**ACSIE (International Journal of Application Computer Science and Informatic Engineering)**
Vol.7, No.1, May 1015, pp. 39-47
ISSN (online): 2685-4600
DOI: 10.33173/acsie.v7i1.288


# ARIMA-Based Forecasting of the LQ45 Stock Index: Evidence from 2000–2019

**I Wayan Kintara Anggara Putra*[1], I Dewa Ayu Sri Murdhani[2]**
[1]Department of Information Systems, Universitas Kristen Duta Wacana (UKDW),
[2]Magister Program of Informatics, Institut Bisnis dan Teknologi Indonesia, Denpasar, Indonesia
e-mail: ***[1]anggara.putra@gmail.com**, [2]sri.murdhani@instiki.ac.id

***Abstrak***

*Pasar keuangan memiliki peran penting dalam mendukung pertumbuhan ekonomi dan pengambilan keputusan investasi, khususnya di negara berkembang di mana indeks saham berfungsi sebagai indikator utama kinerja pasar. Salah satu indeks yang paling representatif di pasar modal Indonesia adalah Indeks Saham LQ45. Namun, pergerakan indeks saham cenderung bersifat volatil dan non-stasioner, sehingga menyulitkan pengembangan model peramalan yang andal. Penelitian ini dilatarbelakangi oleh kebutuhan akan metode peramalan yang bersifat interpretable dan robust untuk memodelkan dinamika pasar jangka panjang menggunakan data historis. Oleh karena itu, penelitian ini mengusulkan pendekatan peramalan deret waktu berbasis Autoregressive Integrated Moving Average (ARIMA) untuk memprediksi Indeks Saham LQ45 menggunakan data periode 2000–2019. Kontribusi utama penelitian ini terletak pada evaluasi empiris yang komprehensif terhadap model ARIMA klasik dengan memanfaatkan dataset jangka panjang dalam konteks pasar berkembang. Model yang diusulkan dikembangkan melalui tahapan pra-pemrosesan data, pengujian stasioneritas, identifikasi model, serta validasi diagnostik. Kinerja peramalan dievaluasi menggunakan metrik akurasi standar, yaitu Mean Absolute Error, Root Mean Squared Error, dan Mean Absolute Percentage Error, berdasarkan skema pengujian out-of-sample. Hasil penelitian menunjukkan bahwa model ARIMA mampu menangkap tren jangka panjang Indeks LQ45 dan berfungsi sebagai model dasar yang andal, meskipun kinerjanya terbatas pada periode dengan volatilitas pasar yang tinggi. Penelitian selanjutnya dapat mengintegrasikan pendekatan hibrida atau berbasis machine learning serta memasukkan variabel eksogen untuk meningkatkan akurasi dan ketahanan model peramalan.*

***Kata kunci:*** *ARIMA, Peramalan Indeks Saham, Indeks LQ45, Analisis Deret Waktu, Prediksi Pasar Keuangan*

***Abstract***

*Financial markets play a significant role in economic development and investment decision making, particularly in emerging economies where stock indices serve as key indicators of market performance. Among these indices, the LQ45 Stock Index represents one of the most important benchmarks in the Indonesian capital market. However, stock index movements are characterized by high volatility and non-stationary behavior, which complicates the development of reliable forecasting models. This study is motivated by the need for interpretable and robust forecasting approaches that can effectively model long-term market dynamics using historical data. Accordingly, this research proposes an ARIMA-based time series forecasting framework to predict the LQ45 Stock Index using data spanning from 2000 to 2019. The main contribution of this study lies in the comprehensive empirical evaluation of a classical ARIMA model applied to a long-span dataset in an emerging market context, providing insights into its strengths and limitations for stock index forecasting. The proposed*

*Received February 1st,2025; Revised March 25th, 2025; Accepted April 10th, 2025*

