Dean&Francis
ISSN 2959-6130

The Test of ARIMA's Stock Price Prediction Capability in Stable and High-Variability Scenarios - A Comparative Study Based on Guizhou Moutai and NVIDIA

Guanting Wu

Shenzhen College of International Education, Shenzhen, 518043, China
s25802.wu@stu.scie.com.cn

Abstract:

The stock market is the "barometer" of the economy. Stock price forecasting is core focus of investors. Time series models like ARIMA are widely used for this purpose. However, the stock market's high volatility challenges the model's universality. This study selects two stocks with distinct price characteristics: Kweichow Moutai (relatively stable stock prices) and NVIDIA (significant price fluctuations driven by major positive news) as research objects. ARIMA is used to predict their price. RMSE is used to assess how accurate the predictions are. Major events that happened to the two companies are examined to explain prediction deviations. Results show that when predicting Kweichow Moutai's stock price, the ARIMA model achieves high accuracy. However, when it comes to the task of predicting NVIDIA's stock price, a striking and persistent gap emerges between the forecasted figures and the actual market outcomes. This pronounced discrepancy underscores a critical weakness in the models employed: they appear to falter precisely when confronted with sudden, high-impact shocks—such as unexpected regulatory announcements, geopolitical flashpoints, or abrupt shifts in consumer demand—that lie outside the historical patterns on which their parameters were trained.

Keywords: ARIMA model; stock price forecasting; prediction accuracy.

ISSN 2959-6130

1. Introduction

1.1 Research Background and Research Topic

The stock market is a "barometer" of the macroeconomy, so stock price forecasting has always been a hot topic in the field of finance. Stock price forecasting allows investors to prepare in advance and achieve higher returns on their investments.

Time series models are commonly used tools for stock price forecasting. ARIMA is one such model. However, the stock market is highly volatile. When sudden news occurs, stock prices can experience significant fluctuations. Based on this, this study uses Kweichow Moutai (with relatively stable stock prices) and NVIDIA (whose stock prices fluctuate significantly due to major positive news) as research subjects to explore the differences in the applicability of ARIMA models in stable environments and under sudden shock scenarios.

1.2 Research Objectives and Significance

The research objectives is to validate the accuracy of the ARIMA model in predicting stock prices of relatively stable stocks (such as Kweichow Moutai) and demonstrate its value; to analyze the prediction biases of the ARIMA model in stocks affected by major positive or negative news (such as NVIDIA) and reveal its limitations; and to summarize the applicability boundaries of the ARIMA model.

The theoretical significance of this study is that this research compares the performance of ARIMA under different market conditions, enriches the application research of time series models in the field of financial forecasting, and simultaneously provides direction for the development of hybrid models in the future.

The practical significance is that this paper provides references for market participants such as investors and financial institutions — in stable markets, ARIMA can be used for short-term forecasting, while in volatile markets, external factors should be incorporated into the model to enhance forecasting accuracy;

1.3 Research Methods and Structure

The research method is the ARIMA model, and this paper uses daily stock price data for Kweichow Moutai (March–June 2023) and NVIDIA (May–June 2024); The root mean square error (RMSE) is used to compare the predictive accuracy of the model across the two datasets; Major events occurring during the same period for each stock are also analyzed to explain the causes of ARIMA prediction deviations.

2. Literature Review

2.1 The Introduction of the ARIMA Model

The Autoregressive Integrated Moving Average (ARIMA) model is a method to forecast and it is based on time analysis [1]. In 1976 George Box and Gwilym Jenkins invented this method, so it is also known as the Box-Jenkins model. The model includes autoregression (AR) [2], differencing (I), and moving average (MA). It identifies trends and cyclical patterns through historical data and is widely applied in areas such as business sales forecasting, inventory management, and financial indicator analysis. Its parameters include the lag order (p), differencing order (d), and moving average window (q), which must be determined by the autocorrelation function (ACF) and partial autocorrelation function (PACF).

The model originated in the 1960s, using differencing to eliminate non-stationarity in the data, establishing a stationary sequence, and then combining AR and MA for forecasting. The implementation process includes data stationarity processing, parameter estimation, model validation, and forecasting applications. The Box-Jenkins method proposes a four-step framework of identification, estimation, testing, and forecasting. Typical cases include retail supply chain optimization and medical supplies procurement forecasting. Research shows that its accuracy rate in pharmaceutical demand forecasting can exceed 89%.

