



Performance Comparison of ARIMA, XGBoost and LSTM Models on the NASDAQ Dataset

Ruize Ma

School of Economics, Shenzhen University, Shenzhen, Guangdong, 518000, China
2023020112@email.szu.edu.cn

Abstract. Based on the historical trading data of the Nasdaq Composite Index, this paper constructs three prediction models—a traditional econometric model (ARIMA), a machine learning model (XGBoost), and a deep learning model (LSTM)—under a unified experimental setting on the same dataset and train/test split. The ARIMA model is trained with rolling one-step-ahead forecasting, while the XGBoost and LSTM models are trained with feature-based regression using fixed hyperparameters. Their performance on the 2023-2024 test set is evaluated with four widely used metrics: MAE, RMSE, MAPE, and R^2 . Results show that the ARIMA model achieves the best short-term forecasting accuracy (MAPE 0.8790%, R^2 0.9941), accurately reflecting short-term fluctuations. The LSTM model ranks second, capturing the overall trend well through long-term temporal dependencies. The XGBoost model is the least accurate, especially on unidirectional trends.

Keywords: Nasdaq Composite Index; Time Series Prediction; ARIMA Model; XGBoost Model; LSTM Model; Performance Comparison

1 Introduction

With the rapid development of global capital markets, the prices of financial assets have become increasingly volatile, and accurately forecasting financial time series has become an important task for investors, financial institutions, and regulatory authorities. The Nasdaq Composite Index, which includes thousands of leading technology companies around the world, reflects both the health of the technology industry and many other factors, such as the domestic and global economy, industry policies, and investor sentiment. It is therefore one of the most representative time-series datasets in financial markets.

Traditional models for financial time series forecasting are generally linear and cannot adequately capture the nonlinearity and non-stationarity of financial indices [1]. In recent years, machine learning and deep learning have provided powerful tools for analyzing complex financial time series. ARIMA is a classic yet simple method [1], widely applied in finance. XGBoost addresses nonlinear problems in high-dimensional data by means of gradient boosting and regularization [2]. LSTM is a recurrent neural network that alleviates the long-term dependency problem and is suitable for extracting deep features from complex time series data [3].

Most existing studies focus on optimizing or improving individual models, or on comparing only a subset of the three paradigms on a given dataset. Fischer and Krauss [4] showed that LSTM networks can outperform memory-free methods in financial market prediction; Siami-Namini et al. [5] compared ARIMA and LSTM and found the latter more accurate on complex data; Zhang [6] demonstrated that hybrid ARIMA-neural network models can improve forecasting accuracy. However, few studies provide a systematic quantitative comparison of ARIMA, XGBoost, and LSTM on the same financial index under a unified experimental setting. This paper fills this gap by developing the three models on the same train/test split.

Regarding originality, the dataset is publicly available from the Nasdaq website; the three models were implemented by the author in Python with statsmodels, xgboost, and TensorFlow, and all experiments were run on the author's personal computer. No experimental results or figures were taken from other sources. Generative AI tools were used only to assist with English language polishing; all technical content, modeling, experiments, and analysis are the author's own work.

2 Prediction Model Construction

2.1 Data and Experimental Setup

Daily historical trading data of the Nasdaq Composite Index were collected from the Nasdaq website [7]. The dataset contains daily observations of the index, including the closing price, open, high, low, volume, and daily percentage change. The training set covers the observations up to 2022-12-31, and the test set consists of the daily observations over 2023-2024. The target variable is the closing price of the next trading day, so the task is formulated as a regression problem of next-day price prediction.

For ARIMA, the order (5, 1, 0) is determined via the ACF/PACF analysis of the first-order differenced series, and the model is re-estimated and rolled forward one step at a time over the test set. For XGBoost, the features include 20 lagged closing prices, several moving averages, volatility and momentum measures, candlestick-style features, volume indicators, and lagged daily changes, with `n_estimators` = 500, `learning_rate` = 0.05, `max_depth` = 6, and early stopping with 30 rounds of patience. For LSTM, the input is a 60-day sliding window of six features normalized with the Min-Max scaler; the network has two LSTM layers of 128 and 64 units with a dropout of 0.2, followed by two dense layers, and is trained with the Adam optimizer ($1e-3$), MSE loss, batch size 32, up to 100 epochs, and early stopping (patience 15). All three models were implemented by the author and run on a personal computer.

Three models—ARIMA, XGBoost, and LSTM—are constructed and compared in this paper. Their parameters are set according to the characteristics of the data and the model, so that each model is used under conditions that match its inductive bias for prediction and stability.

