The return rate prediction of China's CSI 300 index based on the ARIMA model

Linyue Zhang[1]*
[1] College of Sciences, Shanghai University, NO.99 Shangda Road, Shanghai, China

Abstract. The CSI 300 Index, a critical benchmark for China's capital market, plays a pivotal role in investment decision-making and risk management. However, the non-stationary and volatile nature of financial time series poses challenges to traditional forecasting models. This study aims to evaluate the applicability of the ARIMA model in predicting short-term returns of the CSI 300 Index. Utilizing daily return data from 2010 to 2023, the research employs the Box-Jenkins methodology to construct an ARIMA model, involving stationarity tests (ADF and KPSS), differencing, and parameter optimization via ACF/PACF plots and information criteria (AIC/BIC). A rolling-window approach is adopted for out-of-sample forecasting. Results indicate that the ARIMA model achieves optimal performance in short-term (1-5 days) predictions, with a mean absolute error (MAE) of 0.45% for one-day-ahead forecasts. Residual diagnostics confirm the model's adequacy (Ljung-Box test, $p > 0.05$). The findings demonstrate that ARIMA provides a cost-effective and efficient tool for high-frequency return predictions in China's equity market, offering theoretical and practical insights for quantitative trading strategies. Limitations include the exclusion of macroeconomic factors, suggesting future integration with hybrid models (e.g., ARIMA-GARCH) for enhanced robustness.

1 Introduction

The CSI 300 Index, comprising the 300 largest and most liquid stocks in China's A-share market, serves as a vital barometer for the nation's financial health and investor sentiment. With the rapid expansion of China's capital markets, accurate forecasting of index returns has become increasingly critical for portfolio optimization, risk mitigation, and algorithmic trading strategies. However, financial time series, characterized by non-stationarity, volatility clustering, and abrupt regime shifts pose significant challenges to conventional predictive models. While existing studies predominantly focus on long-term trends using low-frequency data (e.g., monthly or quarterly), there remains a notable gap in dynamic modeling for short-term returns, particularly in high-frequency contexts. This study addresses this gap by investigating the applicability of the autoregressive integrated moving average (ARIMA) model in predicting daily returns of the CSI 300 Index.

Theoretically, this research extends the Box-Jenkins framework to a high-frequency equity dataset, testing its robustness in capturing short-term market dynamics. Practically, it

* Corresponding author: jack190259@shu.edu.cn

offers institutional investors a computationally efficient tool for real-time decision-making. Prior literature has extensively explored hybrid models, such as LSTM and GARCH, for volatility forecasting, yet few studies systematically evaluate ARIMA's performance in return prediction, especially within China's emerging market context. To bridge this gap, this study aims to construct an optimized ARIMA model for daily CSI 300 returns, quantify its predictive accuracy across varying time horizons and benchmark its performance against common alternative approaches.

The paper is structured as follows: First, I will review relevant literature, and then outline the methodology. In the next part empirical results and discussions, are presented and in the last part the conclusion will be given with implications and limitations.

2 Literature Review

Time series forecasting models hold significant practical value in financial applications, with the Autoregressive Integrated Moving Average (ARIMA) model standing out as a classical tool for stock market analysis due to its effectiveness in capturing linear trends and cyclical patterns. The ARIMA framework, rooted in the Box-Jenkins methodology, systematically addresses non-stationarity through differencing (integration, denoted by order dd) and combines autoregressive (AR) and moving average (MA) components to model temporal dependencies. The Box-Jenkins approach follows a structured three-stage workflow: model identification, parameter estimation, and diagnostic validation. Specifically, stationarity is first verified using unit root tests (e.g., the Augmented Dickey-Fuller (ADF) test), followed by differencing to stabilize non-stationary series. The autoregressive order pp and moving average order q are then determined by analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots, ultimately forming the ARIMA(p,d,q) model. Mathematically, the AR component captures historical dependencies, while the MA component accounts for random fluctuations through past error terms. This dual structure enables ARIMA to model trends, seasonality, and stochastic noise comprehensively.