2.2 The Application of Time Forecasting in Stock Markets

Stock investment is one of the main investment methods for people. In recent years, more and more people have begun to pay attention to the stock market and join the ranks of stock investors. Thus, being able to forecast the stock price is very tempting for investors. Many researchers have used ARIMA to analyze stock data. They have different emphasis and draw different conclusions. Shimin Huang used ARIMA to analyze the stock price of the CMB, and she concluded that the ARIMA model has a good predictive effect on short-term changes in stock price time, providing investors with reference for stock investment [3]. On the other hand, Zixia weng forecasted the stock price of CBB, and he stated that due to the influence of multiple factors on the stock market, long-term predictions may be subject to significant errors, and it is necessary to explore more accurate stock price prediction models [4]. Zhao Ning pointed out the specific factors that influenced the stock market: Macroeconomic conditions, local government policies, industry development levels, and company operations [5]. Lifang Cao analyzed Muyuan Food and his perspective was quite unique, he focused on the negative impact of forecasting: Short-term stock price forecasts can easily lead to short-sighted behavior among investors, which is detrimental to the long-term stable development of the stock market [6]. Moreover, Haoxiang Yang analyzed a particular kind of stock, ETF, and found out that the prediction results show an error of less than 5% compared to the actual values, indicating that traditional stock price prediction methods are applicable to ETF funds [7]. By contrast Siqi Su used ARIMA to predict the CSI 300 stock price. Her analysis showed that the ARIMA model was not very effective. As the date shortened, the gap between the predicted value and the actual value continued to widen [8]. Chenhao Sun holds a rather moderate opinion. He thought that the model exhibits a certain degree of lag in their prediction results. Although the model has significant deviations in their daily index predictions, the prediction results generally reflect the trend of the index [9].

2.3 Gaps of the Researches

However, there are gaps that previous research haven't fully explored. Researchers such as Zixia Weng did state that the accuracy of ARIMA is strongly influenced by factors, but they did not compare data and strictly prove this. Thus, this article is going to employ the ARIMA model to analyze the stock prices of Kweichow Moutai and NVIDIA and find out whether ARIMA still works when there are strong influence factors.

3. Methodology

3.1 ARIMA and Its Related Models

3.1.1 Autoregressive (AR) Model

The AR model is a linear time series analysis model that establishes a regression equation based on the autocorrelation between the current data and historical data of the variable itself. Its formula is expressed as:

$$Yt = c + \varphi 1Yt\text{-}1 + \varphi 2Yt\text{-}2 + \dots + \varphi pYt\text{-}p + \varepsilon t \quad (1)$$

In this equation, Yt represents the value of the time series at time t; c is a constant term; φ1, φ2, ..., φp are the autoregressive coefficients; Yt-1, Yt-2, ..., Yt-p are the lagged values of at previous time points; and εt denotes the random error term at time t.

3.1.2 Moving Average (MA) Model

This model obtains a moving average equation by weighted summation of white noise sequences in a time series. The formula is as follows:

$$Yt = c + \varepsilon t + \theta 1\varepsilon t\text{-}1 + \theta 2\varepsilon t\text{-}2 + \dots + \theta q\varepsilon t\text{-}q \quad (2)$$

Here, Yt is the value at time t; c is a constant term; εt, εt-1, ..., εt-q are the random error terms at respective time points; and θ1, θ2, ..., θq are the moving average coefficients.

It uses a linear combination of past disturbance terms and the current disturbance term for prediction analysis.

3.1.3 Autoregressive Moving Average (ARMA) Model

The ARMA model, fully known as the Autoregressive Moving Average model, is an organic combination of the AR model and the MA model [2]. Its formula is:

$$Yt = c + \varphi 1Yt\text{-}1 + \varphi 2Yt\text{-}2 + \dots + \varphi pYt\text{-}p - \varepsilon t - \theta 1\varepsilon t\text{-}1 - \theta 2\varepsilon t\text{-}2 - \dots - \theta q\varepsilon t\text{-}q \quad (3)$$

This model is mainly used to describe stationary stochastic processes, integrating the characteristics of the AR(p) model and the MA(q) model.

