Utilizations of the ARIMA Model: Empirical Findings from the Nasdaq Index

Zichu Guo[1,a,*]

[1]*Adam Smith Business School, University of Glasgow, Beijing, China*
a. 1701050130@stu.hrbust.edu.cn
**corresponding author*

Abstract: The Nasdaq Composite is one of the world's most watched stock indices and assumes an important role in the global financing trading process. Forecasting of stocks has always been the focus of scholars' research, but due to factors such as market sentiment fluctuations, it is difficult to obtain effective forecasting results using a macro approach. With the help of technological development, data has become more accessible. Scholars have started using statistical models to make predictions on stock prices. This study uses autoregressive integrated moving average (ARIMA) prediction model to forecast the Nasdaq composite index. After parametric statistics, fitting and residual tests, it was determined that the ARIMA (2,1,2) model produced the most accurate predictions. The results of this study enriched the prediction results of the relevant Nasdaq Composite Index for the ARIMA model, broadened the field of application of the ARIMA model, and provided investors with investment strategy references.

Keywords: Nasdaq, Time series, ARIMA model, Forecast

1. Introduction

Previous research has demonstrated a correlation between the performance of the stock market and the overall economic conditions of the world and the economic conditions of a specific country. These studies have found that stock market movements are unpredictable and it is not possible to foretell future returns [1]. Moreover, the stock market serves the purposes of risk sharing, effective allocation of resources, and minimising information and transaction expenses. The presence of the stock market enables companies to showcase their profitability, thus influencing the economic strategies of the nation and consumers' inclination to invest [2].

Macroeconomic methodologies typically fail to forecast stock market fluctuations due to the influence of numerous unknown factors on market participants' emotions [3]. Global crisis events such as pandemics can lead to worldwide economic contractions, which in turn have a detrimental effect on the stock market [4]. Analysing the future trend of a stock based on macro factors such as stock prices, inflation, the financial condition of the company, and other parameters involves extensive data processing and financial expertise. This significantly raises the complexity of making predictions. While it is acknowledged that predictions relying on these parameters may be unreliable, the accuracy of the results cannot be assured [1,5].

Due to technological advancements, researchers now have easy access to large sample-size data and can observe more detailed stock market data, including daily and even minute-to-minute fluctuations. They have discovered that stock data can be analysed as time series, which exhibit seasonal patterns. As a result, researchers have started using statistical models to attempt to forecast stock market movements [1,6]. The ARIMA prediction model is a widely used statistical model for analysing and forecasting time series data. It utilises autoregression (AR), integration (I), and moving average (MA) techniques to forecast future values by analysing past values of the time series data [7].

This study forecasts future Nasdaq Composite Index values using the ARIMA model. A stock's volatility is influenced by factors such as price, company financial condition, and inflation. The NASDAQ composite index is widely regarded as a benchmark for the technology industry's progress. Since its inception in 1971, it has consistently included shares of technology companies from the United States and abroad. At present, it comprises over 2,500 individual stocks [8]. Forecasting the composite stock index is a significant and difficult task. Previous studies in the literature have rarely utilised the ARIMA model to forecast the time series analysis of the Nasdaq composite index. This study expands the scope and enhances the findings of the ARIMA model concerning stock market-related indices.

This study offers investors valuable insights into investment methods and theoretical guidance for improving global financial investment attitudes, considering the significant role played by the Nasdaq Composite Index in worldwide financing operations [9].

2. Method

2.1. Data Collection

The data utilised in this analysis comprises the final prices of the Nasdaq Composite Index from 1 January 2010 to 1 January 2024. The dataset spans from 4 January 2010 to 29 December 2023, accounting for holiday closures. The data originates from Yahoo Finance (URL: https://finance.yahoo.com/quote/%5EIXIC?.tsrc=fin-srch). The adjusted closing price was selected in the dataset because it considered factors such as dividends and stock splits, which can enhance the accuracy of projections [10, 11]. The original time series has been removed due to nonexistent or missing data.

