Stock Predictions Based on Time Series Models - The Case of Pepsi Research Background

Leyi Liu[1], *Wenhao* Wang[2], *Yijing* Zhao[3*]

[1]Nottingham University Business School China, University of Nottinghan Ningbo China, Zhejiang, 315100, China
[2]Henan University, Shanghai University, Shanghai, 201800, China
[3]Eurasia International School of Henan University, Henan University, Henan, 475004, China

Abstract. A statistical model widely used in time series analysis and prediction is Auto Regressive Integrated Moving Average (ARIMA) model. Through the closing price data of Pepsi's stock during the period from January 1, 2024, to January 1, 2025, this model can be applied to conduct a time order analysis to project Pepsi's closing price in the coming trading time. It provides investors some important information about the future trend of Pepsi's stock price. This research has found that when the data is stable and there are no significant structural changes, ARIMA model performs better for predicting the stock market trend. And investors can also use ARIMA model if they want to avoid the risk of stock fluctuations. The limitations of this research mainly due to the limitations of this model itself (only suitable for short-term prediction), which can be solved through the combination of multiple models in the future study.

1 Introduction

With computers developing swiftly, big data analysis witnesses a tipping point that considerable statistic models can be set to make predictions. In finance, the stock market act as a vital part and means a lot to investors and traders [1]. They are attempting to make precise predictions of stock value utilizing machine learning. Huge profits can be achieved if distinction is diminished [2]. However, the stock market has unavoidable features such as volatility, small signal-noise ratio which decides the complexity of stock price predictions. This paper accesses the time series model (ARIMA) to predict stocks.

In global scope, Engle built the autoregressive conditional heteroskedasticity (ARCH) and then Bollerslev first introduced the generalized autoregressive conditional heteroskedasticity (GARCH) model [3]. It supported a feasibility study of stocks' volatility which was immensely related to risk management, stock price prediction and other finance fields associated with stock market. Wei and Zhou first used GARCH model; meanwhile, they cultivated two nonlinear revised models (QGARCH model and GJR model) to overcome the volatility of stock market in Chinese realm [4]. More research started to conduct in the domestic.

* Corresponding author: 2228200293@henu.edu.cn

Despite the time series models, Artificial Neural Network (ANN) model intimates the neural network from humankind which is capable to learn patterns [5], where the stored structure is used to model the problem [6]. ANN is researched by many scholars to be utilized in stock price prediction studies and prevails increasingly [7]. However, in comparison, autoregressive integrated moving average (ARIMA) models are more extensively used in economics and finance fields, as well as stock forecasting [5].

Frequent and violent fluctuations of stock prices are the inevitable features in the stock market. The time series of stock prices often show autocorrelation to a certain extent. Through using models to analyze and predict Pepsi's stock price, the prediction's accuracy for future trends can be verified [8].

The core content of this paper is to use mathematics model to predict the future stock price. Therefore, the research method is chosen as the time series model (ARIMA). The dataset extracts Pepsi's daily stock price data for one year, ranging from 1.1 2024 to 1.1 2025 where the closing values were only counted, considering that it symbolizes the real closing value of the day [9].

2 Method

2.1 Data collection and preparation

The data is acquired from Investing.com where the stock prices of Pepsi corporations from 2024.1 to 2025.1 are selected. With time series data covering a full cycle of the stock market, the pattern of stock can be better predicted, especially for a short-term prediction. Referring to Table 1, The open price, low price, high price, and closing price are the four components that make up the stock data. This forecast uses the closing price to avoid fluctuations, which is negative to ARIMA predictions. It offers a consolidated view of the stock market of a day.

Table 1 Description of Data

	Min.	Median	Mean	Max.
close	151.3	169.6	169.1	183.1
open	150.5	169.5	169.1	182.6
high	152.4	170.7	170.3	183.4
low	149.7	167.8	167.8	181.8

2.2 Method Introduction

2.2.1 Introduction to ARIMA Model

The ARIMA model stands as a linchpin within the ambit of time series forecasting, particularly entrenched in the economic and financial spheres. This model amalgamates three cardinal components: the AR component, which delineates the current value's dependence on its antecedent values; the I component, instrumental in attaining stationarity; and the MA component, which encapsulates the current value's dependence on past forecast errors. The confluence of these components endows the ARIMA model with the faculty to discern intricate patterns and trends within time series data, thus positioning it as a robust instrument for forecasting future values [10]. The model is expressed as follows:

$$Y_t = \phi_0 + \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{1}$$

Where, Y_t is the actual value and ε_t is the random error at t, φ_i and θ_j are the coefficients, p and q are integers that are often referred to as autoregressive and moving average, respectively [11].