3.1.4 ARIMA Model

The formula of the ARIMA model is the same as that of the ARMA model. The difference is that the Y in the ARIMA formula is replaced by a differencing operator. The purpose of the operator is to make the data stationary. By applying d-order differencing to the non-stationary time series, and then using the ARMA (p, q) model for analysis and prediction, the ARIMA model extends the applicability of the ARMA model to non-stationary time series.

3.2 Application

3.2.1 Raw Data Processing

Stock data for Kweichow Moutai from January 13, 2023, to March 29, 2023, and for NVIDIA from March 1, 2024, to May 9, 2024, was used.

Real-world stock data is generally non-stationary. However, ARIMA can only analyze stationary data. Therefore, the data must be differenced.

ISSN 2959-6130

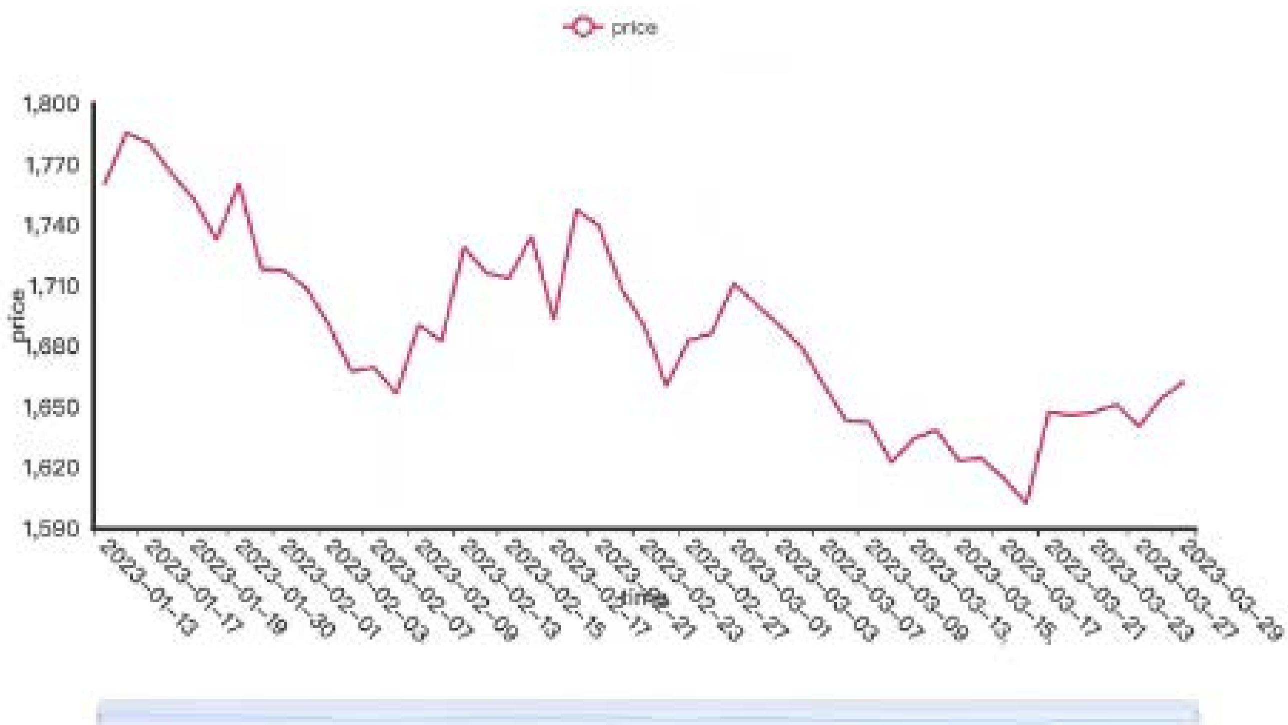

Fig. 1 Kweichow Moutai's original stock price

According to figure 1, the price is not stationary, so it should be difference. Then, we need to determine how much time should differencing be performed. When the difference is 0th order, the P-value is 0.394. The P-value is not significant at this level. Thus, the original hypothesis cannot be rejected, which means that the sequence is a non-stationary time series. When the difference is the first order, the P-value is 0.001. The P-value is now significant. This means that the original hypothesis is rejected, and the sequence is stationary.