Figure 1: Nasdaq performance from 2010 to 2024

Figure 1 depicts a consistently increasing trajectory in the Nasdaq Index from 2010 until the conclusion of 2021, reaching its peak in December 2021. From 2022 to 2023, the Nasdaq Composite

Index underwent a substantial decrease as a result of the pandemics, the Russo-Ukrainian conflict, and the Federal Reserve's decision to increase interest rates [4,12]. However, starting from 2023, the index has seen a gradual recovery and the data has rebounded to a level close to its previous peak. Ultimately, the Nasdaq Composite Index exhibits a pattern of fluctuation and unpredictability, lacking stability as a time series dataset.

2.2. Research Contents

This study utilises ARIMA modelling and forecasting using the methodology introduced by Box and Jenkins in 1970. It seeks to evaluate and forecast time series data using the past information of the Nasdaq Composite Index from the last two decades. ARIMA models for time series analysis require the data to be stable. Nevertheless, the closing prices of the sample raw data stock indexes exhibit a consistent and volatile pattern. This study uses the Augmented Dickey-Fuller (ADF) test to evaluate the non-stationarity of the data. If the p-value generated by the ADF test is less than 0.05, it suggests that the initial hypothesis can be rejected, implying that the data is stable. On the other hand, if the statistic of p exceeds 0.05, differencing becomes necessary to attain stability in the time series. To determine the seasonality of the time series, it is essential to apply a logarithmic transformation to the original data. In the ARIMA (p,d,q) model, the value of p corresponds to the number of autoregressive components, while d represents the smoothing step. On the other hand, q indicates the correlation between current and past residuals. The autocorrelation function (ACF) provides valuable insights into estimating the difference term d. To identify the values of p and q, it is crucial to perform the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) tests, respectively. Several error measures can be useful to evaluate model performance. These include a measure of root mean square error (RMSE), the mean absolute percentage error (MAPE), the mean absolute error (MAE), normalised the Akaike Information Criterion (AIC), and the normalised Bayesian Information Criterion (BIC). After obtaining the best model, the residuals were subjected to the Ljung-Box test. The test yielded a p-value greater than 0.05, suggesting that the residual sequence exhibits characteristics of white noise and validating the effectiveness of the modelling.

2.3. ARIMA Model

The ARIMA(p,d,q) model is expressed as a linear combination of the forecasted values, the original values, and the original errors. The autocorrelation coefficient ACF represents the association between the variable y_t and its lagged value at the kth order y_{t-k} The partial autocorrelation coefficient (PACF) measures the correlation between y_t and the delayed kth order y_{t-k} while accounting for the influence of the preceding values yt-1, yt-2, etc. on y_{t-k}. The formula is as stated:

$$\Delta y_t = c + \alpha_1 \Delta y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p \Delta y_{t-p} + \epsilon_t - b_1 \epsilon_{t-1} - b_2 \epsilon_{t-2} - \cdots b_q \epsilon_{t-q} \tag{1}$$

3. Empirical Results Analysis

In order to verify whether the time series is smooth or not, the Augmented Dickey-Fuller Test Unit Root Test was used in this study and the following results were obtained (see Table 1):

Table 1: ADF test

Value of test-statistic	-3.1926	5.8254	5.0964
t-test	-3.96	-3.41	-3.12

DOI: 10.54254/2754-1169/87/20240892

According to the results of the ADF test, the value of the t-statistic is greater than the t-value at any of the levels of significance, so the time series can be considered non-stationary. This result is the same as that demonstrated by the autocorrelation function ACF and the partial autocorrelation function PACF in Figure (2), i.e., it shows a trailing characteristic. Therefore, the time series is judged to be non-stationary.

