Financial Engineering and Risk Management (2023)
Clausius Scientific Press, Canada

DOI: 10.23977/ferm.2023.060113
ISSN 2523-2576 Vol. 6 Num. 1

Research on the Analysis and Forecast of "Digital Economy" Sector Index Trading Volume Based on Arim (p, q) Model

Chao Liu, Kaijun Li, Shutong Liang, Shixing Han*

College of Engineering, Tibet University, Lhasa, 850000, China
**Corresponding author*

Keywords: Stock, Digital economy, Volume forecast, Arim (p, q) model

Abstract: Stock trading volume is a performance of supply and demand, which refers to the number of transactions in a time unit. The trading volume is an important basis for judging the stock trend, and provides an important basis for analyzing the main behaviors. Therefore, this paper will extract indicators based on the data information of the "digital economy" sector index every five minutes from July 14, 2021 to December 31, 2021, so as to combine ARIMA and multiple linear regression model to obtain the "digital economy" sector index trading volume analysis and prediction model based on ARIM (p, q) model, and then through the comparative analysis of the two models, obtain the best. The example shows that the model has strong applicability and practicability for the prediction of closing price.

1. Introduction

The stock market has the characteristics of high return and high risk, which drives people to explore the volatility of stocks. However, many factors, such as political events and social activities, will affect the changes of stock prices. Predicting the stock trend is a very challenging work [0]. Wang Jingle introduced Markov chain prediction model into stock trading volume [0]. Sun Yanlin and others realized the effective estimation of SV-VOL model through SMC algorithm programming, and found that the trading volume information of the stock market is helpful to the prediction of stock price volatility [0]. Lin Jie et al. established the ARMA model by using the B-J method. With the help of Eviews8 software, they obtained the ARMA (3,2) model that is highly consistent with the actual data. They also predicted the daily trading volume in the next four days and achieved good results [0]. Guo Qinghuang found that the abnormal high (low) trading volume of stocks will affect the return performance in the following months. The investment portfolio with medium turnover and high D.V.R and the investment portfolio with high turnover and low D.V.R can generate 7.71% (annual) premium after 52 weeks of holding [0].

To sum up, it can be seen that the prediction of trading volume and the prediction of trading volume fluctuation are relatively mature, but the "digital economy" sector index trading volume analysis and prediction model proposed in this paper based on ARIM (p, q) model is simpler and faster for the prediction of stock trading volume, and improves the prediction accuracy.

2. Model Establishment

2.1 Establishment of Index Correlation Analysis Model of "Digital Economy" Sector

Pearson correlation coefficient can calculate the correlation between the two parameters Z and Y. Pearson correlation coefficient is widely used to measure the correlation between the two variables, and its value is between - 1 and 1. In this paper, we believe that the turnover, trading volume and opening price in the "digital sector" are the main indicators reflecting the "digital economy" sector. These three indicator variables are simultaneously used with the macro market indicators, domestic stock market indicators Technical indicators and international stock market indicators are analyzed, and the calculation method is as follows:

We set two groups of characteristic parameters $Z\{z_1, z_2, z_3\}$ and $Y\{y_{11}, y_{12}, \cdots y_{ij}\}$, where $i = 1,2,\cdots,5$ represents five categories of macro market indicators, domestic stock market indicators, technical indicators, international stock market indicators and other sector information, and j represents the index in the above five categories of indicators.

$$E(Z) = \frac{\sum_{q=1}^{n} Z_q}{n} \tag{1}$$

$$E(Y) = \frac{\sum_{q=1}^{n} Y_q}{n} \tag{2}$$

The covariance is:

$$Cov(Z,Y) = \frac{\sum_{q=1}^{n} \left(Z_q - E(Z)\right)\left(Y_q - E(Y)\right)}{n} \tag{3}$$

Pearson correlation coefficient is:

$$\rho_{ZY} = \frac{Cov(Z,Y)}{\sigma_Z \sigma_Y} = \frac{\sum_{q=1}^{n} \left(\frac{\left(Z_q - E(Z)\right)\left(Y_q - E(Y)\right)}{\sigma_Z \sigma_Y} \right)}{n} \tag{4}$$