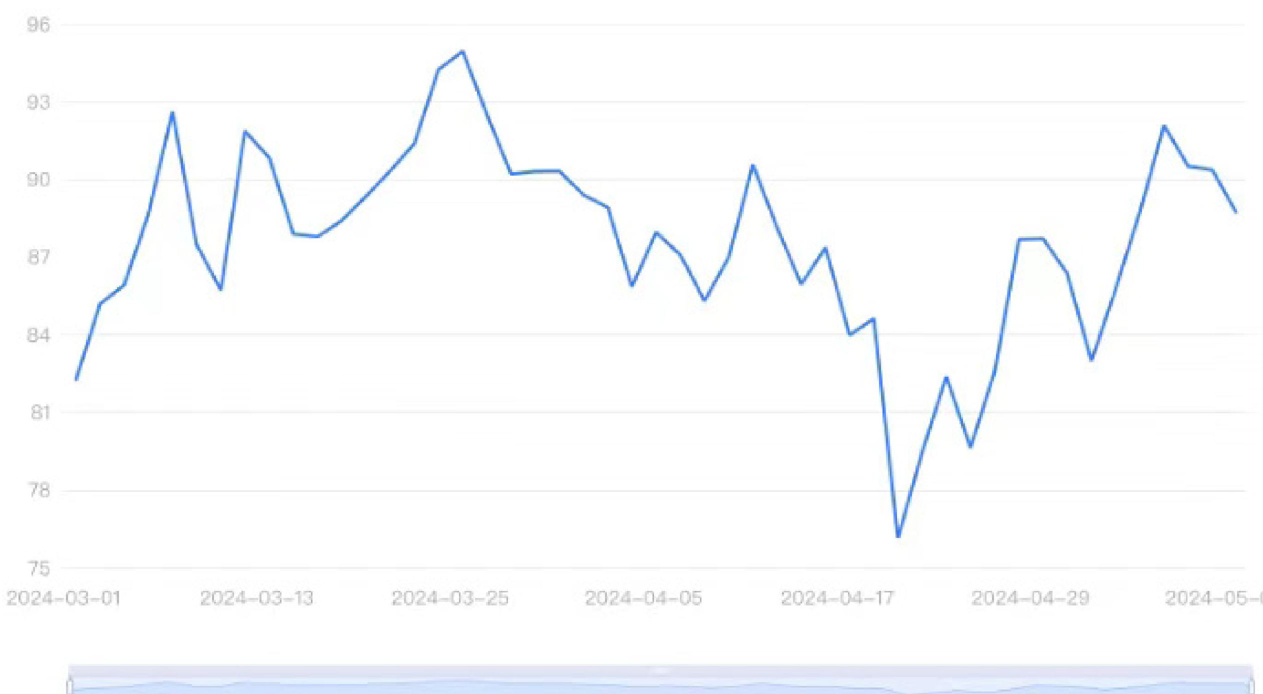

Fig. 2 NVIDIA's original stock price

Table 1: ADF Inspection Sheet

Variables	Order of difference	t	P	AIC	Threshold		
					1%	5%	10%
Price	0	-1.773	0.394	329.336	-3.589	-2.93	-2.603
	1	-4.264	0.001***	328.4	-3.589	-2.93	-2.603
	2	-5.39	0.000***	328.405	-3.597	-2.933	-2.605
Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.							

If the result is statistically significant ($P < 0.5$), it means that the null hypothesis is rejected. Conversely, it indicates that the sequence is a non-stationary time series (Table 1). If the ADF test result is less than 1%, 5%, and 10% simultaneously, it indicates that the hypothesis is well rejected. The AIC value is a standard for measuring the goodness of fit of a statistical model [10]; the smaller the value, the better. The threshold P is a fixed value corresponding to a given significance level. By analyzing the t-value, we can determine whether the null hypothesis of sequence non-stationarity can be significantly rejected.

In the case of NVIDIA, according to figure 2, the stock price is already stationary, so there is no need for differencing (Table 2).

