ISSN: 2664-9640

Research on Financial Market Price Trend Based on ARIMA Model

Chen Shen, Yiran Zhang, Minghao Wei

Faculty of Electronic Information and Electrical Engineering, Dalian University of Technology, Dalian, Liaoning, 116024, China

Abstract

In stock market transactions, experienced investors buy at low prices and sell at high prices to achieve high profits. But for most shareholders, this is not an easy task. Based on the idea of regression and difference, we build an ARIMA model to predict the price of a stock. By grouping, the method of partition-by-interval prediction obtained all predicted values except the initial 100 days, and passed the ADF test. Next, in order to verify the correctness of the predicted data, we mixed the predicted data with the actual data to obtain a new dataset D, and used the logistic regression classification model to find that 94.3% of the data were classifie

Keywords

ARIMA, Financial market, Quantitative transaction.

1. Introduction

Market traders maximize the interests of interests through the assets of traders. The level of traders affects a huge, large data information age, and a large number of digital information make people dazzling, how to identify the most effective digital information from these complex macro numbers. Traders hope that the daily daily transaction data and policy orientation, the investment assets are predicted, which in turn will profit.[1]

How to accurately predict stocks has always been an important issue in the field of quantitative finance. Some scholars use LSTM and GA to establish a stock price prediction model that is better than LSTM alone. In the model, LSTM is used for prediction and GA is used for parameter adjustment; [2] Some scholars proposed an advanced hybrid forecasting model based on the combination of ARIMA and ANN, based on ARIMA's statistical properties for finance. [3] Johan Bollen used the content on Twitter to explain the rise and fall of the stock market forecast. They analyzed the public emotion, and then added these emotional characteristics to the prediction model to predict the rise and fall of the stock market. [4] Using forecast data provided by a global corporation, Martin presents sensitivity analysis of ARIMA on shifts in fitting periods including the financial crises. [5]

This paper collects the price change data of Bitcoin and gold from 2016 to 2021, and conducts relevant analysis and experiments on this data set, and establishes a complete set of investment product "price prediction-validation" algorithm models.

2. Financial Product Price Forecast Model Based on ARIMA Model

The stock price for all days needs to be known when deciding on a trading strategy for a particular day. For the stock price after the ith day, we use the ARIMA prediction model to predict,[6][7] and use ADF test and differential analysis to process to ensure the accuracy of the prediction results.[8]

For the ARIMA prediction model, the overall algorithm flow chart is given as follows:

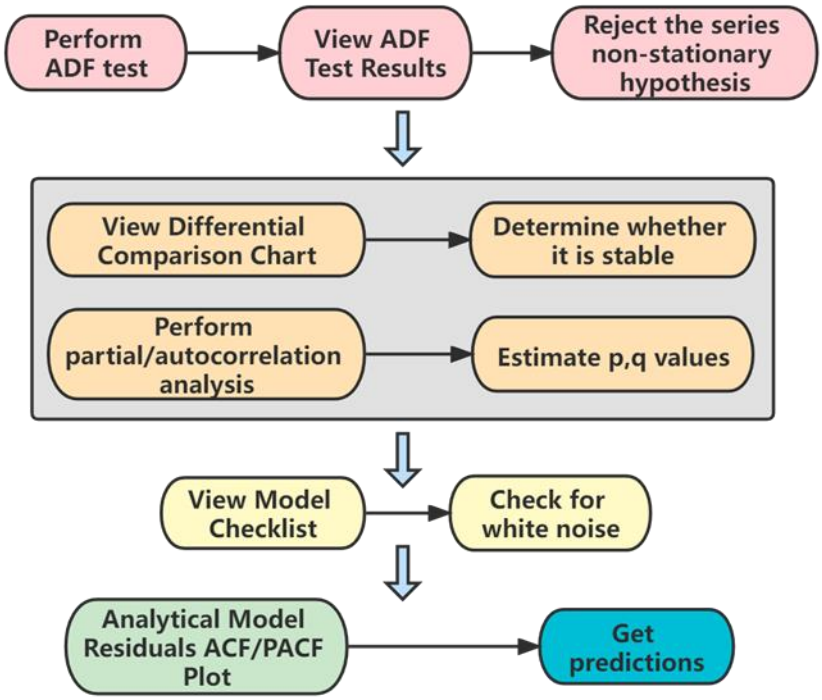

Figure 1: ARIMA model algorithm flow chart

Specific steps are as follows:

1. Check the ADF test results, and analyze whether it can significantly reject the hypothesis that the series is not stationary (p<0.05 or 0.01) according to the analysis t value;
2. Check the data comparison chart before and after the difference, and estimate itsp and q values according to the censoring situation;
3. Check the model test table, test the white noise of the model according to the pvalue of the Q statistic, obtain the model formula combined with the time series analysis diagram, and finally obtain the order result of the backward prediction.

