

Stock Price Prediction based on the AMIAR Model

Yunjie Wei*

School of Mechanical and Power Engineering, East China University of Science and Technology,
Shanghai (200237), China

*Corresponding author email: weiyunjie@mail.ecust.edu.cn

Abstract. Stock price prediction is widely used in the financial industry and holds significant relevance for nations, businesses, and investors. According to research conducted both domestically and internationally, several methods are currently available for forecasting stock prices. This paper focuses on the A-share stock Zhongyuan Env. Protection for research purposes. This paper employs the ARIMA model to predict changes in stock prices. The original data used in this paper is sourced from CSMAR and consists of stock closing prices, with a time range spanning from August 31, 2018, to August 30, 2023. The results indicate that the closing price of the stock follows the ARMA (1,1) model. Furthermore, a comparison between the predicted and actual data for the last three weeks of trading days reveals a consistent trend, affirming the effectiveness of the model. The research findings can provide valuable insights for financial investments. However, this paper also observes that accurate predictions may not be achievable for data with longer time intervals. Therefore, the model needs to consider the timeliness of the data and the influence of other factors.

Keywords: Stock prediction, ARIMA model, time series.

1. Introduction

With the development of economic globalization, the economic activities between various countries are becoming more and more close, the flow of capital in various countries and regions is becoming more and more frequent, and the influence of international stock markets is becoming more and more obvious. As the largest developing country, China is an important part of the global economic system. China's stock market also plays a role in various areas of China's economy to varying degrees. This means that China's stock market has a huge impact on the global economy, so the research and analysis of China's stock market is of great significance.

For different individuals and organizations, the prediction of stock prices has different meanings. For investors, accurately forecasting the trajectory of stock prices and executing trades at opportune moments enables them to achieve maximum returns. Forecast results can be employed to evaluate investment risks, facilitating the selection of strategies in alignment with their risk tolerance. For corporations, stock price forecasting facilitates the analysis of market trends and macroeconomic indicators. By monitoring stock price fluctuations, insights can be gleaned into the market's response to company performance, industry prospects, and the broader economic landscape. With the help of these data, corporations can adjust their short-term and long-term business strategies on time, ensuring the long-term stable development of the company. For countries, the risk on contagion between markets will become more and more obvious. In the process of formulating market rules, the Department of Financial Supervision should pay attention to the linkage between markets and the characteristics of information transmission, to improve the efficiency of supervision and reduce systemic risks [1]. Stock price forecasts can reflect the direction of the market economy and help formulate economic policies and interventions to maintain a stable macroeconomic environment and avoid shocks to the stability of the entire financial system.

This paper provides an overview of the literature on stock price prediction. Currently, there are three methods used for stock price prediction: time series analysis, stochastic process prediction models and machine learning. These methods can predict stock price changes within a certain time frame with relatively good accuracy and hold practical significance. Time series analysis [2] is a method for predicting stock prices, where historical stock prices are used to forecast future price trends. Common models include Generalized Autoregressive Conditional Heteroskedasticity

(GARCH) [3], Autoregressive Moving Average (ARMA) [4] and Autoregressive Integrated Moving Average (ARIMA) [5]. Using time series analysis can help forecast the trend of stock prices. For stocks with seasonal factors, the prediction results will display repetitive patterns occurring during specific periods. Additionally, time series analysis methods are typically used for short-term forecasting, providing relatively accurate predictions for price trends over a few days, weeks, or even months.

The stochastic process prediction model is also a method for stock price prediction, the most used model is the Markov prediction method [6]. Unlike prediction methods that require many sample data, the Markov Chain Analysis can analyze recent and current data, then predict based on the initial state of stock prices and transition probabilities, thus determining the trend of stock price changes. The model needs to consider the seasonality and periodicity of stocks, as well as the volatility of price changes. Predictions typically provide a forecast interval rather than just a single price prediction, to reflect uncertainty. However, due to changes in state transition probabilities over the long term, the accuracy of predictions and potential bias in results may increase with larger sample sizes. Using the Markov prediction method requires the use of an appropriate sample size.

