refine predictions and practices. By addressing the identified limitations and adopting the recommended strategies, stakeholders can contribute to more reliable forecasting and sustainable maize production in Romania.

REFRENCES

1. Dragomir, V., Ioan Sebastian, B., Alina, B., Victor, P., Tanasă, L. and Horhocea, D., 2022. An overview of global maize market compared to Romanian production. *Romanian Agriculture Research*, *39*, pp.535-544.
2. Popescu, M., Teican, D.I., Mocuta, D. and Cristea, S., 2018. The evolution of the recording of maize hybrids in Romania.
3. Petre, I.L., 2017. The effect of maize production and consumption on prices in Romania. In *Agrarian Economy and Rural Development-Realities and Perspectives for Romania. 8th Edition of the International Symposium, November 2017, Bucharest* (pp. 53-59). Bucharest: The Research Institute for Agricultural Economy and Rural Development (ICEADR).
4. Jun, S., 2019. Bayesian structural time series and regression modeling for sustainable technology management. *Sustainability*, *11*(18), p.4945.
5. Scott, S.L. and Varian, H.R., 2013. *Bayesian variable selection for nowcasting economic time series* (No. w19567). National Bureau of Economic Research.
6. Wang, J. and Zivot, E., 2000. A Bayesian time series model of multiple structural changes in level, trend, and variance. *Journal of Business & Economic Statistics*, *18*(3), pp.374-386.
7. Pole, A., West, M. and Harrison, J., 2018. *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC.
8. Osiewalski, J., Wróblewska, J. and Makieła, K., 2020. Bayesian comparison of production function-based and time-series GDP models. *Empirical Economics*, *58*, pp.1355-1380.
9. Steel, M.F., 2010. Bayesian time series analysis. In *Macroeconometrics and time series analysis* (pp. 35-45). London: Palgrave Macmillan UK.