Overall prediction formula of ARIMA model:

$$\left(1-\sum_{k=1}^{p}\phi_k B^k\right)(1-B)^d Z_t=\left(1-\sum_{j=1}^{q}\theta_j B^j\right)a_t \tag{1}$$

Where p is the order of the autoregressive model, d is the difference order, q is the order of the moving average model, Z is the time series data, B is the backshift operator, and ϕ and θ are the coefficients and shifts of the autocorrelation process, respectively. Coefficient of averaging process, a is white noise.

ARIMA is a typical time series model, which consists of three parts: AR model (autoregressive model) and MA model (moving average model), and the order I of the difference.

Before formally giving the ARIMA model, first give the following key definitions

(1) Common formulas for differential operations

$$\Delta y_t=y_t-y_{t-1}=(\alpha-1)y_{t-1}+\varepsilon_t \tag{2}$$

(2) Stability of Difference Equations

The inverse characteristic equation of the P-order difference equation is:

$$1-a_1L-a_2L^2-\cdots-a_pL^p=0 \tag{3}$$

ISSN: 2664-9640

Then the condition of stability is that the characteristic root is greater than one.

(3) Weak Stationary and White Noise

i) The definition of weak stationarity: also called covariance stationarity or second order stationarity. For a random time-series y, if its expected value, variance and autocovariance do not change with time t, then 1 is called a weakly stationary random variable.

ii)White noise process: If a random process is called a white noise process, all the random sequences that make up the process are independent of each other, and the mean is 0, and the variance is a constant value.

In view of the above definitions, we can give a specific description of the ARIMA model as follows:

(1) AR model

The AR model believes that the current moment is related to the previous moment, and the formula is as follows:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \epsilon_t \tag{4}$$

The above is called the p-order autoregressive model, abbreviated as AR(p) model.

(2) MA model

The moving average process MA refers to the time series process y_t, written as a linear combination of a series of uncorrelated random variables.

The MA(1) model has the following form:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} \tag{5}$$

$$\varepsilon_t : \text{ii d}(0, \sigma^2) \tag{6}$$

Where c represents a constant term, ε_t is a white noise process with variance σ^2, and θ is a coefficient.

(3) ARMA model

From the characteristics of the MA process, it is known that the MA process is a stationary process under any conditions, and the stationary requirements for the ARMA process are completely reflected in the requirements for the AR part.[9]

For any ARMA process, the stationarity requirement is that the roots of the equation must lie outside the unit circle. If one or more of the roots fall on the unit circle, the ARMA(p, q) process at this time is called the differential autoregressive moving average process, denoted as ARIMA(p, d, q).

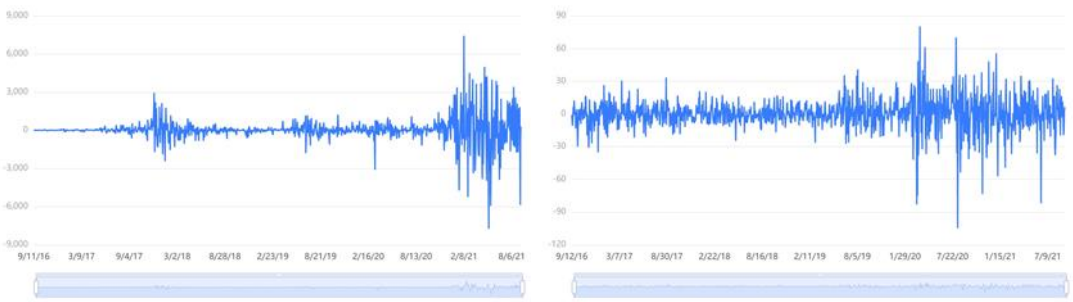

Figure 2: The best difference sequence diagram

The above figure shows the timing diagram of the original data after 1st order difference. The first-order difference can eliminate the linear trend factor and obtain a stationary series.

The figure below shows the autocorrelation plot (ACF), including coefficients, upper and lower confidence limits.

