Proceedings of the 2nd International Conference on Financial Technology and Business Analysis
DOI: 10.54254/2754-1169/64/20231542

Stock Price Undulation Prediction Using ARIMA Model

Yutian Su[1,a,*]

[1]*Institute of Problem Solving, Yangzhou University, Yangzhou, Wuxi, China*
a. 210803126@stu.yzu.edu.cn
**corresponding author*

Abstract: Stock is one of the most important investments nowadays, but it is also changing very rapidly. Its dramatic changes may lead to property damage. The stock price can influence the market situation. If its future price can be predicted, people's investments will be more accurate. To solve the problem quickly, creating models seems to be the best choice, While the influences keep changing, some easy models cannot work well. By comparison between many models, we decided to choose Auto-regressive Integrated Moving Average Model (ARIMA). ARIMA model is a set of AR model and MA model. ARIMA model uses historical information of the data to predict the future. Though ARIMA model cannot respond well to emergencies. Assume that the average month price of the stock is steady. This paper uses this model to predict whether the investor will gain or loss. Through SPSS, the model is built and tested. The model's correct rate is higher than 0.7, so it can greatly help people to make decisions.

Keywords: stock, ARIMA, prediction

1. Introduction

Stock market variations have an enormous economic influence on national and personal consumers' financial conditions [1]. Many widespread economic disturbances might be caused by a stock price collapse which may not attract much attention. Nowadays, many people choose stocks as a way to invest. Hedging an investment plan through diversification is becoming increasingly important due to the high level of global uncertainty and its negative effects on financial markets [2]. Chinese Liquor is one of the Chinese features, and MaoTai is the representative of Chinese Liquor. It is considered to be the most famous liquor representative. Its development level has a reference effect on the development level of other Chinese liquor, so if the stock trend of MaoTai can be predicted, other stocks can also be forecast. This model can be widely used.

Stock prices are influenced by many impacts, as a result, it is difficult for us to find a absolutely easy prediction model to do further research. The accuracy of the forecast data will not be high. Moreover, some influencing factors may decrease or vanish along with time, we cannot correctly estimate only by experience. A fundamental use in the financial rivalry is the early stock price prediction that is based on sentiment research [2]. But the stock data set variations keep changing, many detail prediction are not as simple as people thought. The stock data bank contains a number of traits and qualities [3].

ARIMA is called Auto-regressive Integrated Moving Average Model. It includes AR, MA and I. ARIMA model, a very familiar prediction manner abided by time series analysis. Through unpacking

DOI: 10.54254/2754-1169/64/20231542

the time series of incipient data, ARIMA can propose the short-term forecast correctly. Analysis of time series, in particular, ARIMA model has beis created to solve the disappearing gradient issue.en extensively used in various areas, for example economics and statistics, and it also plays an important role in medical study [4]. AR is called Autogressive model. It is a statistical method of dealing with time series. In the document, in the existence of AR disturbing the denial ability of fictitious goals has not been intensive studied [5]. MA called Moving average model. It is one of the spectral analysis approaches of model parameter approach. MA is a familiar former in the present day's spectrum estimation too. MA has three biggest deficiencies which may cause some problems. The first defect is it may be a few omissions of some records of the beginning and the end through the time series. Secondly, the lack of this model is prospective ability of periodic variation of the initialized information. However, it has not been discovered. Thirdly, the most important lack of the MA is its fragility to the biggest and smallest observational data [6]. ARMA is a combine of AR and MA. There are only few works studying quantized identification of auto-regressive moving average (ARMA) systems [7]. *I* represents differential disposal. Many people believe that ARIMA is only used to learn and forecast some special climate changing issues and appraising climate index. On the contrary, in a board sense, the ARIMA model provides more accurate predictions for certain studies. Specifically, its prediction of intervals seems to be more consistent than other commonly used statistical methods [8]. The basic steps of ARIMA model modeling are three steps, which are stationarity test, model parameter selection and model fitting. On this note, a trustworthy stationarity test that can provide objective results for a specific application is required. As a result, a collection of test reliability data would give the user adequate confidence to choose the right tests [9].

Unlike LSTM, ARIMA is easy to learn and use. LSTM, a typical architecture for recurrent neural networks, is designed to overcome the vanishing gradient problem (Hochreiter, 1998) [10]. The LSTM model's basic abilities is it can remembering some information for a long time. However, the operation of the LSTM is not convenient for beginner and the parameters need to be adjusted for different prediction objects. As a result, this paper chooses ARIMA and to predict.