2.2.2 Stationarity Test

The ARIMA model is a foundational element in the field of time series forecasting, particularly within the disciplines of economics and finance. It integrates three key components: the autoregressive (AR) component, which reflects the dependence of the current value on its past values; the differencing (I) component, which is crucial for achieving stationarity in the time series; and the moving average (MA) component, which captures the dependence of the current value on past forecast errors. To this end, the Augmented Dickey-Fuller (ADF) test is employed as a robust statistical procedure to scrutinize the stationarity of the time series [11].

2.2.3 Model Building and Validation

After establishing the stationarity of the data through the Augmented Dickey-Fuller (ADF) test, the next step was to construct the ARIMA model. The determination of the model's optimal parameters (p, d, q) was conducted by employing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. These visual representations facilitated the discernment of salient lags, which were then instrumental in the preliminary estimation of p and q. A grid search approach was then implemented to explore various combinations of (p, d, q), aiming to minimize the Bayesian Information Criterion (BIC). The BIC balances model complexity with goodness of fit, aiding in the selection of the most parsimonious and effective model. Once the optimal parameters were identified, the ARIMA model was fitted to the data. The model was validated through a series of rigorous statistical tests and diagnostics, including residual analysis and out-of-sample testing. Residual analysis involved a thorough examination of the residuals to ensure their normal distribution and absence of autocorrelation. Out-of-sample testing was conducted to further substantiate the model's predictive power.

3 Result and discussion

3.1 Time series analysis

Fig.1 shows the diagram about the closing price data of Pepsi stock on each trading date from January 1, 2024, to January 1, 2025. It can be found that the seasonal distribution of Pepsi closing price is not obvious. The closing price of the stock fluctuates between 150 and 185 USD, and the overall trend is download.

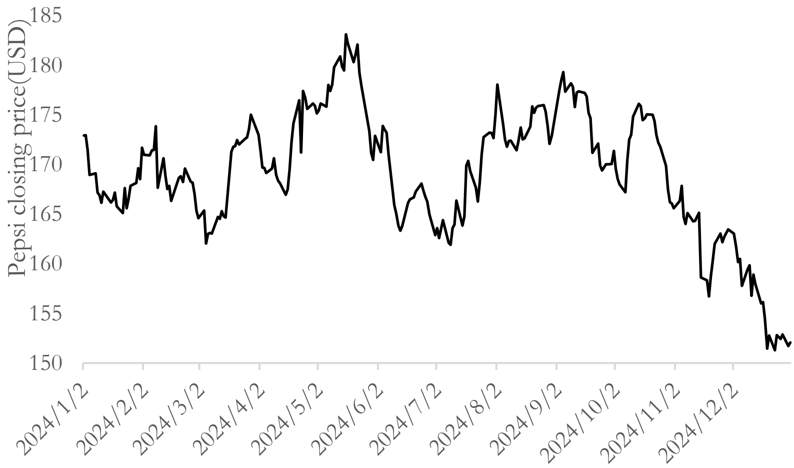

Fig.1 Closing price data time series diagram

Photo credit: Original

3.2 Data preprocessing

To use time series analysis, this section needs to ensure that there are no missing values in the series and to test the stationarity of the data, because time series analysis usually requires the data to be stationary. An Augmented Dickey-Fuller (ADF) test was used to test the data stationarity of 252 Pepsi closing prices. ADF test is a statistical test method used to test whether time series data has a unit root, so as to determine whether the data presents non-stationarity. If the data does not have a unit root, then the data is stationary. In the ADF test, the null hypothesis is that the data has a unit root, so to prove that the data is stationary, it is necessary to reject this null hypothesis [12].

Table 2 Unit root test

Variate	Difference order	t	P	AIC
	0	-0.784	0.824	353.018
Closing price	1	-13.056	0.000***	348.229
	2	-8.6	0.000***	359.016

As shown in Table 2, for the original data, the ADF statistic is -0.784. The original P-value is 0.824. The commonly used significance level is 0.05, so the null hypothesis cannot be rejected. Therefore, the data may be non-stationary.

Therefore, this part identifies the original data as non-stationary data and perform first - and second-order difference processing on it. The P-value of the processed data is all no more than 0.05, and the absolute value of critical value is less than the absolute value of the ADF statistic under all significance levels. Therefore, the null hypothesis should be rejected and the data can be considered stable after processing. In addition, because the AIC value of the data after first-order difference is the smallest, it is determined that d=1 and the first-order difference of the data is stable.

3.3 Model order establishment

Plotting ACF and PACF is a key step in identifying the order of ARIMA models. The ACF plot represents the time series between its own delayed version. The PACF diagram shows the relationship under the condition of removing the intermediate lag. The relationship between two time points. By looking at these shapes, it can get clues about the structure of the time series.

