DOI 10.2478/sbe-2024-0042
SBE no. 19(3) 2024

PREDICTIVE ANALYTICS IN FINANCE USING THE ARIMA MODEL. APPLICATION FOR BUCHAREST STOCK EXCHANGE FINANCIAL COMPANIES CLOSING PRICES

Cristi SPULBAR
University of Craiova, Craiova, Romania

Cezar Cătălin ENE
University of Craiova, Craiova, Romania

Abstract:
Stock markets can be volatile, thus accurate predictions can greatly help investors and stakeholders to make wise financial choices. The main goal of this paper is to test how well the Autoregressive Integrated Moving Average (ARIMA) model can capture and predict changes in closing prices. The ARIMA model is the combination of autoregressive (AR) and moving average (MA) processes of an integrated or differenced time series model. Moreover, the selected model is part of the time series analysis under prediction algorithms, the purpose of the research being to predict the prices of the selected shares. Our analysis compiles daily trading data of financial companies on the BSE using a quantitative methodology. We preprocess the dataset to ensure reliability and accuracy, then conduct an exploratory data analysis to identify underlying patterns and correlations. Next, we use the ARIMA model, carefully optimizing the parameters through selection and a validation process, to forecast the closing prices over a 30-day period, we also evaluate the model's performance. The preliminary results show that the ARIMA model seems to be efficient at predicting closing stock prices accurately. The model appears to understand the patterns and fluctuations in the stock market data, which gives useful information about future price changes. The forecasts we generated show similarities with the real market results, capturing important patterns, and making it a viable option for forecasting market performance.

Key words: *financial forecasting, stock price prediction, time series analysis, ARIMA model*

1. Introduction

The Bucharest Stock Exchange (BSE) is the principal stock market in Romania, playing an important role in the country's evolution of financial markets. It was established in 1995, after almost 50 years the in which BSE activity was suspended by the communist regime. After that the BSE has shown grown and has become a central hub for securities trading in Eastern Europe with 87 listed companies and over 73.267 million USD market capitalization. It operates in a regulatory framework that aligns with European Union standards and it's a financial intermediary between companies who want to raise capital and for investors to trade financial instruments such as equities and bonds. Among the prominent listings on the BSE are BRD – Groupe Société Générale (BRD) and Banca Transilvania (TLV), two of the largest financial institutions in Romania, which are the object of our study.

BRD is part of the global Société Générale Group being identified as one of the largest banks in Romania with over 2.2. million clients, providing a wide array of services from domestic operations to retail banking, corporate banking and investment banking. Its extensive network and substantial customer base make BRD a significant entity on the BSE being a choice for investors to consider in their portfolio.

Banca Transilvania (TLV) is the largest bank in Romania by assets. Its activity is concentrated on 3 main directions: small-medium size enterprises, corporate and retail. TLV offers a wide range of banking services and has a significant impact on the market due to its size and operational reach and we can say that the TLV represents a barometer of Romanian financial sector health.

One frequently debated issue in financial time series analysis is the predictability of holding period returns on risky financial assets. For the efficient market hypothesis (EMH) (Fama, 1970) to hold in its weak form, stock returns must exhibit serial independence, meaning there should be no autocorrelation, an issue that we tacked into account when we applied the methodology. This implies that current stock returns fully reflect all past financial information, leaving no patterns or trends that can be exploited for predictive purposes. The EMH suggests that stock prices follow a random walk, a term widely used in the field, where future movements are unpredictable and past price information is already incorporated into current prices.

Historically, testing the EMH has involved specific model formulations to empirically refute or support the hypothesis. Much of the research has concentrated on predicting common stock returns or indexes of stock returns. These studies have aimed to determine whether stock prices exhibit random walk behaviour or if there are patterns that suggest predictability. The results of these investigations have been mixed, with some studies finding evidence of predictability, while others support the notion of a random walk.

In this context, the Bucharest Stock Exchange (BSE) presents an interesting case for analyzing stock price predictability. The Romanian stock market is characterized by its volatility, influenced by both domestic economic conditions and external factors. Volatility in stock prices is a common phenomenon in emerging markets like Romania, where market dynamics can be affected by macroeconomic policies, political stability, regional stability and global financial trends. Thus, we have employed the Autoregressive Integrated Moving Average (ARIMA) model to forecast the stock prices for the financial companies mentioned above. The ARIMA model is a popular and versatile tool in the field of financial predictions, mostly used on time series analysis and its combining three key components: Autoregressive (AR), Integrated (I), and Moving Average (MA). The AR component captures the relationship between an observation and a number of lagged observations, while the MA component models the relationship between an observation and a lagged residual error. The Integrated component is used to make the time series stationary, which is essential for accurate forecasting.

While extensive research has been conducted on applying ARIMA models to forecast stock prices in various markets, there is a noticeable gap in studies specifically focusing on the Romanian stock market, particularly the stocks of major financial institutions such as BRD and TLV. Most existing studies have concentrated on larger, more developed markets, leaving a gap in understanding the applicability and effectiveness of ARIMA models in emerging markets like Romania. Emerging markets, such as Romania, present unique challenges and opportunities for predictive modeling due to their higher volatility and susceptibility to both domestic economic conditions and external factors. The limited research on applying ARIMA models to these markets means that there is a lack of comprehensive understanding regarding the model's performance in such environments. Furthermore, the Romanian stock market's evolving nature and the significant role of key financial institutions like BRD and TLV necessitate focused studies to provide reliable predictive insights.

The current body of literature shows a significant focus on developed markets, where the market dynamics, regulatory frameworks, and economic conditions differ greatly from those in emerging markets. Emerging markets often experience more pronounced market fluctuations and are influenced by a different set of economic indicators and investor behaviors. Therefore, models that perform well in developed markets may not necessarily yield the same level of accuracy in emerging markets. This underlines the importance of conducting market-specific studies to tailor predictive models to the unique conditions of emerging markets like Romania. Also, the Romanian stock market, being relatively less researched, offers a fertile ground for exploring the nuances of stock price behavior in an emerging market context. The performance of financial institutions like BRD and TLV, which are pivotal to the Romanian economy, can provide deeper findings into the broader market trends and investor sentiment. Taking this into considerantion, this study aims to fill the gap in literature and contribute to a better understanding of stock price predictability in emerging markets.

To address the gap from the existing body of literature, Spulbar and Ene (2024) examine the interplay of macroeconomic variables and the dynamics of the Romanian financial market using a Vector Autoregression (VAR) approach. Their study focuses on the BET index, monetary policy rate, inflation rate, and RON/USD exchange rate, providing critical insights into how these variables interact and influence market behavior. This research emphasizes the significance of the central bank's policy rate and the complex impact of inflationary trends on currency valuation, reinforcing the importance of considering macroeconomic factors in financial forecasting. This aligns with our focus on using ARIMA models to predict stock prices in an emerging market context. While in the mentioned article was employed a VAR approach to understand broader market dynamics, this study narrows down to the predictability of individual stock prices using ARIMA models. This complementary perspective enhances the robustness of our findings, as it situates our model within a broader economic context, emphasizing the interplay between macroeconomic stability and market predictability.

The research questions guiding this study are:

- How accurately can the ARIMA model predict short-term stock prices for BRD and TLV within the context of the emerging Romanian stock market?