Creating a ARIMA model quickly is easy and convenient. There are many fully automatic programs on the network that only need to import data. This paper chooses SPSSAU to create the model. In this software, the operation is easy and the result are accurate.

2. Methodology

2.1. Data Source

Materials all come from Baidu Stock which will update everyday. The data set includes Mao Tai stock prices for each month starting in 2015.The opening stock prices and the closing prices are included in the state. Earlier data may be biased, so they are not taken into account.

2.2. Data Selection

The data used in this paper count a total of 103 months, including some special cases. The data contains 5 variables (Opening price, Transaction volume, Turnover, Closing price, Gain or loss).

According to the different trend on economy, too old data cannot help predict the price accurately and too many data are difficult to deal with, so the data before 2015 are not included. Prices on a daily basis fluctuate greatly and everyday factors cannot be ruled out, so monthly data are used. Some social factors will influence the price of stock, while these influences will not last long and the chance of appear is very low. As a result, these data need to be canceled to increase the correctness of the prediction. To ensure the time series, the canceled data will be replaced by the average.

DOI: 10.54254/2754-1169/64/20231542

2.3. Introduction of Method

This paper uses ARIMA model, closing price is considered as a dependent variable which represented by Y, and Date is considered as a independent variable which represented by X. Creating ARIMA model mainly contains three steps. First is judging the stationarity of the sequence. The stationarity is generally judged by time series diagrams. If it is non-stationarity, difference disposal can be done.

Second, determine the auto-correlation coefficient. The orders of AR, I and MA need to make sure. Generally, the proper p and q orders can be obtained by using ACF and PACF test, and the proper difference order d can be obtained by using ADF test. Third, after making sure of p, d and q, ARIMA model can be created, and test the model. To simplify the process, the paper chooses SPSSAU to analyze the relationship between the effect of X on Y.

3. Results and Discussion

The "Date" is the time term, and the "Closing price" is the actual analysis term. Date increments from top to bottom. The collated data is then uploaded to the SPSS platform.

In order to accurately check the data's stationarity, this paper choose to test in two ways--scatter plot judgment and ADF test. First, scatter plot judgment. It can be seen from Figure 1, the scatter points in the figure have an obvious upward tendency. As a result, the requirement for time series is not met. Therefore, these figures will be done some measures to find the logarithm first. After that, Differential disposal is required and tested at last.

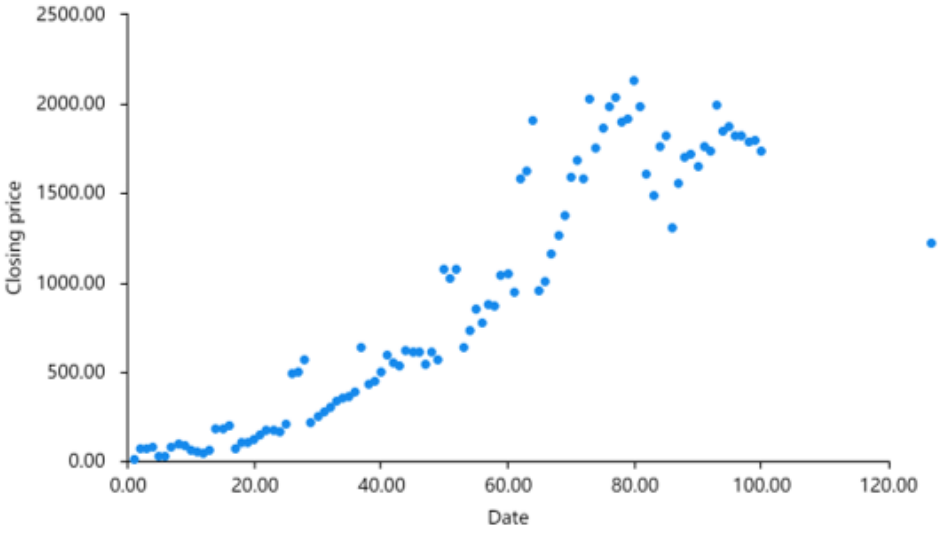

Figure 1: Scatter plot of Date and Closing price.

Secondly, ADF test. According to Table 1, value p=0.002<0.01, so it can refer that the null hypothesis is rejected with over 99% determinacy, and the order is steady. As a result, d is 0.

Table 1: ADF testing.