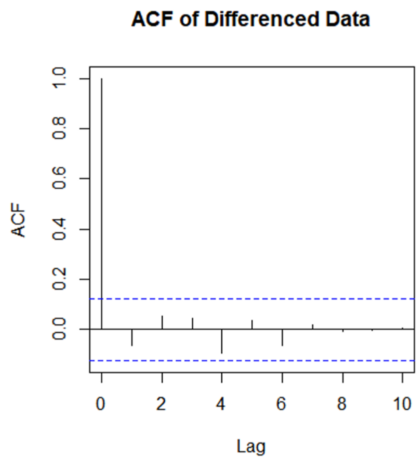

Fig. 2 Autocorrelation function diagram

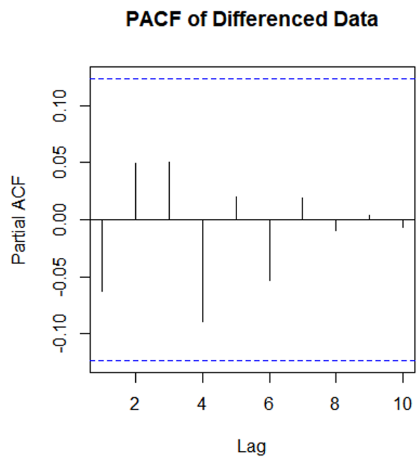

Fig. 3 Partial autocorrelation function diagram

Photo credit: Original

As shown in Fig. 2 and Fig.3, they are all truncated so that it is not very easy to decide the lag order (p, q). Under the circumstances, this paper uses AIC and BIC to determine the lag order of ARIMA.

Table 3 Order identification

ARIMA model（1,1,0）test table		
Information standard	AIC	396.842
	BIC	404.628
Goodness of fit R2	R^2	0.878

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

After reading Table 3, this system automatically searched for optimal parameters based on AIC and BIC information criteria. The model result was a white noise model, and the goodness of fit R2 of the model was 0.878, indicating excellent performance and basically meeting the requirements.

3.4 Model building and Residual testing

The ARI model is created, and the order of the ARIMA model is specified as (1, 1, 0), where (1, 1, 0) indicates that p is 1, d is 1, and q is 0. This part then perform a white noise test on the fitted ARIMA (1, 1, 0) residuals.

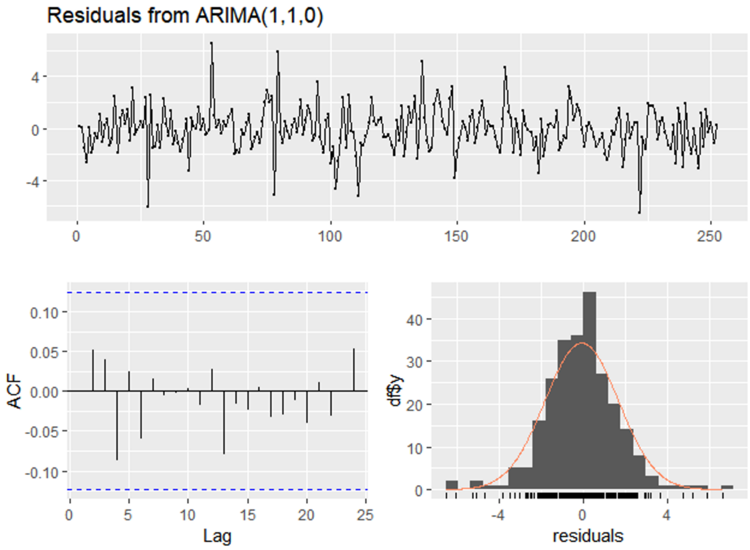

Fig. 4 White noise test diagram

Photo credit: Original

As shown in Fig.4, the residual error of the model passed the white noise test, indicating that the true value and the fitting value were very close and the fitting degree was high.

3.5 Prediction based on ARIMA model

The well-fitted model is utilized for predicting future closing price of Pepsi. This result is shown as the Fig.5, where the shaded part is the range of stock price fluctuations, and the blue line indicates the forecast price.

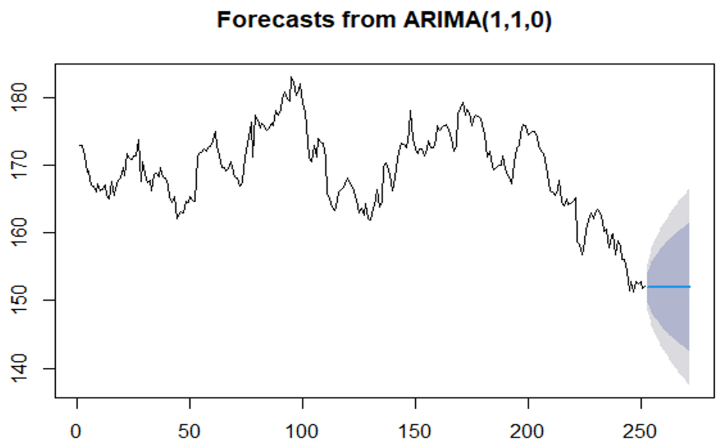

Fig.5 Model prediction

Photo credit: Original

To further substantiate the model's predictive prowess, out-of-sample testing is performed. The predicted closing price data of the next 9 trading days are output and compared with the actual data, and the results can be seen in Table 4. Its maximum relative error is not more than 0.0095. It can be seen that the ARIMA (1, 1, 0) model constructed by the research has high accuracy in predicting the closing price of Pepsi, and can better reflect the future trend and change trend of Pepsi stock price.