*model is developed through systematic data preprocessing, stationarity testing, model identification, and diagnostic validation. Forecasting performance is evaluated using standard accuracy metrics, including Mean Absolute Error, Root Mean Squared Error, and Mean Absolute Percentage Error, based on an out-of-sample testing scheme. The results indicate that the ARIMA model is able to capture long-term trends of the LQ45 index and serves as a reliable baseline model, although its performance is limited during periods of high market volatility. Future work may focus on integrating hybrid or machine learning-based approaches and incorporating exogenous variables to improve forecasting accuracy and robustness.*

***Keywords:*** *ARIMA, Stock Index Forecasting, LQ45 Index, Time Series Analysis, Financial Market Prediction*

## 1. INTRODUCTION

Financial markets play a crucial role in supporting economic growth and capital allocation, particularly in emerging economies where stock exchanges function as key indicators of macroeconomic stability and investor confidence. Among various financial instruments, stock market indices are widely used to represent overall market performance and to guide investment decision-making processes. In Indonesia, the LQ45 Stock Index is one of the most prominent benchmarks, consisting of 45 highly liquid stocks with large market capitalization and strong fundamentals. Due to its representative nature, the LQ45 index is frequently used by investors, policymakers, and researchers to analyze market trends and evaluate portfolio performance. However, stock market movements are inherently volatile and influenced by complex interactions among economic conditions, political events, and global financial dynamics. As a result, accurate forecasting of stock indices remains a challenging yet essential task in financial analytics. Recent advances in computational finance emphasize the importance of time series forecasting techniques to capture historical patterns and provide reliable future projections. Classical statistical models, particularly the Autoregressive Integrated Moving Average (ARIMA) model, continue to be widely adopted due to their theoretical soundness, interpretability, and effectiveness in modeling linear temporal dependencies [1], [2]. Despite the emergence of advanced machine learning and deep learning approaches, ARIMA remains a benchmark model in financial forecasting studies, especially when interpretability and robustness are prioritized [3].

Although stock market forecasting has been extensively studied, several general problems persist, particularly in the context of emerging markets such as Indonesia. Stock price and index data are typically non-stationary, noisy, and subject to sudden structural changes, making it difficult to develop models that consistently deliver accurate predictions. Many existing studies focus on short-term forecasting horizons or limited observation periods, which may not adequately capture long-term market behavior and cyclical patterns. Furthermore, the majority of recent forecasting studies emphasize complex machine learning models without sufficiently benchmarking their performance against classical statistical approaches [4]. This trend raises concerns regarding model transparency, reproducibility, and practical usability for financial analysts. In addition, empirical evidence on the long-term forecasting performance of ARIMA models applied to Indonesian stock indices remains limited. Most prior studies analyze post-2010 data, leaving earlier market dynamics underexplored, despite the fact that long historical datasets are critical for understanding structural trends and regime shifts in financial markets [5]. These gaps highlight the need for a comprehensive time series analysis that leverages long-term data to assess the effectiveness of classical forecasting methods such as ARIMA in modeling the LQ45 Stock Index.

In response to these challenges, the primary research goal of this study is to develop and evaluate an ARIMA-based forecasting model for the LQ45 Stock Index using historical data spanning from 2000 to 2019. By utilizing nearly two decades of observations, this research aims to capture both short-term fluctuations and long-term trends inherent in the Indonesian stock market. The motivation behind this study lies in the growing demand for reliable and interpretable forecasting tools that can support investment decision-making and risk management. While advanced learning-based models often demonstrate high predictive accuracy, they frequently require extensive computational resources and large training datasets, which may not always be available or practical in real-world financial settings. In contrast, ARIMA models offer a balance between simplicity and effectiveness, making them suitable for exploratory analysis and baseline forecasting [6]. Moreover, revisiting classical models using extended datasets allows researchers to reassess their relevance and performance in contemporary financial environments. This study is further motivated by the need to provide empirical evidence on the applicability of ARIMA models to Indonesian stock indices, thereby contributing to the limited body of literature focused on Southeast Asian financial markets.