Table 2: ADF Inspection Sheet

Variables	Order of difference	t	P	AIC	Threshold		
					1%	5%	10%
Price	0	-2.908	0.044**	177.068	-3.575	-2.924	-2.6
	1	-7.347	0.000***	181.145	-3.578	-2.925	-2.601
	2	-4.423	0.000***	190.108	-3.606	-2.937	-2.607

Note: ***, **, and * represent significance levels of 1%, 5%, and 10%, respectively.

3.2.2 Determine the order

Kweichow Moutai (according to figure 3,4):

Autocorrelation Function (ACF) Analysis: The autocorrelation function after first-order differencing exhibits a tailing feature (the autocorrelation coefficient decays slowly within the confidence interval without a sudden truncation point approaching 0). According to the rules, ACF tailing suggests that q may be 0.

Partial autocorrelation function (PACF) analysis: The PACF after first-order differencing truncates at lag 1 (the partial autocorrelation coefficients fall within the confidence interval after lag 1 and approach 0). According to the rules, PACF truncation at lag 1 suggests that $p = 1$.

The system automatically optimizes the model based on the AIC information criterion, ultimately determining the optimal model as ARIMA (1,1,0).

ISSN 2959-6130

Fig. 3 ACF of Kweichow Moutai

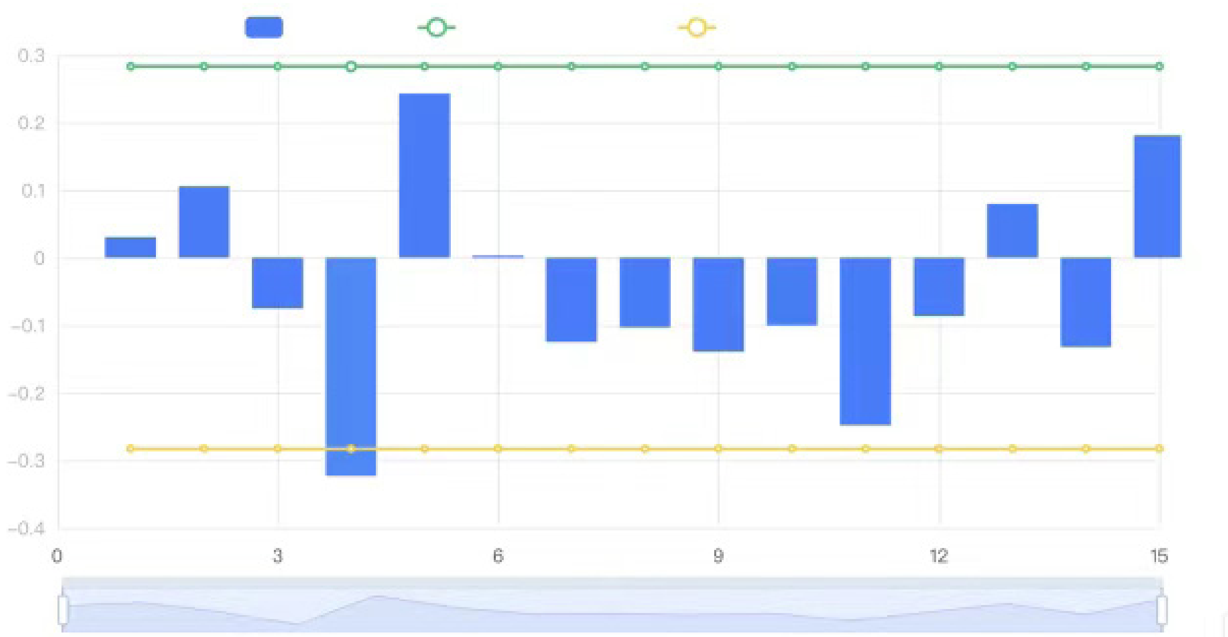

Fig. 4 PACF of Kweichow Moutai

Blue: autocorrelation coefficient. Green: upper limit of confidence interval. Yellow: lower limit of confidence interval.

NVIDIA (according to figure 5,6): Autocorrelation Function (ACF) Analysis: The autocorrelation function after 0th-order differencing exhibits a tailing feature. According to the rules, ACF tailing indicates that q may be 0.

Partial Autocorrelation Function (PACF) Analysis: The partial autocorrelation function after 0th-order differencing truncates at lag 1. According to the rules, PACF truncation at lag 1 indicates that p=1.

