A Case Study of Society General Securities based on ARIMA Model

Danyang Wang *

School of Economics and Management, Shanghai Maritime University, Shanghai, China

* Corresponding Author Email: 202110711394@stu.shmtu.edu.cn

Abstract. Stock price changes have always been a high concern for investors. Stock price modeling can indicate the stock changing trend, which can not only provide references for investors, but also can see the future development of the relevant market by analyzing the changing stock prices. Therefore, this paper establishes an Auto-Regressive Moving Average Model (ARIMA) of the opening price as well as the closing price of the trading day in Industrial Securities Co., Ltd. from November 2021 to August 2023 to predict the opening and closing price of the next seven periods of direction. The results show that the opening price will have a small fluctuation in the next 7 periods and the closing price will decline slowly. The goodness of fit of opening and closing prices are greater than 0.98, so this result is good. The result shows that this prediction of stock price by the ARIMA model is feasible and the result can be used as a basis for investors to predict future stock prices.

Keywords: ARIMA model; stock price; industrial securities.

1. Introduction

Stock prices are changing quickly, it is mainly based on the demand and supply of stocks as well as many other factors that fluctuate. Hence, it is imperative for investors to accurately forecast the fluctuations in stock prices, in this way can they make accurate predictions of stock prices and make accurate decisions of investing. Numerous researchers have employed various models in their studies and attempted to utilize these models for the purpose of forecasting stock prices. Examples of such models include the Capital Asset Pricing Model (CAPM), Price-Earnings Ratio, and Long Short-Term Model (LSTM). The CAPM model is easy to use and using it can have a clear result. It measures the price of an arbitrarily risky security with three factors: the risk-free return, the price of risk and the risky unit of valuation. Also, it makes investment choices on the basis of the mean and the variance. Not to mention, it is a very practical model. However, its assumptions are too rational to realize. The P/E model is more comprehensive as it is simpler to calculate and it covers the effects of growth rate, dividend payout ratio and risk compensation rate. However, if the earnings are negative, the P/E ratio will be meaningless. Also, it is affected by the level of economic prosperity. If the economic market conditions are very bad, there will be anomalies. Although the LSTM model has a long-term memory function, it can only be processed from front to back. Besides, the LSTM model cannot be processed by various methods. Each of the above models has its own advantages and disadvantages. Given that the data pertaining to daily stock prices can be considered as a time series, this research employs the ARIMA model to predict future stock prices by utilizing statistical time series analysis. In recent years, the model has been widely used. Some scholars tend to combine the important issues or phenomena with the result of the ARIMA model. Kenneth Gilbert used the ARIMA model for multi-stage supply chain modeling. The model provided explicit equations for quantifying the bullwhip effect in terms of inventory and orders. The results showed that the greater the effect was, the longer the lead time would be. Besides, the demand was autocorrelated [1]. Anubhav et al. employed the ARIMA model to forecast the performance of hydraulic pumps and developed a technique for predicting the Remaining Useful Life (RUL) of these pumps, which also played a key role in repairing and maintaining off-road hydraulic vehicles [2].

Most scholars combined the ARIMA model with one or multiple models to make up for the shortcomings of the ARIMA model. By doing so they can use the advantage of easy to use, of ARIMA model to make the results better [3]. Wang et al. found that the ARIMA model did not take