With the development of computer technology, machine learning has found valuable applications in various fields. Using machine learning, it's possible to automatically extract useful features from large datasets without the need for manual processing. Moreover, machine learning offers higher predictive accuracy and superior generalization capabilities [7]. Rather than just fitting known data. Therefore, machine learning has become an important method for stock price prediction as well, with representative models such as back propagation (BP) neural networks [8], Elman neural network [9], and support vector machines [10]. Using machine learning, it is possible to predict stock prices for a future period, including opening prices, closing prices, high prices, and low prices. Additionally, machine learning for stock price trend prediction offers fast convergence, high computational efficiency, and yields solutions of high reference value [11].

With the development of the economy and technology, people are increasingly considering the impact of economic growth on the environment. Therefore, this paper takes Zhongyuan Env. Protection (000544), a stock in the environmental protection sector, as the research subject. It uses five years of closing stock prices and employs time series analysis for modeling and prediction.

2. Methodology

2.1. Data source

Zhongyuan Env. Protection, as a Chinese state-owned publicly traded company, was founded in 1996 and is an important representative of environmental protection stocks in the Chinese A-share market. It has valuable reference significance for the development of the environmental protection industry. The company is involved in various environmental services, including wastewater treatment, photovoltaic power generation, and environmental equipment manufacturing, with diversified business operations and stable development trends. In recent years, environmental issues have received increased attention from various countries and regions around the world. Therefore, this paper selects this stock as the research subject.

In this paper, daily closing price data from August 31, 2018, to August 30, 2023, was obtained from CSMAR as a sample. A total of 1213 daily data points were collected. The data were divided into two parts, data from 2018-08-31 to 2023-06-02, were used for modeling, while data from 2023-06-05 to 2023-08-30, were used for testing. Figure 1 displays the trend of the closing prices of the stock over 1213 trading days.

2.2. Variable description

Since the closing price is one of the most common financial data in the market, it is typically easy to obtain and use. Most financial news and data services provide daily closing prices, making it a convenient choice for price estimation and forecasting. Additionally, the closing price often reflects

the overall performance of the market on a given day, as it represents the last trade of the day. It is typically considered representative in pricing models and capable of capturing changes in market sentiment for that day. Analyzing historical data of the closing price can help investors gain a better understanding of an asset's past performance, enabling them to make more informed decisions about the future. This historical data analysis is valuable for both long-term investors and short-term traders. Therefore, this paper chooses the closing price of the stock for modeling and analysis.

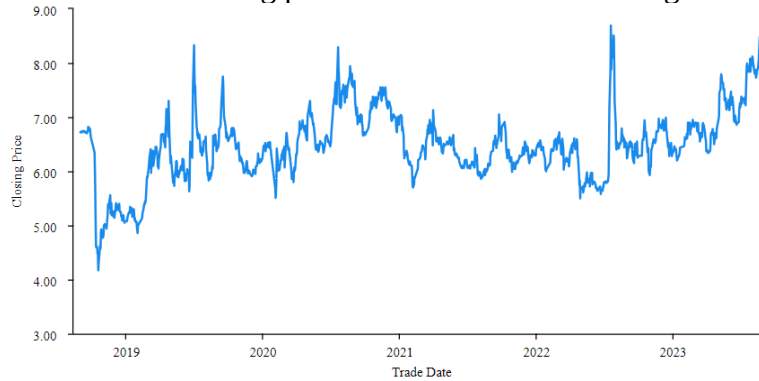


Figure 1. Closing price.

2.3. Mathematical statistics method

The Autoregressive Integrated Moving Average model which is abbreviated as ARIMA (p, d, q), is a well-known time series forecasting model proposed by Box and Jenkins in the early 1970s. In this notation, AR stands for Autoregressive, p represents the order of autoregression, MA stands for Moving Average, q represents the order of the moving average, and d represents the number of differencing operations required to transform the time series into a stationary time series. The essence of the ARIMA (p, d, q) model is the combination of differencing operations and the ARMA (p, q) model. After undergoing differencing operations, the ARMA (p, q) model becomes ARIMA (p, d, q).

$$Y_t = c + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

Where Y_t is a value of a time series, t is the time, c is the constant term, $\varphi_1, \varphi_2, \dots, \varphi_p$ are autoregressive coefficients, ε_t is the disturbance term, and $\theta_1, \theta_2, \dots, \theta_p$ are moving average coefficients.