The system automatically optimizes the model using the AIC information criterion, ultimately determining the optimal model as ARIMA (1,0,0), i.e., p=1 and q=0.

Fig. 5 ACF of NVIDIA

Fig. 6 PACF of NVIDIA

3.2.3 Prediction and Results

The ARIMA is used to predict the price for the next 25 days. According to figure 7, the prediction for Kweichow Moutai is rather accurate, with a RMSE of 27.9. The mean of the actual price data is approximately 1624.63, with a standard deviation of approximately 33.84. In terms of data scale, the RMSE accounts for approximately 1.72% of the mean of the price data (27.90 ÷ 1624.63 × 100%). Therefore, the error is relatively small compared to the

ISSN 2959-6130

scale of the price data.

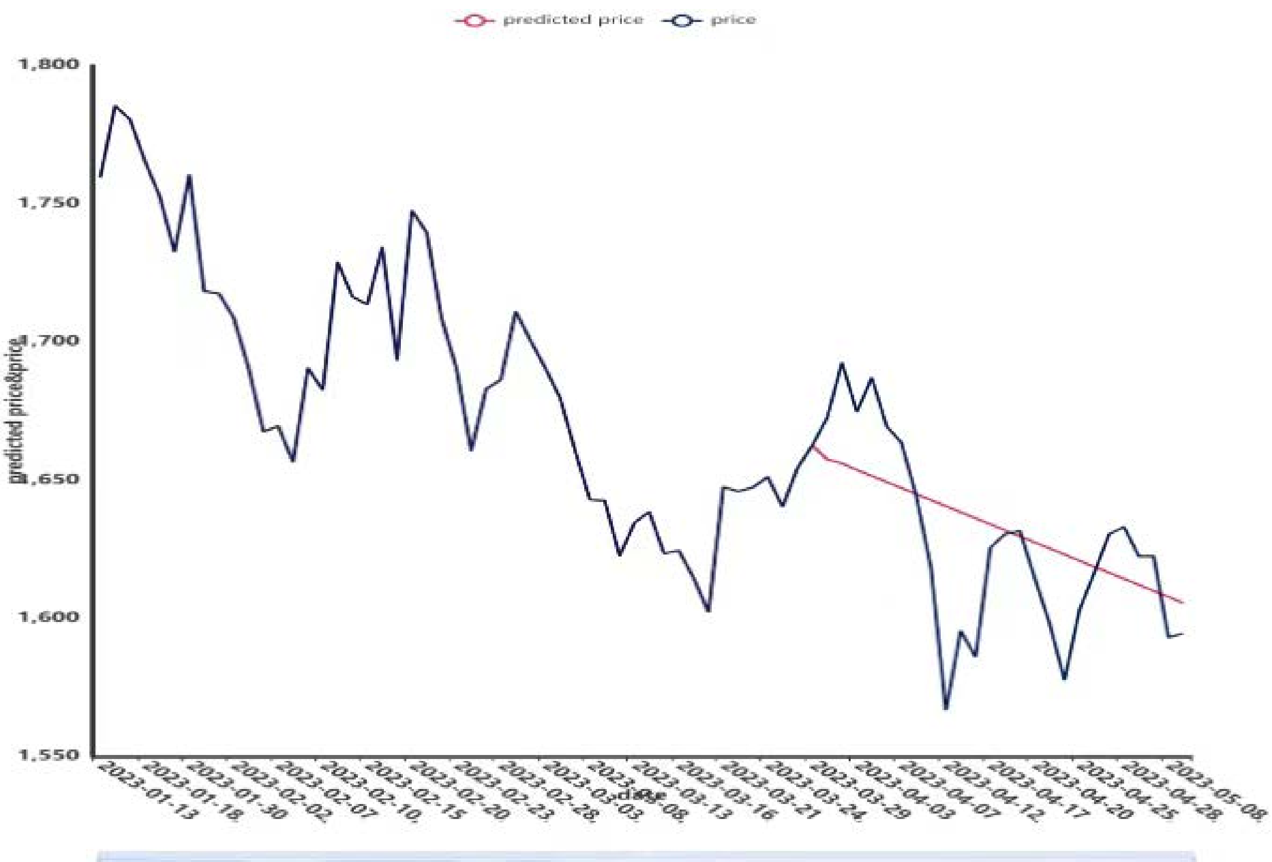

Fig. 7 Predicted price of Kweichow Moutai

As for NVIDIA, according to figure 8, the result is extremely inaccurate, with a RMSE of 25.79. However, the mean of the actual stock prices is approximately 109.64, with a standard deviation of approximately 13.34. The RMSE accounts for approximately 23.52% of the mean value of the price data (25.79 ÷ 109.64 × 100%). Thus, these errors are relatively high compared to the scale of the price data.