Difference order	t	p	Threshold		
			1%	5%	10%
0	-3.959	0.002	-3.496	-2.890	-2.582

DOI: 10.54254/2754-1169/64/20231542

Then we can draw a partial correlation graph to judge the order of p and q. According to Figure 2 and 3, pictures of ACF and PACF are non-truncated, so the data can be used to creating ARIMA model. And p and q are 1.

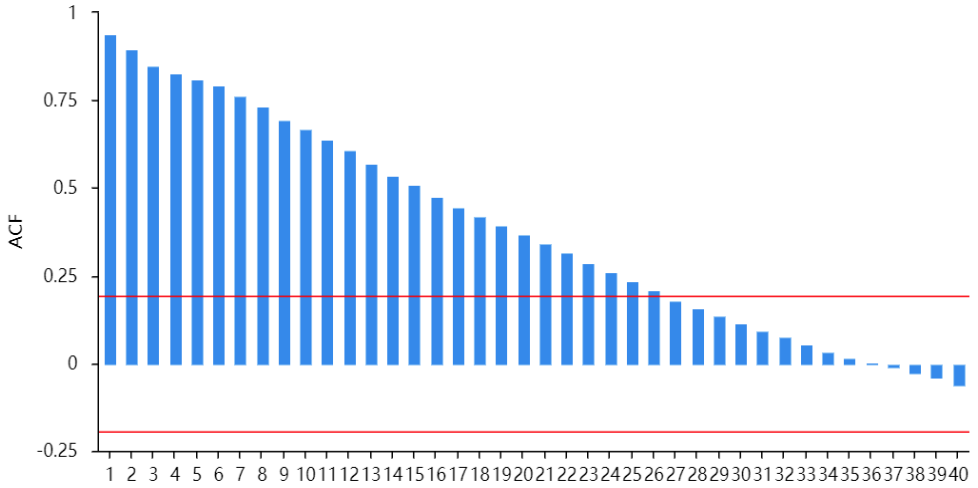

Figure 2: ACF Plot.

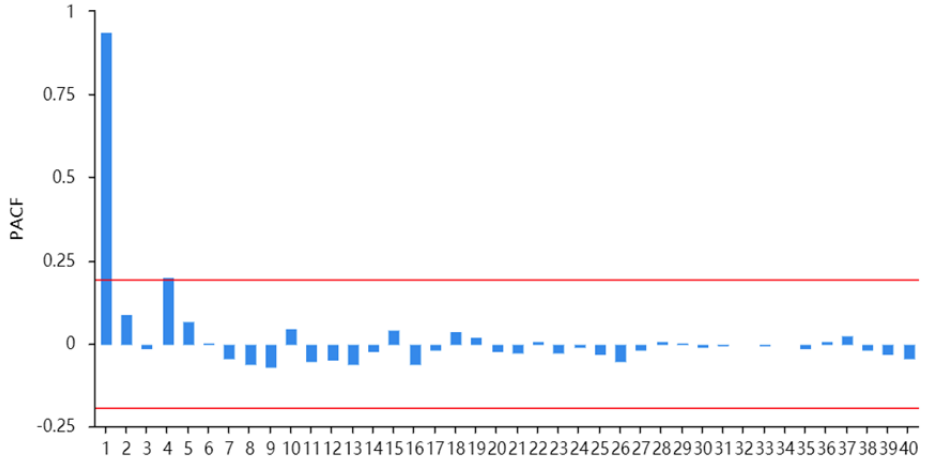

Figure 3: PACF Plot.

After getting all data, we can start to create model. Table 2 and 3 show the results of this model construction. According to Table 3, the p-value of Q18 is 0.286, much bigger than 0.1. As a result, data match white noise test and the model can work well. The ARIMA model is established.

Table 2: Model Results.

	Parameter	Standard error	P-value
Constant term	0.019	0.085	0.828
AR parameter	0.996	0.017	0.000
MA parameter	0.148	0.145	0.306

DOI: 10.54254/2754-1169/64/20231542

Table 3: Model Q statistics.

	Statistical value	P-value
Q1	3.533	0.060
Q6	15.839	0.015
Q12	19.596	0.075
Q18	20.873	0.286

By ARIMA model and Figure 4, the logarithm of price of next month is 7.557, so the prediction price is 1914. The true price is 1847. Though these prices are not same. The two data are close, indicating that the prediction accuracy is high. The ARIMA model can be used to predict the next time's price.

Generally, the ARIMA model is relatively reliable for backward periods 1 and 2, and will be less accurate if there are too many backward periods.