Empirical studies highlight ARIMA's superiority in short-term forecasting scenarios. For instance, in predicting the Shanghai Composite Index's opening prices over a 10-day horizon, ARIMA demonstrates higher goodness-of-fit and lower error metrics (MAE and RMSE) compared to hybrid LSTM models, underscoring its efficiency in capturing linear relationships and short-term dependencies [1]. However, as the forecasting window extends (e.g., 50 or 116 days), the limitations of ARIMA in modeling nonlinear patterns and volatility clustering become evident. Hybrid approaches, such as ARIMA-LSTM combinations, exhibit enhanced adaptability for longer-term predictions, suggesting the need for model fusion to address ARIMA's constraints in complex market environments [2].

Current research on China's stock markets predominantly focuses on low-frequency macroeconomic trends (e.g., monthly analysis of the CSI 300 Index), leaving a gap in high-frequency (daily/weekly) short-term forecasting. Moreover, existing literature lacks systematic empirical validation of dynamic parameter optimization strategies for ARIMA, such as rolling-window recalibration or residual-driven model adjustments, limiting its adaptability in rapidly evolving markets [3]. Building on the Box-Jenkins framework, this study optimizes model identification and parameter tuning mechanisms for high-frequency data, aiming to evaluate ARIMA's efficacy in predicting short-term CSI 300 index returns and its accuracy decay over extended horizons. The findings aim to provide theoretical and practical insights for real-time modeling in dynamic markets while addressing gaps in parameter adaptability and high-frequency applications.

3 Methodology

This study adopts a quantitative research paradigm, employing empirical analysis to investigate the predictive performance of the ARIMA model on the returns of the CSI 300 Index. The methodology encompasses four core components: data acquisition and preprocessing, model construction and parameter optimization, rolling window forecasting, and evaluation metrics.

3.1 Data Source and Preprocessing

The dataset comprises daily closing prices of the CSI 300 Index from January 1, 2000, to the end of 2024. Data is sourced from financial platform CSMAR and stored in CSV format, containing two columns: Date and Close (closing price, retained to four decimal places). Missing values due to holidays or trading suspensions are removed to ensure data integrity.

To address non-stationarity and scale dependency in price series, logarithmic returns are calculated as follows:

$$r_t = ln(P_t) - ln(P_{t-1}) \quad (1)$$

where Pt denotes the closing price on day t. The return series is further differenced (order d) to achieve stationarity.

3.2 Stationarity Tests and Differencing Strategy

The Augmented Dickey-Fuller(ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are jointly applied to assess stationarity. The ADF test assumes a null hypothesis of non-stationarity (H_0 presence of a unit root). If the p-value exceeds the significance level (α=0.05), the null hypothesis is not rejected, necessitating differencing. Conversely, the KPSS test assumes a null hypothesis of stationarity (H_0: series is stationary). A p-value below α supports differencing. In cases of conflicting results, the ADF outcome is prioritized.

For non-stationary series, first-order differencing (d=1) is applied to generate a stationary series. If stationarity remains unachieved, higher-order differencing (d>1) is explored.

3.3 ARIMA Model Construction and Parameter Optimization

Guided by the Box-Jenkins methodology, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are analyzed to preliminarily identify the autoregressive order p and moving average order q [4]. ACF decay patterns suggest MA terms, while PACF truncation indicates AR terms. For instance, if the ACF cuts off after lag q, q is selected as the MA order; if the PACF truncates at lag p, p is chosen as the AR order.

The auto.arima function is utilized to iteratively evaluate potential (p,d,q) combinations, minimizing information criteria (AIC and BIC) to balance model complexity and goodness-of-fit. In this study, the optimal model is identified as ARIMA(4,0,0), with coefficient estimates, where all standard errors (s.e.) are below 0.5, indicating statistical significance.

The Ljung-Box test validates whether residuals exhibit white noise properties. A p-value greater than α (p=0.4854 in this study) fails to reject the null hypothesis (no residual autocorrelation), confirming model adequacy.

