Journal for Business Education and Management
ISSN (Print): 2790-7848, ISSN (Online): 2790-7856
Volume 5, Issue 1, Page 1-22

THE PREDICTIVE ACCURACY OF ARIMA MODELS FOR STOCK MARKET INDICES: A CASE STUDY OF KSE-100

Shoaib Anwer Qambrani
Independent Researcher, Karachi, Pakistan.
qambranishoaib@gmail.com

Israr Ahmed
Associate Professor
Ilma University, Karachi, Pakistan.
dr.israrahmed@ilmauniversity.edu.pk

Abdul Basit
Lecturer, KASBIT, Karachi, Pakistan.
basit@kasbit.edu.pk

Abstract

This study examines the time series predictions of the KSE-100 index using the autoregressive integrated moving average (ARIMA) model. KSE-100, which represents the Stock Exchange of Pakistan. It plays a key role in reflecting the financial status of a country. Predicting stock market indices is important for investors. Policy decision makers and economists, because it is subject to inertial fluctuations and plays a role in economic planning. The study uses daily stock data from January 2012 to December 2023, using the ARIMA model to predict future trends. Two models, ARIMA (1,0,1) and ARIMA (1,1,1), are evaluated and compared according to their performance. The results suggest that although the ARIMA (1,0,1) model is better suited to short-term forecasts, the ARIMA (1,1,1) model captures the long-term uptrend more effectively. The study highlights the limitations of the ARIMA model in capturing the nonlinear nature of stock market movements. It suggests future exploration of hybrid models that combine traditional time series methods with machine learning approaches. These findings provide information on the effectiveness of the ARIMA model in financial forecasting. It has an impact on investors and policymakers who aim to reduce risk and optimize investment strategies**.**

Keywords: Time series forecasting, Stock market prediction, ARIMA model, Pakistan Stock Exchange, Financial modeling

JEL: C32, C53, G17

Introduction

Time series prediction seeks to uncover the underlying relationships within time series data and analyze its evolving patterns based on historical observations. Due to extensive data sources across several domains, time series forecasting tasks have been widely implemented and have swiftly evolved into a crucial aspect of time series research. Theoretical studies and applications in several disciplines, including financial market analysis, traffic flow forecasting, human behavior prediction, and climate forecasting, define current research on time series prediction. Numerous predictive strategies have been developed during the past few decades from various modeling concepts and points of view (Jones, 2017; Kolambe & Arora, 2024; Zheng & Huang,

This work is licensed under a Creative Commons Attribution 4.0 International License.

2020). The autoregressive integrated moving average (ARIMA) model is a basic method. The ARIMA model integrates many techniques derived from the autoregressive moving average (ARMA) model to address non-stationarity issues. Notwithstanding some limitations on parameter estimation and strict assumptions on the noise term, notable improvements in predictive accuracy have been obtained. Using its online optimization process, the ARIMA model efficiently estimates parameters, decreases training time, and increases its applicability to large-scale data, therefore addressing these difficulties (He et al., 2023; Ho & Xie, 1998; Khan & Alghulaiakh, 2020). Nonetheless, the inconsistency in prediction accuracy with noise interference remains inadequately addressed. In the presence of substantial noise, the variations between preceding and subsequent points in a time series may be markedly divergent and misleading. Nonstationary behavior significantly affects the precision of the ARIMA model. Differencing techniques struggle to reveal the original trend information amongst significant noise contamination. The influence of the online ARIMA may be more pronounced, and the extraction of sequential information is not sensitive in the one-pass training mode (Li et al., 2023).

Research indicates that stock prices do not strictly adhere to the principles of random wandering or the efficient market hypothesis; rather, their price fluctuations exhibit a long-term memory effect in the stock market. The stock market is defined by significant risk and potential for substantial returns. Both securities institutions and investors exhibit a strong interest in the predictive analysis of the stock market, actively seeking to identify trends in stock price fluctuations. Stock price forecasting seeks to predict future trends in stock prices by analyzing historical price movements. This approach assists institutions and investors in understanding the dynamics of the stock market, thereby enabling individuals to mitigate risks and enhance returns during stock investment activities (Zhao et al., 2023). Time series forecasting is extensively utilized across multiple domains, including finance, climate science, and energy, among others. Accurate time series prediction presents challenges due to intricate characteristics such as seasonal patterns, trends, and volatility. Several techniques, including Autoregressive Integrated Moving Average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH), have been employed to predict univariate time series data. Nevertheless, these methods primarily concentrate on linear trends, which may not consistently correspond to real-world data that frequently demonstrate nonlinear and nonstationary characteristics (Mutinda & Langat, 2024). Stocks serve as a significant avenue for investment in the securities market, drawing the interest of numerous investors. Developing a robust stock investment strategy to forecast future prices and stock trends remains a pressing challenge in the investment market. The trend of stock series, as a typical time series data, is influenced by numerous factors, including policies, economic development, interest rates, capital flows, and investor sentiment, which present significant challenges in accurately predicting stock movements (Song et al., 2023). Over the past twenty years, the forecast for stock returns has emerged as a significant area of research. In many instances, researchers have sought to establish a linear relationship between the input macroeconomic variables and stock returns. However, the discovery of nonlinear trends in stock market index returns has led to a significant shift in researchers' focus on the nonlinear prediction of stock returns. Although much literature has emerged in nonlinear statistical modeling of stock returns, most of it necessitates that the nonlinear model be specified before the estimation process, due to the inherent noise, uncertainty, chaos,

ISSN (Print): 2790-7848, ISSN (Online): 2790-7856
Volume 5, Issue 1, Page 1-22

and nonlinearity of stock market returns. In the literature, various sets of input variables are utilized to forecast stock returns. Various input variables are utilized to forecast the same set of stock return data. Some researchers utilized input data from a single time series, while others took into account the incorporation of diverse market information and macroeconomic variables (Zaffar & Hussain, 2022).

The stock market is an essential element of the financial sector that significantly influences a nation's progress. The stock market serves as the channel for the redistribution of financial resources across diverse economic entities. Progress in the stock market is considered a gauge of a nation's economic growth. The stock market serves as an effective mechanism for capital acquisition for people, organizations, and governments. Due to its high liquidity, it facilitates seamless stock trades among interested parties. The development of capital and financial markets in any country is mostly influenced by the volume of individuals participating in stock market transactions and the consistency of stock market trends. (Jadoon et al., 2024). The Pakistan Stock Exchange (PSX) combined three already-existing stock exchanges from Islamabad, Karachi, and Lahore in January 2016. Later, in May 2017, it became part of the Emerging Market Index MSCI. PSX has registered members totaling 400 brokerage companies and 21 asset management firms at present. The prediction of stock returns has become a major focus of study throughout the past two decades. Most of the time, researchers aimed to find a linear link between stock returns and the macroeconomic factors of input. The discovery of nonlinear trends in stock market index returns, however, has lately attracted much more attention among researchers to the nonlinear prediction of stock returns. (Zaffar & Hussain, 2022).