2.2 ARIMA Model

ARIMA, in its full name Autoregressive Integrated Moving Average, is an ARIMA(p, d, q) model: p is the order of autoregression, d is the order of differencing to make a non-stationary series stationary, and q is the number of moving-average terms. The basic idea is to difference the series by order d, then determine the autoregressive and moving-average parts to forecast the future [1]. ARIMA has a simple principle and is suitable for short-term prediction of mildly non-stationary series.

2.3 XGBoost Model

XGBoost (Extreme Gradient Boosting) is an ensemble learning method that uses gradient boosting trees: many weak classifiers (decision trees) are trained repeatedly and combined to produce a final output [2]. Regularization terms (L1 and L2) reduce overfitting, and parallel computing accelerates training via gradient descent. XGBoost is well suited to high-dimensional non-linear regression.

2.4 LSTM Model

LSTM (Long Short-Term Memory) is a recurrent neural network (RNN) modified to solve the long-term dependency problem of conventional RNNs. Three gates—input, forgetting, and output—decide whether to add new information, alter the importance of previous cell states, or export the modified cell state [3]. LSTM extracts deep features from time series and handles nonlinearity and non-stationarity in financial data without manual feature engineering.

3 Experimental Results and Performance Comparison

This paper uses MAE, RMSE, MAPE, and R^2 to assess the accuracy and stability of the three models from complementary perspectives, following standard practice in the forecasting literature [8]; a model is judged to have better predictive performance when it yields smaller MAE, RMSE, and MAPE together with a larger R^2 [8]. It should be noted that, because the models are evaluated on the absolute level of the index, R^2 tends to be high for all models since the index exhibits a strong upward trend over the test period. R^2 should therefore be interpreted together with the scale-dependent MAE and RMSE and the percentage-based MAPE.

3.1 Results of the Single Model

ARIMA Model Prediction Results. The ARIMA model uses rolling one-step-ahead forecasting, suitable for short-term prediction of the Nasdaq index. The forecast curve is close to the actual values without systematic bias; the residual volatility is small and evenly distributed, with no apparent trend in the residuals.

The ARIMA results are presented as a global trend and residual plot, together with an enlarged view of the test set, providing an intuitive demonstration of the model's predictive performance. As shown in Figs. 1 and 2, the x-axis represents time; the y-axis in Fig. 1 is the index level, while that in Fig. 2 is the residual. The blue solid line is the actual value, the red solid line is the ARIMA prediction, gray dashed lines or bars are the residual series, and the zero line denotes zero deviation.

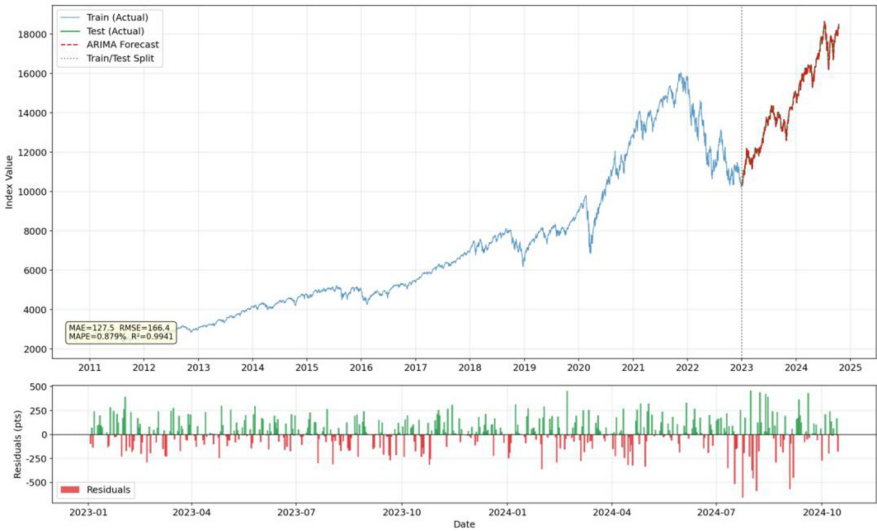


Fig. 1. ARIMA Model Prediction Results for the Nasdaq Index (Global Trend + Residuals) (Photo/Picture credit: Original).