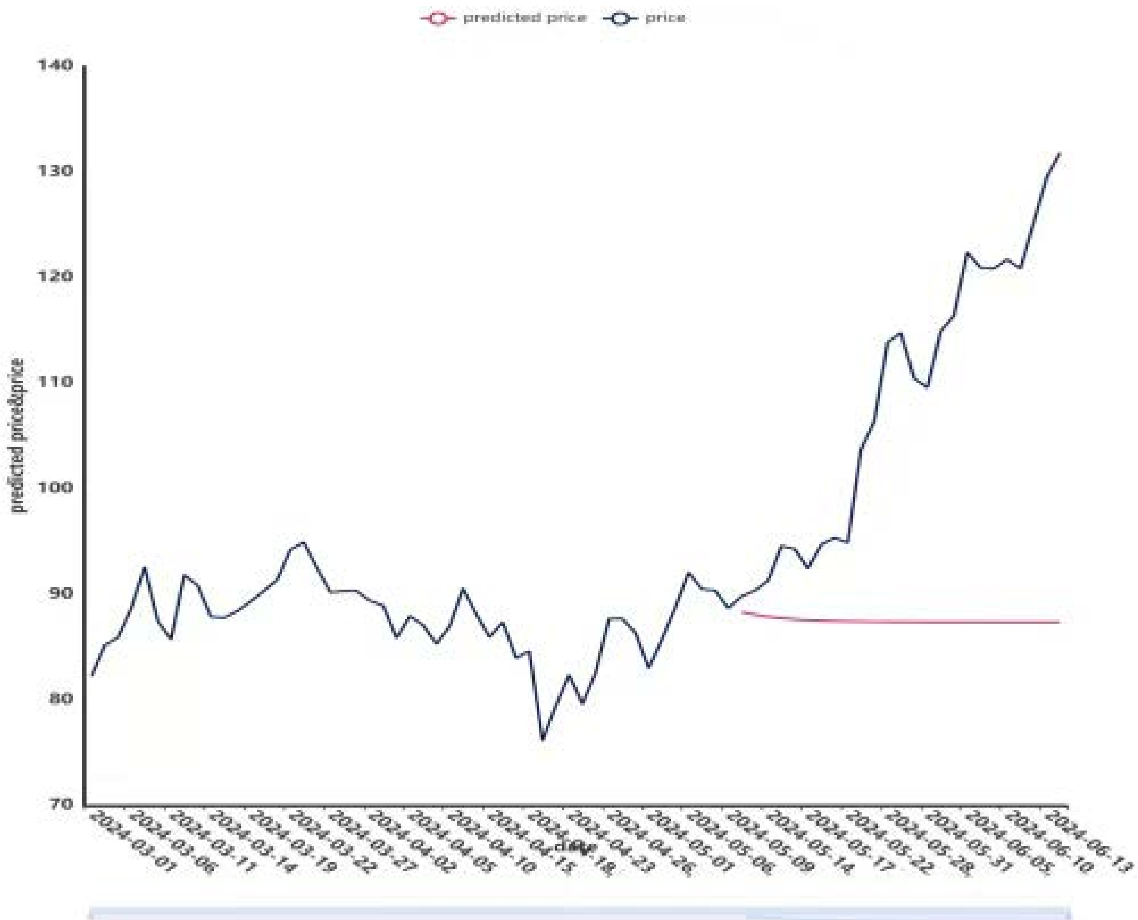

Fig. 8 Predicted price of NVIDIA

The reason why there are such difference between accuracy is that the stock can be influenced by multiple factors that are not considered in the ARIMA model. The external market for baijiu is relatively stable, and Guizhou Maotai's position and sales volume in the baijiu industry have not changed significantly. Thus, the stock price of Moutai follows the established pattern. On the other hand there are many factors pushing the NVIDIA upward. The financial results for the first fiscal quarter of fiscal year 2025, announced on May 23, were impressive, with a revenue of US$26 billion, up 262% year-on-year. This result far exceeded investors' expectations. The net profit was US$14.81 billion, up 628% year-on-year. Moreover, the earnings per share was US$5.98, higher than the expected US$5.59. The results directly demonstrated the company's profitability. This made investors confident about its future prospects and drove the stock price higher. The artificial intelligence industry was in a phase of rapid development, which means that NVIDIA's GPU will be in high demand. As the world's largest provider of AI computing power and a leading AI chip manufacturer, the market continued to raise its expectations for NVIDIA's future performance growth. NVIDIA's upcoming Blackwell graphics processor was gaining significant attention. Although the product has not yet entered mass production, the market was anticipating it, believing it would further raise NVIDIA's position in the AI chip market and generate additional revenue for the company. This positive market outlook was driving the stock price higher.

4. Conclusion

All in all, the result shows that when there are no strong bullish or bearish factors, the ARIMA model can give a rather accurate prediction of the stock price.

This limitation is because of ARIMA's fundamental reliance on historical data, as it fails to incorporate sudden external shocks—such as policy changes, technological breakthroughs, or earnings announcements—that disrupt market. These events cannot be predicted by historical data, which means that the model's extrapolation of past trends ineffective.

The results concluded that ARIMA is appropriate for short-term forecasting in mature and stable markets. This gives indications for researchers and practitioners. The ARIMA should be used in conjunction with other methods, such as sentiment analysis or machine learning models. Thus it will be able to consider external shocks.