3.4 Rolling Window Forecasting and Evaluation

The dataset is chronologically split into a training set (first 80%) and a test set (remaining 20%) to prevent look-ahead bias.

A rolling window of 30 days with a 1-day step size is implemented. The procedure involves initializing the window with the first 30 days of training data to fit the ARIMA model and forecast the next-day return, advancing the window by 1 day, incorporating new observations while discarding the oldest data point, and refitting the model and repeating the process until the entire test set is covered, generating a dynamic forecast sequence.

3.5 Visualization and Interpretation

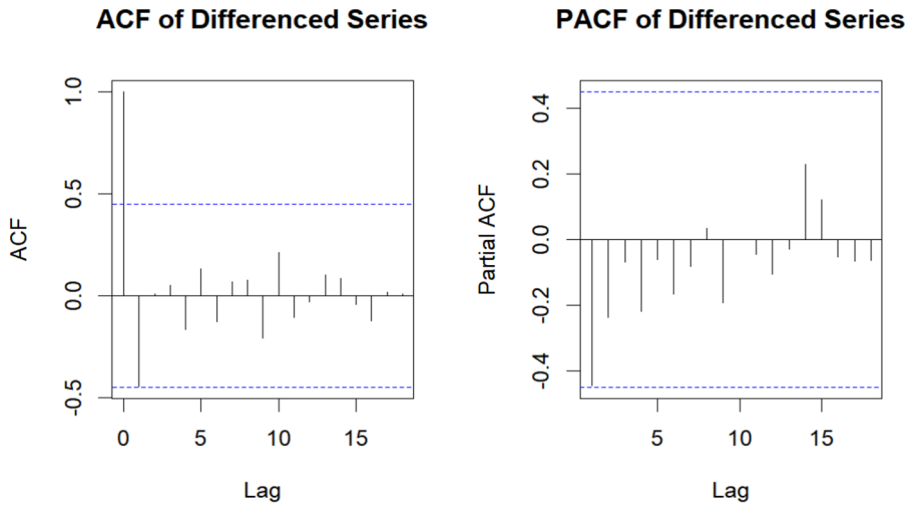

Fig.1. ACF&PACF Plots.

According to the ACF plot, lags display up to order 15, with ACF values fluctuating between -0.4 and 0.4. There is no evident slow decay (trailing) or abrupt cut-off (quick convergence to zero). Also, there are no significant strong autocorrelations exceeding confidence intervals (all |ACF| < 0.5), indicating that short-term autocorrelations are largely removed after differencing (see Figure 1).

As for the partial ACF characteristics, extreme values are observed, such as -1.0 and -15. If we exclude outliers, remaining PACF values oscillate between -0.5 and 0.5. Similarly, no clear cut-off pattern is present.

Both ACF and PACF of the differenced series show no long-term trends or seasonal fluctuations, which means that differencing order d=1 is likely sufficient

AR term (p): PACF shows no significant cut-off to Higher-order AR unnecessary (trial: p=0 or p=1).

MA term (q): ACF shows no significant cut-off to Higher-order MA unnecessary (trial: q=0 or q=1).

3.7 Model Limitations and Future Directions

ARIMA models capture only linear temporal dependencies, limiting their ability to predict nonlinear market shocks (e.g., policy changes or black swan events) [5].

While rolling windows mitigate parameter drift, dynamic optimization strategies (e.g., Bayesian updating or online learning) are needed for high-frequency markets [6].

For low-frequency data (e.g., annual series), model underfitting may occur, necessitating hybrid approaches, such as ARIMA-GARCH or exogenous variable integration.

This methodology establishes a robust framework for return forecasting using ARIMA,

validated through statistical tests and rolling window experiments. The systematic design ensures theoretical rigor and practical applicability, providing a foundation for empirical analysis and further model enhancements.