nonlinearities in the time series into account, but the ANN model was more widely used in the problem of nonlinearities [4]. In this study, the researchers conducted a comparative analysis of the ARIMA and ARIMA-ERNN models for predicting the occurrence of pertussis in mainland China from 2004 to 2021. The findings indicated that the ARIMA-ERNN model exhibited superior performance in terms of predictive accuracy [4]. Additionally, Valenzuela and other scholars developed a novel approach for constructing an enhanced ARIMA model by incorporating artificial neural networks, fuzzy systems, and evolutionary algorithms into its framework. He found that the accuracy of the results after using the final hybrid model exceeds the accuracy of using any of these individual models alone [5]. While Durdu combined neural networks with ARIMA model to forecast water quality time series, it is found that based on the hybrid model's ability to recognize the pattern of time series and nonlinearities, such a hybrid model provided higher accuracy when it studied problems with nonlinear data [4]. In terms of stock price forecasting, scholars such as Wang et al. developed a stock forecasting model with a mixture of three models, ESM, ARIMA, and BPNN. The result of this hybrid model was considered to be the best among the three models alone [5]. In addition, scholars such as Tsenga and others scholars optimized the ARIMA model itself. Researchers developed a fuzzy ARIMA model and conducted a comparative analysis with the conventional ARIMA model. Their findings indicated that the fuzzy ARIMA model yielded the most favorable and unfavorable outcomes with a reduced number of observations compared to the ARIMA model [6]. Additionally, the optimization of the ARIMA model was identified as a potential area of research. Ong et al. proposed a genetic algorithm-based approach to enhance the model identification process of ARIMA [7].

The ARIMA model is commonly employed in practical scenarios for the purpose of predicting the trajectory of time series data and examining the periodic patterns within such data. Also, it is said to be the most widely used time series forecasting model because is easy to use and adjust [8]. It only processes time series data and can be used in forecasting future trends in any field without doing overcomplicated processes [9]. Besides, short-term data can also be used in the analysis of stock price forecasting and the results are good. Dariusz and other researchers have discovered that the ARIMA model outperforms the LSTM model in situations where a single feature (historical price values) is utilized to predict multiple time periods [10]. Kumneth et al. argued that although artificial neural networks are more accurate in value prediction than the standard ARIMA model, they are more complex at the same time. So if there are only movements of stock need to be predicted, it is better to choose the ARIMA model [8]. As is known to all that Industrial Securities stock is the stock issued by the brokerage firm. Therefore, it is very sensitive to market fluctuations. When the market is good in financial time, trading and investment activities will be active. Hence, the construction of an ARIMA model can provide a certain level of assistance in predicting the trajectory of stock prices in the domestic financial market, thereby aiding investors in making informed and logical investment choices.

2. Methods

2.1. Data Sources and Description

The data this research used comes from the Tom Financial Terminal. It is a multi-functional securities information platform that allows access to various stock data.

2.2. Election and Description of Indicators

The focus of this study revolves around forecasting the volatility of stock prices, so this paper takes the necessary variables into consideration, the data mainly includes the opening, high, low, and closing prices of Industrial Securities for each trading day from November 2021 to August 2023. This study mainly uses the opening and closing prices to predict the stock price volatility (Figure 1).



Fig. 1 Industrial Securities stock price volatility chart (2021/11/12~2023/8/11).

2.3. Introduction to the Methodology

This study aims to employ the ARIMA model for the purpose of forecasting the stock price of Industrial Securities. The ARIMA model formula is:

$$X_t = \mathrm{c} + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \cdots + \alpha_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \cdots + \theta_q \varepsilon_{t-q} \quad (1)$$

In the formula p, q are non-negative integers, ε_t, ε_{t-1}, $\varepsilon_{t-2} \ldots \varepsilon_{t-q}$ are white noise terms. The random wandering time series of these white noise terms is not a reliable indicator of future trends. ARIMA model is made up of AR and MA models. It is a forecasting model that predicts future data based on observations at the past p time points and error terms at the past q time points. The differencing process is represented by the letter I. Since one of the prerequisites of ARIMA is using a smoothed series, before using the AR and MA model to process the data, if the time series is not a smoothed series, a d-th order differencing is required to change the data from non-smooth to smooth. The AR term, determined by p, describes the values in the previous p periods in the model; the MA term indicates how much error there is in past predictions when predictions are made about future values, and is determined by q.

The first step in modeling is to check if the series is smooth, to be specific, the first step is to confirm whether the series' statistical properties such as variance and mean are constants. This part can be determined by making an ADF plot. The ADF test is a statistical procedure used to assess the presence of a unit root in a time series. Its purpose is to determine whether the series exhibits a unit root or not. In the case of a smooth series, the ADF test will indicate the absence of a unit root, whereas a non-smooth series will be characterized by the presence of a unit root. The initial hypothesis of the test posits the presence of a unit root, which is subsequently evaluated through the significance of the test. If the significance test statistic yielded is less than three degrees of confidence (10%, 5%, 1%), the initial hypothesis will almost certainly be rejected (90%, 95, 99% of the time). If it is not smooth, the first step is to differentiate a number of times until this time series becomes a stable time series, the order d can also be determined by the ADF test. Differencing is usually used to calculate the difference between neighboring data points in a set of numerical series.