- How does the volatility characteristic of emerging markets, such as Romania, affect the predictive performance of the ARIMA model for BRD and TLV stocks?
- What insights can the application of ARIMA models to BRD and TLV stocks provide for investors and stakeholders regarding market dynamics and investment strategies in emerging markets?

In our study, we utilize the ARIMA model to short-term forecast the closing prices of BRD and TLV stocks. The analysis involves preprocessing the dataset to ensure data quality and accuracy, followed by exploratory data analysis to uncover underlying patterns and correlations. We then apply the ARIMA model, carefully selecting and validating the parameters to optimize forecasting accuracy.

The rest of our article is organized as follows Section 2. The Literature Review examines previous studies on ARIMA models applied to stock price predictions, in which we highlight methodologies and findings from various researchers. Section 3, the methodology details the data sources, preprocessing steps, and the ARIMA modeling approach, including model identification, estimation, and diagnostic checking. The Section 4, findings and discussion section presents the results of our ARIMA models for BRD and TLV, comparing predicted and actual stock prices and discussing the model's performance. Finally, the last section number 5, the conclusion summarizes the study's findings, emphasizes the effectiveness of the ARIMA models in financial forecasting, and suggests directions for future research.

2. Literature review

Our study used daily stock price data for BRD and TLV from January 2018 to December 2023, sourced from TradingView. We applied the Box and Jenkins (1970) methodology for ARIMA modeling, which involves three stages: Identification, Estimation, and Diagnostic Checking. For stationarity testing, we used the Augmented Dickey-Fuller (ADF) test, revealing non-stationary series that required differencing to achieve stationarity. Similarly, Adebiyi et al. (2014) applied ARIMA models to stock price data from the New York Stock Exchange (NYSE) and Nigeria Stock Exchange (NSE). They also emphasized the importance of stationarity, using differencing to transform non-stationary series. The methodology employed by Adebiyi et al. aligns closely with ours, reinforcing the robustness of our approach. Our forecast for BRD and TLV stocks demonstrated high accuracy, with predicted values closely matching the actual observed data. For instance, the forecasted closing prices for BRD in January 2024 show a minimal deviation from the actual prices, indicating the model's precision. This finding is consistent with Adebiyi et al. (2014), who reported that their ARIMA models provided satisfactory short-term forecasts. They illustrated this with graphs showing close alignment between predicted and actual stock prices for both Nokia and Zenith Bank. This close match underscores the reliability of ARIMA models for short-term financial predictions.

In their study, Jarrett and Schilling (2008) also applied the ADF test to determine the presence of unit roots in the daily stock returns for firms listed on the Frankfurt Stock Exchange. Their methodology involved differencing the data to achieve stationarity, similar to our approach. They then developed ARIMA models for prediction, concluding that these models could effectively forecast stock returns. The alignment in methodologies, particularly in the use of ADF tests and differencing, reinforces the validity of our approach. The authors reported similar success in forecasting stock returns for firms listed on the Frankfurt Stock Exchange. They found that their ARIMA models provided reliable predictions, evidenced by close alignment between forecasted and actual stock returns.

To forecast the Indian stock market, specifically the S&P BSE Sensex, monthly closing stock indices data from 2007 to 2012 were used and tested for stationarity using the Durbin-Watson test Banerjee (2014). The necessary data transformations were applied to achieve stationarity. Banerjee (2014) found success with the ARIMA(1, 0, 1) model, where the predicted values for 2013 closely matched the observed values. This methodological alignment, particularly the focus on stationarity and the use of differencing, supports the robustness of our approach.

Using data from 2012 to 2018, in the study made by Alhwatmeh (2022), the ARIMA model was applied to forecast stock prices on the Amman Stock Exchange (ASE). The study, which focused on 15 well-known banks, employed the Box-Jenkins methodology to identify, estimate, and validate the ARIMA models. Stationarity was achieved through differencing, ensuring the robustness of the time series data. Alhwatmeh (2022) reported a similar success with the ARIMA(1, 1, 0) model for ASE, where the predicted values closely matched the observed values. The study highlighted the ARIMA model's strong potential for short-term stock price prediction, aligning with our findings for BRD and TLV, and further validating the efficacy of ARIMA models for accurate financial forecasting.

Prakhar et al. (2022) applied ARIMA models to forecast stock prices on the National Stock Exchange (NSE) of India, using data from January 2018 to February 2021. Their study involved multiple forecasting models, including ARIMA, Facebook Prophet, and ETS (Error, Trend, Seasonal) models, to identify the most effective method. The ARIMA model they employed required differencing to transform non-stationary series into stationary ones, aligning closely with our methodological approach. In our analysis, we identified ARIMA(1, 2, 2) for BRD and ARIMA(7, 1, 56) for TLV as the best-fit models based on the Akaike Information Criterion (AIC), Schwarz Criterion (SC), and Hannan-Quinn Criterion (HQC). We validated these models through diagnostic checks, including the Ljung-Box Q-statistic and the analysis of autocorrelation (ACF) and partial autocorrelation functions (PACF). Prakhar et al. (2022) also used criteria such as AIC and the Ljung-Box Q-statistic to select and validate their ARIMA models. They compared the ARIMA model's performance with that of the Facebook Prophet and ETS models. The ARIMA model's selection and validation process in their study, which involved similar diagnostic checks, supports the reliability of our approach and the robustness of our chosen ARIMA models for BRD and TLV. Our forecasts for BRD and TLV stocks demonstrated high accuracy, with predicted values closely matching the actual observed data for January 2024.

The ARIMA(4, 1, 0) model was selected for the NSE dataset after conducting similar criteria and diagnostic checks (Mahadik et al., 2021). This selection process, which mirrors our own, highlights the reliability and robustness of our chosen ARIMA models for BRD and TLV. Mahadik et al. (2021) found that the ARIMA model provided accurate short-term forecasts for their datasets, with high accuracy confirmed through various performance metrics. They also noted that ARIMA outperformed LSTM in some cases, especially in short-term predictions where the data was more suitable for linear models. The alignment in forecasting accuracy between our study and Mahadik et al.'s findings further supports the efficacy of ARIMA models for accurate financial forecasting.

In another research study, the authors Khan and Alghulaiakh (2020) also applied ARIMA models to forecast stock prices using Netflix historical data from April 2015 to April 2020. Their study involved testing both auto ARIMA and custom ARIMA(p, d, q) models to find the most accurate forecasting model. The dataset's non-stationary characteristics required differencing to achieve stationarity, aligning with our methodological approach. They employed similar criteria and diagnostic checks to select and validate their ARIMA models and found that ARIMA(1, 1, 33) provided the best results, with high accuracy, reporting similar success with their ARIMA(1, 1, 33) model for Netflix, achieving high accuracy as indicated by MAPE values close to 99%. Their findings that ARIMA models deliver precise short-term forecasts align with our results for BRD and TLV stocks, further validating the efficacy of ARIMA models for accurate financial forecasting.

Historical data of Asian Paints Ltd. spanning the past 20 years was used by Lohar et al. (2024) to apply ARIMA models for forecasting stock prices. Their methodology included differencing non-stationary data to achieve stationarity, which closely parallels our approach. This consistency in methodology reinforces the robustness of our ARIMA models for BRD and TLV. The ARIMA(5, 2, 0) model for Asian Paints Ltd. was selected based on criteria similar to ours, involving iterative testing and validation using AIC and the Ljung-Box Q-statistic. Lohar et al. (2024) reported high accuracy in short-term forecasts with their model, as demonstrated by the close alignment between predicted and actual stock prices. This successful application of ARIMA models further supports the reliability of our approach and confirms the effectiveness of ARIMA models for accurate financial forecasting.