To obtain an appropriate model, this paper follows the process depicted in Figure 2. First, confirm the stationarity of the sequence, and if the sequence is not stationary, perform differencing on it. Second, determines the autocorrelation and moving average coefficients. Then establish an ARIMA model by these parameters and optimize these parameters. Finally, evaluate the model, and forecast the closing prices in the future by the model.

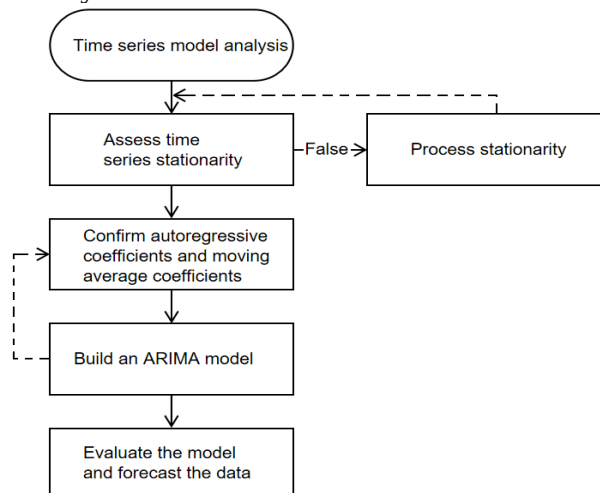


Figure 2. Modeling analysis process

2.4. Model principle

Autoregressive considers the relationship between the current value and past values. It's similar to performing a linear regression of the time series against its lagged values and Moving Average models the dependency between the current value and past white noise error terms. ARIMA comprehensively considers the influence of past values and forecast errors on the current value while using differencing to eliminate trends and seasonal factors in the time series.

3. Results and Discussion

3.1. Parameter determination

First, it is necessary to determine the parameters p , d , and q and establish the corresponding ARIMA (p , d , q) model. For the closing price, the ADF test shows that the test statistic of the ADF test for this time series data is -4.521 and a p -value is 0.000. The original data has no unit root, and the original data series is smooth. Therefore, the parameter $d=0$, and the model becomes ARMA (p , d).

3.2. Estimated values

For the ARMA (p , q) model, the values of p and q are determined through the analysis of correlation and partial correlation plots. Figure 3 displays the autocorrelation coefficient (ACF) of the closing prices, and Figure 4 displays the partial correlation coefficient (PACF) of the closing prices. From the ACF and PACF, recommended parameter values can be obtained as $p=q=1$.

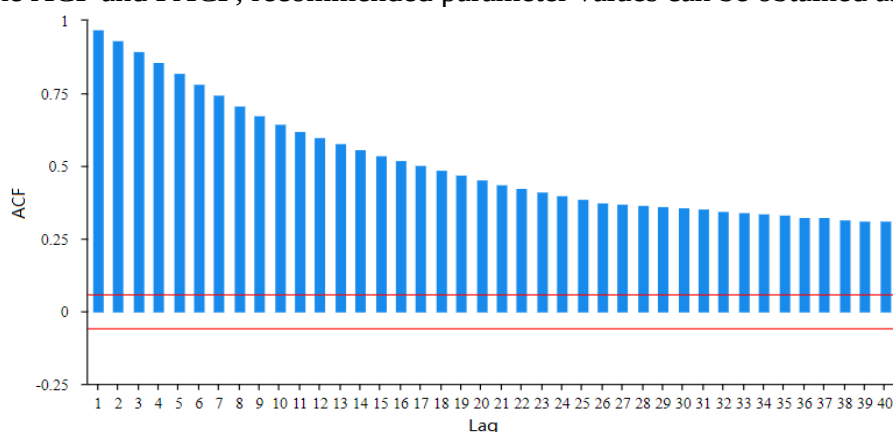


Figure 3. ACF of closing price.