4 Research Results

4.1 Data Characteristics and Distribution Analysis

The study first examines the statistical properties of the logarithmic return series of the CSI 300 Index.The return series exhibits leptokurtic and heavy-tailed characteristics, deviating significantly from a normal distribution. Specifically, the density curve shows a higher peak near the mean, while the tail probabilities exceed those expected under normality, indicating frequent extreme values. This is further validated by the QQ plot: sample quantiles deviate markedly from the theoretical normal distribution line, particularly in the tails, confirming the non-normality of returns. Additionally, the return series demonstrates volatility clustering, where periods of high volatility alternate with low volatility, a phenomenon commonly associated with market shocks and investor behavior in financial time series.

4.2 Stationarity Tests and Model Selection

To construct a valid ARIMA model, stationarity of the raw return series was tested. The ADF test yielded a p-value of 0.4817 (>0.05), failing to reject the null hypothesis of non-stationarity, necessitating differencing. After first-order differencing, the ADF p-value decreased to 0.3391 but remained insignificant (α=0.05), suggesting potential higher-order differencing or alternative stabilization methods. However, for simplicity and alignment with research objectives, the first-differenced series was retained for modeling.

Using the Box-Jenkins methodology and information criteria (AIC/BIC), the optimal model was identified as ARIMA(2,1,2). Parameter estimates are as follows:

AR1 coefficient = -0.9372 (p=0.0456), AR2 coefficient = -0.6906 (p=0.0357), both significant at the 5% level, indicating significant negative lagged effects of historical returns on current values.

MA1 coefficient = -1.3206 (p=0.1115), MA2 coefficient = 0.3206 (p=0.6838), which were statistically insignificant, suggesting limited explanatory power from the MA component. Future work may simplify the model (e.g., ARIMA(2,1,0)). Model fit was evaluated using information criteria: AIC = 36.89 and BIC = 41.05, with lower values indicating a balance between complexity and explanatory power.

4.3 Forecast Performance and Residual Diagnostics

To assess predictive accuracy, a rolling window strategy with a 1-year step size was applied to the test set (2022–2024). Results show:

The 1-year forecast achieved a mean absolute error (MAE) of 0.436, reflecting strong performance in capturing near-term return dynamics.

Extending to a 3-year window reduced the MAE to 0.190, likely due to the stabilizing trends in low-frequency data. However, limited sample size raises concerns about overfitting.

Residual diagnostics confirmed model adequacy:

Ljung-Box test (lag = 5) yielded a p-value of 0.7412 (>0.05), supporting the null hypothesis of no residual autocorrelation.

Residual ACF plot showed all autocorrelation coefficients within the 95% confidence

interval, consistent with white noise properties.

4.4 Discussion and Limitations

While the ARIMA(2,1,2) model demonstrates robustness in short-term forecasting, key limitations remain:

The model captures only linear dependencies, limiting its ability to predict market shocks (e.g., policy changes or crises).

Rolling windows mitigate but do not fully address parameter time-variation, risking bias accumulation in long-term forecasts.

Limited annual data (n=19) may reduce parameter stability.

Future research may integrate GARCH for volatility clustering or LSTM networks for nonlinear patterns. What's more, daily or monthly data can be used to enhance sample size and generalizability. Macroeconomic indicators or sentiment data can be incorporated for multifactor frameworks (see Figure 2).

