ISSN 2959-6157

ARIMA Model-Based Research on Stock Price Prediction

Dong Liang

School of Mathematics and Applied Mathematics, China Agricultural University, Beijing, 100083, China
Corresponding author: dongliang@ldy.edu.rs

Abstract:

Securities trading has always been a high-risk, high-return domain. Investors seek high returns while endeavoring to minimize risks as much as possible. Therefore, stock price prediction has become a popular and immensely valuable research topic. This paper will use the ARMA model to forecast stock prices. Firstly, an analysis was conducted on selected stock, determining that the price sequence exhibits no seasonal effects but does display volatility effects. The trend is essentially linear, and the relationship between volatility effects and trends fits an additive model. Based on this, preprocessing was conducted by taking the three-day moving average sequence of the series to eliminate the volatility effects, yielding a clean sequence trend. Then, the trend was differenced once to obtain a stationary sequence. Subsequently, the appropriate ARIMA model order was determined by the (partial) autocorrelation plot of this stationary sequence, and the model was fitted to the stock for prediction, yielding satisfactory results. This indicates that the model can accurately forecast long-term trends, but the filtering of volatility effects prevents the prediction results from sensitively reflecting short-term fluctuations.

Keywords: ARIMA; stocks price; prediction.

1. Introduction

Since the emergence of stocks in the 19th century, many researchers, both domestically and abroad, have been devoted to the study of stock prediction, proposing some traditional technical analysis theories. These theories, trendline theory, wave theory, pattern theory, and Dow theory, are still widely circulated among the public. However, these methods are too crude and increasingly difficult to cope with the complex phenomena in the stock market, such as fat tails, volatility clustering, and asymmetric volatility. Later, researchers applied various statistical methods to stock market modeling, such as multiple regressions, exponential smoothing, Autoregressive Moving Average (ARMA) models, Autoregressive Conditional Heteroskedasticity (ARCH) models, etc. For example, Li used multiple linear regression to predict various stocks. He selected the opening price, highest price, lowest price, closing price, trading volume, turnover, and next day's opening price of these stocks as independent variables to predict the next day's closing price. He ultimately reduced the variables to two [1]. Luo predicted the closing price of the Shanghai Composite Index using exponential smoothing [2]. Exponential smoothing recognizes that the impact of past data on current data decays from recent to distant, but it assumes that this decay is exponential, which is somewhat rigid. The ARMA model uses (partial) autocorrelation tests to specifically analyze the magnitude of the influence of past values on current values, thereby determining the model's independent variables. ARMA can also be extended to Autoregressive Integrated Moving Average (ARIMA) to deal with non-stationary sequences. Many domestic researchers have attempted to analyze and predict stock prices using ARIMA models [3]. However, the ARMA model has a prerequisite, namely, that the time series does not exhibit heteroscedasticity. However, in actual stocks, stock volatility often exhibits clustering effects, namely, heteroscedasticity. Therefore, researchers invented the ARCH model to address this phenomenon. Later, on the basis of this model, a series of model families, such as Generalized Autoregressive Conditional Heteroskedasticity (GARCH), was developed and widely applied [4].

Due to the influence of numerous factors on stock prices and the complexity of internal patterns in stock price sequences, it is difficult to model them with simple models. Thus, traditional forecasting methods generally fail to achieve ideal results. With the rapid development of artificial intelligence, neural networks have provided a new approach to stock prediction. In the early 1990s, Matsuba proposed that artificial neural networks could be used for stock prediction, believing that the stock price process is a complex nonlinear function that neural networks can simulate to predict stock prices. Since then, researchers have begun to apply various neural networks to stock prediction. Yu and Guo used the closing price of China International Marine Containers (Group) Ltd. as the forecasting target. They proposed a stock price prediction model based on Radial Basis Function (RBF) neural networks, which was feasible and effective for short-term stock price prediction [5]. Zhang and Yuan leveraged the memory advantages of Elman neural networks to propose a GA-Elman dynamic regression neural network stock price prediction model, which was fast and stable. They had higher accuracy, making it feasible and effective for stock price prediction [6]. In addition, there are Back Propagation (BP) neural networks, Support Vector Machine (SVM), and so on [7, 8].