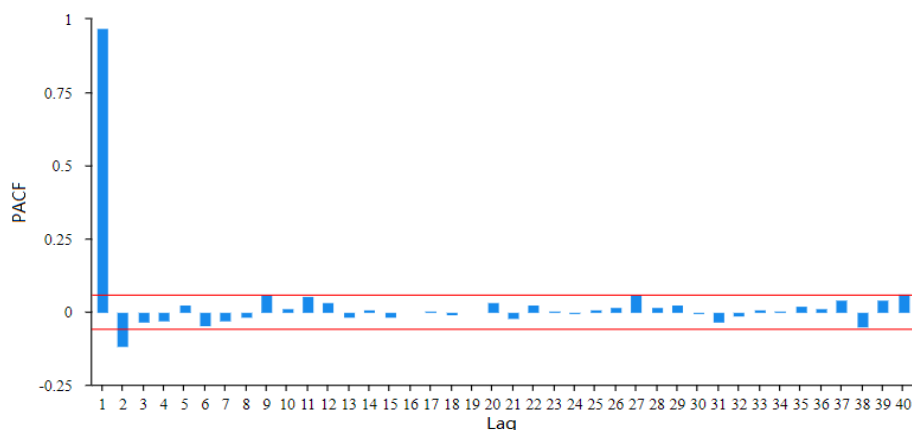


Figure 4. PACF of closing price.

To obtain the most suitable parameters, various values of p and q were attempted, as shown in Table 2. By comparing the Akaike Information Criterion (AIC) and Bayesian Information Criterion

(BIC) values obtained under different parameter combinations, it can be observed that the model achieves the lowest AIC and BIC values when $p=q=1$. Therefore, an ARMA (1, 1) model is selected for the analysis. Table 2 shows the parameters of the ARMA (1, 1).

Table 1. AIC and BIC of different model

d	p	q	AIC	BIC
0	1	1	-1149.73	-1129.54
0	1	2	-1149.35	-1124.11
0	1	3	-1149.30	-1119.01
0	2	1	-1150.60	-1125.35
0	2	2	-1148.80	-1118.50
0	2	3	-1149.08	-1113.73
0	3	1	-1148.57	-1118.28
0	3	2	-1135.58	-1100.23
0	3	3	-1146.87	-1106.48

Table 2. ARMA (1, 1) Parameter Estimation.

Variable	Coefficient	Std. Error
C	0.230	0.043
AR (1)	0.964	0.006
MA (1)	0.111	0.016

3.3. Prediction results

Table 3 shows the predicted closing prices for the next three weeks' trading days. The maximum relative error is 2.7%, but not more than 5%. The average relative error is 1.6%, so the proposed model can be considered valid.

Table 3. Predicted closing prices.

Date	Predict	Actual	Absolute error	Relative error
2023-06-05	7.291	7.47	0.179	0.024
2023-06-06	7.261	7.23	0.031	0.004
2023-06-07	7.231	7.26	0.029	0.004
2023-06-08	7.203	7.36	0.157	0.021
2023-06-09	7.176	7.37	0.194	0.026
2023-06-12	7.150	7.12	0.030	0.004
2023-06-13	7.124	6.94	0.184	0.027
2023-06-14	7.100	6.99	0.110	0.016
2023-06-15	7.077	6.91	0.167	0.024
2023-06-16	7.054	7.03	0.024	0.003
2023-06-19	7.032	6.89	0.142	0.021
2023-06-20	7.011	6.86	0.151	0.022
2023-06-21	6.991	6.88	0.111	0.016

3.4. Test model

To assess the reliability of the ARMA (1, 1) model, residual analysis is performed to check for white noise behavior. Figure 5 shows the residual series of the model. The ACF and PACF of the residual series are calculated and presented in Figures 6 and 7. Table 4 shows the Ljung-Box Q statistic of the residual series.