Table 4 Predicted data, actual data and relative error

Time	Forecast data	Actual data	Relative error
2024/12/20	152.34	152.79	0.2818*10-2
2024/12/23	152.12	151.29	0.549*10-2
2024/12/24	151.89	152.81	0.602*10-2
2024/12/26	151.67	152.44	0.505*10-2
2024/12/27	151.45	152.89	0.942*10-2
2024/12/30	151.22	151.72	0.329*10-2
2024/12/31	151.00	152.06	0.697*10-2
2025/1/2	150.77	150.21	0.532*10-2
2025/1/3	150.55	149.65	0.601*10-2

4 Conclusion

Through the analysis and prediction of Pepsi ARIMA model, a series of useful findings have been made. First, after time series analysis and BIC evaluation, the best model parameters are selected on the given data set, and the ARIMA (1, 1, 0) model is constructed. The model shows good performance in fitting training data and can accurately reflect the historical trend of Pepsi stock closing price.

The time series analysis shows that the stock as a whole showed a slight downward trend during the period from January 1, 2024, to January 1, 25. For investors, this provides valuable information to help develop a more informed investment strategy. At the same time, the analysis of ACF and PACF also reveals some possible autocorrelations and partial autocorrelations, which provide clues for in-depth study of market behavior and trend changes.

However, some limitations can be found in this study. When considering time order data, ARIMA model requires the data should be stationary, which may not be sufficient for some financial time series data. Future research could further explore more complex models to better capture the volatility of the stock market. In addition, external factors such as the macroeconomic environment, competitive enterprises, etc. have not been taken into account. All of them may cause a significant impact on the volatility of stock prices.

In further study, more factors, such as other stock market indicators and macroeconomic data, can be considered to build a more comprehensive model. At the same time, similar analyses can be performed on other stocks to compare the predictive performance between different stocks to obtain a more comprehensive understanding of the dynamics of the stock market.

In conclusion, this paper provides a foundation for using ARIMA model to analyze and forecast stock prices, and provides some suggestions for future related research.

References

1. I. R. Parray, S. S. Khurana, M. Kumar, A. A. Altalbe. Time series data analysis of stock price movement using machine learning techniques. Soft Computing, **24**(21): 16509–16517 (2020)

2. A. Harel, G. Harpaz. Why stock analysts may make wrong predictions? Economics Letters, **244**: 111956 (2024)
3. J. Ye, T. Huang. Long-term risk management and general economic fundamentals of the stock market. Economics, **40**(2): 199-235 (2012)
4. W. Wei, X. Zhou. A Nonlinear GARCH Prediction Model of Stock Market Volatility in China. Forecast, (5): 47-50 (1999)
5. A. A. Ariyo, A. O. Adewumi, C. K. Ayo. Stock price prediction using the ARIMA model. In 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation: 106-112 (2014)
6. A. S. Zamani, N. S. AL-Arifi, S. Khan. Response prediction of earthquake motion using artificial neural networks. International Journal of Applied Research in Computer Science and Information Technology, **1**(2) (2012)
7. D. Selvamuthu, V. Kumar, A. Mishra. Indian stock market prediction using artificial neural networks on tick data. Financial Innovation, **5**(1): 16 (2019)
8. J. Zhou. Stock Forecasting Based on ARIMA Model - A Case of Bank of China. Modern Computer, **30**(14): 89-92 (2024)
9. S. Shakir Khan, H. Alghulaiakh. ARIMA Model for Accurate Time Series Stocks Forecasting. International Journal of Advanced Computer Science and Applications (IJACSA), **11**(7) (2020)
10. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, G. M. Ljung. Time Series Analysis: Forecasting and Control (5th ed.). Wiley (2015)
11. M. Areef, S. Rajeswari, N. Vani, G. M. Naidu. Lasalgaon onion market price forecasting-An application of arima model. Indian Journal of Agricultural Marketing, **33**(3s): 122-130 (2019)
12. J. Xie. Sales forecast of new energy vehicles and SUVs based on ARIMA promotes the formation of new quality productivity. China Economic and Trade Guide, (10): 136-138 (2024)