Due to the complexity of stock prices, using a single prediction model often does not achieve ideal results, so many researchers combine multiple models to model and predict stock prices. Xiong and Che combined ARMA and GARCH models for stock price prediction. They compared the results with those of standalone ARMA models, with experimental results showing that the combined model's forecasting results were better [9]. Li combined ARMA models with Generalized Regression Neural Networks (GRNN) to predict stock prices, and experimental results showed that the combined model's forecasting results were better than those of individual ARMA models and GRNN neural networks, demonstrating better nonlinear approximation capabilities [10]. Lin et al. combined time series models with SVM models to predict stock prices, obtaining very good forecasting results and proving that the hybrid model's performance was superior to using time series models or SVM models alone [11]. Li and Bai proposed a combined forecasting model based on ARIMA models and BP neural network models, applied the model to analyze and predict stock price time series, and achieved satisfactory results [12]. This paper will analyze and forecast a domestic stock based on the ARIMA model.

2. Methods

2.1 Data Source

This article's data is sourced from Tongdaxin and includes the daily closing prices of Zhongmin Energy from April 30, 2021, to September 30, 2021. The data for the last 20 days will be used to validate the model, while the preceding data will be used for model fitting.

2.2 Method Introduction

Before selecting a model with suitable orders, it's crucial to assess the stationarity of the time series. Stationarity can be determined through direct observation or the autocorrelation function test. If non-stationary, the data needs to be stabilized, typically by differencing once or multiple times to eliminate trends or periodicities. When selecting a model, the decision can be informed by the truncation or tailing of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the time series, helping choose among Autoregressive (AR), Moving Average (MA), or ARMA models and their respective orders. This step can be done through direct observation or using the F-test. Once the orders are determined, parameter estimation can proceed, which is commonly done using the least squares method. With this, the model is established, but it requires validation for fitness and parameter significance before being deployed for predictions.

3. Results and Discussion

3.1 Stationary Processing

Below (Fig. 1) is the time series plot of the daily closing prices of Zhongmin Energy stock. It can be observed that the sequence exhibits sporadic volatility. The volatility does not increase with the rise in stock prices, so judging the relationship between trend and volatility conforms to an additive model.

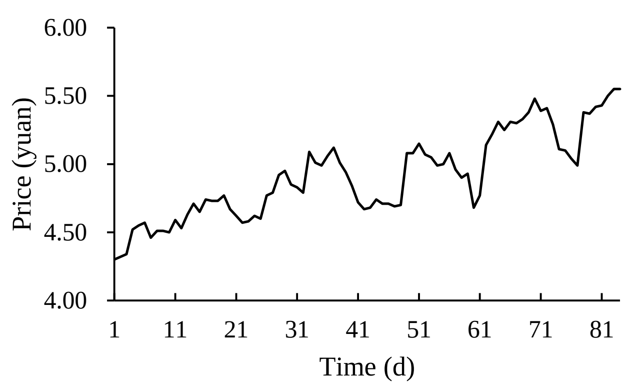

Fig. 1 Daily closing plot

Now, the volatility of this series needs to be eliminated. The specific approach is to compute the 3-day moving average of the series, resulting in the following plot (Fig 2).

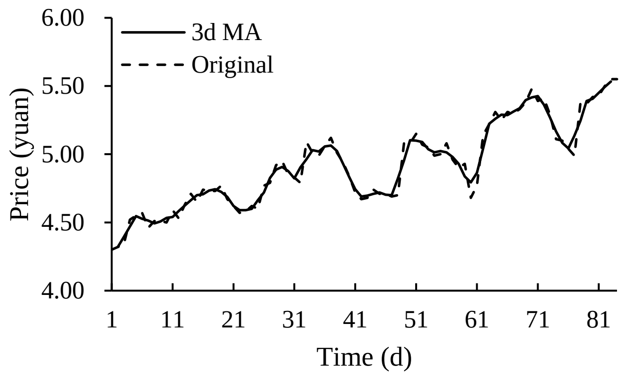

Fig. 2 Three-day MA plot

It can be observed that the trend of this series is linear. Therefore, the series needs to be applied first-order differencing to stabilize, resulting in the following plot (Fig 3).