Among them, σ_Z and σ_Y is the standard deviation of Z and Y respectively, where the expression is:

$$\sigma_Z = \frac{\sqrt{\sum_{q=1}^{n} \left(Z_q - E(Z)^2\right)}}{n} \tag{5}$$

$$\sigma_Y = \frac{\sqrt{\sum_{q=1}^{n} \left(Y_q - E(Y)^2\right)}}{n} \tag{6}$$

2.2 Establishment of Trading Volume Multiple Linear Regression Model

Part I: Multiple linear regression model refers to a linear regression model with multiple explanatory variables, which is used to explain the linear relationship between the explanatory variables and other explanatory variables. The mathematical model is:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{xi} + \mu_i \; i = 1,2,\ldots,n \tag{7}$$

Where k is the number of explanatory variables, $\beta_j(j = 1,2,\ldots,k)$ is called regression coefficient. The above formula is also called random expression of population regression function or partial regression coefficient. Its non random expression is:

$$E(Y|X_{1i}, X_{2i}, \ldots, X_{ki}) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{xi} \tag{8}$$

Part II: Explain the part of y change caused by random variable σ As part of the random error, the parameter in the formula is $\beta_0, \beta_1, \ldots, \beta_{\mathrm{v}}$. Both are unknowns of the equation, which can be expressed as partial regression constant and regression constant, so the regression equation of the multivariate linear regression model is:

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{\mathrm{v}} \beta_{\mathrm{v}} \tag{9}$$

Multivariate linear regression uses the domestic market trading volume, information trading volume of the digital economy sector, opening price, closing price, highest price and lowest price analyzed by the results of the first question as independent variables, and the "digital economy" sector index trading volume every five minutes as dependent variables to solve the prediction as a multivariate linear equation.

2.3 Volume Arim (p, q) Model Establishment

ARIMA model (autoregressive moving average model) is often used to fit series whose properties do not change with time, that is, stable time series. ARIMA model is proposed to model, estimate, test and forecast stationary time series. ARIM (p, q) model is a mixture of the two models. AR model is a linear prediction. For a stationary time series y_t. Think y_t.

Related to the previous results, it can be expressed as a p-order autoregressive model, which is recorded as $AR(p)$, and the formula is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t \tag{10}$$

In equation (10), y_t is a stationary time series, β_0 is a constant term, $\beta_i(i = 1,2,\cdots,p)$ is the model parameter of AR model, representing the order of AR model, ε_t Error. If the current result y_t is related to the previous disturbance, it becomes a q order moving average model, recorded as $MA(q)$, and the formula is:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{11}$$

In equation (11), $\theta_i(i = 1,2,\cdots,p)$ is the model parameter of MA model, q represents the order of MA model, ε_t is the error. ARIMA model is the most commonly used model for fitting stationary time series. The value of y_t is not only related to the sequence value of period p, but also related to the disturbance term of period q. The formula is:

$$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{12}$$

In equation (12), $\beta_i(i = 1,2,\cdots,p)$ and $\theta_i(i = 1,2,\cdots,p)$ is the model parameter of ARIMA model $\{\varepsilon_t\}$ Is a white noise sequence, p, q are non negative integers. The d in ARIMA indicates that the time series data needs several order difference differentiation to be stable.

Under the assumption of weak stationarity, the autocorrelation coefficient with interval m is independent of time t, but only related to interval n μ_n means, therefore:

$$\mu_n = \frac{Cov(r_t, r_{t-n})}{\tau_{r_{t-n}}} = \frac{Cov(r_t, r_{t-n})}{\tau_{r_i} \tau_{r_i}} = \frac{r_m}{r_0} \tag{13}$$

That is, covariance and variance stability. It is not difficult to see from this definition that when

k=0, there are:

$$\mu_0 = \frac{r_0}{r_0} = 1 \tag{14}$$

This means that the autocorrelation coefficient of r_t with an interval of 0 is always 1.
In addition:

$$\mu_m = \mu_{-m}, -1 \le \mu_m \le 1 \tag{15}$$

Among them, μ_m is the statistical property of the population. In practice, we can still calculate the statistical property of the sample through limited sample data. If φ_m is the sample statistic corresponding to μ_m, then:

$$\varphi_m = \frac{a_m}{a_0}, a_m = \frac{1}{k}\sum_{t=1}^{k-m}(r_t - \bar{r})(r_{t+m} - \bar{r}) \tag{16}$$

In the above formula, a_m is the sample autocovariance with interval m of r_t; φ_m is the sample autocorrelation coefficient of r_t with an interval of m.

