Amazon's Stock Trends Prediction based on ARIMA Model

Stacy Garner[1], Yingxian Pan[2, *] and Meijia Shi[3]

[1] Hunan Hongyi International High School, Changsha, 410131, China

[2] Guangdong University of Finance & Economics, Guangzhou, 510000, China

[3] Henan University of Economics and Law, Zhengzhou, 450046, China

*Corresponding author: 1707030209@stu.hrbust.edu.cn

Abstract. The rapid development of e-commerce in recent years has made stock market prediction a challenging yet important task for investors and traders. The ARIMA model has been extensively applied for forecasting in financial time series data. This study applied the ARIMA model to predict the Amazon stock price, utilizing historical stock price data from a period of a decade. The results of the analysis indicated that the ARIMA model can effectively predict short-term stock price movements with a certain level of accuracy. The model managed to encapsulate the predominant trend in the variations of Amazon's stock prices, providing insights into potential future price movements. The predicted stock prices closely matched the actual prices, with a reasonable level of precision. Furthermore, the ARIMA model can be useful for investors and traders in making informed decisions. By utilizing the model's predictions, investors can better assess the potential risks and returns associated with their investment strategies. It should be noted that the ARIMA model is not without its limitations. The model may not perform well during periods of structural changes or when there are significant events affecting the stock price. Besides, it can also be influenced by other factors such as policies, market sentiment and corporate fundamentals. In conclusion, the ARIMA model can assist in predicting short-term Amazon stock price movements, serving as a valuable tool for investors and traders. However, it is essential to integrate the model's forecasts with other pertinent information and analysis to form comprehensive investment decisions.

Keywords: Stock price forecasting; ARIMA model; autoregression; prediction accuracy.

1. Introduction

The stock market, characterized by its complicatedness and dynamism, has been a focal point for thorough research and analysis. Predicting the direction of future stock trends and the exact values of stock prices is one of the most difficult challenges in the financial markets. Stock market prediction has mesmerized investors and analysts for a quite long time, playing an essential role in increasing the opportunity to obtain maximum profit, reduce risks and provide basis for decision-makers. In the last few decades, abundant studies have investigated whether the stock could be predicted [1].

Stock trends prediction is of great importance for both investors and stakeholders. The desire of decision-makers for precise predictions of the stock market became a motivation for inventing new accurate models. It turns out that if predicted successfully, the investment can create a huge windfall. However, the volatility and information asymmetry of financial market make it challenging to forecast the stock trends. In recent decades, methods of stock price analysis have been emerging and upgrading. Methods like qualitative method, causal or behavioral models [2], neural networks [3] and ANFIS-induced ordered weighted average weighted average (OWAWA) [4] have been widely used by scholars in price forecasting. The objective of this study is to investigate the utility of the Auto-Regressive Integrated Moving Average (ARIMA) model for forecasting specific stock market trends, taking Amazon's stock data as an example. The performance of ARIMA can be assessed by Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), etc.

Time series model has been broadly used and practiced in science, engineering and other fields, and has been developed and reformed by lots of scholars [5]. The history of using ARIMA model to predict stock price can date back to the 1990s [6]. ARIMA model is an improvement of the traditional single time series model auto-regressive moving average (ARMA), which takes differencing into account [7]. Hence, ARIMA can deal with non-stationary time series data while ARMA assumes that the mean and variance of the time series stay unchanged with time. By de-trending or suppressing quasi-periodic smooth components [8], ARIMA uses differencing to solve the non-stationarity of time series, such as periodicity and seasonality [9], to obtain stationary time series. It focuses mainly on the autocorrelations in the data [10].

AR models, or auto-regressive models, which using a linear combination of past value of the variable [10], have the advantage of capturing long-term trends in data and making predictions based on those trends. Nevertheless， the AR model struggles with certain types of time series data, particularly those featuring temporary fluctuations and unexpected shifts. The AR model largely overlooks the complexity and unpredictability of real-world influences. On the contrary, MA model, i.e. moving average model, could handle those time series data with temporariness or abrupt change better. However, the MA model may fail to account for longer-term historical trends in the data. MA model assumes that time series is relatively stable while it is difficult to maintain the assumption of stability in real world. Based on the above, ARIMA combines the advantages of both AR and MA models, and by using differencing, it can deal with more complex time series situations.