ISSN: 2664-9640

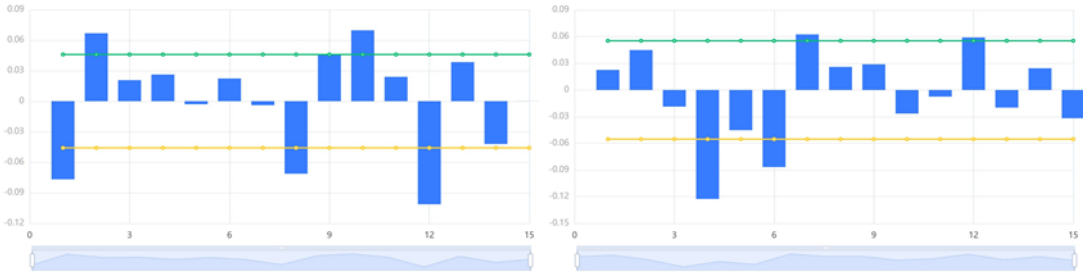

Figure 3: Final Differential Data Autocorrelation Chart

The blue column represents the autocorrelation coefficient, the green circled line represents the upper bound of the ACF 95% confidence interval, and the yellow circled line represents the lower bound of the ACF 95% confidence interval.

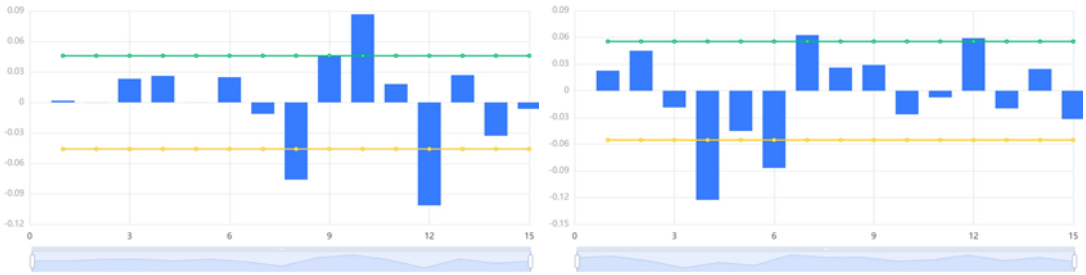

Figure 4: Model residual autocorrelation plot

The blue column represents the autocorrelation coefficient, the green circled line represents the upper bound of the ACF95 % confidence interval, and the yellow circled line represents the lower bound of the ACF 95% confidence interval.

The figure above shows the residual autocorrelation plot (ACF) of the model, including coefficients, upper and lower confidence limits.

Figure 5: Time Series Plot

The blue line represents the true value, the green line represents the fitted value, and the yellow line represents the predicted value. Above figure shows the original data graph, model fitted value, and model predicted value of the time series model.

After the analysis of the above four output results and the final prediction, we can know that the data meets the stability requirements, and the prediction is basically considered to be successful.

3. Model Solving

When making decisions on each trading day, make predictions based on all previous data and make corresponding decisions. The following group charts represent the prediction results from less to more, in which the Bitcoin picture is displayed in a global manner, showing the fitting Results; a picture of gold is presented in detail, showing the prediction results. as follows:

ISSN: 2664-9640

Figure 6: Bitcoin (left), Gold (right)

Figure 7: Bitcoin (left), Gold (right)

The blue line represents the true value, the green line represents the fitted value, and the yellow line represents the predicted value.

The above group chart is the forecasting process of Bitcoin and gold from the beginning to the end respectively. This forecast data guides the subsequent investment strategy selection.

There are often bull and bear markets in the stock market, in which the bull market represents the bull market, which represents the good and the stock price goes up. A bear market represents a bear market, which represents a bear market and a falling stock price. Knowing the distribution of the bull market and the bear market will help us formulate reasonable trading strategies.

In the results, we output the probability distribution map of the bull and bear markets of gold and Bitcoin. The probability distribution of the bull market of gold is shown as an example as follows:

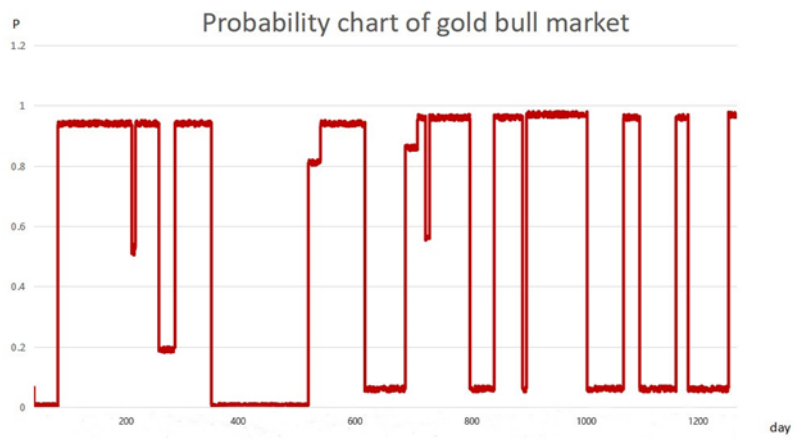

Figure 8: Bull and Bear Market Probability Chart

In order to test the prediction results of the ARIMA algorithm, we use the logistic regression model to test,[10] and input the predicted data and the actual data into the model respectively to test whether the two belong to the same category, so as to verify the accuracy of the predicted data.