With the purpose to forecast stock prices for Sun Pharmaceutical Industries Ltd., the ARIMA(24, 1, 47) model was identified by Meher et al. (2021) , using similar criteria to those used in our study such as AIC and the Ljung-Box Q-statistic to ensure model adequacy. Meher et al. (2021) reported that this approach yielded high accuracy in short-term forecasts, with a close match between predicted and actual stock prices. The use of comparable model selection and validation techniques in both studies supports the reliability of our approach and the robustness of our chosen ARIMA models for BRD and TLV. This alignment in methodologies further validates the effectiveness of ARIMA models in financial forecasting.

Kumar et al. (2021) selected ARIMA(3, 1, 3) as the most appropriate model for forecasting Areca nut prices. They used similar criteria, including AIC and BIC, to select and validate their ARIMA models. Kumar et al. (2021) reported similar success with their ARIMA(3, 1, 3) model for Areca nut prices, achieving high accuracy in short-term forecasts as evidenced by the close match between predicted and actual prices. Their findings that ARIMA models provide precise short-term forecasts align with our results for BRD and TLV stocks, further validating the efficacy of ARIMA models for accurate financial forecasting.

By using criteria such as AIC and BIC for model selection and validation, Kumar et al. (2023) selected the ARIMA(1, 1, 4) model as the most suitable for forecasting RSS-1 rubber prices. The authors reported high accuracy in short-term forecasts with this model, as evidenced by the close match between predicted and actual prices. Their success with the ARIMA(1, 1, 4) model underscores the reliability of our approach in predicting BRD and TLV stock prices, further validating the effectiveness of ARIMA models across different markets and data types.

3. Methodology

The stock data for BRD – Groupe Société Générale (BRD) and Banca Transilvania (TLV) was sourced from TradingView, covering the period from January 3, 2018, to December 29, 2023. TradingView was chosen as the data source because it is the platform used by the Bucharest Stock Exchange (BSE) to display historical stock prices of listed companies, ensuring consistency and reliability in the data. The dataset includes daily frequency data, capturing the opening, maximum, minimum, and closing prices for both BRD and TLV. The objective is to predict the stock prices for January 2024 and compare the predicted values with the actual ones collected from TradingView.

This dataset includes four key elements: open price, low price, high price, and close price. For the purpose of this research, the closing price is chosen as the primary variable for prediction. The closing price is considered because it encapsulates all trading activities of the day, reflecting the final consensus value of the stock for that trading session. While various factors such as open, high, low, close, adjusted close, and volume can be used to predict stock prices, the closing price is particularly significant. It provides a global snapshot of the stock's performance for the day, incorporating the effects of all intra-day fluctuations and trading volumes.

To perform the ARIMA modelling and other statistical analyses, we used EViews software version 12. EViews is a powerful econometric and statistical software that facilitates the estimation, forecasting, and simulation of time series data. It provides a vast suite of tools for model building, diagnostic testing, and result visualization, making it an ideal choice for conducting time series analysis.

The Autoregressive Integrated Moving Average (ARIMA) model was employed to forecast the stock prices. The ARIMA model, as proposed by Box and Jenkins (1970), involves a systematic methodology divided into three main stages: Identification, Estimation, and Diagnostic Checking and Forecasting.

The first stage involves identifying the appropriate model for the time series data. This requires analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine the values of the autoregressive (p) and moving average (q) parameters. Additionally, stationarity of the time series is checked using the Augmented Dickey-Fuller (ADF) test. If the series is non-stationary, differencing is applied to achieve stationarity, denoted by the parameter (d). The resulting model is specified as ARIMA(p, d, q).

The mathematical formulation of the ARIMA model is typically denoted as ARIMA(p, d, q), where:

- p represents the number of lag observations in the model (autoregressive part);

- d is the number of times that the raw observations are differenced (integrated part) to achieve stationarity;
- q denotes the size of the moving average window (moving average part).

In the estimation stage, the parameters of the ARIMA model are estimated using maximum likelihood estimation (MLE) or other suitable methods. The goal is to find the values of p, d, and q that best fit the historical data. The model's performance is evaluated using criteria such as the Akaike Information Criterion (AIC), Schwarz Criterion (SC), and Hannan-Quinn Criterion (HQC) to ensure the optimal model is selected.

Once the model parameters are estimated, diagnostic checks are performed to validate the model. This involves analysing the residuals of the model to ensure they behave like white noise, meaning they are randomly distributed with no significant autocorrelation. The Ljung-Box Q-statistic is used to test for autocorrelation in the residuals. If the model passes these diagnostic checks, it is deemed adequate for forecasting. The ARIMA model is then used to generate forecasts for the stock prices of BRD and TLV for January 2024. These forecasts are compared with the actual observed values to evaluate the model's predictive accuracy.

To ensure the quality and accuracy of the data, preprocessing steps were undertaken. We did not have any missing values in the dataset so there was no need for any interpolation methods to maintain continuity in the time series. The dataset was then divided into training and testing sets: the training set was used to build the ARIMA model, while the testing set was reserved for evaluating the model's performance.

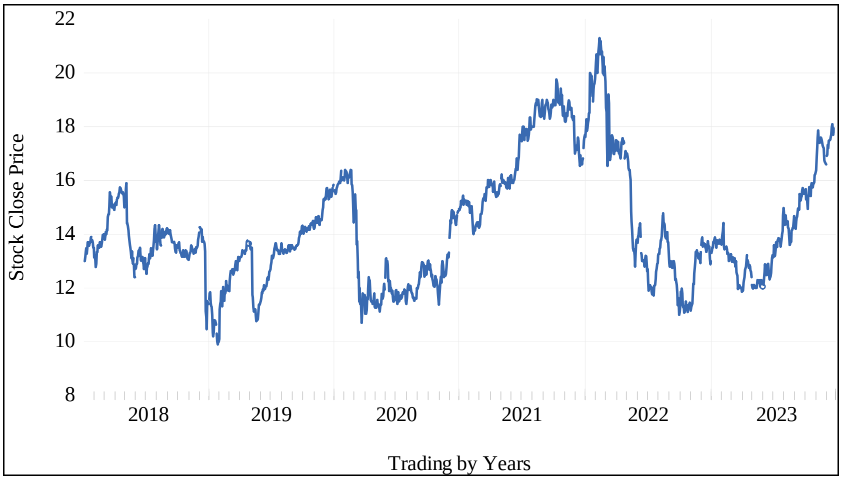
3.1. ARIMA (p, d, q) for the BRD – Groupe Société Générale stock

In time series analysis, most economic variables exhibit non-stationarity, meaning their statistical properties such as mean and variance change over time. Non-stationary data often contains trends, seasonality, or other systematic patterns that need to be removed to make the data suitable for modelling and forecasting. The process of differencing is commonly used to transform non-stationary data into a stationary form. For stationary time series data, the Autoregressive Moving Average (ARMA) model, denoted as ARMA(p, q), can be applied. However, if the data is non-stationary, the Autoregressive Integrated Moving Average (ARIMA) model, denoted as ARIMA(p, d, q), is used. In ARIMA, the 'Integrated' component (d) represents the number of differences required to make the time series stationary.