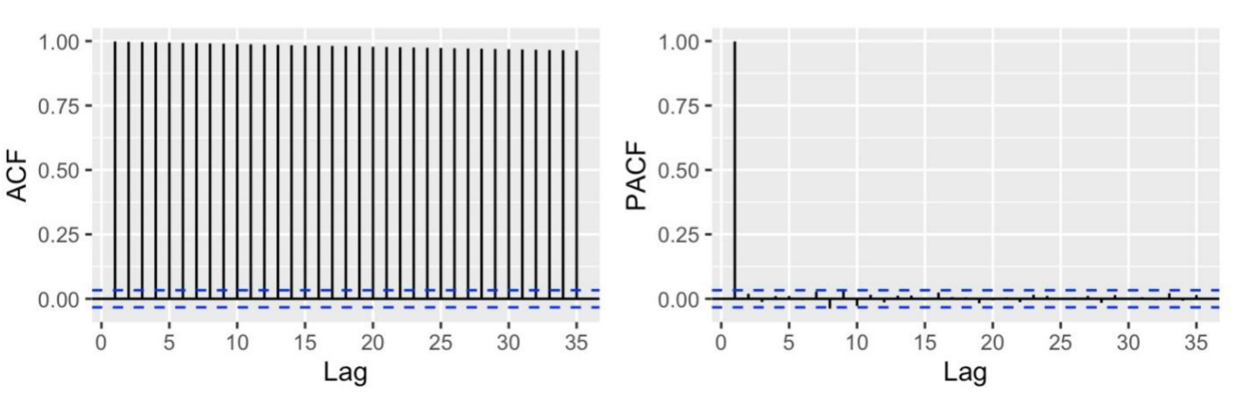

Figure 2: Autocorrelation function and partial autocorrelation function

After passing the unsteady time series through logarithmic transformation and first-order differencing, a smooth time series is obtained where the upward trend has disappeared and seasonality is evident, as shown in Figure 3:

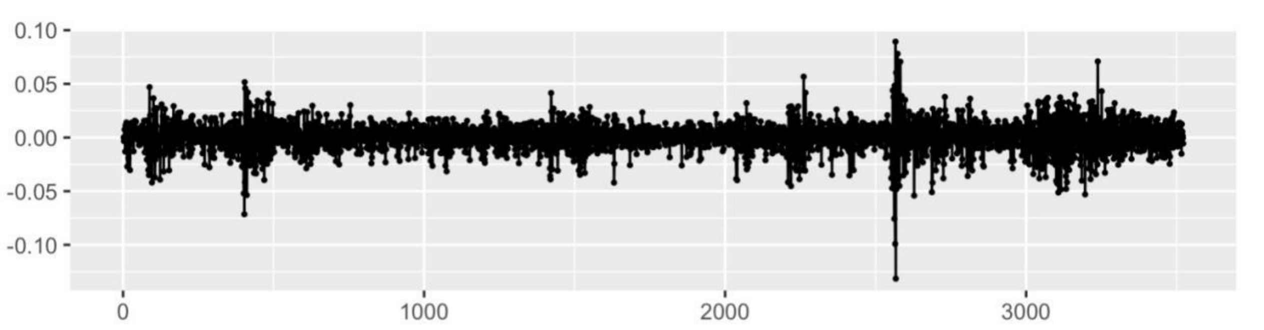

Figure 3: First Differences of log (Nasdaq)

The parameter estimation can be done here using auto.arima command and the results show that the best model is ARIMA (2, 1, 2) and the model performance is measured and the results are shown in Table 2:

Table 2: The model performance

RMSE	MAE	MPE	MAPE	MASE	AIC
0.01269519	0.008852969	-0.000104578	0.1019059	0.9947497	-20744.17

The ARIMA parameters are (2, 1, 2) By fitting test, the model with the smallest AIC value performs best, so the best model output is ARIMA (2, 1, 2). The parameters are shown in Table 3.

Table 3: ARIMA (2, 1, 2) parameters

AR (1)	AR (2)	MA (1)	MA (2)
-0.2505	0.0367	0.1603	-0.0031

Then, the feasibility of the ARIMA model was tested by performing the Ljung-Box test on the residual series of ARIMA (2, 1, 2), which yielded a p-value = 0.9605, which is greater than 0.05. Therefore, the residual series could not reject the original hypothesis, and it was considered to be white noise, and the modelling was successful.

Finally, this study needs to test whether there are outliers in the residuals, which is verified by Q-Q plots, and from the results in Figure 4, it can be observed that the residuals are consistent with normality and there are not a large number of outliers, which are almost always on a straight line, so that the ARIMA (2, 1, 2) proposed in this study is more than adequate to fit the data and can be used to make the next step of prediction.