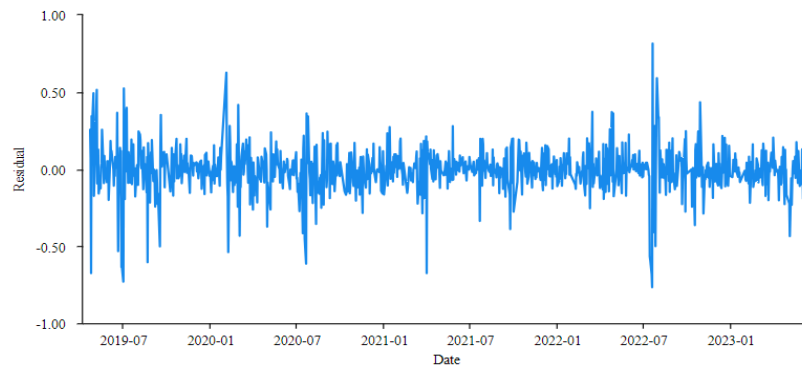


Figure 5. Residual series.

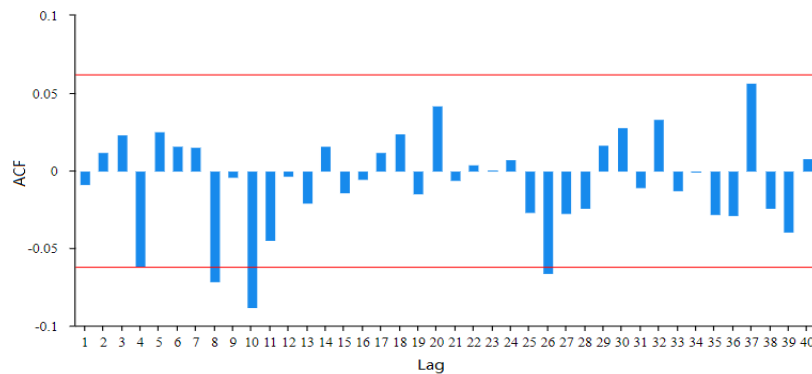


Figure 6. ACF of residual series.

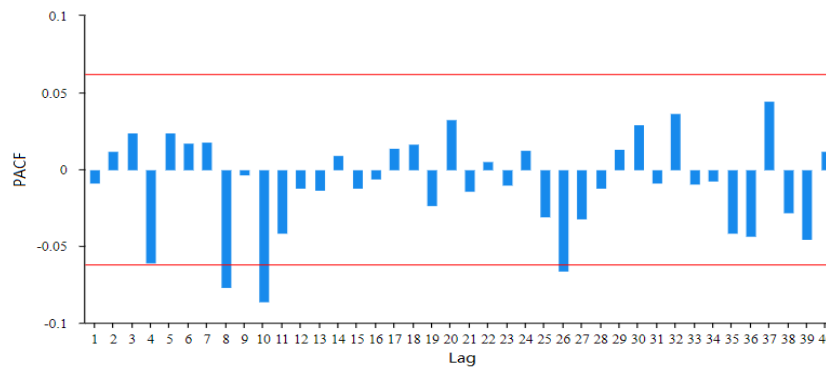


Figure 7. PACF of residual series.

Most of the ACF of the residuals and PACF fall within the confidence interval. Even though some of them exceed two times the standard deviation, this may be because there are a few large values in the residuals. P-value also shows that the residual series is white noise.

Table 4. Ljung-Box Q statistic

Lag	Q-statistic	P-value
1	0.026	0.873
2	1.87	0.393
3	3.148	0.369
4	3.964	0.411
5	5.578	0.349
6	6.155	0.406
7	6.319	0.503
8	10.247	0.248
9	11.02	0.274
10	15.343	0.120
11	17.668	0.09
12	17.846	0.120

3.5. Trading strategies

Making buy and sell decisions based on the trend of stock price movements is a commonly used stock trading strategy. By assessing the stock's trend, informed decisions are made to generate profits.

As shown in figure 8, the stock exhibited an overall upward trend until mid-May, followed by a sustained downtrend for over a month. Based on the prediction results, it can be inferred that the stock price is likely to continue declining in early June. Therefore, investors can choose to liquidate their holdings promptly and buy in at a later turning point.