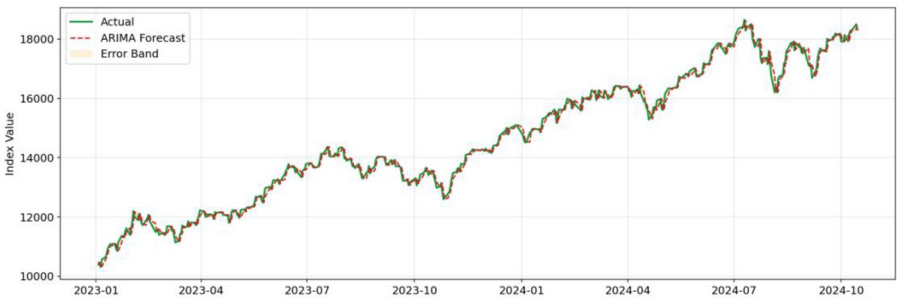


Fig. 2. Details of the ARIMA model test set (2023-2024) (Photo/Picture credit: Original).

The ARIMA model achieves $MAE = 127.53$, $RMSE = 166.45$, $MAPE = 0.8790\%$, $R^2 = 0.9941$, showing excellent accuracy and stability for short-term time series forecasting.

XGBoost Model Prediction Results. The prediction curve of the XGBoost model shows a large deviation from the actual values, especially during the continuous rise

of the Nasdaq Index in 2024, where the predictions were significantly delayed and dampened. As shown in the residual plot, the residuals exhibit considerable fluctuation and a noticeable systematic bias, suggesting that the model struggles to adapt to changes in the underlying trend and tends to reproduce the most recent level rather than to extrapolate the trend.

The XGBoost results are shown in Figs. 3 and 4, with the x-axis as time; Fig. 3 shows the index level, and Fig. 4 shows the residual. The blue solid line is the actual value, the red solid line is the XGBoost prediction, and gray dashed lines or bars are the residual series.

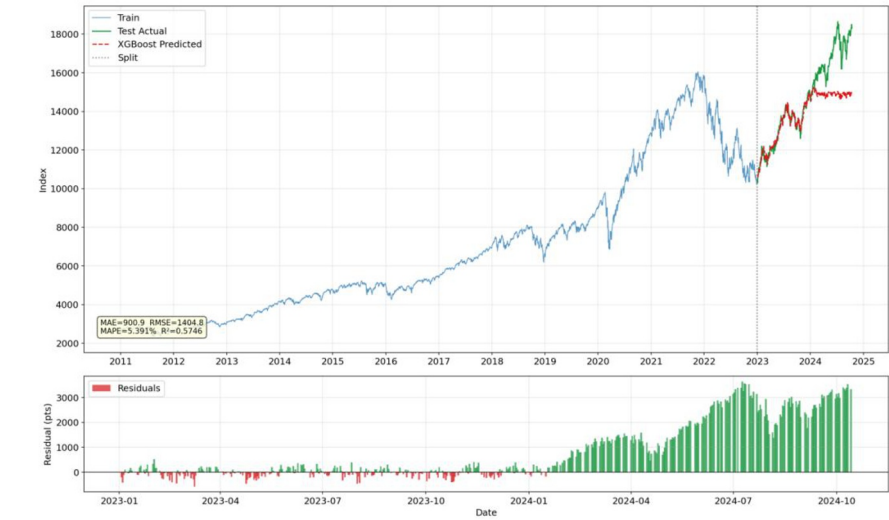


Fig. 3. XGBoost model's prediction results for the Nasdaq Index (global trend+residuals) (Photo/Picture credit: Original).



Fig. 4. Details of the XGBoost model on the test set (2023-2024) (Photo/Picture credit: Original).

The XGBoost model achieves $MAE = 900.91$, $RMSE = 1404.85$, $MAPE = 5.3909\%$, $R^2 = 0.5746$, all relatively high; its generalization ability is weak, especially for unidirectional trend prediction.

LSTM Model Prediction Results. The LSTM model is able to capture the general trend of the Nasdaq index better than the XGBoost model, and its prediction curve shows relatively good consistency with the actual values. However, there is a systematic bias at the peaks and troughs, and the fluctuations around the extremes are not fully reflected. As shown in the residual plot, the residual volatility remains relatively stable without large trend deviation, indicating good generalization ability and adaptability to complex time-series fluctuations.

The LSTM results are shown in Figs. 5 and 6, with the x-axis as time; Fig. 5 shows the index level, and Fig. 6 shows the residual. The blue solid line is the actual value, the red solid line is the LSTM prediction, and the gray dashed line is the residual series.



Fig. 5. LSTM model's prediction results for the Nasdaq Index (global trend + residuals) (Photo/Picture credit: Original).



Fig. 6. Details of the LSTM model's test set (2023-2024) (Photo/Picture credit: Original).