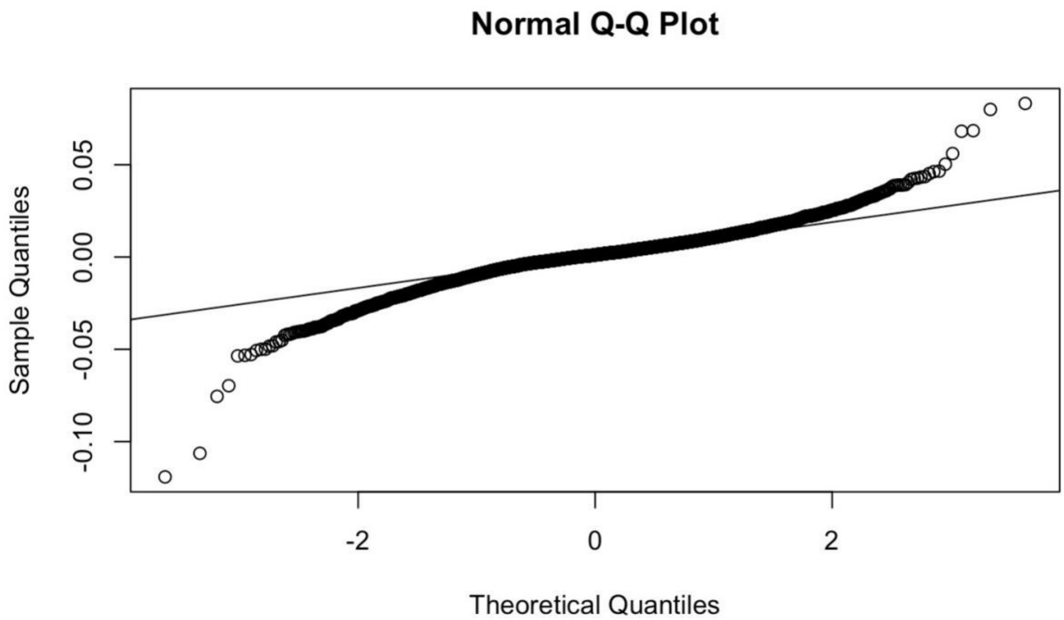

Figure 4: Q-Qplot of the residuals

Above all, through the Ljung-Box test and the residual Q-Q plot test for many times, the experimental results show that the model is feasible and has a good expression, and this study can be concluded that the Nasdaq Composite Index can be represented by ARIMA (2, 1, 2).

In the prediction session, this study predicts the adjusted closing price for the 10 trading days after 1 January 2024, and the prediction results are shown in Figure 5 below, and Table 4 shows the comparison between the predicted closing price and the actual closing price. Although there is a difference in the values, the upward trend is the same and the actual values are within the 95% forecast limit. It demonstrates the accuracy of ARIMA (2, 1, 2) proposed in this study on the prediction of Nasdaq Composite Index.