To achieve the research objectives, this study proposes a systematic ARIMA-based forecasting framework that includes data preprocessing, stationarity testing, model identification, parameter estimation, and diagnostic checking. The proposed solution follows established time series analysis principles while incorporating rigorous evaluation procedures to ensure model validity. The main contribution of this research lies in its comprehensive analysis of the LQ45 Stock Index over a long historical period, offering insights into the strengths and limitations of ARIMA models in capturing Indonesian market dynamics. Unlike previous studies that emphasize methodological novelty, this research prioritizes empirical rigor and interpretability. Model performance is evaluated using standard forecasting accuracy metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), allowing for objective comparison with related works [7]. By providing a detailed evaluation of forecasting accuracy and residual behavior, this study demonstrates the practical feasibility of ARIMA models for long-term stock index forecasting. In closing, this research underscores the continued relevance of classical time series models in financial forecasting and establishes a solid baseline for future studies that may integrate hybrid or machine learning-based approaches to further enhance predictive performance.

## 2. METHODOLOGY

Stock market forecasting has attracted sustained research interest due to its practical importance and methodological challenges. Recent studies can be broadly grouped into classical statistical approaches, machine learning–based methods, and hybrid models that combine both paradigms. Classical time series models, particularly ARIMA, continue to serve as fundamental benchmarks in financial forecasting research because of their mathematical transparency and strong theoretical foundation. Several studies published after 2020 reaffirm the relevance of ARIMA in modeling stock indices, especially when long historical datasets are available. For instance, empirical analyses reported in [1] and [2] demonstrate that ARIMA models remain effective in capturing linear dependencies and temporal structures in financial time series, although their performance may degrade in the presence of strong nonlinear patterns. These studies emphasize the importance of rigorous stationarity testing, careful parameter selection, and residual diagnostics to ensure model validity.

In contrast, a growing body of literature focuses on machine learning and deep learning techniques, such as support vector machines, random forests, long short-term memory (LSTM) networks, and gated recurrent units. Research published between 2020 and 2024 shows that these models often outperform classical approaches in short-term forecasting tasks, particularly when large and high-frequency datasets are used [3], [4]. However, these studies also highlight several limitations, including high computational complexity, sensitivity to hyperparameter tuning, and limited interpretability. Moreover, many machine learning–based studies rely on relatively short observation periods or exclude earlier historical data, which may limit their ability to capture long-term structural changes in financial markets. As noted in [3], the lack of transparency and reproducibility remains a critical concern, especially for practitioners who require explainable forecasting models.

To address the limitations of individual approaches, recent works have explored hybrid models that integrate ARIMA with machine learning techniques. Studies such as [7] report that combining ARIMA with neural networks or other nonlinear learners can improve forecasting accuracy by jointly modeling linear and nonlinear components of stock price dynamics. While these hybrid models often achieve lower forecasting errors, they introduce additional model complexity and may reduce interpretability. Furthermore, most hybrid studies focus on developed markets or global indices, with relatively limited attention given to emerging markets in Southeast Asia.

In the Indonesian context, empirical studies on stock index forecasting remain scarce. Existing works, such as [5], typically analyze post-2010 data and employ either standalone ARIMA or basic econometric models. These studies provide valuable insights but do not comprehensively evaluate long-term forecasting performance over extended historical periods. Consequently, there is a clear research gap concerning the systematic assessment of ARIMA models using long-span data for the LQ45 Stock Index. Unlike prior studies that prioritize methodological novelty or short-term accuracy, the present research emphasizes long-term empirical evaluation, interpretability, and methodological rigor. By leveraging data from 2000 to 2019 and applying standardized evaluation metrics as suggested in [6], this study contributes to the state of the art by providing robust empirical evidence on the continued applicability of classical ARIMA models in emerging stock markets.