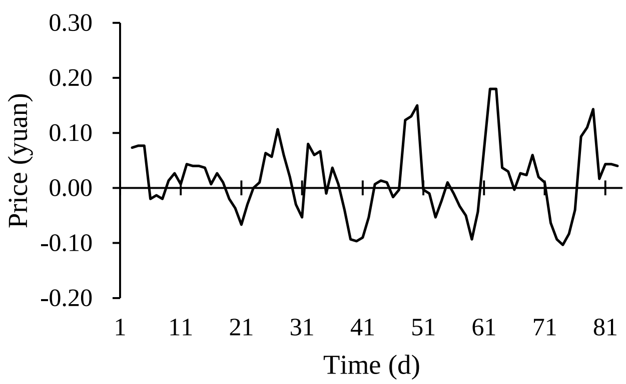

Fig. 3 First-order difference of 3d MA plot

Following the ADF test, the time series data yielded an ADF test statistic of -5.412, with a p-value of 0.000, which is less than 0.01. With more than 99% confidence, the null hypothesis is rejected, indicating the stationarity of the series.

3.2 Model Identification

Now, the orders of the ARIMA model need to be determined. The specific method is to assess the following (partial) autocorrelation plot of the series (Fig 4, Fig 5).

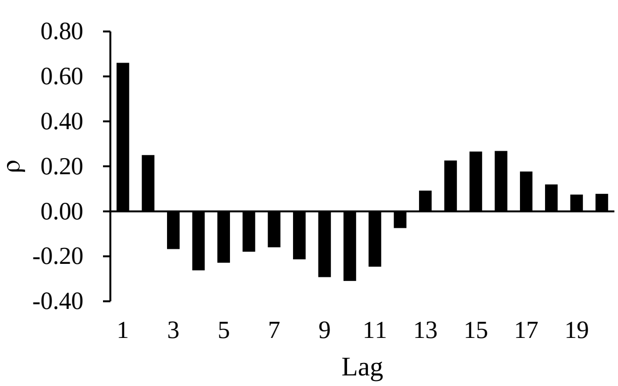

Fig. 4 Autocorrelation plot

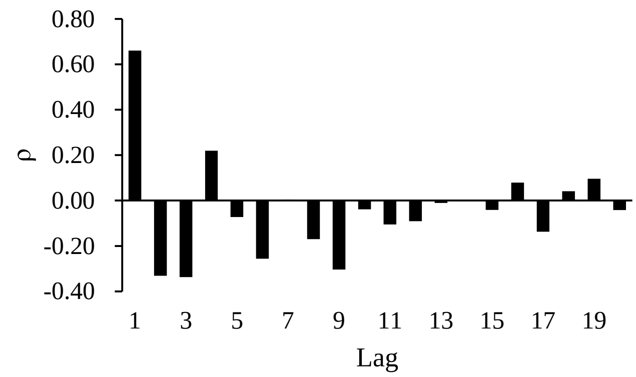

Fig. 5 Partial autocorrelation plot

Observing a trailing pattern in the autocorrelation plot and a third-order truncation in the partial autocorrelation plot, the preliminary model is tentatively identified as ARIMA(3,1,0). Below (Table 1) are the fitting results of the model.

Table 1. ARIMA(3,1,0) models parameter list

Term	Coefficient	value
c	0.009	0.118
y_{t-1}	0.749	0.000
y_{t-2}	-0.005	0.973
y_{t-3}	-0.348	0.004

After testing, it is found that the residuals of the model are white noise, indicating that the information of the sequence has been fully extracted, so there is no need to increase the order of the model. Additionally, from Table 1, it can be seen that the coefficients of the highest-order term are significantly non-zero, so there is no need to decrease the order of the model either. Therefore, the final model is determined to be ARIMA(3,1,0).

3.3 Parameter Estimation

It should be noted that the coefficients of the second-order term and the constant term in Table 1 are not significantly non-zero. Therefore, these two terms need to be omitted, and a new fitting is performed, resulting in Table 2 below.