Previous literature in the field of financial prediction has primarily concentrated on the development of novel models, with less emphasis on the validation of their suitability for every unique individual stock. This gap in research is addressed by this paper, which seeks to investigate the efficacy of the ARIMA model in forecasting the stock price trends of Amazon, a leading player in the e-commerce industry. By applying the ARIMA model to historical Amazon stock data, this study aims to provide valuable insights that can assist stakeholders, investors, and market analysts in making more informed decisions regarding their investments. The findings are expected to contribute to a deeper understanding of the complex dynamics of the stock market, specifically in the context of Amazon, and potentially aid in enhancing investment strategies.

2. Methodology

2.1. Data Source

Preprocessing historical daily stock price data of Amazon from January 2010 to December 2019 from Choice Financial Terminal was the first stage in this investigation, including 2516 observations with an increasing trend of price overall. The dataset, which includes elements like date, open, high, low, close, adjusted close, and volume, has been collated for additional analysis. To focus the prediction efforts, the 'Open' price was identified as the primary feature of interest. Min-Max scaling was applied to standardize the data and ensure uniformity in both time series analysis and LSTM (Long Short-Term Memory) neural networks.

2.2. Indicators Selection

The stock price is a long-term statistic, therefore research on various topics concludes with the entire statistic. Plotting below in Fig. 1, the whole and entire statistics are available for the study of time series and segmented time series, determining the primary rules of how these trends were evolving.

Observation of the daily time series plot of the stock price reveals some simple conclusions, overall the Fig 1 shows an increasing trend with fluctuations.

Fig. 1 Time-sequence-closing Price

2.3. Model Selection

A key element of this study was Time Series Analysis, a well-known forecasting technique. The investigation of Time Series Analysis began with a careful assessment of dataset stationarity. The consistency of statistical features like mean and variance across time is ensured by stationarity, a predetermined condition that is essential for time series modeling. The data was tested for stationarity or non-stationarity using the Augmented Dickey-Fuller (ADF) test, which initially revealed non-stationarity. Differencing was utilized—a crucial step in this analysis—to overcome non-stationarity. By essentially deducting each value from its previous time step, the process of differencing entailed calculating the first difference of the "Open" price. The seasonality and trends included in time series data can be effectively mitigated with this strategy. The ADF test was repeated after differencing, verifying the achievement of stationarity.

Box and Jenkins presented the ARIMA model in 1970, which is also known by another name, the Box-Jenkins methodology, ARIMA models can be determined, approximated, and diagnosed through the use of time series data. Financial forecasting frequently employs the model as one of its techniques. The ARIMA models have proven to be effective in producing short-term projections. Short-term prediction was consistently superior to intricate structural models. Model identification, parameter checking, and diagnostic testing form part of the process of constructing an ARIMA predictive model.

The autoregressive integrated moving average (ARIMA) model is a commonly used statistical model for time series analysis and forecasting. It is a powerful tool for capturing patterns and trends in data as they change over time. Therefore, we chose the ARIMA model to establish a forecasting model for Amazon's stock price, which combines three main components: autoregressive (AR), integral (I), and moving average (MA).

3. Results and Discussion

3.1. ADF Test

The Augmented Dickey-Fuller test is employed to test the stationarity of a time series, where the null hypothesis assumes that the series is non-stationary. Firstly, usually, if the p-value is less than 0.1 (or 0.05 as a standard), it signifies that the null hypothesis is rejected at the 0.1 level, implying the series exhibits stationary. Secondly, if the series is not stationary, one can perform first-order or second-order differencing and then conduct the ADF test until the series becomes stationary.

As can be seen from the table 1, for the Closing Price, the time series data ADF test yields a t-statistic of 0.706, with a p-value of 0.990. The critical values at 1%, 5%, and 10% are -3.433, -2.863, and -2.567, respectively.