The stock market represents the economic vitality of the actual industry. (Li, 2023). The Efficient Market Hypothesis (EMH) continues to be esteemed within the academic community. This hypothesis posits that stock prices fully incorporate all accessible information, while in inefficient stock markets, stock prices exhibit random fluctuations. Furthermore, the Random Walk Hypothesis posits that past share prices do not serve as a foundation for predicting future values. Consequently, share prices exhibit a random walk. Numerous investigations have been undertaken on the subject since the foundational work referenced previously. (Ahmad et al., 2018). Many statisticians and economists have argued historically about the random walk theory regarding stock market pricing. Advocates of the Random-walk hypothesis usually claim that the major stock exchanges across the world represent "efficient" markets. In an efficient market, the actions of many competing players produce actual prices that already integrate the consequences of current information, hence varying the price of a security from its inherent value. A market marked by independent sequential price swings in individual securities is, by definition, a random-walk market. According to the random walk hypothesis, existing stock prices reflect all the information at hand; therefore, a short period is required for new information to be incorporated into prices, which provides market participants with limited chances to profit from this knowledge. The homogeneity of information systems causes market players to act in a like manner, therefore reducing the above-average profit potential of new knowledge to the nominal average returns given by the market. Stock prices reflect all the relevant information that defines an efficient market. Consistent, unusual returns cannot be obtained from random walk prices.

In debates on investments, future predictions are essential. Given the natural uncertainty of the future, investors sometimes wonder about the best moment to commit their money. Forecasting the stock market index is like forecasting any investments, since many elements affect the results. A change in the stock market index reveals the direction of a certain economy. Investors thus keep an eye on the index to evaluate changes in the stock market. Two different ways are possible to forecast a time-series variable. One to predict the path of the variable in line with all pertinent impacting factors. The other approach uses its lagging observations to project future values. Researchers in the modern day rely increasingly on the subsequently suggested approach for variable prediction; this study has also made use of lagged values of the variable of interest. Forecasting time series in the current work has made use of the Autoregressive Integrated Moving Average (ARIMA) approach. (Afeef, Ali, & Khan, 2018). Extensive in scale, high in dimensionality, and always needing updating, the time series consists of chronologically recorded historical data. Time series observations are applied cumulatively, starting with substantial research and enhancement to advance the study of time series. A fundamental research area involving the gathering and analysis of past values of the same variable is modeling and prediction. The prediction of the next events is grounded on the present and past observable events. Although stock market predictions have been extensively recorded in the literature, they are still a subject of the current study. Since its discovery in 1930, the effect of the day of the week has attracted great attention. There is already enough research on the effect of the day of the week. (Marri & Talpur, 2021).

The present research contribution to the area of financial forecasting is a conclusive appraisal of the effectiveness of ARIMA models to forecast the KSE-100 Index. By comparing two configurations, ARIMA (1,0,1) and ARIMA (1,1,1), the research stresses their different strengths. The former permits good forecast precision in the immediate term, bolstered by an enhanced adjusted R-squared value (0.99892), so they can deal well with immediate market movements. On the other hand, ARIMA (1,1,1) fits long-term upward trends better, such that it agrees with the general growth path of the index. These empirical results accentuate the great value of determining the approaches needed concomitant with the horizon of the forecast. Also, the study enumerates drawbacks of ARIMA models, especially their inability to define non-linearities and high volatility associated with financial markets. These findings echo previous studies that indicated the same challenges faced while using ARIMA in volatile environments. (Zotteri et al., 2005)

Additionally, the research paves the way for future hybrid modeling approaches. By demonstrating the unfortunate shortcomings of ARIMA, it calls for combining an advanced machine learning technique, such as Long Short-Term Memory (LSTM) networks, to amplify predictive accuracy, particularly in volatile markets. Other works showed that these hybrid models outperform ARIMA in handling complex nonlinear patterning (Albeladi et al., 2023; Nasir et al., 2023). The practical implications for investors and policymakers from these findings are tremendous. ARIMA (1,0,1) provides a reasonable short-term forecast and therefore helps investors make an on-the-spot decision. On the other hand, the ARIMA (1,1,1) models accompany the Long-term trend and offer policymakers useful insights, which go quite a way in influencing strategic action in future regulatory interventions. Lastly, focusing this study on the KSE-100 Index implies an understanding of financial forecasting in emerging markets. It stresses the need to apply macroeconomic factors and market-

ISSN (Print): 2790-7848, ISSN (Online): 2790-7856
Volume 5, Issue 1, Page 1-22

specific characteristics in any study. Therefore, these contributions provide a basis for exploring further research to consider a hybrid approach and further applying the forecasting analysis to other indices and macroeconomic variables.

Rationale

The purpose of undertaking this research is to see how accurate ARIMA models are in predicting stock market indices, especially the KSE-100 Index of Pakistan (Afeef, Ali, & Khan, 2018; Marri & Talpur, 2021). The rationale for this is the need for efficient forecasting measures in a volatile and uncertain financial market. Predictions of the stock market indices are critical to investors, policymakers, and financial analysts as they will assist them in reducing risk and maximizing return on investments. Besides, the research studies the potential of applying ARIMA models for making precise short-term and long-term forecasts and increasing the pool of financial time series analysis literature.

Problem Statement

The instability, as well as the underlying unpredictory of financial markets, challenges the accurate forecast of stock indices (Li et al., 2023; Wang et al., 2023). While ARIMA models are widely used in the prediction of financial data, the problem of the model's limits for nonlinearity and nonstationarity remains. The study concerns the predictive capability of the ARIMA models for the KSE-100 index with two configurations. ARIMA (1,0,1) and ARIMA (1,1,1). The purpose of this research is to establish whether ARIMA models produce reliable forecasts for both short-term and long-term trends of the KSE-100 index and to determine the limitations and areas for improvement.

Research Questions

This part is constructed based on the available literature regarding the ARIMA models in financial forecasts (Bektemyssova et al., 2022; Mochurad & Dereviannyi, 2024).

1. How well do the ARIMA (1,0,1) and ARIMA (1,1,1) models model the short-term and long-run trends of the KSE-100 index?

2. What are the strengths and weaknesses of ARIMA models in forecasting stock market indices through volatile circumstances?

3. Can hybrid ARIMA models combined with a machine learning algorithm enhance the predictive accuracy of the KSE-100 index?

Research Objectives

The objectives coincide with the methodological approaches used in the earlier research studies aimed at stock market forecasting using the ARIMA models (Arashi & Rounaghi, 2022; Singh et al., 2024).

- To assess the short-term predictive validity of the ARIMA (1,0,1) model of the KSE-100 index.

- To determine the long-term predictive power of the ARIMA (1,1,1) model on the trend pattern of the KSE-100 index.
- To determine the limitations of the ARIMA models for nonlinear and volatility data in stock market forecasting.
- To suggest some potential hybrid modelling strategies to enhance predictive accuracy.

Theoretical Framework

The theoretical basis of this study is rooted in time series analysis and, specifically, the Autoregressive Integrated Moving Average (ARIMA) model (Box & Jenkins, 1976; Shumway & Stoffer, 2011).