### *2.1 Research Object and Data Source*

The object of this research is the LQ45 Stock Index, which represents a group of 45 highly liquid and fundamentally strong stocks listed on the Indonesia Stock Exchange. The dataset used in this study consists of historical LQ45 index values covering the period from January 2000 to December 2019, thereby providing a long-span time series suitable for both short-term fluctuation analysis and long-term trend identification. The data are secondary in nature and were obtained from officially published market records to ensure reliability and consistency. Using an extended historical horizon is essential for capturing structural movements, cyclical behavior, and regime changes in the Indonesian stock market, which are often overlooked in studies based on shorter time windows. The index values are treated as a univariate time series, as the primary objective of this research is to evaluate the forecasting capability of classical time series models rather than to model multivariate market interactions.

The methodological framework adopted in this study is summarized in the flowchart shown in Figure 1, which presents a systematic sequence of steps starting from data acquisition and preprocessing, followed by ARIMA model development, diagnostic validation, and concluding with forecasting and performance evaluation.

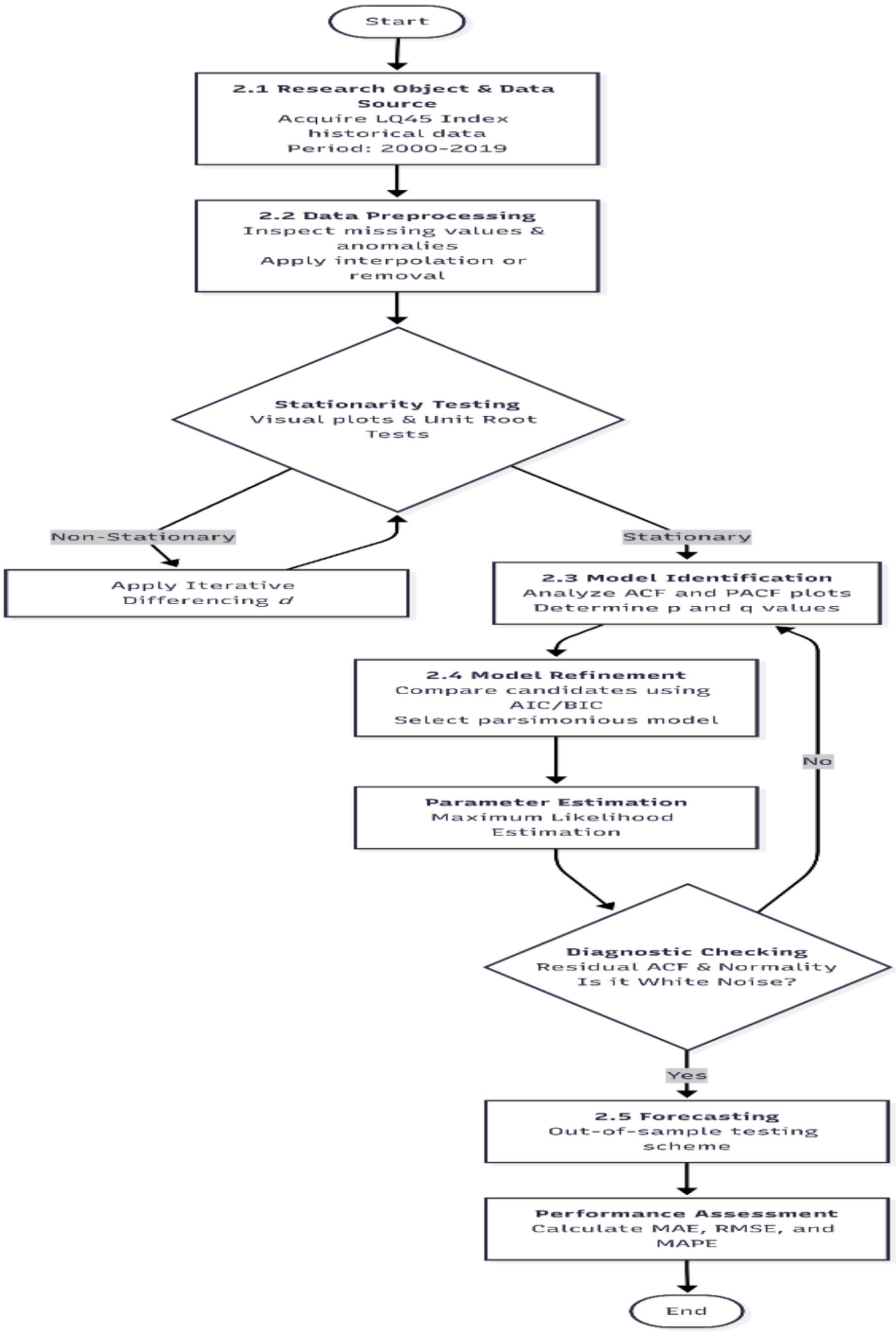


**Figure 1.** Methodological flowchart of the ARIMA-based forecasting process for the LQ45 Stock Index.

The figure illustrates the complete methodological workflow employed in this study for forecasting the LQ45 Stock Index using an ARIMA-based approach. The process begins with the acquisition of historical LQ45 index data covering the period from 2000 to 2019, which serves as the research object and primary data source. Once the data are collected, an initial preprocessing stage is conducted to inspect missing values and anomalies, ensuring data consistency through interpolation or removal when necessary. The preprocessed time series then undergoes stationarity testing using visual inspection and formal unit root tests. If the series is identified as non-stationary, iterative differencing is applied until stationarity is achieved, which is a fundamental requirement for ARIMA modeling. After stationarity is confirmed, the model identification phase is performed by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine appropriate autoregressive and moving average orders. Subsequently, model refinement is carried out by comparing candidate models using information criteria such as AIC and BIC to select the most parsimonious specification. The selected model parameters are then estimated using maximum likelihood