The stock data for BRD – Groupe Société Générale (BRD) used in this study it's in the interval of January 3, 2018, to December 29, 2023, with a total of 1,498 daily observations. To determine whether the time series is stationary, we start by plotting the original data to observe its pattern. Figure 1 below shows the time series plot of the closing prices of BRD.

From the graph below, it is evident that the time series follows a random walk pattern, which is indicative of non-stationarity. The series does not exhibit a constant mean or variance over time, confirming the need for differencing to achieve stationarity.

Fig. 1. Historical close price of BRD



Source: Author's computations based on historical close price data of BRD

To further investigate the stationarity of the BRD time series, we analyze its stationarity using the Augmented Dickey-Fuller (ADF) Test. The results of the ADF test for the BRD closing price series are presented in Table 1.

Table 1. ADF Test Results for BRD Closing Prices

Null Hypothesis: CLOSE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 1 (Automatic - based on SIC, maxlag=36)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.242834	0.4647
Test critical values:	1% level	-3.964190	
	5% level	-3.412817	
	10% level	-3.128391	

*MacKinnon (1996) one-sided p-values.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CLOSE)
Method: Least Squares
Sample (adjusted): 1/05/2018 12/29/2023
Included observations: 1496 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-0.006629	0.002956	-2.242834	0.0251
D(CLOSE(-1))	0.081216	0.025818	3.145761	0.0017
C	0.086892	0.041357	2.101042	0.0358
@TREND("1/03/2018")	1.45E-05	1.53E-05	0.947123	0.3437

R-squared	0.009491	Mean dependent var	0.003289
Adjusted R-squared	0.007499	S.D. dependent var	0.247896
S.E. of regression	0.246965	Akaike info criterion	0.043526
Sum squared resid	90.99930	Schwarz criterion	0.057726
Log likelihood	-28.55770	Hannan-Quinn criter.	0.048817
F-statistic	4.765252	Durbin-Watson stat	1.992992
Prob(F-statistic)	0.002605		

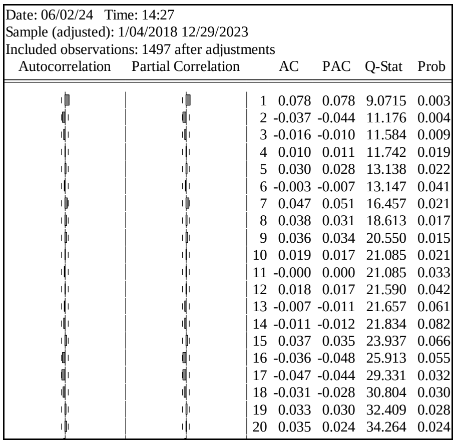
Source: Author's computations based on historical close price data of BRD

The p-value for the constant (C) is 0.0358, which is less than 0.05, indicating that the constant is significant and should be included in the model. The p-value for the trend is 0.3437, which is greater than 0.05, suggesting that the trend may not be significant in this context. However, for the purpose of the test, both the intercept and the trend were included. The ADF test statistic is -2.242834.

To determine if the series is stationary, we compare this statistic to the critical values at the 1%, 5%, and 10% levels. The test statistic is not lower than the critical values at the 1%, 5%, or 10% significance levels (-3.964190, -3.412817, -3.128391 respectively). This means we cannot reject the null hypothesis that the series has a unit root. Since the ADF test statistic does not fall below the critical values, and the p-value associated with the test statistic is 0.4647, which is greater than 0.05, we fail to reject the null hypothesis. This indicates that the BRD closing price series is non-stationary, so we have a non-stationary variable, and we are working with an AR(I)MA model (p, d, q).

After identifying the non-stationarity of the original BRD time series, we apply the first differencing to transform it into a stationary series. The differenced series, denoted as "DCLOSE," represents the changes in the closing prices from one day to the next. Figure 3 below displays the line graph of the differenced series "DCLOSE" for the BRD closing prices. This graph provides a visual representation of how the first differencing affects the time series, helping us to assess the elimination of trends and seasonality. Figure 2 presents the correlogram (ACF plot) of the differenced series "DCLOSE." The correlogram helps to verify the stationarity by showing the autocorrelations of the differenced series at various lags.

Fig. 2. Correlogram at first difference of BRD close price

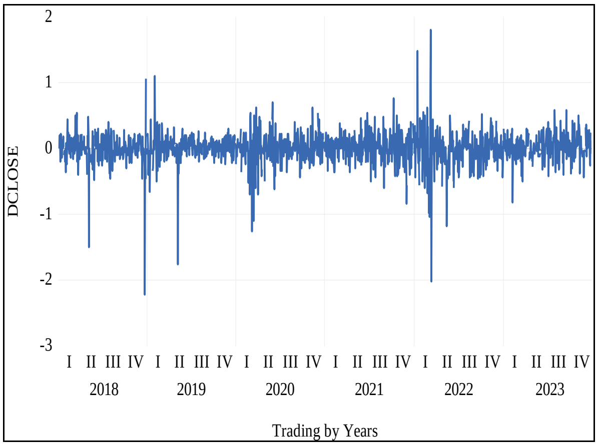


Date: 06/02/24 Time: 14:27
Sample (adjusted): 1/04/2018 12/29/2023
Included observations: 1497 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.078	0.078	9.0715	0.003
		2	-0.037	-0.044	11.176	0.004
		3	-0.016	-0.010	11.584	0.009
		4	0.010	0.011	11.742	0.019
		5	0.030	0.028	13.138	0.022
		6	-0.003	-0.007	13.147	0.041
		7	0.047	0.051	16.457	0.021
		8	0.038	0.031	18.613	0.017
		9	0.036	0.034	20.550	0.015
		10	0.019	0.017	21.085	0.021
		11	-0.000	0.000	21.085	0.033
		12	0.018	0.017	21.590	0.042
		13	-0.007	-0.011	21.657	0.061
		14	-0.011	-0.012	21.834	0.082
		15	0.037	0.035	23.937	0.066
		16	-0.036	-0.048	25.913	0.055
		17	-0.047	-0.044	29.331	0.032
		18	-0.031	-0.028	30.804	0.030
		19	0.033	0.030	32.409	0.028
		20	0.035	0.024	34.264	0.024

Source: Author's computations based on historical close price data of BRD

Fig. 3. Graphic of first difference of BRD close price



Source: Author's computations based on historical close price data of BRD

To identify the best model, we compared models using three key metrics: Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC), and Hannan-Quinn Criterion (HQC). These metrics helped us balance model complexity against fit, with smaller values indicating better models. After many iterations, we identified the ARIMA(1, 2, 2) model as the optimal choice for our data set, BRD. The model includes one autoregressive term, differencing the data twice to achieve stationarity, and two moving average terms. This structure effectively captures the data's underlying patterns and trends. The results of the tests for ARIMA(1, 2, 2) are the studied as follow.

First, we ensure that the model satisfies the requirements for a stable univariate process. This involves checking that the residuals of the model are white noise. To do this, we examine the Ljung-Box Q-statistic, where the Null Hypothesis is that the residuals are white noise. In our case, as shown in Figure 4, the absence of significant spikes in the Autocorrelation (AC) and Partial Autocorrelation (PAC) plots confirms that the residuals of the selected ARIMA model are white noise.