This study relies on the theoretical framework, which is based on time series analysis, and in its concentration, is applied to the Autoregressive Integrated Moving Average (ARIMA) model. The assumption behind the ARIMA model is that the previous observations of the time series and stochastic errors linearly determine the future observations for a certain time series. The study also largely depends on the Efficient Market Hypothesis (EMH), along with the Random Walk Theory, which claims that the stock prices represent all the available information. As such, their prices cannot be predicted in efficient markets. The study aims to explore how machine learning methods can be incorporated into the process, such as Long Short-Term Memory (LSTM) networks, to avoid ARIMA's disadvantages in detecting non-linear trends and volatility in the market.

Conceptual Framework

The conceptual framework entails the configuration of ARIMA models, as shown in previous studies (Afeef, Ihsan, & Zada, 2018; Liu, 2024).

The conceptual framework involves a comparison of two constructs of an ARIMA model, the short-term ARIMA(1,0,1) and the long-term ARIMA (1,1,1), which are used for trend analysis. This is a study on how effective these models are in their ability to forecast the KSE-100 index and the impact of volatility and market changes on forecasting. Also, the presented study accounts for the integration of machine learning techniques to eliminate the discussed shortcomings, hence developing a combination model that brings both statistical and machine learning techniques together to optimize predictive accuracy.

Significance of Study

This research is important to both investors and policymakers, as well as financial analysts who wish to increase the accuracy of the prediction of the stock market, particularly in emerging economies such as Pakistan (Zaffar & Hussain, 2022; Zhao et al., 2023).

For investors, policymakers, and financial analysts who would like to increase the predictability of the stock markets, especially in emerging nations like Pakistan, this research work is very crucial. The research under consideration reveals the relative

advantages and disadvantages of classically oriented time series modeling during periods of sharp market developments by evaluating the ARIMA model's forecasting ability on the KSE-100 index. Furthermore, the study's conclusions advance the field of hybrid modeling techniques by suggesting that advanced machine learning techniques and ARIMA could be combined to increase forecasting accuracy..

Literature Review

Named for the Box-Jenkins method, the ARIMA model was first presented by George Box and Gwilym Jenkins in 1970. This method, which was first presented by Box and Jenkins in 1976, was further refined for time series forecasting in the 1990s. According to Box and Jenkins' research from the 1970s, ARIMA is applied with less volatile data to create a systematic class of models for managing time-correlated modeling and forecasting (Shumway & Stoffer, 2011). This model can be used to predict future values and define a time series using observed values. This method consists of a sequence of steps for identifying, estimating, and diagnosing ARIMA models based on time series data. When applied to any time series, these models reveal non-seasonal trends devoid of random white noise. To create an accurate stock forecasting model, the accuracy of an auto ARIMA model is compared to two customized ARIMA models. We need to be completely present right now and foresee future changes if we want to invest safely. Understanding the ARIMA model's applicability to time series forecasting and forecast accuracy is also beneficial. By contrasting an Auto ARIMA model with a bespoke ARIMA model, this study seeks to develop an accurate stock forecasting model. This emphasizes how time series forecasting methods need to be improved. In response to their use, econometric models have been upgraded. (Zotteri & colleagues, 2005).

Li (2023) looks at the NASDAQ index, which shows a historical pattern with a trajectory of long-term growth even during volatile times. Nonetheless, the expansion during the last two decades has been noteworthy, particularly amid the economic downturn, which included the 2008 financial crisis and the stock market crash. This result suggests that the NASDAQ index model's trajectory will be steady and smooth according to ARIMA (1,1,0). However, ARIMA (0, 2, 1) forecasts a higher growth on the NASDAQ index due to current market conditions. Li et al. (2017) used the ARIMA model to forecast the volatility of the Shanghai stock index in November and December of 2016. The results show that the Shanghai stock index has been slightly rising in the last two months of 2016, giving investors a strong base for assessing the future market. It enables them to lower any losses resulting from the unknown future hazards in the stock market. A focus and difficulty in the financial industry is predicting the Shanghai stock index. The ARIMA model is a leading method with unambiguous advantages in Shanghai stock index prediction. The ARIMA model shows optimum performance and effectively catches the linear properties of the data for short-term forecasts. In the framework of forecasting the Borsa Istanbul Banks Index (XBANK), Armagan (2023), Artificial neural networks (ANN) were compared against conventional models, including ARIMA. Using historical data, the study evaluates their performance in modeling trends, seasonality, and volatility. The results show that ARIMA works well for stationary and linear data, but it suffers from non-linear patterns and market swings. Conversely, ANN shows a great capacity to capture complex, non-linear interactions and always provides better predicted accuracy with lowered error measures, including RMSE and MAE. The study emphasizes how

flexible ANN is in fitting with changing and fluctuating market conditions, which makes it more suitable for financial projection. Armağan emphasizes their increasing importance in modern financial markets and advises the application of machine learning approaches, more especially, artificial neural networks, to improve prediction accuracy.



Figure 1 KSE-100 Historical Trend

Afeef, Ali and Khan (2018) Stated that the ARIMA (1, 1, 1) model was the most appropriate for our variables in the analysis. This specific ARIMA configuration has consistently been favored in literature, with other studies previously identifying the same model as suitable for their variables. Research has indicated that larger models, characterized by an increased number of delays, are more effective in forecasting their respective variables. Generally, ARIMA models are regarded as fairly effective in forecasting the future values of a specific variable. Afeef, Ihsan and Zada (2018) Stock prices were examined utilizing the Box-Jenkins methodology, and the ARIMA method was also applied in the present study. The study's findings indicate that ARIMA possesses a strong capability to predict future values in the short term. Certainly, long-term predictions utilizing lagged values of a variable will have minimal significance. It was noticed that all ARIMA configurations assessed in this study exhibited robust predictive capabilities, affirming the principle of allowing the data to convey its narrative. The autoregressive integrated moving average model may serve as a valuable tool for stock market investors to predict prospective upward or downward movements in their invested companies in the near future.

Arashi and Rounaghi (2022) Examined the daily return series of the NASDAQ stock index from 2000 to the end of 2016. After running the model, they found the best one for the stock index's daily returns time series. They then projected the daily return series of stock index values for 2017 in the next phase, showing that the ARMA-GARCH model can fairly project at a 1% error level. The results show a relationship between indices of stock prices throughout several periods. Their results on the efficiency and fractal properties of the NASDAQ stock exchange led us to infer that the market is non-fractal but efficient.