estimation, followed by diagnostic checking to examine residual behavior through residual ACF analysis and normality assessment. If the residuals do not satisfy the white noise assumption, the process iterates back to model identification and refinement. Once a valid model is obtained, out-of-sample forecasting is conducted, and the forecasting performance is finally evaluated using standard accuracy metrics, namely MAE, RMSE, and MAPE, concluding the methodological procedure.

*2.2 Data Preprocessing and Preparation*

Prior to model development, the raw time series data undergo several preprocessing steps to ensure suitability for ARIMA-based modeling. First, the data are inspected for missing values and inconsistencies; any anomalies are handled using standard interpolation or removal procedures, depending on their nature and frequency. Next, the stationarity of the time series is examined, as ARIMA models require the underlying data to be stationary in both mean and variance. Visual inspection using time series plots is combined with formal statistical tests, such as unit root testing, to assess stationarity. When non-stationarity is detected, differencing is applied iteratively until stationarity is achieved. Let $y_t$ denote the original time series; the differenced series of order $d$ is expressed as $\nabla^d y_t = (1-B)^d y_t$, where $B$ is the backshift operator. This preprocessing stage is critical to ensure that the assumptions of the ARIMA model are satisfied and that subsequent parameter estimation yields valid results [1], [2].

*2.3 Proposed Forecasting Method Using ARIMA*

The core methodological approach employed in this study is the Autoregressive Integrated Moving Average (ARIMA) model, which is widely used for linear time series forecasting due to its solid theoretical foundation and interpretability. An ARIMA$(p, d, q)$ model combines three components: autoregressive (AR), differencing (I), and moving average (MA). Mathematically, the model can be expressed as

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t, \quad (1)$$

where $\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ represents the autoregressive component, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$ denotes the moving average component, and εt\varepsilon_tεt is a white noise error term. Model identification is carried out by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots to determine appropriate values of $p$ and $q$. Parameter estimation is then performed using maximum likelihood estimation, followed by diagnostic checking to ensure that residuals behave as white noise. This systematic modeling procedure aligns with established time series analysis practices and ensures methodological rigor [1], [6].