3. Example Verification

3.1 Results of the "Digital Economy" Sector Index Correlation Analysis Model

To analyze macro market indicators, domestic stock market indicators, technical indicators, international stock market indicators. The indicators in the "digital economy" sector information and other sector information have a significant impact on the "digital economy. "For the main indicators related to the sector, first of all, we will make statistics on the data indicators provided. Among them, there are four macro market indicators, namely: stock market: total market price at the end of the period; Hong Kong, China: deposit interest rate of less than 100000 yuan: 12 months: average during the period; proportion of the RMB loan interest rate range of financial institutions: equal to LPR; GDP: current price; 13 international stock market indicators and 12 domestic stock market indicators, respectively: turnover: Shanghai Composite Index; Transaction amount: Shanghai Composite Index; Shanghai Stock Exchange: stock: circulating market value; Shenzhen Stock Exchange: stock: circulating market value, etc; There are 6 pieces of information in the "digital economy" sector, 14 technical indicators, and 5 pieces of information in other sectors, as shown in Table 1:

Table 1 Distribution of Investment Indicators

Index category	Number	Index category	Number
Macro market	4	International stock market	13
Domestic stock market	12	Information of "Digital Economy"	6
Technology	14	Information of other sectors	5

Through correlation analysis, this paper finds that the main indicators related to technical indicators in the "digital economy" sector are the total retail sales of consumer goods and the purchasing manager's index; The main indicators related to the technical indicators of the "digital economy" sector include the market capitalization of stocks, the Shanghai Composite Index, the China Securities 500 Index, the Shanghai A-share Index and the GEM Index; The main indicators related to technical indicators include VMA, VMACD, OBV, BBI, MA, EXPMA, BOLL; It is found that the main indicators related to the "digital economy" sector include the S&P 500 index, the American Stock Exchange index, the London Financial Times 100 index, the CAC40 index in Paris, France, the AEX index in the Netherlands and the Hang Seng index in Hong Kong; The main

indicators related to the information of the “digital economy” sector and other sectors include the “Internet e-commerce” sector index, the “digital twin” sector index and the “Kwai concept” sector index.

3.2 Results of Multiple Linear Regression Model of Trading Volume

3.2.1 Data Analysis

Assuming that the trading volume of every five minutes is Y, the domestic market trading volume is X_1, the highest price is X_2, the lowest price is X_3, the closing price is X_4, and the information trading volume of the digital economy sector is X_5, the multiple regression equation can be obtained as follows:

$$Y = -8.099x_1 + 0.412x_2 + 0.724x_3 + 1.334x_4 + 0.859x_5 \quad (17)$$

R^2 is 0.988, which indicates that the fitting degree of the data is quite good. According to the DW test value, the model has no autocorrelation because the DW value is about 2.

3.2.2 Trading Volume Forecast

By solving the five minute trading volume of the “digital economy” sector index with the multiple linear regression model, we can get some forecast results on January 4, 2022 in Table 2:

Table 2 Trading Volume Prediction Results Based on Multiple Linear Regression Model