The LSTM model achieves MAE = 436.98, RMSE = 507.91, MAPE = 2.9605%, $R^2 = 0.9448$, all between ARIMA and XGBoost; it fits the trend reasonably well, with short-term accuracy slightly below ARIMA.

3.2 Comparison of the Comprehensive Performance of the Three Models

Quantitative Indicator Comparison. Table 1 compiles the quantitative evaluation indicators of all three models so that their performance can be compared directly on the same test set.

Table 1. Comparison of Performance Metrics for the Three Models

Model	MAE(Index Point)	RMSE(Index Point)	MAPE(%)	R^2
ARIMA	127.53	166.45	0.8790	0.9941
LSTM	436.98	507.91	2.9605	0.9448
XGBoost	900.91	1404.85	5.3909	0.5746

As shown in Table 1, the ARIMA model has the best performance across all four indicators, with a MAPE as low as 0.8790% and an R^2 of about 0.9941, indicating excellent short-term forecasting accuracy. The LSTM model is next, with a MAPE of 2.9605% and an R^2 of 0.9448, showing that it can grasp the overall trend well. The XGBoost model performs the worst, with a MAPE of 5.3909% and an R^2 of only 0.5746, suggesting that it is not as accurate or stable as the other two.

Discussion. There are still some deficiencies in this study, and in the future, further research will be conducted in the following directions:

Several observations can be drawn from the results. First, the ARIMA model achieves the highest accuracy on the test set: for a mildly non-stationary series that is made stationary by first-order differencing, the parsimonious (5, 1, 0) specification re-estimated in a rolling manner adapts quickly to the local dynamics of the index. Second, the XGBoost model performs worst on the sustained upward trend of 2024: because its features are mainly lagged values of the price, the model tends to reproduce the most recent level and thus systematically lags behind a persistent trend, leading to large errors during strong upward movements. Third, the LSTM model strikes a balance between trend fitting and generalization: it follows the overall trajectory well, but its point-wise accuracy at peaks and troughs is limited by the smoothing effect of sequence-based learning. These findings suggest that no single model dominates in all settings, and the choice of model should depend on the forecasting horizon and the expected dynamics of the market. Moreover, because the current models are evaluated on absolute price levels, future work will reformulate the forecasting target as log returns or relative returns, following common practice in financial time series studies [4], to provide a more stringent and economically meaningful evaluation.

Only the historical trading data of the Nasdaq Composite Index is used; other factors, such as economic changes, industry policies, and public opinion, are not

incorporated, even though they may be informative for improving prediction accuracy.

All three models are basic; future extensions include adding non-linear components to ARIMA, attention mechanisms to LSTM, or time-series attention to XGBoost to enhance its ability to capture temporal dependence.

At present, the paper aims to forecast only daily closing prices; it will be expanded to cover high-frequency intraday data and multi-step predictions of 7- or 30-day trends.

Only the prediction accuracy of the three models is compared; future work may also add training efficiency, computational cost, and interpretability for a more comprehensive model-selection reference.

4 Conclusion

In this paper, three prediction models — ARIMA, XGBoost, and LSTM — are developed for the historical trading data of the Nasdaq Composite Index based on different modeling paradigms. Through quantitative indicators and visual analysis, a full comparison of the prediction abilities of the models is performed. The results show that the ARIMA model performs best for short-term forecasting on this dataset, followed by the LSTM model, while the XGBoost model is relatively weak, especially on unidirectional trends. These findings reflect the compatibility of each modeling paradigm with the characteristics of different data: ARIMA is suitable for short-term stationary time series, LSTM is well suited to complex nonlinear time series, and XGBoost is generally more appropriate for non-time-series or weakly trending data. The work provides practical references for model selection in financial forecasting tasks.

References

1. Box, G.E.P., Jenkins, G.M.: Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco (1970)
2. Chen, T., Guestrin, C.: XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785-794. ACM, New York (2016)
3. Hochreiter, S., Schmidhuber, J.: Long Short-Term Memory. *Neural Computation* 9(8), 1735-1780 (1997)
4. Fischer, T., Krauss, C.: Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research* 270(2), 654-669 (2018)
5. Siami-Namini, S., Tavakoli, N., Siami Namin, A.: A comparison of ARIMA and LSTM in forecasting time series. In: Proceedings of the 17th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 1394-1401. IEEE, Orlando (2018)

6. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* 50, 159-175 (2003)
7. Nasdaq Composite Index Historical Data, <https://www.nasdaq.com/market-activity/index/comp>, last accessed 2024/12/31
8. Hyndman, R.J., Athanasopoulos, G.: *Forecasting: Principles and Practice*, 3rd edn. OTexts, Melbourne (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