With $p = 0.990 > 0.1$, the failure to reject the null hypothesis suggests that the series is non-stationary. A first-order difference is then applied to the series, and the ADF test is conducted again.

After the first-order differencing, the ADF test results show that $p = 0.000 < 0.01$, offering over 99% confidence to reject the null hypothesis. At this point, the series is deemed stationary.

Table 1. Closing Price-ADF Inspection Table

Difference order	t	p	Threshold		
			1%	5%	10%
0	0.706	0.990	-3.433	-2.863	-2.567
1	-9.024	0.000	-3.433	-2.863	-2.567

3.2. ACF and PACF Test

According to the Closing Price, combined with the Fig 2 and 3, it can be preliminarily judged that the p value of the autoregressive order is 1, and the q value of the moving average order is 1.

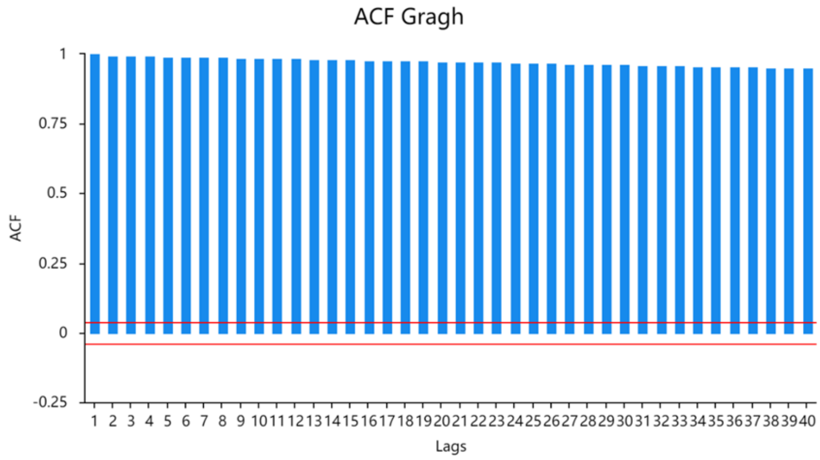

Fig. 2 ACF plot

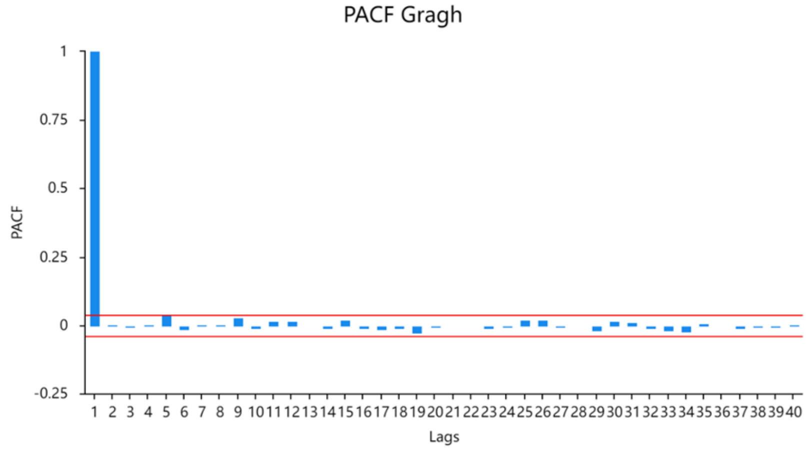

Fig. 3 PACF plot

3.3. Model Results

We tried different combinations of p and q values to compare RMSE and AIC indices of different models (Table 2).

Table 2. ARIMA model results

Model	RMSE	AIC
ARIMA(0,1,0)	15.8284	21033.153
ARIMA(1,1,3)	15.8148	21036.827
ARIMA(1,1,0)	15.8254	21034.180
ARIMA(1,1,1)	15.8105	21031.464

After comparison, we find that when p is 1, q is 1, and d is 1, RSME and AIC values are the lowest, which is the optimal model. According to the model of ARIMA (1,1,1) parameter list, available model formula to y(t)=0.109+0.840*y(t-1)-0.866*ε(t-1).