Date	Turnover	Date	Turnover
2022/1/4 9:35	73860357.8	2022/1/28 13:40	52359259.17
2022/1/4 9:40	62959895.38	2022/1/28 13:45	46624923.64
2022/1/4 9:45	35062315.23	2022/1/28 13:50	47580574.88
2022/1/4 9:50	54283969.73	2022/1/28 13:55	52191957.87
2022/1/4 9:55	54794178.54	2022/1/28 14:00	43657120.71
2022/1/4 10:00	88849092.87	2022/1/28 14:05	34955591.63
2022/1/4 10:05	42801259.8	2022/1/28 14:10	33161112.08
2022/1/4 10:10	49690503.72	2022/1/28 14:15	49369308.43
2022/1/4 10:15	58427523.81	2022/1/28 14:20	34684072.87
2022/1/4 10:20	32533742.17	2022/1/28 14:25	39529810.82
2022/1/4 10:25	32655009.14	2022/1/28 14:30	46998371.09
2022/1/4 10:30	55491387.28	2022/1/28 14:35	50647258.15
2022/1/4 10:35	31523617.23	2022/1/28 14:40	42117735.4
2022/1/4 10:40	44705134.52	2022/1/28 14:45	30097216.64
2022/1/4 10:45	40143803.32	2022/1/28 14:50	43586716
2022/1/4 10:50	42300360.28	2022/1/28 14:55	40120153.7
2022/1/4 10:55	36452705.6	2022/1/28 15:00	47744603.04

This result can be said to be very close to the real stock price. The prediction model has basically passed the test and the econometric test of the model. The statistical test has passed significantly, and the model establishment effect is relatively good.

3.3 Arim (p, q) Model Solution of Trading Volume

3.3.1 Inspection of Stability

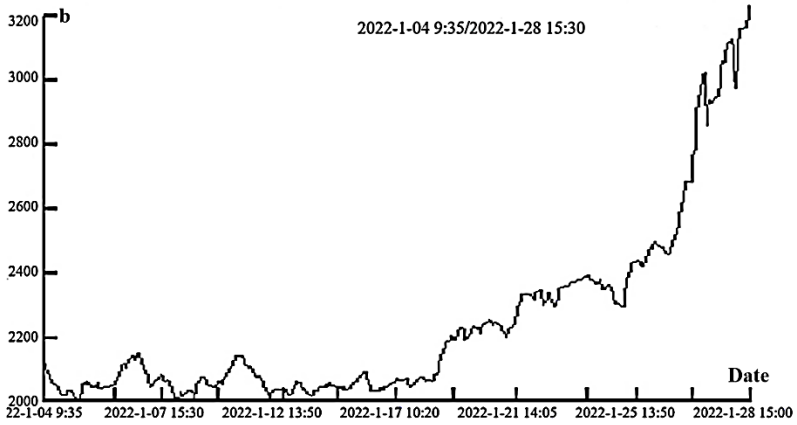


Figure 1 Time Sequence Diagram of Data

The "digital economy" sector index every five minutes from July 14, 2021 to December 31, 2021 is selected as the training set, and the "digital economy" sector index every five minutes from January 4, 2022 to January 28, 2022 is selected as the test set for time series model analysis. Before establishing the prediction model, data needs to be preprocessed, that is, to test whether the series is stable and random. Only through stationarity and randomness tests can the data be further studied and analyzed. Therefore, the data sequence diagram is drawn, and the results are shown in Figure 1:

It can be seen from the time series diagram that the data of each observation day is not stable, and there is an obvious trend of gradual increase with time change, and this trend has a long-term characteristics. Therefore, the data is divided into first-order difference. The results are shown in Figure 2:

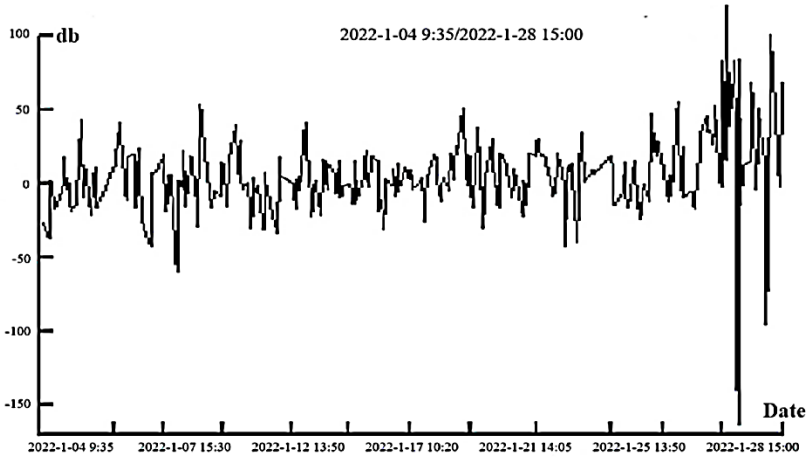
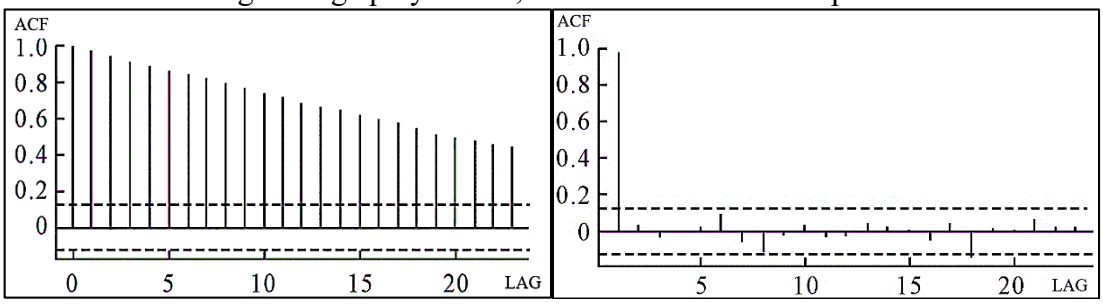