The values of p and q are then generally estimated from the maximum lag points of the PACF and ACF graphs. Next, by using the AIC and BIC criteria to compare the values of different orders, the values of p and q are further determined. After that, an ARIMA model (p, d, and q) can be fitted, and finally, future values can be predicted.

$$AIC = -2In(L) + 2K \quad (2)$$

$$BIC = -2I_n(L) + KIn(n) \quad (3)$$

3. Results and Discussion

3.1. Model Results

To determine if the series is smooth or not, the ADF test is first run. The t-statistic of the ADF test is calculated in the study using SPSS. As shown in Table 1, the opening data result is -1. 662, the p-value is 0. 451, and the critical values for 1%, 5%, and 10% are, respectively, -3. 446, -2. 869, and -2. 570.; the p-value is 0. 451> 0. 1(Table 1). In general, the p-value is typically set at a threshold of less than 0.1 (or sometimes 0.05 as a commonly used standard). This implies that if the calculated p-value exceeds 0.1, the initial hypothesis can be rejected, indicating that the observed sequence is considered to be smooth. The series is not smooth, but the initial theory still holds true at this point. The next step is to perform the first-order differences and then rerun the ADF test. The data's ADF test result after first-order differencing indicates that p=0. 0000.1. The original hypothesis may be ruled out with greater than 99% certainty at this stage, according to the smooth series and sequence of differencing d.

Table 1. ADF test table for opening data.

Variable	Differential order	t	P	AIC	Critical value		
					1%	5%	10%
Opening	0	-1.662	0.451	-521.387	-3.446	-2.869	-2.57
	1	-8.937	0.000***	-513.509	-3.446	-2.869	-2.57
	2	-9.053	0.000***	-477.518	-3.447	-2.869	-2.571

The ADF test for the closing data has a t-statistic of -1. 664 with a p-value of 0. 450, as reported in Table 2. Critical values are -3. 446, -2.868, and -2.570, respectively. p=0. The original hypothesis cannot be rejected due to p=0.450 > 0. 1(Table 2), and the series is not smooth. First-order differencing is necessary for the series. The data's ADF test result after first-order differencing indicates that p=0. 000 <0.1 so the original hypothesis can be rejected with more than 99% certainty as stated in paragraph 1 so there is more than 99% certainty to reject the original hypothesis. The series is smooth at this time, and the order of differencing d is 1.

Table 2. ADF test table for closing data.

Variable	Differential order	t	P	AIC	Critical value		
					1%	5%	10%
Close	0	-1.664	0.450	-484.817	-3.446	-2.868	-2.57
	1	-21.193	0.000***	-485.149	-3.446	-2.868	-2.57
	2	-8.875	0.000***	-442.63	-3.447	-2.869	-2.571

The p-value and q-value are then roughly estimated using the maximum lags of the partial autocorrelation coefficient (PACF) and autocorrelation coefficient (ACF) plots, respectively, to further enhance the ARIMA model.

The autocorrelation coefficient (ACF) is a tool for measuring the degree of correlation between two events during different periods of time. In a figurative sense, it is a measure of the effect of a person's past actions on their current situation. Partial autocorrelation coefficient (PACF) refers to the coefficient obtained by calculating the influence or correlation of one element. At this time, the influence of other elements is considered as constant, that is to say, the PACF only studies the closeness of the relationship between those two elements without considering the influence of other elements.

The opening and closing price ACF and PACF results were obtained by SPSSPRO. If the series becomes very small from a certain time period, such a phenomenon is called truncation; if the series decays slowly, such a performance is called torus tailing. Figure 2 shows that p=1 and q=1 for the opening price; p=1 and q=1 for the closing price.