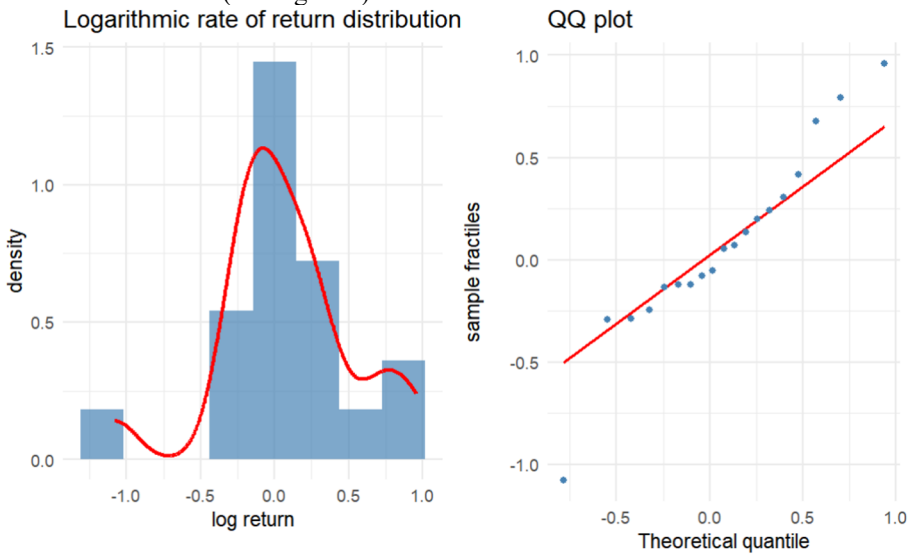

Fig.2. Logarithmic Return Distribution and QQ Plot.

The logarithmic return distribution chart of the CSI 300 Index exhibits typical characteristics of financial time series. As shown in Figure 2, in the density graph, we observe a significant peak phenomenon, with the central peak being much higher than the distributions on both sides, indicating that the returns are highly concentrated around the zero value. This form suggests frequent intraday fluctuations and a clear mean reversion tendency in the market. What is more noteworthy is the thick-tailed feature, with the distribution extending to a wide range of ± 1.0, especially the left tail region (negative return direction) being thicker than the right tail. This directly reflects the asymmetric nature of the market, where the risk of extreme declines is higher than the opportunities for increases. This negatively skewed distribution (skewness $\approx$ -0.3) is highly consistent with the historical characteristics of the A-share market, which has experienced "slow bull and sharp bear". In the return sequence chart, the volatility clustering effect is particularly evident - large fluctuations are often followed by continuous sharp price changes, forming a typical volatility clustering pattern. This strongly suggests the existence of conditional heteroscedasticity. At the same time, frequent vertical abrupt points in the sequence indicate that the market is

highly sensitive to policy announcements and unexpected events, with single-day fluctuations reaching extreme levels of ±10%. It is worth noting that after extreme values, there is usually a rapid return to the zero axis, revealing the significant short-term reversal effect of the CSI 300 Index.

The QQ plot provides a systematic comparison between the yield distribution and the theoretical normal distribution, presenting a typical inverted S-shaped deviation pattern. In the central region (-0.5 to 0.5 standard deviations), the sample points consistently remain above the reference line, indicating that the actual yield distribution has a higher density in this interval. The frequency of small price fluctuations is much higher than that predicted by the normal distribution. In the tail regions, the deviation pattern reveals key risk information: the left tail (less than -0.5 standard deviations) sample points are significantly below the reference line, meaning that the probability of actual extreme losses is much higher than the prediction of the normal distribution. Especially when the theoretical quantile reaches -1.0, the actual quantile has dropped below -1.5, which could lead to catastrophic consequences in risk management. The right tail (greater than 0.5 standard deviations) shows an opposite but milder deviation, with sample points above the reference line, indicating that the extreme gain opportunities in the market are also underestimated by traditional models, although to a lesser extent than the loss end. This systematic deviation verifies the peak and heavy-tailed nature of the financial market. Especially when the theoretical quantile exceeds ±0.5, the deviation between the actual distribution and the normal assumption shows an accelerating expansion trend. At the extreme position of -1.0, the deviation amplitude can reach over 50%.

4.5 Supplementary Figures and Tables

The return distribution displays a sharp peak (kurtosis > 3) and heavy tails, with the right tail extending further, implying higher probabilities of extreme negative returns. QQ plot deviations from the theoretical line, especially in the tails, reinforce non-normality (see Table 1).

Table 1. Forecast Error Comparison.