Figure 2 Sequence Diagram after Difference

After the first order difference, the trend of the time series is basically eliminated and becomes stable, all of which fluctuate around 0. It can be judged that the sequence after the first order difference is stable. Therefore, the time series model fitting of the "digital economy" sector index within 5 minutes trading volume can be performed according to the indicators extracted in 3.1.

3.3.2 Determination of Prediction Model

After drawing the difference chart, select the candidate model by analyzing the ACF and PACF charts: that is, judge whether the sample ACF and PACF are tailed or truncated, specify the parameters p and q, where p is the highest order of the autoregressive polynomial and q is the highest term of the moving average polynomial, and then determine the prediction model.



(a) ACF Diagram (b) PACF Diagram

Figure 3 Acf and Pacf Diagram

In order to select the model, it is necessary to determine the parameters p, d and q in ARIMA model. From the previous step, we can get d=1. Since the ACF graph gradually decreases to 0 and the PACF graph decreases to 0 after the first order, ARIMA (1, 1, 0) model can be considered.

3.3.3 Rim (p, q) Model Turnover Forecast

The output result of the drawn Q-Q normal graph is shown in Figure 3. When the points in the data fall on the line in the graph, it means that the normal distribution is satisfied. As can be seen from the figure, the data can be fitted well.The results of the model fit evaluation are shown in Figure 4.

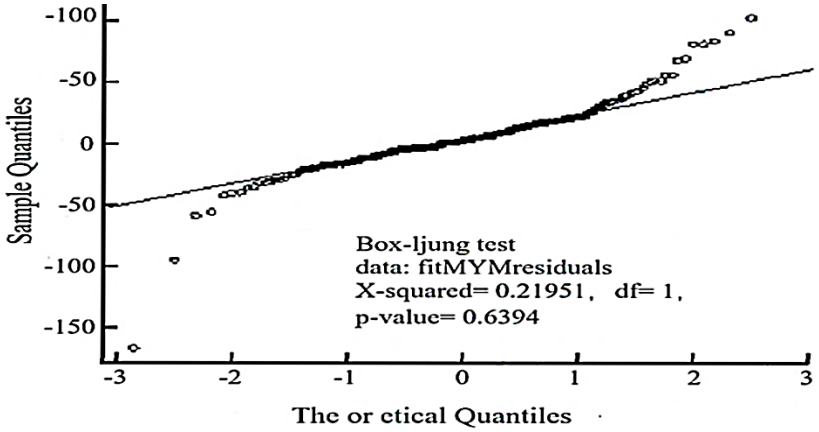


Figure 4 Model Fitting Evaluation

The Box. test function checks whether the autocorrelation coefficients of the residuals are all zero. It can be seen from the results that the P value is 0.6394. If the significance test is not passed, the autocorrelation coefficient of the residual is considered to be zero. It shows that ARIMA model can fit this data well.