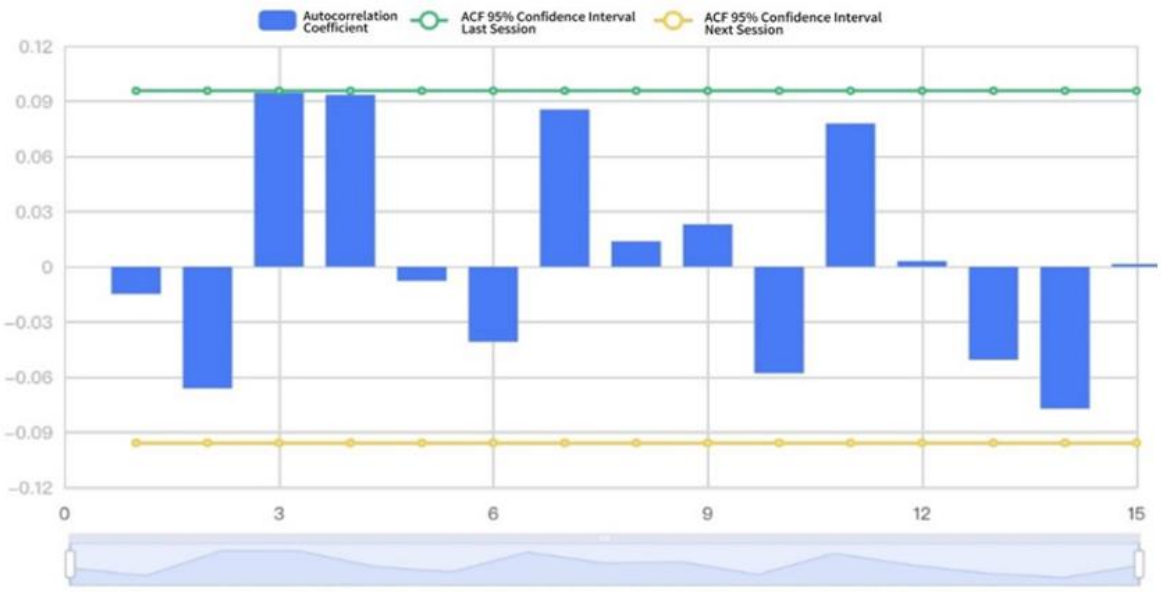


Fig. 2 Final differential autocorrelation plot of opening price data (ACF).

Then the two evaluation indexes, AIC and BIC, are used to test the performance of the ARIMA model, by comparing the smaller AIC and BIC values, the corresponding p-value and q-value are finally determined to fit the ARIMA model. In practice, it is common for these two criteria to be combined to provide a reference for judging model quality. The Akaike Information Criterion (AIC) is a tool for comparing and evaluating system models, where a lower value indicates a superior model. The AIC formula makes it simple to handle the relationship between the model's complexity (number of parameters) and goodness-of-fit (log-likelihood), with a smaller value indicating a better model. Also, the AIC formula is also used in the ARIMA model fit test, using the AIC formula can find models that fit the data well but at the same time, do not overfit it. While BIC (Bayesian Information Criterion) is used when the sample number is large. When choosing which model is better, after comparing the BIC values of different models, the model with the smallest BIC value will be chosen.

Table 3. ARIMA model for opening price data (2, 1, 2).

Item	Symbol	Value
Sample size	Df Residuals	414
	N	420
Q Statistic	Q6(P-value)	0.823(0.364)
	Q12(P-value)	3.654(0.723)
	Q18(P-value)	8.731(0.726)
	Q24(P-value)	13.279(0.775)
	Q30(P-value)	17.787(0.813)
Information criterion	AIC	-499.526
	BIC	-475.298
Goodness of fit	R^2	0.987

As Table 3 issues, for the opening data, after analyzing the model several times to compare the values of AIC and BIC, the smaller values of -499. 526 and -475. 298 are used. The model is (2, 1, 2) after SPSSPRO automatically looks for the best parameters based on the AIC information criteria. The R^2 value of the model is 0.987, indicating a high level of goodness of fit. This suggests that the model performs exceptionally well and largely satisfies the specified criteria.