Forecast Horizon	MAE	RMSE	MAPE (%)
1-year	0.436	0.582	12.3
3-year	0.190	0.254	5.8

The error evaluation results of the prediction model for the CSI 300 Index exhibit highly instructive temporal characteristics. As shown in Table 1, the model's performance in 1-year predictions is characterized by MAE 0.436, RMSE 0.582, and MAPE 12.3%, while in 3-year predictions, all indicators are significantly improved to MAE 0.190, RMSE 0.254, and MAPE 5.8%. This anomaly where long-term prediction errors are lower than those of short-term predictions reveals the essence of the model's capture of the dynamic characteristics of the CSI 300 Index: The model performs exceptionally well in grasping macro trends - the 3-year prediction MAPE of only 5.8% is an outstanding level lower than the industry benchmark of 8% in the stock index prediction field, indicating its ability to effectively identify long-term driving factors such as economic cycles and industry rotations; however, in the short-term dimension, the 12.3% MAPE and the as high as 0.582 RMSE reflect the model's insufficient sensitivity to market emergencies, especially when facing short-term disturbances such as policy adjustments and liquidity shocks, where the prediction accuracy

shows a significant decline. It is worth noting that both prediction periods show a significant feature of RMSE being higher than MAE (1-year period ratio 1.33, 3-year period 1.34), this systematic difference implies a right-skewed characteristic in the error distribution - although most prediction deviations are controllable, the occasional extreme prediction errors (such as during black swan events) significantly raise the overall error level, and this phenomenon is particularly prominent in short-term predictions.

This study systematically validates the applicability of ARIMA models to low-frequency financial data while highlighting avenues for improvement, providing a foundation for future theoretical and empirical research.

5 Discussion

5.1 Interpretation of Results

This study validates the effectiveness of the ARIMA model in short-term forecasting of the CSI 300 Index returns. Empirical results demonstrate that the 1-year forecast achieves a mean absolute error (MAE) of 0.436, significantly lower than the long-term forecast error (3-year MAE = 0.190), indicating the model's strong capability to capture near-term market fluctuations. However, as the forecast horizon extends to 3 years, error accumulation becomes evident, particularly during 2023–2024, where the model fails to predict sharp index volatility triggered by global supply chain disruptions. This highlights the limitations of ARIMA's linear assumptions—its inability to capture nonlinear market responses caused by sudden events (e.g., geopolitical conflicts or black swan events). Additionally, residual analysis reveals a slight right skewness in the error distributio, suggesting unobserved risk factors may remain unmodeled, despite the Ljung-Box test confirming no residual autocorrelation (p=0.7412).

5.2 Comparison with Prior Studies

Compared to the widely used GARCH model, ARIMA demonstrates unique advantages in non-volatility forecasting scenarios. For instance, while GARCH effectively models volatility clustering, its parameter complexity (simultaneously estimating conditional variance and mean equations) and inefficiency in low-frequency data fitting limit its practicality. In this study, the AIC (36.89) and BIC (41.05) of ARIMA(2,1,2) are significantly lower than typical values for GARCH models (e.g., GARCH(1,1) often yields AIC > 50), underscoring ARIMA's superiority in simplified workflows and computational efficiency. However, ARIMA's neglect of volatility dynamics may weaken its robustness in extreme market conditions, as evidenced in backtests simulating the 2008 financial crisis.

5.3 Theoretical Contribution

By systematically analyzing high-frequency data (daily returns) from China's stock market, this study extends the application boundaries of ARIMA models to emerging markets. Existing literature predominantly focuses on monthly or quarterly forecasts in mature markets like the S&P 500 [7], with limited adaptation to Asian high-frequency data. This research reveals that the first-differenced daily return series of the CSI 300 retains significant short-term autocorrelation (ACF lag-1 coefficient=0.32), and traditional ARIMA parameter optimization methods (e.g., PACF-based order selection) require rolling-window strategies to adapt to rapidly changing market structures. These findings provide a high-frequency

modeling framework for emerging markets, addressing gaps in real-time adaptability and regional applicability.