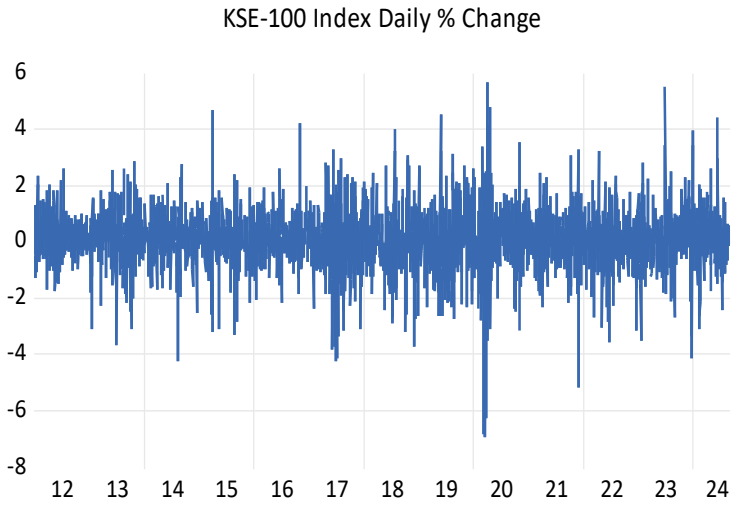


Figure 2: KSE-100 Returns

Forecasting stock market movements remains a challenging task considering their chaotic and dynamic character. The complex and irregular time series data of the market challenged conventional linear models such as AR, ARMA, and ARIMA, which have been widely applied. The development of deep learning models, especially deep neural networks, significantly advances the effort to address these challenges. These models provide a more all-encompassing and adaptable method for identifying the fundamental complex trends in financial data. Deep neural networks could alleviate linear model flaws by their adaptability and handling of non-linear data patterns. Still, using these models comes with several challenges, including the possibility of overfitting and the need for large computational resources. Determining success also depends much on the caliber and volume of the accessible data. Deep learning methods must be constantly refined and adapted to improve their predictive accuracy as the financial markets change. (Chowdhury et al., 2024).

Mochurad et al. (2024) introduced a novel ensemble methodology for financial market forecasting. They were unable to determine the current state of the financial market and whether LSTM or ARIMA made a more accurate prediction at the time of forecasting (the accuracy of the prediction is contingent upon the suitability of the data for each model).Mochurad and Dereviannyi (2024). Consequently, they generalized to identify a compromise that, at the very least, is a slight deviation from the best prediction and, at the very worst, corrects the predicted value to the true value, thereby reducing the final error. Their proposed method outperforms the LSTM recurrent neural network, the optimized deep recurrent neural network with long short-term memory, and the use of transformers in the construction of financial market forecasts, as evidenced by the simulation results and three real datasets based on the RMSE criterion. Specifically, the error is reduced by approximately 30.3%, 98.08%, and 66.80%, respectively. As previously mentioned, the pace of forecasting is also a critical factor to consider when forecasting financial markets. Liu (2024) found that the ARIMA model demonstrates significant efficacy in data analytics and time series forecasting. Its capacity to absorb and analyze data, generate forecasts, and facilitate informed decision-making has rendered it an indispensable resource across multiple sectors, especially in finance (Liu (2024).

The ARIMA model gives insights into economic indicator assessment, risk management techniques, stock price prediction, and so on. With informed investment choices, traders, investors, and financial analysts understand the complexities behind the financial markets and act on judgments that would be straightforward for them. Its content serves to analyze factors of credit risk, understand market dynamics, and build automatic trading plans. ARIMA remains the primary model, but one should know that it is a very large model with certain limits. Financial data are meant for an easy analysis with ARIMA, but this is not always the case when nonlinear components like human emotions and chance events are considered. Extreme events and structural changes could cause the model predictions to become dubious. Moreover, it may be that the choices regarding the parameters of the ARIMA model are, by definition, capricious and thus render uncertain. Besides, it may not be applicable to natural variables or for long-range forecasts. On the other hand, under these limitations, both scholars and financial analysts ought to act judiciously. There would be boosted accuracy and reliability of tests and forecasts when other methodologies and models are used. Even though aware that it remains a tool, acknowledging its limitations and adjusting it to fit financial data problems would go a long way in informing decision-making in the current dynamic financial environment. Bektemyssova et al. (2022) explain the usage of the Auto-Regressive Integrated Moving Average (ARIMA) model headings in areas concerning the anticipated movement of networks concerning the Internet of Things. It sheds light on the distinctive ability of ARIMA to have various methodologies in achieving control of non-stationary time series, which is in line with the traditional Box-Jenkins methodology (Bektemyssova et al., 2022).

The study has shown the model to be computationally efficient and applicable even for moderate-sized datasets, specifically that of the 18,000 observations for MPEG packet intensities over a backbone network. The authors created and validated an ARIMA (0, 2, 1) model using rigorous statistical analysis that included the Augmented Dickey-Fuller (ADF) test as well as autocorrelation and partial correlation plots. The procedures adopted resulted in a stabilization of the series and an accurate identification of model parameters. Short-term estimates produced through the ARIMA (0, 2, 1) model were accompanied by a confidence interval of 95%, with an enormous degree of reliability. According to Bektemyssova et al. (2022), the model is particularly useful in network traffic management due to its ability to generate predictions of very great accuracy and at the same time with little computational complexity. Their findings show how important ARIMA is as a trustworthy tool for forecasting in modern network environments, because of its ability to carry out a dependable manner of stream control and reduce risks associated with network congestion. This research aims to bring forth ARIMA's importance in solving real-life problems in telecommunication, especially those advancing with the rapidly changing dynamics of network traffic.

Wang et al. (2023) extended the application of the Auto-Regressive Integrated Moving Average (ARIMA) models to ultra-long time series. However, when applied to large volumes of data, ARIMA loses its scalability and processing efficiency, notwithstanding its generally known ability to produce seasonal and non-stationary time series. To solve these problems, they separated the incredibly long series into manageable subseries through a divide-and-conquer strategy (Wang et al., 2023). Accurate and consistent forecasts are ensured by fitting local ARIMA models to each subseries and integrating their parameter estimates by weighted least squares to minimize a global loss function. The DARIMA framework harnesses the advantages

ISSN (Print): 2790-7848, ISSN (Online): 2790-7856
Volume 5, Issue 1, Page 1-22

of ARIMA, especially the ability to address non-stationarity in patterns; it tries to minimize computing inefficiencies associated with conventional ARIMA in processing large datasets. DARIMA supports the parallel processing concept with a MapReduce architecture integration with distributed systems, which significantly reduces computing complexity. The study proves that DARIMA is more computationally economical and has greater predictive accuracy compared to normal ARIMA models, particularly on long-term projections. Their studies proved that ARIMA is the backbone in time series forecasting and also indicate its versatility in meeting ultra-long time series requirements in distributed systems. It also discusses how adaptations like DARIMA can be used to improve the applicability of the approaches in big data environments.

Singh et al. (2014) analyzed the value of predicting Bitcoin prices by machine learning models and time series in view of ARIMA limitations with its opportunities in this context. They stated that the non-linear and non-stationary characteristics of Bitcoin render considerable difficulties for ARIMA, unlike stationary and seasonal data, for which time series modeling using ARIMA is more suitable. Generally, ARIMA is a simple and interpretable model for financial forecasting; however, the present study shows that with a mean absolute percentage error (MAPE) of 11.86%, ARIMA is rendered useless for the highly volatile price data of Bitcoin (Singh et al., 2024). They combine ARIMA with modern deep learning models, especially Long Short-Term Memory (LSTM) networks, to address this problem. Performance results showed that LSTM significantly outperformed ARIMA with an MAPE of 1.40%. This excellent performance can be attributed to LSTM's ability to capture long-term dependencies in time-series data with great regulation over complex, non-linear patterns. The authors profess that ARIMA is the basic model of time series forecasting, but it has little application in volatile and non-linear data like Bitcoin. They advocate applying machine learning techniques, particularly LSTM, for accurate and reliable forecasting in cryptocurrency markets.