Table 3. ARIMA(1, 2, 2) Estimation Output – BRD

Dependent Variable: D(CLOSE,2)
Method: ARMA Maximum Likelihood (OPG – BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.32E-05	4.90E-05	0.270079	0.7871
AR(1)	-0.981003	0.005531	-177.3748	0.0000
MA(2)	-0.987523	0.006253	-157.9203	0.0000
SIGMASQ	0.061364	0.000843	72.76429	0.0000

R-squared	0.458215	Mean dependent var	8.02E-05
Adjusted R-squared	0.457125	S.D. dependent var	0.336657
S.E. of regression	0.248049	Akaike info criterion	0.055407
Sum squared resid	91.79998	Schwarz criterion	0.069606
Log likelihood	-37.44436	Hannan-Quinn criter.	0.060697
F-statistic	420.6196	Durbin-Watson stat	1.867527
Prob(F-statistic)	0.000000		

Inverted AR Roots	-.98	
Inverted MA Roots	.99	-.99

Source: Author's computations based on historical close price data of BRD

3.2. ARIMA (p, d, q) for the Banca Transilvania - TLV

For the TLV close price prediction we followed the same pattern as for the BRD. The stock data for TLV it’s in the same interval as for BRD, from January 3, 2018, to December 29, 2023, with a total of 1,496 daily observations. To determine whether the time series is stationary, we plotted the original data to observe its pattern. Figure 5 below shows the time series plot of the closing prices of TLV.

From the graph above, it is also evident that the time series follows a random walk pattern, which is indicative of non-stationarity, just like TLV stock close price.

Fig. 5. Historical close price of TLV

Sourced: Author's computations based on historical close price data of TLV

To further investigate the stationarity of the TLV time series, we analyze its stationarity using the Augmented Dickey-Fuller (ADF) Test. The results of the ADF test for the TLV closing price series are presented in Table 4.

Table 4. ADF Test Results for TLV Closing Prices

Null Hypothesis: CLOSE has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 0 (Automatic - based on SIC, maxlag=36)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.425841	0.3658
Test critical values:	1% level	-3.964194	
	5% level	-3.412819	
	10% level	-3.128392	

*MacKinnon (1996) one-sided p-values.
Augmented Dickey-Fuller Test Equation
Dependent Variable: D(CLOSE)
Method: Least Squares
Sample (adjusted): 1/04/2018 12/29/2023
Included observations: 1495 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
CLOSE(-1)	-0.008837	0.003643	-2.425841	0.0154
C	0.111873	0.047163	2.372061	0.0178
@TREND("1/03/2018")	5.74E-05	2.59E-05	2.214816	0.0269

R-squared	0.004089	Mean dependent var	0.008314
Adjusted R-squared	0.002754	S.D. dependent var	0.266570
S.E. of regression	0.266202	Akaike info criterion	0.192884
Sum squared resid	105.7286	Schwarz criterion	0.203539
Log likelihood	-141.1808	Hannan-Quinn criter.	0.196854
F-statistic	3.062970	Durbin-Watson stat	1.922788
Prob(F-statistic)	0.047043		

Source: Author's computations based on historical close price data of TLV

The p-value for the constant (C) is 0.0178, which is less than 0.05, indicating that the constant is significant and should be included in the model. The p-value for the trend is 0.0269, which is also less than 0.05, indicating that the trend is significant and should be included in the test.

The ADF test statistic is -2.425841. To determine if the series is stationary, we compare this statistic to the critical values at the 1%, 5%, and 10% levels. The test statistic is not lower than the critical values at the 1%, 5%, or 10% significance levels (-3.964194, -3.412819, -3.128392 respectively). This means we cannot reject the null hypothesis that the series has a unit root. Since the ADF test statistic does not fall below the critical values, and the p-value associated with the test statistic is 0.3658, which is greater than 0.05, we fail to reject the null hypothesis. This indicates that the TLV closing price series is non-stationary.

Given the non-stationary nature of the TLV time series, it is necessary to apply differencing to achieve stationarity. This confirms that we are working with an ARIMA model (p, d, q), where the integrated component (I) accounts for the differencing required to stabilize the series.

After identifying the non-stationarity of TLV time series, we apply the first differencing to transform it into a stationary series. The differenced series, denoted as "DCLOSE," represents the changes in the closing prices from one day to the next. Figure 7 below displays the line graph of the differenced series "DCLOSE" for the TLV closing prices, Figure 6 presents the correlogram (ACF plot) of the differenced series "DCLOSE." The correlogram helps to verify the stationarity by showing the autocorrelations of the differenced series at various lags.

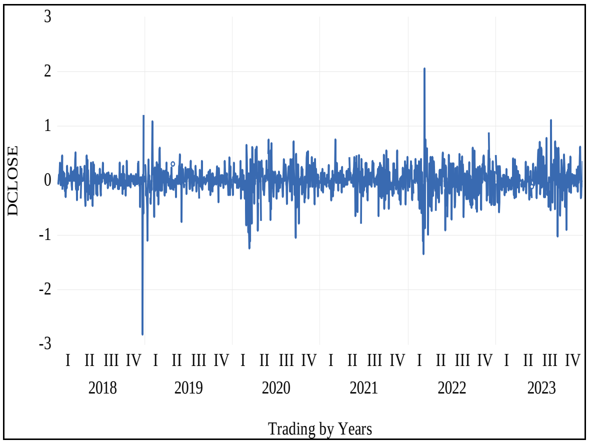
Fig. 6. Correlogram at first difference of TLV close price

Date: 06/02/24 Time: 16:25
Sample (adjusted): 1/04/2018 12/29/2023
Included observations: 1495 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.033	0.033	1.6768	0.195
		2	0.030	0.029	3.0556	0.217
		3	-0.030	-0.032	4.4197	0.220
		4	-0.011	-0.010	4.6098	0.330
		5	-0.036	-0.033	6.5122	0.260
		6	-0.007	-0.005	6.5848	0.361
		7	-0.043	-0.041	9.3551	0.228
		8	-0.013	-0.013	9.6252	0.292
		9	0.043	0.046	12.420	0.191
		10	0.002	-0.004	12.425	0.258
		11	0.000	-0.004	12.425	0.333
		12	0.025	0.025	13.378	0.342
		13	0.008	0.006	13.477	0.412
		14	0.016	0.016	13.886	0.458
		15	0.029	0.029	15.170	0.439
		16	-0.020	-0.019	15.785	0.468
		17	-0.057	-0.054	20.681	0.241
		18	-0.001	0.004	20.683	0.296
		19	-0.024	-0.018	21.544	0.308
		20	0.003	0.004	21.560	0.365

Source: Author's computations based on historical close price data of TLV

Fig. 7. Graphic of first difference of TLV close price



Source: Author's computations based on historical close price data of TLV

To identify the best model, we have used the same algorithm as for BRD namely 3 key metrics: Akaike Information Criterion (AIC), Schwarz Bayesian Information Criterion (BIC), Hannan-Quinn Criterion (HQC) and we identified the ARIMA(7, 1, 56) as the most optimal model for TLV. The model includes 7 autoregressive term, differencing the data once, and 56 moving average terms. The results of the tests for ARIMA(7, 1, 56) are the studied as follow.