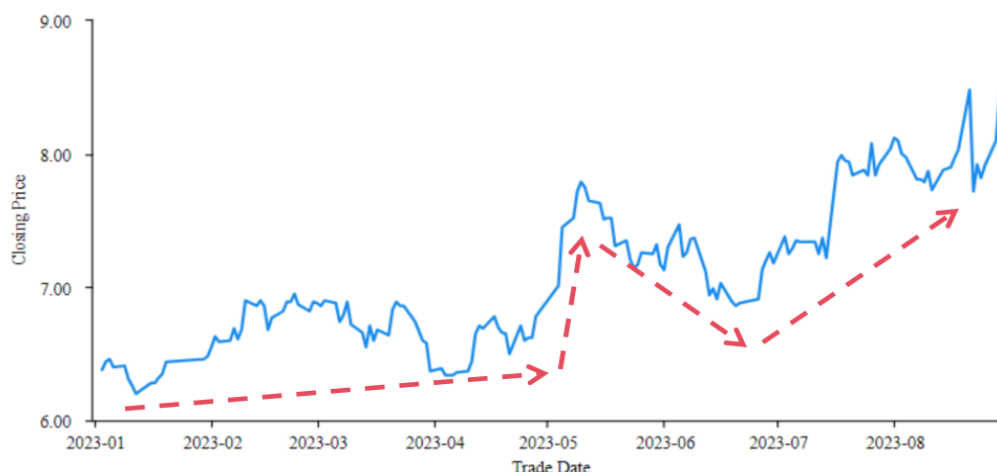


Figure 8. Closing price trend chart for the year 2023.

4. Conclusion

In this paper, a time series forecasting method of the financial market based on differenced moving average is proposed. The data is divided into two parts, with the closing price data from 2018-08-31 to 2023-06-02, used for modeling, and data from 2023-06-05 to 2023-08-30, used for validation. The prediction model is ARMA (1,1), with prediction accuracy errors of less than 5%, indicating that the model is effective and suitable for practical applications.

During the research process, several issues were identified. Like other financial data, closing prices exhibit some randomness. Short-term predicted prices closely match actual prices, but there are differences in actual prices and predicted prices over longer periods. Additionally, the model only considers the closing price of the stock and does not account for external factors such as policy changes and overall economic developments, which can lead to disparities between predicted and actual stock prices. During the research process, several issues were identified. Like other financial data, closing prices exhibit some randomness. Short-term predicted prices closely match actual prices, but there are differences in actual prices and predicted prices over longer periods. Additionally, the model only considers the closing price of the stock and does not account for external factors such as policy changes and overall economic developments, which can lead to disparities between predicted and actual stock prices. Therefore, in the actual stock prediction process, it is necessary to use the latest data for modeling to obtain a more accurate model and recent stock price forecasts. At the same time, during the modeling process, one can consider incorporating other variables that may influence stock prices.

References

- [1] Zhang, B., "Research on the Linkage of Chinese and American Stock Markets," *Economic Research* (11), 141-151 (2010).
- [2] Hao, B., "Research on Stock Prediction Models Based on Time Series Analysis," *UESTC*, (2011).
- [3] Awartani, B. M. A. and Corradi, V., "Predicting the volatility of the S&P-500 stock index via GARCH models: the role of asymmetries," *International Journal of Forecasting* (21), 167-183 (2005).

- [4] Fen, P., "Empirical Study of Stock Price Analysis and Forecasting Based on ARMA Model," *Mathematics in Practice and Theory* (22), 84-89 (2011).
- [5] Ariyo, A. A., et al. "Stock Price Prediction Using the ARIMA Model," 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation, 106-112 (2014).
- [6] Li, H., "Using Markov Prediction Method to Forecast Stock Prices," *Decision Reference* (05), 25-26 (2002).
- [7] Cai, H., "Stock Price Prediction Research Based on PCA-BP Neural Network," *Computer Simulation* (03), 365-368 (2011).
- [8] Yu, J., "Stock Market Modeling and Decision-Making Based on BP Neural Network," *Systems Engineering Theory and Practice* (05), 15-19 (2003).
- [9] Naeini, M. P., Taremian, H. and Hashemi, H. B., "Stock Market Value Prediction Using Neural Networks," 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 132-136 (2010).
- [10] Peng, L., "Application of Time Series-Based Support Vector Machine in Stock Prediction," *Computing Technology and Automation* (09), 89-91 (2006).
- [11] Lin, C., "Stock Price Prediction Research Based on Elman Neural Network," *Computer Applications* (02), 476-477 (2006).