Method

The KSE-100 Index of the Pakistan Stock Exchange was selected as the sample stock index for the study. Over twelve years and seven months, from January 2012 to September 2024, secondary data were gathered from the Pakistan Stock Exchange, comprising daily values of the KSE-100 index. Trading occurs on five days per week. A total of 3,139 observations were employed in the investigation. The ARIMA model, Box and Jenkins (1976), is the most widely used time series approach. Empirical finance literature has primarily concentrated on forecasting returns and volatility, while it has overlooked the aspect of volume. The existing research on this topic is limited, with few contributions focused on volume forecasting. Certain studies employ neural networks and ARIMA models to predict monthly futures trading volume, highlighting the practical implications for the functioning of financial exchanges. Apart from its importance for estimating transaction costs and liquidity, volume forecasting is also amazing, as the time series traits of volatility and volume are so similar. Given the data of similar long-term dependency in both series, it is crucial to look at whether models with this quality can fairly project future volume and volatility (Lux & Kaizoji, 2007). It is characterized by the premise that the future values of a variable are a linear function of multiple antecedent observations, which are further supported by a stochastic error component. Three parametric elements are employed to construct the

linear function. The ARIMA (p,d,q) equation is a combination of auto-regression (AR), integration (I), and moving average (MA). The parameter p represents the count of auto-regressive components, d denotes the number of non-seasonal differences, and q represents the number of deferred forecast errors in the predictive model. The ARIMA stochastic model is contingent upon historical data for the autoregressive component (p), the moving average component (q), and a random error term ɛt. Consequently, the entire autoregressive moving average model can be expressed as

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \mu - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} + \epsilon_t$$

The identification of p, d, and q in an ARIMA model is a critical process that is frequently repeated until an optimal model is selected based on the minimal values of the Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC). As well as the maximal Log-Likelihood values and the minimal Mean Absolute Percentage Error and Mean Absolute Error (Marri & Talpur, 2021).

Result

Time series data are not distributed in most cases. Therefore, it may not be consistent. The stability of the series was checked using the unit root test. This is because many previous studies have reported that it is necessary to check the stability of the data before using statistical methods. It is also necessary to estimate the seasonality of data trends. In the present analysis, two separate techniques are used to check the stability of the index through the unit root test. To check consistency in our data, we use the Augmented Dickey-Fuller (ADF) test and the Phillips-Perron (PP) test. Table 1 presents the results of the unit root test for the KSE-100 index using the Augmented Dickey-Fuller (ADF) test statistic and Phillips-Perron (PP) both tests reject the null hypothesis of unit root, with an ADF statistic of 0.3551 (p-value 0.891) and a PP statistic of 0.3057 (p-value 0.9786), indicating non-stationarity. However, considering the first difference in both tests, the null hypothesis of a unit root is strongly rejected. ADF statistic is -49.5255 (p-value 0.001), PP statistic is -49.7330 (p-value 0.001), which indicates the series is stable after making a difference. Therefore, the KSE-100 index fluctuates in its level but is stable after the first difference.

Table 1: Unit Root Test

KSE-100 Index	Max Lag	Augmented Dickey-Fuller		Phillips-Perron test statistic	
		level	Difference	level	Difference
	28	0.3551 (0.891)	-49.5255 (0.001)	0.3057 (0.9786)	-49.7330 (0.001)

The KSE-100 Index's descriptive statistics reveal several important insights presented in Table 2. The meaning of the index is 37,945.61, median 39,919.95, indicating that the mean and median are quite similar, although the distribution is slightly skewed to the right. The index reached a maximum value of 81,932.92 and a minimum value of 10,909.63, respectively. The standard deviation of 12,880.47 reveals the extreme volatility in the market during the period analyzed. This also reflects this prominent level of volatility. A skewness value of 0.46 indicates a slight positive

skewness. The tail is longer on the right side of the distribution. Additionally, a kurtosis of 4.38 indicates that the distribution is leptokurtic. It has more tails compared to the normal distribution. The Jarke-Bera statistic of 359.52 indicates a significant deviation from normality. This means that the index's returns are not normally distributed. Overall, these data indicate that the KSE-100 Index is highly variable. It tends to have positive skewness and non-normal distribution patterns.

Table 2: Descriptive Statistics

KSE-100 Index	
Mean	37945.61
Median	39919.95
Maximum	81932.92
Minimum	10909.63
Std. Dev.	12880.47
Skewness	0.457739
Kurtosis	4.382281
Jarque-Bera	359.5203

Table 3 shows the results of the autocorrelation function (ACF) and the partial autocorrelation function (PACF) over correlation for the KSE-100 indices. Both the autocorrelation and the partial autocorrelation show positive values. It has a significance of 0.122, a Q-statistic of 46.824, and a p-value of 0.000, indicating significance. The following regressions show smaller values for both autocorrelation and partial autocorrelation, not exceeding 0.04. Although autocorrelation fluctuates slightly, Q-Stat remains statistically significant in all regressions. This indicates the existence of appropriate autocorrelation across data sets. When taken together, the results indicate that while the first row shows a significant correlation, subsequent rows show little correlation. This indicates that the differentiation process has removed autocorrelation from the series.

Table 3: ACF and PACF Through Correlogram at Difference

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
|* |	|* |	1	0.122	0.122	46.824	0.000
| |	| |	2	-0.032	-0.048	50.107	0.000
| |	| |	3	0.002	0.012	50.118	0.000
| |	| |	4	0.039	0.036	54.830	0.000
| |	| |	5	0.023	0.014	56.434	0.000

| |	| |	6	-0.006	-0.008	56.550	0.000
| |	| |	7	0.026	0.029	58.642	0.000
| |	| |	8	0.037	0.029	62.976	0.000
| |	| |	9	0.013	0.006	63.533	0.000
| |	| |	10	0.011	0.012	63.934	0.000

Table 4 shows the estimates for two ARIMA models: ARIMA (1,0,1) and ARIMA (1,1,1). The error variance (SIGMASQ) is slightly reduced for the ARIMA (1,0,1) 179,208.8 and. 179,710.6 for ARIMA (1,0,1), both models are statistically significant with a significance level of 0. However, the ARIMA (1,0,1) model has a much higher adjusted R-squared of 0.99892. This indicates that it explains all the differences given to us. Regardless of ARIMA (1,1,1), the model has a very low adjusted R-squared value of 0.01605, which means minimal explanatory power. The AIC values for both models are the same, with 14.94224 for ARIMA (1,0,1) and 14.94445 for ARIMA (1,1,9), indicating a slightly better fit than for ARIMA (1,0,1). The Schwarz Criterion (SC) values also showed a slight difference, with ARIMA (1,1,1) having a slightly lower SC of 14.93951 compared to 14.94995 for ARIMA (1,0,1). In general, ARIMA (1,0,1) provides a better fit in terms of adjusted R-squared, and the two models presented similar performance in terms of AIC and SC.