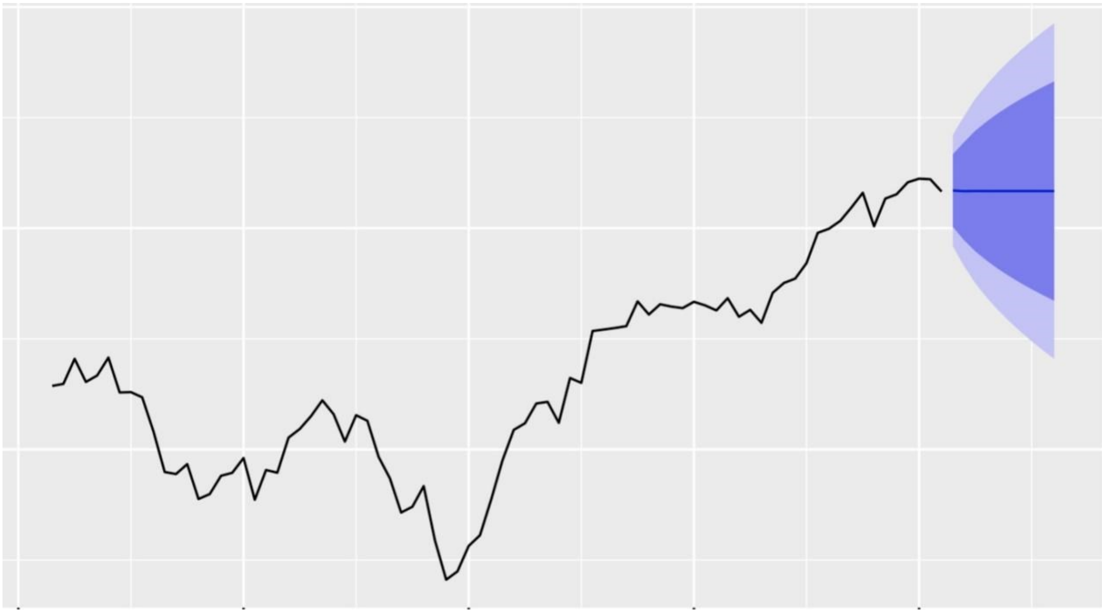

Figure 5: forecasting figure

Table 4: Predicted results

Date	predicted close	actual close
2/1/2024	15,400.74	14,765.94
3/1/2024	15,532.72	14,592.21
4/1/2024	15,655.54	14,510.30
5/1/2024	15,754.32	14,524.07
8/1/2024	15,843.89	14,843.77
9/1/2024	15,924.69	14,857.71
10/1/2024	15,999.63	14,969.65
11/1/2024	16,069.64	14,970.19
12/1/2024	16,135.71	14,972.76
16/1/2024	16198.45263	14,944.35

4. Conclusion

This study analysed the daily closing prices of Nasdaq Composite Index during the last 14 years and applied the ARIMA model to predict the movement of the stock index for the next ten trading days. The ARIMA (2,1,2) model proposed in this study is effective in predicting the Nasdaq Composite Index. The results of this study extend the sample range of the ARIMA model prediction, it is very difficult to forecast the time series of the financial data in the Nasdaq Composite Index, which is watched by the whole world, and there is still room to continue to improve this study, such as the use of the integrated model to avoid the linear dependence. This study also provides a basis for investors in their investment strategies related to the NASDAQ and aids in investment decisions.

References

[1] Shah, J. S., Vaidya, D., & Shah, M. (2022). A comprehensive review on multiple hybrid deep learning approaches for stock prediction. Intelligent Systems With Applications, 16, 200111.
[2] Tripathi, V., & Seth, R. (2014). Stock market performance and Macroeconomic Factors: The Study of Indian Equity Market. Global Business Review, 15(2), 291–316.
[3] Xing, F. Z., Wang, Z., & Welsch, R. E. (2018). Intelligent asset allocation via market sentiment views. IEEE Computational Intelligence Magazine, 13(4), 25–34.
[4] Li, W., Chien, F., Kamran, H. W., Aldeehani, T. M., Sadiq, M., Nguyen, V. C., & Taghizadeh-Hesary, F. (2021). The nexus between COVID-19 fear and stock market volatility. Ekonomska Istrazivanja-economic Research, 35(1), 1765–1785.
[5] Thakkar, A., & Chaudhari, K. (2020). CREST: Cross-Reference to Exchange-based Stock Trend Prediction using Long Short-Term Memory. Procedia Computer Science, 167, 616–625.
[6] Kobiela, D., Krefta, D., Król, W., & Weichbroth, P. (2022). ARIMA vs LSTM on NASDAQ stock exchange data. Procedia Computer Science, 207, 3836–3845.
[7] Geurts, M. D., Box, G. E. P., & Jenkins, G. M. (1977). Time Series Analysis: Forecasting and control. Journal of Marketing Research, 14(2), 269.
[8] Chen, J. (2023, August 2). What does the Nasdaq Composite Index measure? Investopedia.
[9] Hussein Neama, N. (2013). Accumulated remark forecasting for American NASDAQ stock market by using artificial neural network models from 2006-2015. Global Journal of Finance and Management., 8, 143–151.
[10] Chen, L. (2020). Using machine learning algorithms on prediction of stock price. Journal of Modeling and Optimization, 12(2), 84–99.
[11] Indikadulle, P., Hendahewa, D., Perera, N., D, D. M., Perera, S., & Rathnayake, S. (2021). Selecting a suitable variable combination to predict stock prices using support vector machines.
[12] Breeden, D. T., & Litzenberger, R. H. (2022). Central Bank policy impacts on the distribution of state prices for future interest rates, 2003–2022. The Journal of Fixed Income, 32(2), 64–92.