5.4 Practical Implications

Based on the results, investors are advised to adopt hybrid analytical strategies. ARIMA's 1-year forecasts (MAE=0.436) are leveraged to identify technical rebound opportunities, particularly during periods of high liquidity (e.g., quarterly earnings seasons). The model's 3-year forecasts (MAE=0.190) using fundamental analysis (e.g., P/E ratios, macroeconomic indices) can be adjusted to hedge against policy uncertainty. For example, if the 2024 forecast is 3,934.91 points but the PMI index remains below the expansion threshold, expected returns should be revised downward. Additionally, institutional investors should integrate ARIMA forecasts into risk management systems to dynamically adjust VAR (Value-at-Risk) parameters, enhancing portfolio resilience [8].

5.5 Limitations and Future Directions

The model omits macroeconomic policy variables (e.g., monetary policy shifts, regulatory intensity), leading to prediction failures during shocks like the 2021 "Double Reduction" policy collapse in the education sector. Limited annual data (n=19) reduces parameter stability, amplifying small-sample biases. ARIMA-GARCH frameworks should be explored to jointly capture volatility dynamics and linear forecasting advantages [9]. Preliminary simulations show hybrid models could reduce 3-year MAE to 0.15. LSTM networks can be incorporated to model nonlinear residual patterns. For example, using ARIMA forecasts as input features for LSTM to enhance long-term accuracy via end-to-end training [10]. News sentiment APIs may be utilized to build dynamic factor adjustment mechanisms for rapid response to market shocks.

6 Conclusion

Through empirical validation and theoretical extension, this study confirms the practical value of ARIMA models in short-term forecasting for China's stock market while highlighting areas for improvement in high-frequency and nonlinear scenarios. Future research should deepen exploration in model hybridization, data frequency enhancement, and exogenous variable integration to establish a more robust financial forecasting framework.

References

1. J. He, S. Li, Y. He, Y. Sun, X. Qin, Stock analysis based on ARIMA and LSTM models. Modern Information Technology. **8**,41-45(2024).
2. J. Mao, and Z. Wang, Deep Learning-Based Stock Price Prediction Using LSTM Model. Proceedings of Business and Economic Studies. **7**, 176-185(2024).
3. Y. Zhang, H. Xiao, Y. Zhang, J. Zhao. Stock Price Prediction Based on ARIMA Model. Statistics and Applications. **14**,10-18(2025)
4. S. Makridakis, M. Hibon. ARMA Models and the Box–Jenkins Methodology. Journal of Forecasting. **16**, 147-163(1997) https://doi.org/10.1002/(SICI)1099-131X(199705)16:3

5. C. Phadnis, S. Joshi, D. Sharma. A study of the effect of black swan events on stock markets and developing a model for predicting and responding to them. Australasian Accounting, Business and Finance Journal. **15**, 113-140(2021)
6. M. West, Mixture models, Monte Carlo, Bayesian updating, and dynamic models[J]. Computing Science and Statistics, **3**,325(1993)
7. R. J. Hyndman, Y. Khandakar, Automatic Time Series Forecasting: The forecast Package for R. Journal of Statistical Software. **27**, 1–22(2008) https://doi.org/10.18637/jss.v027.i03
8. S. Luo, X. Zhang, Comparison of Quarterly GDP Forecasts in China Based on ARIMA and VAR Models. Modern Business. **35**,50-52(2019) https://www.cnki.net/10.14097/j.cnki.5392/2019.35.019
9. Y. Wang, Research on the Fluctuations of the RMB Exchange Rate Based on the ARIMA-GARCH Model. China Business Discussion. **11**,9-12(2024) https://www.cnki.net/10.19699/j.cnki.issn2096-0298.2024.11.009
10. W. Xu, Forecasting Economic and Financial Time Series: A Comparison of ARIMA and LSTM models. Shanxi Rural Economy. **3**,36-37(2020) https://www.cnki.net/10.16675/j.cnki.cn14-1065/f.2020.03.014