Table 4: Estimated ARIMA Models

	1,0,1	1,1,1
SIGMASQ	179710.6	179208.8
Significance	0	0
Adj R2	0.99892	0.01605
AIC	14.94224	14.94445
SC	14.94995	14.93951

Figure 3 contains the autocorrelation and partial autocorrelation plots for the ARIMA model applied to the KSE-100 index. The diagram on the left represents the ARIMA (1,0,1) model. The diagram on the upper left shows the function autocorrelation (ACF), where the autocorrelation value of real continues to decrease as the delay increases. However, the value is still positive. This indicates a continued relationship at higher lags, which indicates the existence of autocorrelation in the data. The lower left plot shows the partial autocorrelation function (PACF), with only the first lag showing significant correlation. This is followed by the lag that shows the least partial autocorrelation. This pattern supports the appropriateness of the first-order autoregression procedure for the data. On the right, the plot represents the ARIMA (1,1,1) model including the differences. The top right ACF plot shows no significant autocorrelation at any lag. This confirms that the difference removed autocorrelation. In the effective output, the bottom right PACF plot shows a lack of significant partial

autocorrelation after the first delay, which confirms that the various series behave more like white noise. The ARIMA (1,0,1) model reveals continuous autocorrelation, while the ARIMA (1,1,1) model, after differentiation, will automatically end the relationship; this indicates a more stable series.

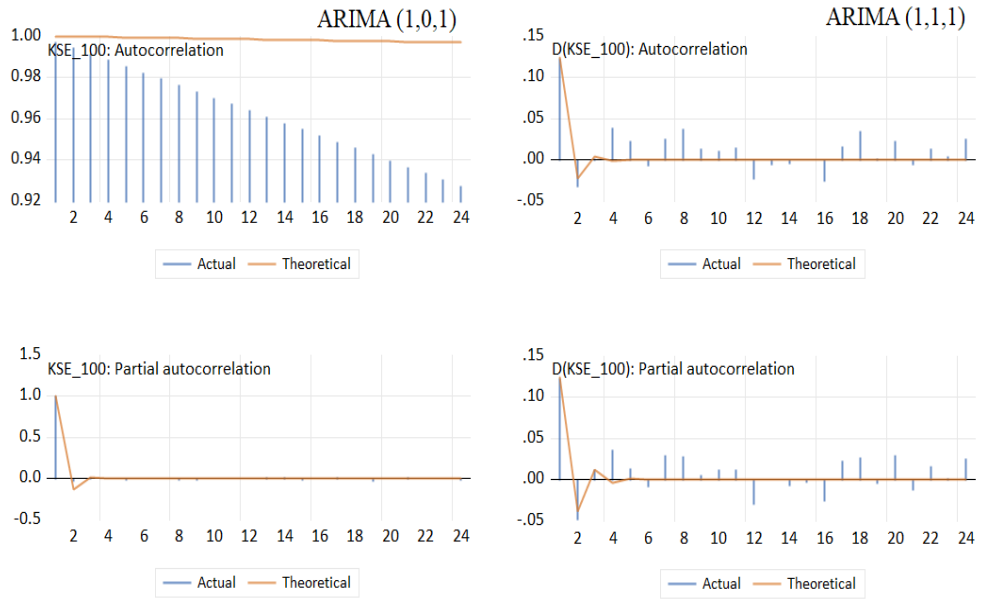


Figure 3 ACF and PACF of ARIMA Models

Figures 4 and 5 compare the forecasted values of the KSE-100 Index using two distinct ARIMA models: ARIMA (1,0,1) and ARIMA (1,1,1), plotted against the real values of the index. The forecast from the ARIMA (1,0,1) model indicates a flat, constant prediction past the observed length, indicating that the model expects no significant future movement within the index. This shows that the model does not always capture the upward trend found in the actual data. The forecast from the ARIMA (1,1) model reflects a rising trend, which is more aligned with the upward motion of the actual index in later durations. This model captures the trend higher than the ARIMA (1,0,1), showing a sluggish increase because the index remains on an upward trend. Overall, the ARIMA (1,1,1) model affords a more realistic forecast by accounting for the upward trend. However, at the same time, the ARIMA (1,0,1) fails to capture any significant future adjustments, predicting a constant value instead.

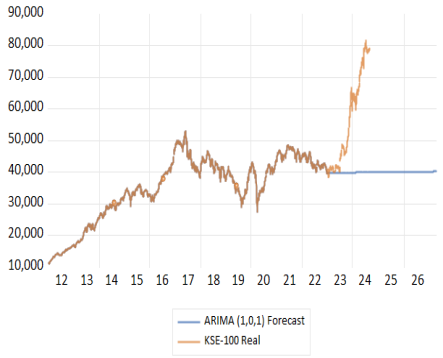


Figure 4 ARIMA (1,0,1) Forecast

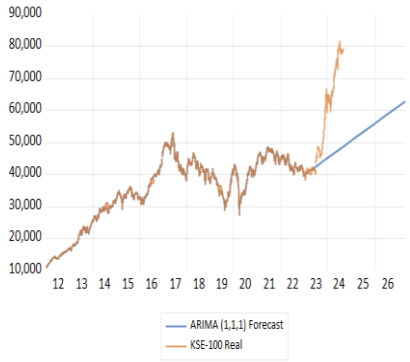


Figure 5 ARIMA (1,1,1) Forecast

Discussions

The study's conclusions offer a more comprehensive grasp of the financial market time series forecasting. Our analysis using ARIMA on the advantages and disadvantages of this statistical forecasting methodology on stock market movements, particularly that of the KSE-100, is reviewed in depth. The key conclusions of our empirical results also contradict previous works in the area. The main results of our analysis indicate that with a higher fitted R-squared value of 0.99892 and 0.01605, ARIMA (1,0,1) provides a better fit than ARIMA (1,1,1). This implies that both were true. The forecasts by ARIMA (1,0,1) are preferred in terms of accuracy. This goes well with the past research performed by Afeef et al. (2018) related to the ARIMA model's robustness in stock market index prediction, especially concerning the short term. Our results of the ARIMA (1,0,1) forecast, however, show that ARIMA (1,0,1) has limited capability to mostly capture the overall upward trend of the KSE-100 Index, which seems evident. The ARIMA (1,1) model shows the best representation of the rising trend of the index. This result is similar to Li et al. (2017), who found that ARIMA models are unable to depict long-term trends in highly volatile markets. The ARIMA (1,0,1) model's incapacity to forecast significant changes in the future is derived from its linear structure. Non-linear deviations induced by stock market data are ignored here.