First, we ensure that the model satisfies the requirements for a stable univariate process. This involves checking that the residuals of the model are white noise. So, we examine the Ljung-Box Q-statistic, where the Null Hypothesis is that the residuals are white noise.

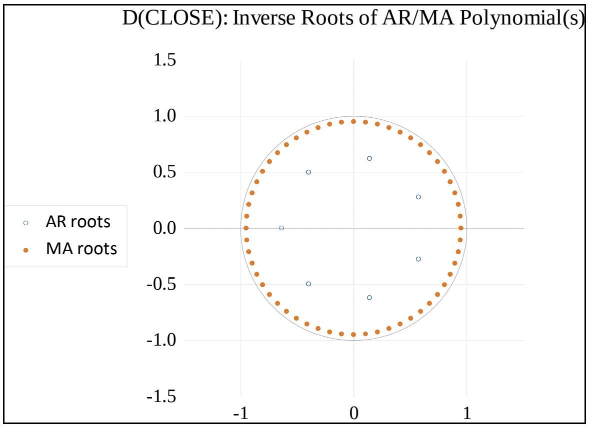
Fig. 8. Ljung-Box Q-statistic Test ARIMA(7, 1, 56)

Date: 06/02/24 Time: 16:43
Sample (adjusted): 1/04/2018 12/29/2023
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.032	0.032	1.5022	
		2	0.030	0.030	2.8949	
		3	-0.033	-0.035	4.5404	0.033
		4	-0.011	-0.010	4.7358	0.094
		5	-0.034	-0.031	6.4295	0.092
		6	-0.005	-0.003	6.4615	0.167
		7	0.000	0.002	6.4617	0.264
		8	-0.012	-0.014	6.6655	0.353
		9	0.041	0.041	9.1765	0.240
		10	-0.001	-0.004	9.1780	0.328
		11	0.005	0.002	9.2146	0.418
		12	0.021	0.024	9.9056	0.449
		13	0.011	0.009	10.081	0.523
		14	0.016	0.017	10.452	0.576
		15	0.026	0.027	11.503	0.569
		16	-0.019	-0.021	12.071	0.601
		17	-0.057	-0.054	17.044	0.316
		18	0.003	0.009	17.059	0.382

Source: Author's computations based on historical close price data of TLV

Fig. 9. Inverse Roots of AR/MA for ARIMA(7, 1, 56)



Source: Author's computations based on historical close price data of TLV

In our case, as shown in Figure 8, the absence of significant spikes in the Autocorrelation (AC) and Partial Autocorrelation (PAC) plots confirms that the residuals of the selected ARIMA model are white noise.

Second, we verify whether the estimated AR/MA process is stationary by checking if the AR and MA roots lie inside the unit circle. The Figure 9 below shows that both the AR and MA roots are within the unit circle, confirming the stationarity of the ARIMA(1, 2, 2) model.

The model also returned the smallest Bayesian or Schwarz information criterion of 0.207692 and relatively smallest standard error of regression of 0.266176 as shown in the Table 5, below.

Table 5. ARIMA(7, 1, 56) Estimation Output – TLV

Dependent Variable: D(CLOSE)
Method: ARMA Maximum Likelihood (OPG - BHHH)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.008178	0.006467	1.264552	0.2062
AR(7)	-0.042634	0.017819	-2.392625	0.0169
MA(56)	-0.056645	0.025920	-2.185385	0.0290
SIGMASQ	0.070660	0.000970	72.84602	0.0000

R-squared	0.004951	Mean dependent var	0.008314
Adjusted R-squared	0.002949	S.D. dependent var	0.266570
S.E. of regression	0.266176	Akaike info criterion	0.193485
Sum squared resid	105.6371	Schwarz criterion	0.207692
Log likelihood	-140.6300	Hannan-Quinn criter.	0.198778
F-statistic	2.472849	Durbin-Watson stat	1.935276
Prob(F-statistic)	0.060114		

Inverted AR Roots	.57-.28i	.57+.28i	.14-.62i	.14+.62i
	-.40-.50i	-.40+.50i	-.64	
Inverted MA Roots	.95	.94-.11i	.94+.11i	.93-.21i
	.93+.21i	.90+.31i	.90-.31i	.86-.41i
	.86+.41i	.80-.51i	.80+.51i	.74+.59i
	.74-.59i	.67+.67i	.67-.67i	.59+.74i
	.59-.74i	.51+.80i	.51-.80i	.41-.86i
	.41+.86i	.31+.90i	.31-.90i	.21-.93i
	.21+.93i	.11+.94i	.11-.94i	.00+.95i
	-.00-.95i	-.11+.94i	-.11-.94i	-.21+.93i
	-.21-.93i	-.31-.90i	-.31+.90i	-.41+.86i
	-.41-.86i	-.51+.80i	-.51-.80i	-.59+.74i
	-.59-.74i	-.67-.67i	-.67-.67i	-.74-.59i
	-.74+.59i	-.80+.51i	-.80-.51i	-.86-.41i
	-.86+.41i	-.90+.31i	-.90-.31i	-.93-.21i
	-.93+.21i	-.94+.11i	-.94-.11i	-.95

Source: Author's computations based on historical close price data of TLV

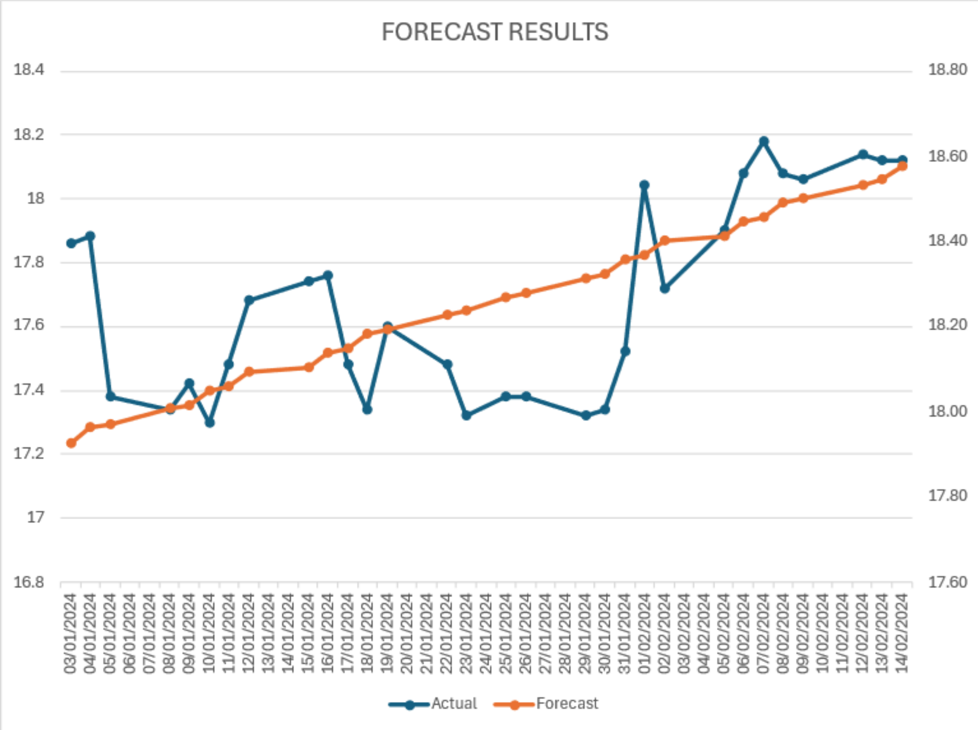
4. Findings and discussion

In this section, we present the results of our study on the ARIMA models applied to the BRD and TLV datasets. For BRD, we utilized the ARIMA(1, 2, 2) model, and for TLV, the ARIMA(7, 1, 56) model.