ISSN 2959-6130

There are limitations of the research. The data has a short time span and a small sample size. This research only used short-term stock price data. This makes it difficult to capture long-term cyclical fluctuations and may also lead to fault decision on incorrect parameter determinations in the ARIMA model. These all affects the generalizability of the conclusions.
Future studies could focus on hybrid models that combine ARIMA with event-driven factors, enabling the incorporation of information (e.g., news sentiment, policy announcements) into time series forecasting.

References

[1] Box, George E. P., and Gwilym M. Jenkins. Time Series Analysis: Forecasting and Control. Holden-Day, 1970.
[2] Yule, George Udny. “On a Method of Investigating Periodicities in Disturbed Series, with Special Reference to Wolfer’s Sunspot Numbers.” Philosophical Transactions of the Royal Society of London. Series A, vol. 226, 1927, pp. 267–298.
[3] Huang, Shimin. “Stock Price Analysis and Prediction Based on ARIMA Model—Taking China Merchants Bank as an Example.” Small and Medium-Sized Enterprise Management and Technology, no. 11, 2022, pp. 184–187.
[4] Weng, Zixia. “Stock Price Analysis and Prediction Based on ARIMA Model—Taking Construction Bank as an Example.” Modern Information Technology, 2023, pp. 137–141. doi:10.19850/j.cnki.2096-4706.2023.14.029.
[5] Zhao, Ning. “Short-Term Prediction of Wanzhou Stock Price Based on ARIMA Model.” Finance Guest, Jan. 2023, pp. 34–36.
[6] Cao, Lifang, and Zhang Ruolin. “Stock Price Prediction Based on ARIMA Model—Taking Man Yuan Co., Ltd. as an Example.” Wealth Times, no. 2, 2024, pp. 54–56.
[7] Yang, Haoxiang. “Verification of the Applicability of Traditional Short-Term Prediction Methods for ETF Onshore Funds Based on ARIMA Model.” 2022. Huazhong University of Science and Technology, MA thesis.
[8] Su, Siqi. “Comparative Study of Stock Price Prediction Models Based on LSTM and ARIMA.” 2023. Chongqing University, MA thesis.
[9] Sun, Chenhao, and Wang Lin. “Prediction and Analysis of Shanghai Composite Index Based on ARIMA and LSTM.” Information and Computer (Theoretical Edition), vol. 35, no. 2, 2023, pp. 29–31.
[10] Akaike, Hirotugu. “A New Look at the Statistical Model Identification.” IEEE Transactions on Automatic Control, vol. 19, no. 6, 1974, pp. 716–723.