Different researchers have mentioned the forecasting challenges of financial time series data. On this issue, Zotteri et al. (2005) and Zotteri and Kalchschmidt (2007) argued that volatility has nothing to do with stock market indices and that ARIMA uses a rather subtle approach of time series modeling. This is reflected in our findings. The values of the sample deviation pattern of 12,880.47 and the Jarque-Bera statistic of 359.52, which show a large departure from normality, confirm the acute volatility that we observe in the KSE-100 index. There are several examples of possibly more complex models that take into account the nonlinearity and clustering features of volatility, such as the conditional autoregressive diversity (GARCH) model which was studied by Arashi and Rounaghi (2022), Bauwens et al. (2006), Lin (2018), and Shaik and Rehman (2023).

The work is worth mentioning from this perspective. Albeladi and colleagues (2023); Cao and colleagues (2019); Fang and colleagues (2023). It integrates non-

linear methods for financial time series forecasting, including long-term and short-term memory (LSTM) networks. According to their findings, in extremely volatile markets, machine learning models perform better than conventional statistical techniques like ARIMA. Nonetheless, ARIMA models continue to be widely used because of their interpretability and simplicity. To increase prediction accuracy, machine learning approaches could be used in future studies. This is particularly true for nonlinearities. (Nasir and others, 2023).

Conclusion

Finally, this study demonstrates the performance of the ARIMA model on the KSE-100 index, where the ARIMA (1,0,1) model provides superior adaptation to short-term shocks and the ARIMA (1,1,1) model catches the long-term trend and remains a valuable reference for time series forecasting. Limitations in capturing non-linear movements and market fluctuations must be acknowledged. Future research should explore the integration of machine learning techniques and macroeconomic variables. To increase the accuracy and robustness of both prediction models. The findings of this study have serious implications for investors and decision-makers, focusing on the need for timely forecasts. For those investors who would like to predict market movement, the ability of the ARIMA (1,1,1) model to detect uptrends in the KSE-100 Index provides valuable insights. Forecasts of ARIMA (1,1,1), although appearing to be a good fit for short-term periods, appear to be more accurate in predicting medium-term behavior that conforms to long-term trends. The findings indicate that understanding market volatility and the limitations of conventional forecasting models are essential for policymakers. In their articles, Che et al. (2024), Kehinde et al. (2023), and Xu and Zhang (2023) insisted that precise forecasting models may be of use to traders and policymakers alike in making informed decisions regarding market intervention and regulation. High volatility on the KSE-100 index suggests that authorities should closely monitor market movements and prepare for possible interventions during severe instabilities.

With the reserves of this study, we certainly have given good bases in terms of the ARIMA model application in the KSE-100 index. First, regarding the robustness of this work, the straight-line aspect of these models limits it to the ARIMA models, thus forsaking the very complexity of stock markets. The very high volatility of the KSE-100 index and its anomalies may imply the applicability of some other models, such as GARCH or even LSTM networks (Mualifah et al., 2024; García-Medina & Aguayo-Moreno, 2024), to increase forecast accuracy, particularly in long-term patterns (Marri & Talpur, 2021; Afeef et al., 2020). Second, the historical data of the KSE-100 index were solely utilized in this study, which may have restricted the generalizability of the findings in the stock market. Future studies can then expand upon this investigation with the addition of other models and comparisons of ARIMA with different indices to see if any similar trend appears. Additionally, external factors such as interest rates and macroeconomic variables like inflation may impede the forecasting power of the model (Liyanage, 2023). Another limitation of the ARIMA model may be addressed regarding its adjustments for non-linear oscillations within the financial markets. Future studies should consider the possible synergies between these utilized machine learning approaches and classical time series modeling. A combination of GARCH or ARIMA with an LSTM might provide a comprehensive prediction structure. To increase efficiency and clarity, these datasets will be calculated

using both linear and non-linear methods. Further research may explore the effect of two macroeconomic variables on the stock market indices for the sake of augmenting the forecasting model's capacity. Both taxes based on inflation and taxes based on the growth of GDP include changes in exchange rates. According to Jadoon et al. (2024, this could give an in-depth understanding of two variables affecting stock market behaviour. In the end, future studies may investigate how stock market predictions correlate with investor sentiment. Here is an example of how sentiment analysis can cause movements in the stock market. Through news, combining sentiment data and forecasting models will ease the market's volatility, hence giving timely insights to investors.

Recommendations

1. Machine learning techniques would typically be LSTM and GARCH modeling (Mualifah, Soleh, & Notodiputro, 2024) to compensate for ARIMA's downside and extract nonlinear trends from stock market data.

2. Expand the study to include more stock indices and macroeconomic variables to have a more solid ground for predicting analysis (Li et al., 2023).

3. Develop hybrid models that will integrate advanced machine learning techniques with statistical time series methods to enhance prediction accuracy in the unpredictable world of finance (Nasir et al., 2023).

4. To make up for ARIMA's failure to detect the nonlinear patterns in the stock market data, combine different machine learning techniques, such as LSTM and GARCH models.

5. Extend the study area to include multiple stock indices and macroeconomic variables for a more robust forecasting analysis.

6. Create hybrid models that will integrate statistical time series with state-of-the-art machine learning techniques to enhance the accuracy of prediction in the volatile stock markets.

Limitations

1. The study is focused only on the KSE-100 index and does not relate to any other stock indexes or financial markets, as per Ahmad et al. (2018).

2. The major disadvantage of the ARIMA model is its linear assumptions, which are not adequate to take into account the nonlinearity of stock market data (Mochurad & Dereviannyi, 2024).

3. The study is based solely on past information, which will hamper the model's capability to foresee new market shocks or black swans (Singh et al., 2024).

4. The KSE-100 index restricted the study's findings, and these findings may not apply to other stock indices or even other financial markets.

5. The ARIMA model's linear assumptions prove to be a limitation in capturing the nonlinearity of data such as that in the stock markets.

References

Afeef, M., Ali, N., & Khan, A. (2018). Tracing stock returns on a quarterly basis: The case of KSE-100 index. *Global Social Sciences Review*, 3(3), 466–476. https://doi.org/10.31703/gssr.2018(III-III).26

Afeef, M., Ihsan, A., & Zada, H. (2018). Forecasting stock prices through univariate ARIMA modeling. *NUML International Journal of Business & Management*, 13(2), 130–143.

Afeef, M., Shah, F. A., & Kakakhel, S. J. (2020). Stock index prediction using the variance modeling approach. *City University Research Journal*, 10(4).

Ahmad, M. I., Hasan, M., Rafiq, M. Y., Naeem, M. A., & Naseem, M. A. (2018). Time series modeling of KSE-100 index. *SMART Journal of Business Management Studies*, 14(2), 8–16.

Albeladi, K., Zafar, B., & Mueen, A. (2023). Time series forecasting using LSTM and ARIMA. *International Journal of Advanced Computer Science and Applications*, 14(1), 313–320.

Arashi, M., & Rounaghi, M. M. (2022). Analysis of market efficiency and fractal feature of NASDAQ stock exchange: Time series modeling and forecasting of stock index using ARMA-GARCH model. *Future Business Journal*, 8(1), 14.

Armagan, I. U. (2023). Price prediction of the Borsa Istanbul banks index with traditional methods and artificial neural networks. *Borsa Istanbul Review*, 23, S30–S39.