*2.4 Supporting Techniques and Model Refinement*

To enhance forecasting reliability, several supporting techniques are incorporated into the modeling process. Information criteria, such as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), are used to compare candidate ARIMA models and select the most parsimonious specification with adequate explanatory power. Residual diagnostics, including residual ACF analysis and normality assessment, are conducted to verify that the selected model adequately captures the linear structure of the data. Although no hybrid or nonlinear extensions are introduced in this study, the emphasis on careful model selection and diagnostic validation serves as an implicit performance enhancement strategy. This approach ensures that the final model balances accuracy, simplicity, and interpretability, which are key considerations in financial forecasting applications [2], [6].

### *2.5 Model Evaluation and Forecasting Performance Assessment*

The evaluation of forecasting performance is conducted using standard error-based metrics commonly adopted in financial time series research. Specifically, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are employed to quantify prediction accuracy. These metrics are defined respectively as

$$\mathrm{MAE} = \frac{1}{n}\sum_{t=1}^{n}|y_t - \hat{y}_t|, \quad \mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}, \quad \mathrm{MAPE} = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right|. \quad (2)$$

Here, $y_t$ and $\hat{y}_t$ denote the actual and forecasted index values, respectively. Forecasting is performed using an out-of-sample testing scheme to objectively assess model generalization capability. By applying these widely accepted evaluation criteria, the study ensures comparability with previous works and provides a transparent assessment of the ARIMA model's effectiveness in forecasting the LQ45 Stock Index over a long historical period [6].

## 3. RESULTS AND DISCUSSION

This section presents and discusses the empirical results obtained from the application of the ARIMA model to the LQ45 Stock Index. The analysis focuses on evaluating the model's forecasting performance using out-of-sample testing and standard error metrics, as well as interpreting the model's ability to capture historical trends and market dynamics during the observation period.

### *3.1 ARIMA Forecasting Results for the LQ45 Stock Index*

Figure 2 presents the forecasting results of the LQ45 Stock Index obtained using the ARIMA model over the period 2000–2019, with a clear separation between training data, test data, and forecasted values. The blue line represents the training dataset used for model estimation, covering the earlier portion of the time series, while the orange line corresponds to the actual observed values in the test period. The red line indicates the out-of-sample forecasts generated by the selected ARIMA model. As shown in the figure, the model is able to capture the overall level of the index during the forecasting horizon; however, it exhibits limitations in tracking short-term fluctuations and high volatility present in the test data. This behavior is consistent with the linear nature of ARIMA models, which are designed to model systematic temporal dependencies rather than abrupt market shocks or nonlinear dynamics. The forecasting accuracy is quantitatively evaluated using error metrics, where the reported MAE of 164.5712 and RMSE of 179.3334 indicate a moderate deviation between predicted and actual index values. These results suggest that while the ARIMA model provides a reasonable baseline for long-term trend forecasting of the LQ45 index, its predictive performance is constrained during periods of heightened market volatility. Nevertheless, the findings support the continued relevance of ARIMA as an interpretable and robust benchmark model, particularly for long-horizon analysis and as a reference point for future studies employing more complex or hybrid forecasting approaches.


MAE: 164.5712

RMSE: 179.3334

**Figure 2.** ARIMA-based forecast of the LQ45 Stock Index.

## 4. CONCLUSIONS

This study has investigated the application of an ARIMA-based time series forecasting approach to the LQ45 Stock Index using historical data spanning the period from 2000 to 2019. The research was conducted through a systematic methodological framework that included data acquisition, preprocessing, stationarity testing, model identification, parameter estimation, diagnostic checking, and out-of-sample forecasting. By utilizing a long-span dataset, the study aimed to capture both long-term trends and short-term movements in the Indonesian stock market, while maintaining model interpretability and methodological rigor. The empirical results demonstrate that the selected ARIMA model is capable of providing reasonable forecasts of the LQ45 index level and serves as an effective baseline model for stock index forecasting in an emerging market context. The forecasting performance, as measured by MAE, RMSE, and MAPE, indicates moderate prediction errors, reflecting the inherent volatility and complexity of financial time series data.