The forecast function is used to forecast the trading volume of the “digital economy” sector index every five minutes, and some forecast results on January 4, 2022 in Table 3 are obtained:

Table 3 Trading Volume Forecast Result Data Based on Arim (p, q)

Date	Turnover	Date	Turnover
2022/1/4 9:35	69175346.08	2022/1/28 13:15	40829450.72
2022/1/4 9:40	73364513.29	2022/1/28 13:20	38677512.45
2022/1/4 9:45	98120388.88	2022/1/28 13:25	64895186.89
2022/1/4 9:50	57367019.85	2022/1/28 13:30	50441340.65
2022/1/4 9:55	60465825.52	2022/1/28 13:35	34471680.09
2022/1/4 10:00	72511149.76	2022/1/28 13:40	57820400.98
2022/1/4 10:05	73468317.44	2022/1/28 13:45	40832740.61
2022/1/4 10:10	50833358.73	2022/1/28 13:50	43257951.22
2022/1/4 10:15	76059880.13	2022/1/28 13:55	34684883.62
2022/1/4 10:20	44770319.56	2022/1/28 14:00	41655481.49
2022/1/4 10:25	50189136.31	2022/1/28 14:05	30036883.88
2022/1/4 10:30	49412652.95	2022/1/28 14:10	45855023.17
2022/1/4 10:35	45111973.31	2022/1/28 14:15	51521293.29
2022/1/4 10:40	60971868.78	2022/1/28 14:20	48070456.81
2022/1/4 10:45	52687518.63	2022/1/28 14:25	37279950.71

3.4 Comparison between Arim (p, q) Model and Multiple Linear Regression Model

The above results show that the model performs well and is relatively balanced in volume prediction. The prediction object of this paper is the trading volume of the "digital economy" sector indicator every five minutes. In order to more comprehensively reflect the prediction effect of ARIM (p, q) model on the stock trading volume data, this paper uses two different models to predict the trading volume data every five minutes. The prediction results are compared by calculating the two kinds of errors of MAPE and MAE.

MAPE:

$$MAPE = \sum_{t=1}^{n} \left| \frac{observed_t - predicted_t}{observed_t} \right| \times \frac{100}{n} \tag{18}$$

MAE:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{19}$$

The MAPE and MAE values of the two models are calculated as shown in Table 4:

Table 4 Comparison of Two Models

	ARIM(p,q)	Multivariate linear regression prediction
MAPE	0.026	0.031
MAE	0.819	0.123

Therefore, ARIM (p, q) model is more practical than multiple hieroglyphic regression model in predicting stock turnover.

4. Conclusion

Based on the multiple linear regression model and ARIM (p, q) model, this paper forecasts the trading volume of the "digital economy" sector index every five minutes from January 4, 2022 to January 28, 2022, respectively. Among them, the multiple linear regression model can take the domestic market trading volume, the highest price, the lowest price, the closing price, and the information trading volume of the digital economy sector as independent variables, and then take

the trading volume as a dependent variable to solve; Therefore, ARIMA and multiple linear regression model can be combined to obtain ARIM (p, q) model solution, and then the MAPE and MAE values of the two models can be calculated, so as to obtain the optimal model. At the same time, the volume of every five minutes can be predicted and calculated to obtain the results of the volume prediction.

References

[1] Liu Min, Hou Junhua. Research on the forecast of crude oil stock turnover based on GS-ARIMA model [J]. Foreign Trade and Economic Cooperation, 2021 (12): 38-40.

[2] Wang Jingle. Markov Chain Analysis and Prediction of Stock Trading Volume [J]. Career Circle, 2007 (13): 13-14.

[3] Sun Yanlin, Chen Shoudong, Liu Yang. Prediction of Stock Price Fluctuation Based on Volume Information of Stock Market and Foreign Exchange Market [J]. System Engineering Theory and Practice, 2019, 39 (04): 935-945.

[4] Lin Jie, Zhu Jiaming, Chen Fuyuan. Analysis and prediction of trading volume of stock "Qingdao Haier" based on ARMA model [J]. Journal of Natural Science of Harbin Normal University, 2016, 32 (06): 6-10.

[5] Guo Qinghuang Predictive ability of dynamic changes in trading volume to future stock returns [D]. Nanjing University of Technology, 2014.