4.1. Findings for BRD – Groupe Société Générale stock

We have introduced in Table 5 the predicted values generated by the ARIMA(1, 2, 2) model for the BRD – Groupe Société Générale stock. Figure 10 visually represents the comparison between the predicted stock prices and the actual stock prices, allowing for a clear demonstration of the model's performance. From the graph in Figure 10, we observe that the ARIMA model's performance is quite impressive.

Fig. 10. Forecast and actual of close price - BRD



Source: Author's computations based on historical close price data of BRD

Table 5. Values of Forecast and actual of close price - BRD

Date	Actual	Forecast ARIMA(1, 2, 2)	Date	Actual	Forecast ARIMA(1, 2, 2)
03/01/2024	17.86	17.93	25/01/2024	17.38	18.27
04/01/2024	17.88	17.96	26/01/2024	17.38	18.28
05/01/2024	17.38	17.97	29/01/2024	17.32	18.31
08/01/2024	17.34	18.01	30/01/2024	17.34	18.32
09/01/2024	17.42	18.01	31/01/2024	17.52	18.36
10/01/2024	17.3	18.05	01/02/2024	18.04	18.37
11/01/2024	17.48	18.06	02/02/2024	17.72	18.40
12/01/2024	17.68	18.09	05/02/2024	17.9	18.41
15/01/2024	17.74	18.10	06/02/2024	18.08	18.44
16/01/2024	17.76	18.14	07/02/2024	18.18	18.46
17/01/2024	17.48	18.15	08/02/2024	18.08	18.49
18/01/2024	17.34	18.18	09/02/2024	18.06	18.50
19/01/2024	17.6	18.19	12/02/2024	18.14	18.53
22/01/2024	17.48	18.23	13/02/2024	18.12	18.55
23/01/2024	17.32	18.24	14/02/2024	18.12	18.58

Source: Author's computations based on historical close price data of BRD

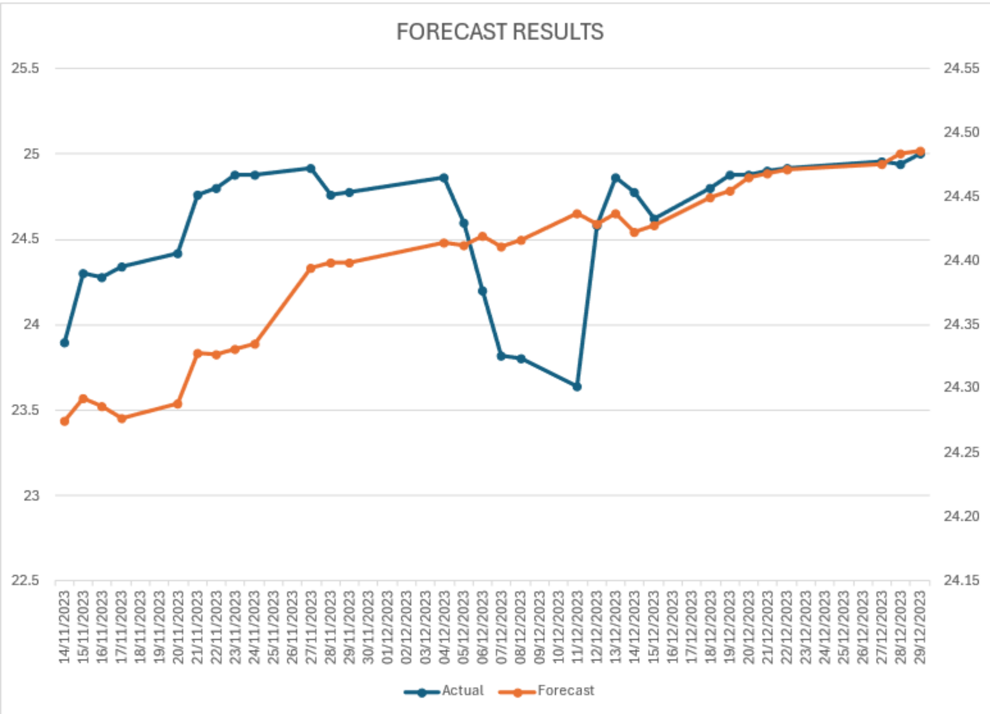
We can see that overall the actual values suggest a positive trend, which our ARIMA model managed to capture, suggesting a strong correlation and high accuracy in the model's forecasts. The ability of the model to predict these values accurately indicates its effectiveness in capturing the underlying patterns and trends within the BRD stock data.

Moreover, the residuals analysis, as discussed earlier, supports the stability and reliability of this model. The residuals' white noise characteristic confirms that the model has effectively captured the significant patterns in the data, leaving only random noise unexplained.

4.2. Findings for Banca Transilvania – TLV

We have presented in Table 6 the predicted values generated by the ARIMA(7, 1, 56) model, which was identified as the best model for the TLV stock index. Figure 11 provides a graphical illustration comparing the predicted stock prices with the actual stock prices, showcasing the model's accuracy and performance. From the graph in Figure 11, it is evident that the ARIMA(7, 1, 56) model performs satisfactorily. The predicted values closely align with the trend.

Fig. 11. Forecast and actual of close price - TLV



Source: Author's computations based on historical close price data of TLV

Table 6. Values of Forecast and actual of close price - TLV

Date	Actual	Forecast ARIMA(7, 1, 56)	Date	Actual	Forecast ARIMA(7, 1, 56)
03/01/2024	23.9	24.28	25/01/2024	23.82	24.41
04/01/2024	24.3	24.29	26/01/2024	23.8	24.42
05/01/2024	24.28	24.29	29/01/2024	23.64	24.44
08/01/2024	24.34	24.28	30/01/2024	24.58	24.43
09/01/2024	24.42	24.29	31/01/2024	24.86	24.44
10/01/2024	24.76	24.33	01/02/2024	24.78	24.42
11/01/2024	24.8	24.33	02/02/2024	24.62	24.43
12/01/2024	24.88	24.33	05/02/2024	24.8	24.45
15/01/2024	24.88	24.34	06/02/2024	24.88	24.45
16/01/2024	24.92	24.39	07/02/2024	24.88	24.47
17/01/2024	24.76	24.40	08/02/2024	24.9	24.47
18/01/2024	24.78	24.40	09/02/2024	24.92	24.47
19/01/2024	24.86	24.41	12/02/2024	24.96	24.48
22/01/2024	24.6	24.41	13/02/2024	24.94	24.48
23/01/2024	24.2	24.42	14/02/2024	25	24.49

Source: Author's computations based on historical close price data of TLV

The objective of this analysis was to find the most optimal model to forecast the trend in the TLV stock index, and our results confirm that the ARIMA(7, 1, 56) model meets this goal.

6. Conclusions

In our study, we applied ARIMA models to predict the closing prices of BRD – Groupe Société Générale and Banca Transilvania (TLV) stocks on the Bucharest Stock Exchange (BSE). For BRD, the ARIMA(1, 2, 2) model was selected, while for TLV, the ARIMA(7, 1, 56) model was identified as the optimal choice.