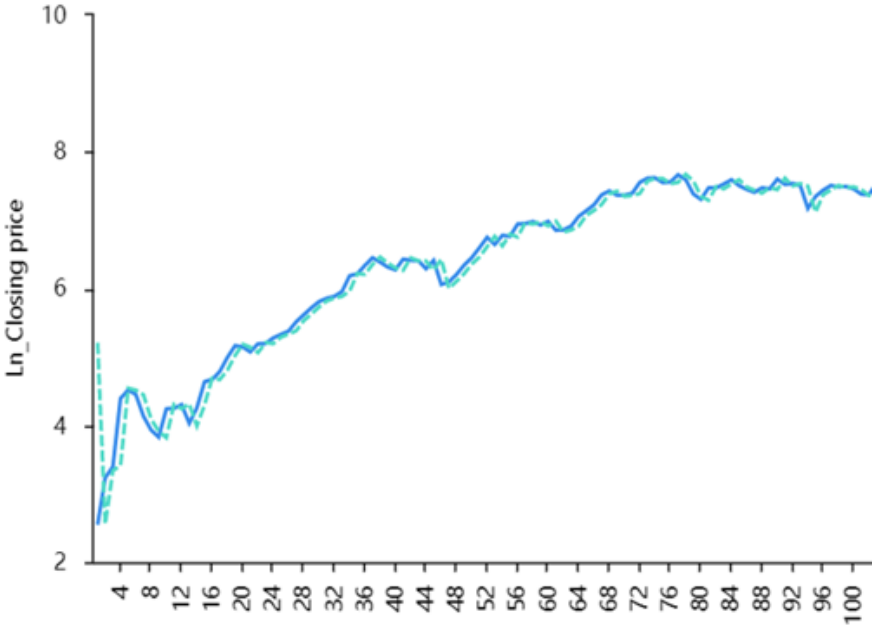

Figure 4: Ln_Closing price model prediction and fitting.

4. Conclusion

The current study selects diverse data and focus on influencing factors that may be associated with the price of stock. It was concluded that Opening price, Transaction volume and Turnover all will influence the price. Using these data to create a ARIMA model to predict. Through testing, the reliability of the model is very high. There is no doubt that these data are limited and may have some faults. However, these errors are inevitable. And the accuracy of the model is still up to a high level. As a result, people can assume that these errors do not greatly affect the accuracy of the model. The model can still work well. Only the data need people to find in person that may cost some time, other steps can be finished by computer. If more and corrected data is provided, the model can predict more accurately. Moreover, if the data is replaced by other model's data, this model can still work well to predict. Most stocks can be predicted by ARIMA models. Different stock should consider different factors, so people need to identify which data to consider. This method works for most stocks. As a

DOI: 10.54254/2754-1169/64/20231542

result, this model has high economic value. With these predictions, people can make a better decision when they are hesitated whether buy more stock or sell it, so they can gain more money and avoid loss. The development of MaoTai can also be forecast.

References

[1] *Jaimin, S., Darsh, V. and Manan, S. (2022) A comprehensive review on multiple hybrid deep learning approaches for stock prediction. Intelligent Systems with Applications, 16.*

[2] *Walid, M., et al. (2023) Extreme dependencies and spillovers between gold and stock markets: evidence from MENA countries. Financial Innovation, 9(1).*

[3] *Yining, J., et al. (2022) ARIMA model for predicting chronic kidney disease and estimating its economic burden in China. BMC Public Health, 22(1).*

[4] *Tareek, P. and Dinesh, J. (2022) Stock prediction analysis by customers opinion in Twitter data using an optimized intelligent model. Social Network Analysis and Mining, 12(1).*

[5] *Mehdi, D. and Reza, M.T. (2021) A new parametric adaptive detector with mismatched signals rejection capability in AR model interference. Digital Signal Processing, 109.*

[6] *Chen, R.J.C., et al. (2003) An evaluation of alternative forecasting methods to recreation visitation. J leis Res, 35(4), 441–454.*

[7] *Ting, W., et al. (2023) Identification of ARMA models with binary-valued observations. Automatica, 149.*

[8] *Muhammad, A., et al. (2022) Analysis of Temperature Variability, Trends and Prediction in the Karachi Region of Pakistan Using ARIMA Models. Atmosphere, 14(1).*

[9] *Amol, A.B. and Rajanarayan, B.P. (2022) Selection of Stationarity Tests for Time Series Forecasting Using Reliability Analysis. Mathematical Problems in Engineering.*

[10] *Nath, H.B., et al. (2022) Predicting stock market index using LSTM. Machine Learning with Applications, 9.*