Table 2. Parameter list (simplified)

Term	Coefficient	value
y_{t-1}	0.746	0.000
y_{t-3}	-0.361	0.000

It can be observed that all coefficients in the table are significantly non-zero. The model simplification is completed, and the final expression is as follows (based on first-order differencing).

$$y_t = 0.746 y_{t-1} - 0.361 y_{t-3} \tag{1}$$

3.4 Trend Prediction

Finally, using equation (1) to forecast the daily closing prices for the next 20 periods and comparing them with the actual prices, the results are shown below (Fig 6). It can be observed that the overall trend is relatively consistent, but it does not reflect short-term fluctuations well.

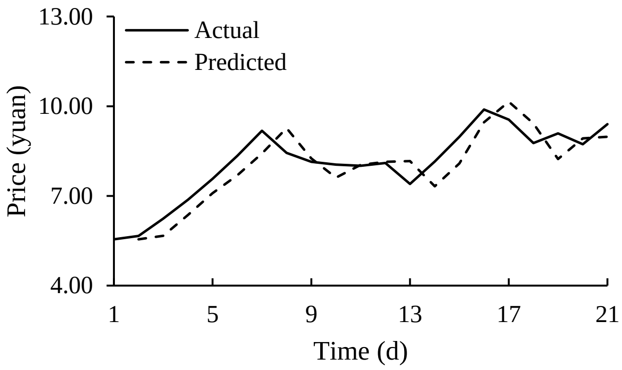

Fig. 6 Comparison between actual and predicted

4. Conclusion

This article fitted a stock using the ARIMA model and predicted its future trends over a certain period. Through the analysis process and the expected results, it can be observed that the ARIMA model performs well in fitting and forecasting stock time series data that have undergone appropriate trend extraction. In actual market conditions, a stock not only exhibits long-term trends but also short-term cyclical effects. These cyclical effects do not have fixed periods, making them challenging to extract and fit compared to the seasonal impacts. However, if not separated, they can seriously interfere with the examination of (partial) autocorrelations in the sequence. Thus, they must be eliminated, leaving only the trend. For instance, in this article, the 3-day moving average of the daily closing price was used to eliminate the cyclical effects of the time series, resulting in a purer trend. However, because of this, the model's prediction results lost information on cyclical effects, reducing the short-term sensitivity of the forecasts. Therefore, this model may not be suitable for guiding short-term investments by investors, but it still holds significant importance for guiding long-term investments.

References

[1]Li Jun. Factor analysis-based multiple linear regression method and the application in prediction to stock price. Nanjing University, 2014.

[2]Luo Jiawen. Prediction to closing price of Shanghai securities composite index based on HP filtering and exponential smoothing. Time Finance, 2016, 21: 159-160.

[3]Shi Jia. Analysis and forecast of stock price law based on ARIMA model. Statistics and Application, 2020, 9(1): 101-114.

[4]Luo Jiawen. Regularity exploration to daily closing price of Shanghai securities composite index based on GARCH(1,1). Time Finance, 2016, 29: 138-139.

[5]Yu Jian, Guo Ping. Financial time series prediction based on RBF network. Journal of Hunan College of Engineering (Natural Science Edition), 2007, 4: 46-48.

[6]Zhang Lijun, Yuan Di. GA-Elman dynamic regression neural network-based research on stocks prediction model. East China Economic Management, 2008, 9: 79-82.

[7]Chu Wenhua. The application of BP neural network in stock prediction. Journal of Harbin University of Commerce (Natural Science Edition), 2019, 35(2): 237-240.

[8]Junyoung H, Yong J Y. Stock price prediction based on financial statements using SVM. International Journal of Hybrid Information Technology, 2016, 9(2): 57-66.

[9]Xiong Zheng, Che Wengang. The application of ARIMA-GARCH-M model in short-term stock prediction. Journal of Shaanxi University of Technology (Natural Science Edition), 2022, 38(4): 69-74.

[10]Li Weimin. ARMA-GRNN-applied research on stocks prediction. Shandong University of Science and Technology, 2004.

[11]Pai P F, Lin C S. A hybrid ARIMA and support vector machines model in stock price forecasting. Omega, 2005, 33(6): 497-505.

[12]Li Chunxing, Bai Jiandong. Combined prediction model-based research on the method of stocks prediction. Journal of Qingdao University of Technology, 2008, 2: 82-85.