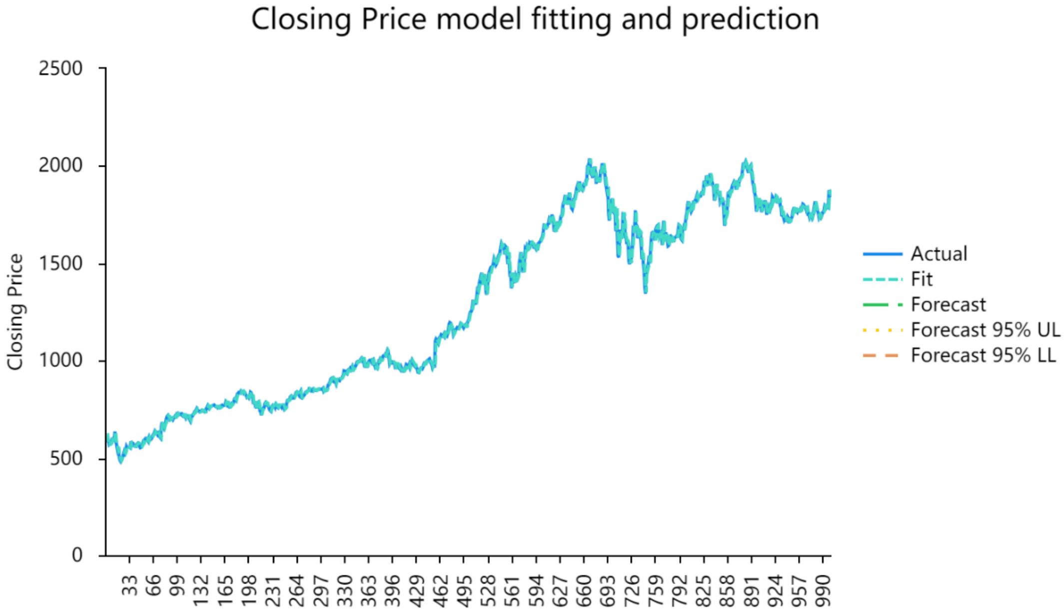

Fig. 4 Closing Price model fitting and prediction

The model shows the fit of the data and the prediction of the data for the last 12 periods. Fig 4 shows the actual closing price, the fit, the forecast for the upper 95% confidence interval, and the forecast for the lower 95% confidence interval. It can be observed that the forecasts broadly align with the actual trajectory of the closing prices, and the 95% confidence interval's upper and lower bounds suggest prices that are marginally above or below the actual closing values.

3.4. Discussion

The results have demonstrated that the model is of high calibre in predicting Amazon's stock trends, showing relatively similar stock trends comparing to recent years. This proves that the ARIMA model can predict individual stock trend to some extent. However, the trend of stock market is influenced by numerous factors, for example, macroeconomic environment, corporations' management performance, interest rates, exchange rates etc. Unpredictable factors like covid-19 in 2021 hit global financial market, leading to a steep decline in global stock market, shows that it is difficult to fully predict the trend of stocks using linear model solely.

More sophisticated and detailed mathematical models are therefore needed in order to predict individual stock trend more accurately. Many scholars combine time series model with other models to enlarge the range of variables and form more precise prediction. Artificial intelligence could be considered as a main research interest in further study, since advanced science and technology develop rapidly nowadays. The application of artificial intelligence in financial sector still remains an extensive development space.

Artificial intelligence can build complicated machine learning models that learn from historical data to make more accurate predictions about future stock trends. In comparison with singular prediction model like ARIMA, artificial intelligence is capable of capturing and modeling non-linear relationships between intricate external factors. Moreover, artificial intelligence is able to concurrently analyze a plenitude of data sources including textual and visual data that traditional models struggle to capture.

While artificial intelligence can improve stock trend prediction, it is crucial to remember that the stock market is naturally susceptible to a wide range of external events and variables. Distinct categories of stocks will result in different changes under the same factor. Therefore, focusing on training the applicability of artificial intelligence to different stocks and make self-adaptive adjustment according to the characteristics of stocks will be essential in making precise predictions.