Bauwens, L., Laurent, S., & Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics*, 21(1), 79–109.

Bektemyssova, G., Ahmad, A. R., Mirzakulova, S., & Ibraeva, Z. (2022). Time series forecasting by the ARIMA method. *Scientific Journal of Astana IT University*, 14–23.

Box, G. E., & Jenkins, G. M. (1976). *Time series analysis: Forecasting and control*. San Francisco, CA: Holden-Day.

Cao, J., Li, Z., & Li, J. (2019). The financial time series forecasting model based on CEEMDAN and LSTM. *Physica A: Statistical Mechanics and Its Applications*, 519, 127–139.

Che, C., Huang, Z., Li, C., Zheng, H., & Tian, X. (2024). Integrating generative AI into financial market prediction for improved decision making. *arXiv Preprint*, arXiv:2404.03523.

Chowdhury, M. S., Nabi, N., Rana, M. N. U., Shaima, M., Esa, H., Mitra, A., Mozumder, M. A. S., Liza, I. A., Sweet, M. M. R., & Naznin, R. (2024). Deep learning models for stock market forecasting: A comprehensive comparative analysis. *Journal of Business and Management Studies*, 6(2), 95–99.

Fang, Z., Ma, X., Pan, H., Yang, G., & Arce, G. R. (2023). Movement forecasting of financial time series based on an adaptive LSTM-BN network. *Expert Systems with Applications*, 213, 119207.

García-Medina, A., & Aguayo-Moreno, E. (2024). LSTM–GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios. *Computational Economics*, 63(4), 1511–1542.

He, K., Yang, Q., Ji, L., Pan, J., & Zou, Y. (2023). Financial time series forecasting with the deep learning ensemble model. *Mathematics*, 11(4), 1054.

Ho, S. L., & Xie, M. (1998). The use of ARIMA models for reliability forecasting and analysis. *Computers & Industrial Engineering*, 35(1-2), 213–216.

Jadoon, A. K., Mahmood, T., Sarwar, A., Javaid, M. F., & Iqbal, M. (2024). Prediction of stock market movement using long short-term memory (LSTM) artificial neural network: Analysis of KSE-100 index. *Pakistan Journal of Life & Social Sciences*, 22(1).

Jilhajj, K. (2023). Forecasting lending interest rate and deposit interest rate of Bangladesh using the autoregressive integrated moving average model. *International Journal of Economics and Financial Issues*, 13(3), 169–177.

Jones, N. (2017). How machine learning could help improve climate forecasts. *Nature*, 548(7668).

Kehinde, T., Chan, F. T., & Chung, S. H. (2023). Scientometric review and analysis of recent approaches to stock market forecasting: A two-decade survey. *Expert Systems with Applications*, 213, 119299.

Khan, S., & Alghulaiakh, H. (2020). ARIMA model for accurate time series stock forecasting. *International Journal of Advanced Computer Science and Applications*, 11(7).

Kolambe, M., & Arora, S. (2024). Forecasting the future: A comprehensive review of time series prediction techniques. *Journal of Electrical Systems*, 20(2s), 575–586.

Li, C., Yang, B., & Li, M. (2017). Forecasting analysis of the Shanghai stock index based on the ARIMA model. *MATEC Web of Conferences*.

Li, Y., Wu, K., & Liu, J. (2023). Self-paced ARIMA for robust time series prediction. *Knowledge-Based Systems*, 269, 110489.

Liu, J. (2024). Navigating the financial landscape: The power and limitations of the ARIMA model. *Highlights in Science, Engineering and Technology*, 88, 747–752.

Liyanage, D. (2023). Inflation forecasting using an automatic ARIMA model in Sri Lanka. *International Journal of Economic Behavior and Organization*, 11(2).

Lux, T., & Kaizoji, T. (2007). Forecasting volatility and volume in the Tokyo stock market: Long memory, fractality and regime switching. *Journal of Economic Dynamics and Control*, 31(6), 1808–1843.

Marri, A. A., & Talpur, M. G. (2021). Forecasting using neural networks and stochastic models on the day of the week effect: A case study of the KSE 100 index. *Journal of Contemporary Issues in Business and Government*, 27(5).

Mochurad, L., & Dereviannyi, A. (2024). An ensemble approach integrating LSTM and ARIMA models for enhanced financial market predictions. *Royal Society Open Science*, 11(9), 240699.

Mualifah, L. N. A., Soleh, A. M., & Notodiputro, K. A. (2024). Comparison of GARCH, LSTM, and hybrid GARCH-LSTM models for analyzing data volatility. *International Journal of Advanced Soft Computing Applications*, 16(2).

Mutinda, J. K., & Langat, A. K. (2024). Stock price prediction using combined GARCH-AI models. *Scientific African*, 26, e02374.

Nasir, J., Aamir, M., Haq, Z. U., Khan, S., Amin, M. Y., & Naeem, M. (2023). A new approach for forecasting crude oil prices based on stochastic and deterministic influences of LMD using ARIMA and LSTM models. *IEEE Access*, 11, 14322–14339.

Shaik, M., & Rehman, M. Z. (2023). The dynamic volatility connectedness of major environmental, social, and governance (ESG) stock indices: Evidence based on DCC-GARCH model. *Asia-Pacific Financial Markets*, 30(1), 231–246.

Shumway, R. H., & Stoffer, D. S. (2011). *Time series and its applications: With R examples*. Springer.

Singh, S., Pise, A., & Yoon, B. (2024). Prediction of Bitcoin stock price using feature subset optimization. *Heliyon*, 10(7).

Song, G., Zhao, T., Wang, S., Wang, H., & Li, X. (2023). Stock ranking prediction using a graph aggregation network based on stock price and stock relationship information. *Information Sciences*, 643, 119236.

Wang, X., Kang, Y., Hyndman, R. J., & Li, F. (2023). Distributed ARIMA models for ultra-long time series. *International Journal of Forecasting*, 39(3), 1163–1184.

Xu, X., & Zhang, Y. (2023). A high-frequency trading volume prediction model using neural networks. *Decision Analytics Journal*, 7, 100235.

Zaffar, A., & Hussain, S. A. (2022). Modeling and prediction of KSE–100 index closing based on news sentiments: An application of machine learning model and ARMA (p, q) model. *Multimedia Tools and Applications*, 81(23), 33311–33333.

Zhao, C., Wu, M., Liu, J., Duan, Z., Shen, L., Shangguan, X., Liu, D., & Wang, Y. (2023). Progress and prospects of data-driven stock price forecasting research. *International Journal of Cognitive Computing in Engineering*, 4, 100–108.

Zheng, J., & Huang, M. (2020). Traffic flow forecast through time series analysis based on deep learning. *IEEE Access*, 8, 82562–82570.

Zotteri, G., & Kalchschmidt, M. (2007). A model for selecting the appropriate level of aggregation in forecasting processes. *International Journal of Production Economics*, 108(1-2), 74–83.

Zotteri, G., Kalchschmidt, M., & Caniato, F. (2005). The impact of aggregation level on forecasting performance. *International Journal of Production Economics*, 93, 479–491.