The results for BRD showed that the ARIMA(1, 2, 2) model performed impressively, with predicted values closely matching the actual stock prices. The absence of significant autocorrelation in the residuals confirmed that the model captured the essential patterns in the data, indicating that the residuals were white noise. This suggests that the model is both stable and reliable for forecasting purposes.

Similarly, the findings for TLV indicated that the ARIMA(7, 1, 56) model provided satisfactory predictions. The model's performance was evident from the close alignment of predicted values with the actual stock prices, confirming its capability to accurately capture the trend. The model passed all diagnostic checks, including the Ljung-Box Q-statistic, and demonstrated stationarity.

The ARIMA models selected for BRD and TLV effectively captured and predicted the trends in their respective stock prices. The successful application of these models showed the potential of ARIMA methodologies in financial forecasting. Future research could explore extending the model to include additional variables or to mix the ARIMA model with other forecasting techniques such as artificial neural networks to enhance predictive accuracy further. To sustain this idea, Ene (2024) conducted a bibliometric analysis and found in the existing body of literature that artificial neural networks (ANN) and deep learning (DL) are widely used and accepted in financial forecasting due to their ability to capture complex, non-linear patterns in financial data. These advanced models, including Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Long Short-Term Memory (LSTM) networks, have demonstrated superior performance in various forecasting tasks, integrating these insights, future research could benefit from combining ARIMA model with these advanced models resulting a hybrid approach that could leverage the strengths of ARIMA in capturing linear trends and the ability of ANN and DL models to handle complex, non-linear relationships in the data.

The results of this study are particularly useful for various stakeholders in the financial market. Investors, both individual and institutional, can utilize the findings to make more informed decisions regarding the buying or selling of financial company's stocks listed on BSE. The ability to predict stock prices accurately can significantly enhance investment strategies, helping to mitigate risks and optimize returns. Financial analysts and advisors, who analyse market trends and provide investment advice, can leverage the ARIMA model's demonstrated accuracy in forecasting stock prices trends. This can enhance the quality of their recommendations and support more robust financial planning. Additionally, portfolio managers responsible for managing investment portfolios can benefit from the predictive insights provided by the ARIMA models. Accurate forecasts trends of stock prices can aid in adjusting portfolio allocations to maximize gains and minimize losses.

In behalf of the research questions mentioned in the introduction section, the following answers were identified. The ARIMA models applied in this study were able to accurately predict the short-term trends of the closing prices for both BRD and TLV stocks. Despite the inherent higher level of volatility in the emerging Romanian stock market, the ARIMA(1, 2, 2) model for BRD and the ARIMA(7, 1, 56) model for TLV successfully captured the essential patterns and provided reliable forecasts. This indicates that the ARIMA model, when correctly specified and validated, can effectively handle the volatility characteristic of emerging markets and offer accurate short-term predictions.

The volatility of the Romanian stock market, amplified by specific events affecting BRD and TLV, posed challenges to the predictive performance of the ARIMA models. For BRD, speculation about the potential sale by Groupe Société Générale significantly influenced the stock price, leading to increased volatility. Similarly, TLV experienced heightened volatility due to customer reluctance towards increased bank fees announced at the beginning of 2024. Despite these challenges, the ARIMA models were able to estimate the correct trend, demonstrating robustness in the face of market volatility. This suggests that while volatility impacts the precision of point predictions, the overall trend estimation by ARIMA models remains reliable. The application of ARIMA models to BRD and TLV stocks provides several key insights for investors and stakeholders. First, the ability of ARIMA models to capture and predict trends even in a highly volatile market underscores their utility in emerging markets like Romania. Investors can leverage these models to anticipate market movements and make informed decisions. For BRD, understanding the impact of speculative events on stock prices can help investors mitigate risks associated with sudden market reactions. For TLV, recognizing the effects of operational changes, such as fee increases, on customer sentiment and stock volatility can guide strategic decisions. Overall, the use of ARIMA models in this context highlights the importance of trend analysis in investment strategies, enabling stakeholders to navigate the complexities of emerging markets with greater confidence.

7. References

Ariyo, A.A., Adewumi, A.O., Ayo, C.K. (2014). Stock Price Prediction Using the ARIMA Model. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, pp. 106–112

Alhawatmeh, O. (2022). Share price forecast: using ARIMA model special study on Amman Stock Exchange (ASE). *Italian Journal of Pure and Applied Mathematics*, 47, pp. 173–181

Box, G.E.P., Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day

Banerjee, D. (2014). Forecasting of Indian stock market using time-series ARIMA model. *2014 2nd International Conference on Business and Information Management (ICBIM)*, pp. 131–135

Ene, C.C. (2024). An overview over the impact of neural networks and deep learning in financial forecasting. *Annals of the „Constantin Brâncuşi" University of Târgu Jiu*, *Economy Series*, Vol. 1/2024, pp. 138–149

Fama, E.F. (1970). Efficient capital markets: a review of theory and empirical work, *Journal of Finance*, Vol. 38, pp. 34–105

Jarrett, J.E., Schilling, J. (2008). Daily variation and predicting stock market returns for the frankfurter börse (stock market). *Journal of Business Economics and Management*, Vol. 9(3), pp. 189–198

Khan, S., Alghulaiakh, H. (2020). ARIMA model for accurate Time series stocks Forecasting. *International Journal of Advanced Computer Science and Applications*, Vol. 11(7), pp. 524–528

Kumar, K.A., Pinto, P., Hawaldar, I.T., Spulbar, C., Birau, R., Minea, L.E. (2021). Forecasting Areca Nut Market Prices Using the ARIMA Model: A Case Study of India. *Annals of the „Constantin Brâncuşi" University of Târgu Jiu, Economy Series*, Vol. 2/2021, pp. 4–18

Kumar, K.A., Pinto, P., Spulbar, C., Birau, R., Hawaldar, I.T., Vishal, S., Bărbăcioru, I.C. (2023). ARIMA model to forecast the RSS-1 rubber price in India: A case study for textile industry. *Industria Textilă*, Vol. 74(02), pp. 238–245

Lohar, K., Pahwa, P., Asthana, N., Gautam, D. (2020). Stock Market Analysis Using Time Series with ARIMA Model. *(IJACSA) International Journal of Advanced Computer Science and Applications*, Vol. 11(7), pp. 524–528

Mahadik, A., Vaghela, D., Mhaisgawali, A. (2021). Stock Price Prediction using LSTM and ARIMA. *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*

Meher, B.K., Hawaldar, I.T., Spulbar, C., Birau, R. (2021). Forecasting stock market prices using mixed ARIMA model: a case study of Indian pharmaceutical companies. *Investment Management and Financial Innovations*, Vol. 18(1), pp. 42–54

Prakhar, K., Sountharrajan, S., Suganya, E., Karthiga, M., Sathis, B.K. (2022). Effective Stock Price Prediction using Time Series Forecasting. *6th International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1636-1640

Spulbar, C., Ene, C.C. (2024). The Empirical Approach of the Interplay of Macroeconomic Variables and the Dynamics of the Financial Market in Romania. *Revista de Științe Politice. Revue des Sciences Politiques*, Vol. (82/2024), pp. 55-69