However, artificial intelligence is highly possible to be prone to overfitting, which occurs when a model gets too tightly matched to the training data and struggles to generalize effectively to new and unseen data. Excessive optimism during training but subpar performance in real-world might result from overfitting. To reduce this phenomenon, scholars must balance the complexity and quantity of data of the model.

To sum up, ARIMA model has shown excellent ability in predicting the trends of Amazon's stock, but it can not utterly foresee the movements of individual stock due to the volatility and complexity of the stock market. Future research should address the limitations of linear models, broaden the types of collected data, and predict the result more scrupulously.

4. Conclusion

In this study, the author utilized ARIMA model to predict Amazon's stock to evaluate the accuracy of the model in a specific context. In practical application, the ARIMA model has shown great capability and accuracy. When compared to typical AR or MA models, this model is capable of dealing with more complex time series circumstances, as well as handling long-term patterns in data and make predictions based on those trends. Additionally, it is able to capture long-term trends in data. The volatility and complexity of the stock market, on the other hand, make it necessary to realize the limitations of the ARIMA model in terms of its ability to fully predict the movements of individual stocks. In addition to the macroeconomic environment, the performance of company management, interest rates, exchange rates, and unforeseeable occurrences like the COVID-19 epidemic, it is impacted by a wide range of factors.

Future research should aim to overcome the constraints of linear models such as ARIMA by expanding the range of acquired data and integrating advanced and intricate mathematical models. Artificial intelligence has demonstrated significant potential in forecasting stock patterns. AI models can utilize historical data to enhance the accuracy of predictions regarding future stock trends, identify and model complex non-linear relationships among various external factors, and simultaneously analyze a wide range of data sources, including textual and visual data that conventional models find challenging to handle. It is important to find a balance between the complexity and amount of data in the AI model to avoid overfitting, which may cause too optimistic results during training but poor performance in real-world scenarios.

As with any mathematics model, it is vital to note that there are inherent uncertainties and risks involved in stock market forecasting. While ARIMA provides a solid framework for analyzing stock trends, it is crucial for investors to consider other factors such as market conditions, economic indicators, and company-specific news when making investment decisions.

In summary, investors seeking to sort through the intricacies of the stock market may find the insights derived from our ARIMA analysis of Amazon's stock fluctuations to be an invaluable resource. The integration of qualitative research and quantitative analysis empowers investors to enhance the quality of their decision-making processes and potentially improve investment outcomes in the stock trading prediction.

Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

References

[1] Ang A, Bekaert G. Stock return predictability: Is it there. The Review of Financial Studies, 2007, 20(3): 651-707.

[2] Dooley G, Lenihan H. An assessment of time series methods in metal price forecasting. Resources Policy, 2005, 30(3): 208-217.

[3] De Faria E L, Albuquerque M P, Gonzalez J L, et al. Predicting the Brazilian stock market through neural networks and adaptive exponential smoothing methods. Expert Systems with Applications, 2009, 36(10): 12506-12509.

[4] Hussain W, Merigó J M, Raza M R. Predictive intelligence using ANFIS-induced OWAWA for complex stock market prediction. International Journal of Intelligent Systems, 2022, 37(8): 4586-4611.

[5] Büyükşahin Ü Ç, Ertekin Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. Neurocomputing, 2019, 361: 151-163.

[6] Xiao D, Su J. Research on stock price time series prediction based on deep learning and autoregressive integrated moving average. Scientific Programming, 2022.

[7] Siami-Namini S, Namin A S. Forecasting economics and financial time series: ARIMA vs. LSTM. Papers, 2018.

[8] Adamowski J, Fung Chan H, Prasher S O, et al. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. Water Resources Research, 2012, 48(1).

[9] Ramos P, Santos N, Rebelo R. Performance of state space and ARIMA models for consumer retail sales forecasting. Robotics and computer-integrated manufacturing, 2015, 34: 151-163.

[10] Hyndman R J, Athanasopoulos G. Forecasting: principles and practice. OTexts, 2018.