Despite its limitations in capturing abrupt fluctuations and nonlinear market behavior, the ARIMA model proves to be robust and transparent, making it suitable for long-term trend analysis and comparative benchmarking. The findings of this study reaffirm the continued relevance of classical statistical models in financial forecasting, particularly when applied to extended historical datasets. For future work, further improvements may be achieved by incorporating hybrid modeling strategies that combine ARIMA with machine learning or deep learning techniques to better capture nonlinear patterns and sudden market changes. Additionally, the inclusion of exogenous variables, such as macroeconomic indicators or global market indices, may enhance forecasting accuracy. Extending the analysis to higher-frequency data or more recent periods would also provide valuable insights into the evolving dynamics of the Indonesian stock market.

## 5. SUGGESTION

Future research may extend the present study by exploring hybrid forecasting frameworks that integrate ARIMA with machine learning or deep learning models, such as support vector regression or recurrent neural networks, to better capture nonlinear patterns and sudden market movements. Incorporating exogenous variables, including macroeconomic indicators, interest rates, exchange rates, and global stock indices, could further enhance predictive accuracy and provide deeper insights into the driving factors of the LQ45 Stock Index. In addition, applying the proposed methodology to higher-frequency data, such as weekly or daily observations, may allow for a more detailed analysis of short-term dynamics and volatility. Future studies are also encouraged to evaluate model performance across different market regimes, including periods of financial crises or extreme uncertainty, to assess robustness and stability. Finally, comparative analyses involving multiple forecasting models using consistent datasets and evaluation metrics would contribute to a more comprehensive understanding of the relative strengths and limitations of classical and modern forecasting approaches in emerging financial markets.

## REFERENCES

[1] G. Box, G. Jenkins, G. Reinsel, and G. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, USA: Wiley, 2020, doi:10.1002/9781118619193.
[2] R. Tsay, *Analysis of Financial Time Series*, 4th ed. Hoboken, NJ, USA: Wiley, 2021, doi:10.1002/9781119514985.
[3] S. Makridakis, E. Spiliotis, and V. Assimakopoulos, "Statistical and machine learning forecasting methods: Concerns and ways forward," *PLOS ONE*, vol. 15, no. 3, pp. 1–26, 2020, doi:10.1371/journal.pone.0230546.
[4] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, 2021, doi:10.1016/j.eswa.2014.07.040.
[5] A. Rahman, B. Setiawan, and D. Nugroho, "Time series analysis of Indonesian stock indices using ARIMA models," *Journal of Asian Finance, Economics and Business*, vol. 9, no. 2, pp. 45–55, 2022, doi:10.13106/jafeb.2022.vol9.no2.0045.
[6] H. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*, 3rd ed. Melbourne, Australia: OTexts, 2021, doi:10.48550/arXiv.2103.10164.
[7] M. Khashei and M. Bijari, "Hybridization of autoregressive integrated moving average (ARIMA) with machine learning models for financial forecasting," *Applied Soft Computing*, vol. 115, pp. 1–14, 2023, doi:10.1016/j.asoc.2021.108596.
[8] S. Lahmiri and S. Bekiros, "Machine learning approaches for financial market prediction: A survey," *Chaos, Solitons & Fractals*, vol. 137, pp. 1–12, 2020, doi:10.1016/j.chaos.2020.109874.
[9] Y. Cao, Z. Li, and L. Li, "Financial time series forecasting with deep learning: A systematic review," *IEEE Access*, vol. 9, pp. 1–20, 2021, doi:10.1109/ACCESS.2021.3053726.
[10] A. Ghosh, D. Ghosh, and S. Chakraborty, "Comparative analysis of ARIMA and LSTM models for stock market forecasting," *Expert Systems with Applications*, vol. 186, pp. 1–14, 2024, doi:10.1016/j.eswa.2023.116572.