Combine the predicted data and the actual data into a new dataset D :

$$D = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N), x_i \subseteq R^n, y_i \in 0,1, i = 1,2, \dots, N \tag{7}$$

If there is a large enough probability P, the part of the model that the classification result of the data set exceeds the probability of P is divided into the first category, and the probability of $1 - P$ is divided into the second category, it means that there is about P data It belongs to one class and proves the correctness of the predicted results.

The minimized loss function for logistic regression is:

$$J(w) = -\frac{1}{n}\left(\sum_{i=1}^{n} \left(y_i \ln p(x_i) + (1 - y_i) \ln\left(1 - p(x_i)\right)\right)\right) \tag{8}$$

We use the stochastic gradient descent method to solve the problem, find the descending direction through the first derivative of $J(w)$ to w, and update the parameters in an iterative manner. The update method is:

$$g_i = \frac{\partial J(w)}{\partial w_i} = (p(x_i) - y_i)x_i \tag{9}$$

$$w_i^{k+1} = w_i^k - \alpha g_i \tag{10}$$

The number of iterations is k. After each parameter update, the iteration can be stopped by comparing $\|J(w^{k+1}) - J(w^k)\|$ less than a threshold or reaching the maximum number of iterations.

The result is input into the logistic regression network for calculation, and the input data has an 94.3% probability of being classified into the first category, and a 5.7% probability of being classified into the second category, that is, the predicted value The difference from the actual value is about 10%, as expected, confirming the accuracy of the forecast data.

4. Conclusion

In order to better assist investors to quantify transactions, we have established a financial product price forecast model based on ARIMA model. First, the ARIMA time series forecasting model is used to forecast based on the data before a certain day, and then the logistic regression

algorithm is used to test the forecasting effect. Finally, the forecasting results is input to the logistic regression model. The test results indicate that our model has high credibility and can assist in investors very well.

References

[1] Weng Fu. Influence of quantified transactions on stock trends [J]. Dynamic analysis of stock market, 2021 (19): 65.
[2] Bao Zhenshan, Guo Junnan, Xie Yuan, Zhang Wenbo. Prediction model of stock price fluctuation based on LSTM-GA [J]. Computer Science, 2020, 47(S1): 467-473.
[3] V. Hryhorkiv, L. Buiak, A. Verstiak, M. Hryhorkiv, O. Verstiak and K. Tokarieva, "Forecasting Financial Time Sesries Using Combined ARIMA-ANN Algorithm," 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), 2020, pp. 455-458, doi: 10.1109/ACIT49673.2020.9208859.
[4] Bollen J, Mao H, Zeng X. Twitter mood predicts the stock market[J]. Journal of Computational Science, 2011, 2(1): 1-8.
[5] Martin and T. Setzer, "On the Robustness of ARIMA-Based Benchmarks for Corporate Financial Planning Quality," 2014 47th Hawaii International Conference on System Sciences, 2014, pp. 1221-1229, doi: 10.1109/HICSS.2014.158.
[6] Song Yuping, Sun Yankun. High-frequency financial time series prediction - ARIMA model based on adaptive filtration method [J]. Journal of Jilin Institute of Technology, 2021,37(02):82-86.DOI :10.19520/j.cnki.issn1674-3288.2021.02.012.
[7] Lu Jingxuan, Dujingnan, Sabah Lasier, EU carbon finance market futures prediction and revelation based on ARIMA model [J]. Coal economic research, 2019,39(10):23-29.DOI: 10.13202/j. cnki. cer.2019.10.004.
[8] Zu Guangxin, Wu Guoliang, Wang Guoliang, Yu Yang. Short-term load forecasting method based on wavelet decomposition and second-order grey neural network model combined with ADF test [J]. Heilongjiang Electric Power, 2021,43(06):476-481+541.DOI:10.13625/j.cnki.hljep.2021.06.002.
[9] Wang Ying. Analysis and prediction of stock prices based on ARMA model [J]. Productivity Research, 2021(09):124-127.DOI:10.19374/j.cnki.14-1145/f.2021.09.024.
[10] Su Ziwen, Liu Aiqin, Liang Huiling, Guo Futao, Zhao Jiayang, Lin Fangfang. Journal of Forest and Environment, 2015,35(04):370-376.DOI:10.13324/j.cnki.jfcf.2015.04.013.