For the closing data, after analyzing the model as well as comparing the values of AIC and BIC, the research takes the smaller values of -478. 656 and -470. 581. The model ends up being (0, 1, 0) after SPSSPRO automatically determines the best parameters based on the AIC information criterion. The model's goodness of fit $R^2$is 0. 986, which means the model performs excellently and the model basically meets the requirements, as Table 4 issues.

Table 4. ARIMA model for closing price data (0, 1, 0).

Item	Symbol	Value
Sample size	Df Residuals	418
	N	420
Q Statistic	Q6(P-value)	0.727(0.394)
	Q12(P-value)	4.372(0.626)
	Q18(P-value)	6.259(0.902)
	Q24(P-value)	10.583(0.911)
	Q30(P-value)	14.687(0.930)
Information criterion	AIC	-478.656
	BIC	-470.581
Goodness of fit	R^2	0.986

By drawing the residual autocorrelation plot and the biased autoregressive autocorrelation plot, a white noise test of the ARIMA model can be done. From the residual autocorrelation plot (ACF) of the opening model, it can be seen that the value of the autoregressive model (AR) residuals explains that they are white noise series this study needs.

From reading the partial autocorrelation plot (PACF) of the model residuals for the opening data, it can be seen that the residuals of MA perform a shape of white series that meets the needs of doing the MA model successfully. From reading the partial autocorrelation plot (ACF) of the model residuals for the opening data, it can be seen that the correlation coefficients and the residuals of AR are the ideal value that the time series requires, so the AR model is successfully applied.

As Figure 3 issues, the value of correlation coefficients and the residuals of MA shows that the MA model is successfully applied.

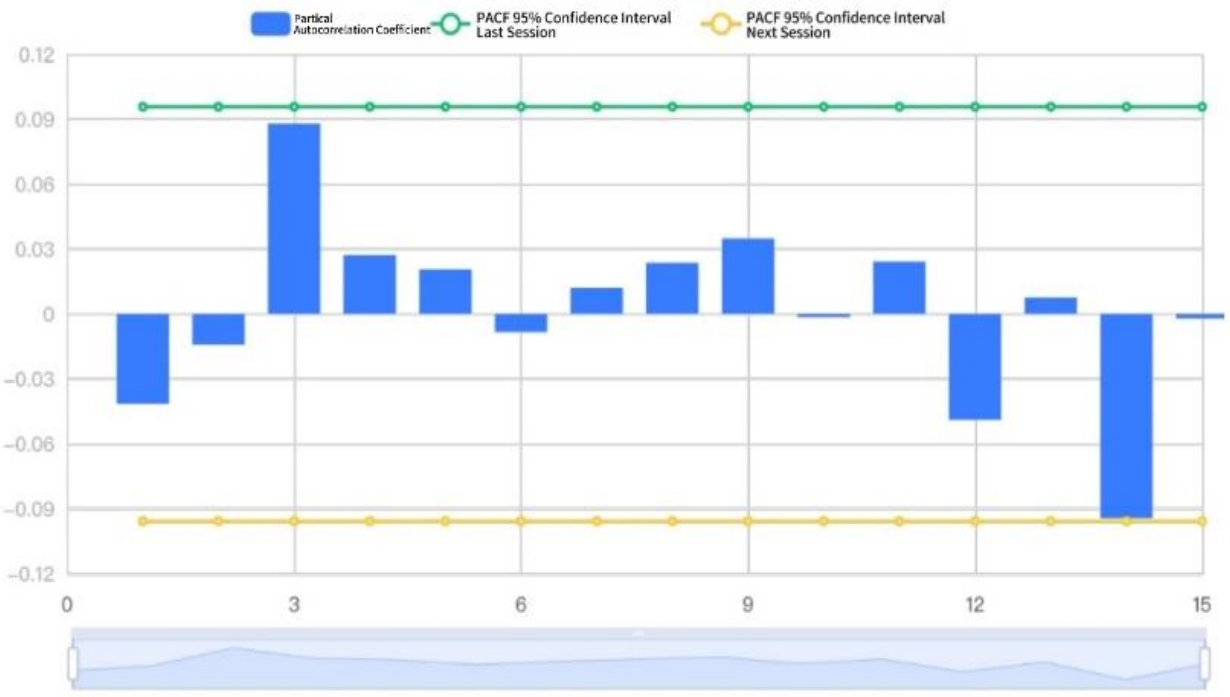


Fig. 3. ARIMA model residual bias autocorrelation of closing price data (PACF).

Finally, the parameter results of the opening data model are given as Table 5 shows, the model results are ARIMA model (2, 1, and 2). After 1st order differencing, using a multi variate linear combination, the model equation can be derived as $y_{(t)}$=-0.004+0.122* $y_{(t-1)}$-0.911*$y_{(t-2)}$-0.182*$\varepsilon_{(t-1)}$+0.894*$\varepsilon_{(t-2)}$.

Table 5. ARIMA model parameter test table for opening price data.

	Coefficient	SE	t	P>|t|	0.025	0.975
Constant	-0.004	0.006	-0.636	0.525	-0.016	0.008
ar.L1.D. opening	0.122	0.072	1.697	0.09	-0.019	0.263
ar.L2.D. opening	-0.911	0.065	-14.04	0	-1.039	-0.784
ma.L1.D. opening	-0.182	0.075	-2.418	0.016	-0.33	-0.035
ma.L2.D. opening	0.894	0.075	11.972	0	0.748	1.041

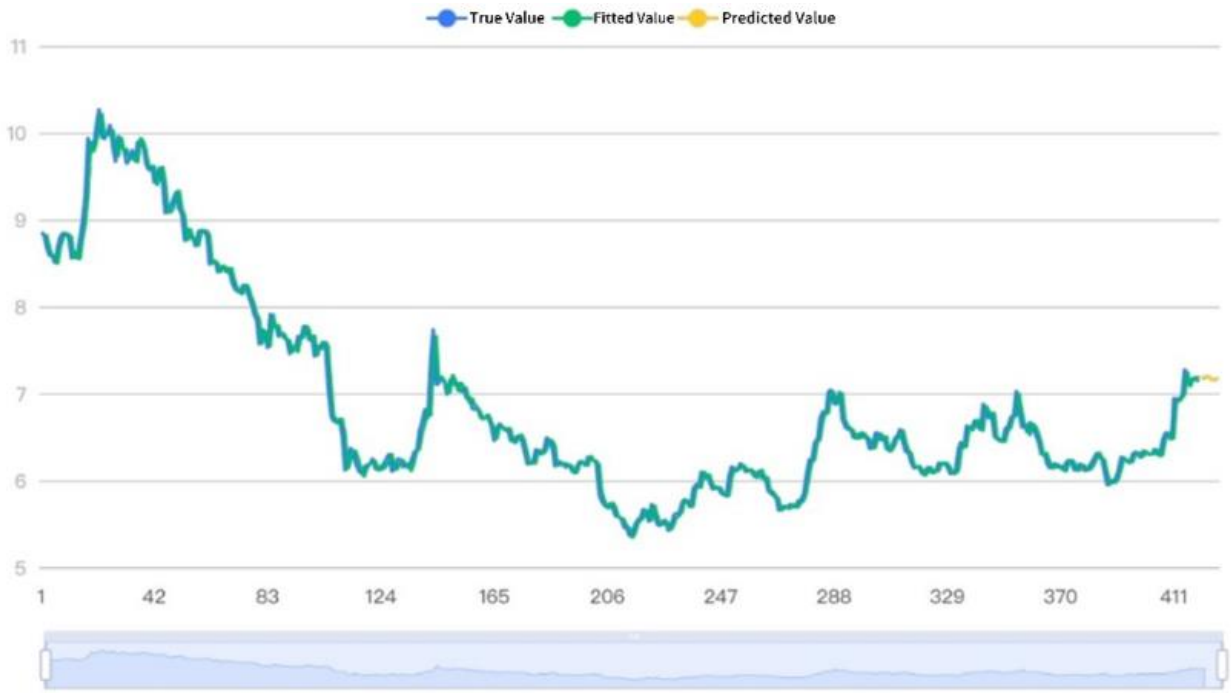


Fig. 4 Time series plot of opening price forecast data.

Figure 4 issues the forecast data of stock price in the next seven periods. This plot shows that the predicted opening stock price will have a slight increase, but it will soon come down and in the last period it tends to increase again. As Table 6 issues in detail, the future stock price will increase to 7.2 in the 3rd period and it has a trend to increase to 7.2 again.

Table 6. Opening price time series forecast table.

Order (time)	Forecast Result
1	7.1744
2	7.1827
3	7.2001
4	7.1876
5	7.1633
6	7.1647
7	7.1801

Table 7 displays the outcomes of the closing data parameter results. The model's parameter combination is the ARIMA model (0, 1, and 0). After the first-order differencing and the use of multivariate linear combination, the model formula turns out to be$y_{(t)} = -0.004$.

Table 7. ARIMA model parameter test table for closing price data.

	Coefficient	SE	t	P>|t|	0.025	0.975
Constant	-0.004	0.007	-0.661	0.509	-0.017	0.009



Fig. 5 Time series plot of closing price forecast data.

Figure 5 shows the forecast data of closing price in the next seven periods. It issues that the predicted closing price shows a nearly downward trend. And table 8 shows the forecast value of closing price, which performs a decreasing trend.

Table 8. Closing price time series forecast table.

Order (time)	Forecast Result
1	6.9956
2	6.9912
3	6.9868
4	6.9824
5	6.9780
6	6.9736
7	6.9692

3.2. Discussion

The ARIMA model stock price prediction by other scholars mostly combines the ARIMA model with other models to compare whether using the ARIMA model alone performs better or using the hybrid model can have a better result. While the study in this paper focuses on one stock and uses the ARIMA model alone, which is a lack of diversity.

4. Conclusion

This study employs the ARIMA model to forecast the opening and closing movements of stock prices over the course of the next seven periods. These predictions can serve as a guide for Industrial Securities Company and the entire stock market. The analysis of the data reveals that the opening and closing values are 0. 987 and 0. 986, respectively, which can indicate the predictions of stock price trend is accurate. From the results, in the next seven periods, the opening price fluctuates relatively small above and below the historical level, but the closing price maintains a continuous slow and small downward trend. Since the ARIMA model predicts better for short-term time series, this paper speculates that the reason for the better fit is related to the shorter time series this study selected.

References

[1] Kenneth Gilbert. An ARIMA Supply Chain Model, ManagementScience, 2005, 51: 305-310.

[2] Sharma A K, et al. Lifetime Prediction of a Hydraulic Pump Using ARIMA Model. Arab J Sci Eng, 2023.

[3] Wang M, et al. ARIMA and ARIMA-ERNN models for prediction of pertussis incidence in mainland China from 2004 to 2021. BMC Public Health, 2022, 22: 1447.

[4] Durdu Ömer Faruk. A hybrid neural network and ARIMA model for water quality time series prediction, Engineering Applications of Artificial Intelligence, 2010, 23: 586-594.

[5] JuJie Wang, et al. Stock index forecasting based on a hybrid model, Omega, 2012, 40: 758-766.

[6] FangMei Tseng, et al. Fuzzy ARIMA model for forecasting the foreign exchange market, Fuzzy Sets and Systems, 2001, 118: 9-19.

[7] Chong Shyong Ong, et al. Model identification of ARIMA family using genetic algorithms, Applied Mathematics and Computation, 2005, 885-912.

[8] Sim S K. et al.Wind Speed Modeling by Nested ARIMA Processes. Energies, 2019, 12.

[9] Dariusz Kobiela, et al. ARIMA vs LSTM on NASDAQ stock exchange data. Procedia Computer Science, 2022, 207: 3836-3845.

[10] Valenzuela O, et al. Hybridization of intelligent techniques and ARIMA models for time series prediction, Fuzzy Sets and Systems, 2008, 159: 821-845.