ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

Research on ARIMA Based Quantitative Investment Model

Junchao Zhang, Yisheng Huang, Chengbiao Huang, Wei Huang

Guilin University of Technology at Nanning, Chongzuo, Guangxi, China, 532200

Abstract: *The stock market is the most active and main component of the stock market. The operating situation of the stock market is a thumbnail of a country's macroeconomic development. With the continuous development of China's market economy, the increasingly perfect financial market, and the soaring enthusiasm for investment in society, stocks, as a way and means of financing, play an increasingly important role in promoting investment and financing, improving the turnover rate and efficiency of funds, promoting the rational allocation of scarce resources, stabilizing the capital chain of the real economy, and thus improving the welfare of the whole society. It is of theoretical and practical significance to study the relationship between stock turnover and macroeconomic variables, explore the influencing factors of stock turnover, so as to make more accurate analysis and prediction in the future, grasp the market situation more accurately, and better serve the economy and society. Based on the requirements given by the title: the data information given by the "Digital Economy" section every five minutes from July 14, 2021 to January 28, 2022 will be filtered out of the given attachments, and the unnecessary time data will be deleted to obtain a new attachment table. According to the macro market indicators, domestic market indicators, technical indicators, international market indicators 46 sub categories of indicators in the exchange rate category use the clustering method of taking the average value to express the indicators of the month, use the method of multiple linear regression to construct a multiple linear regression model to obtain the correlation coefficient, and analyze to get the more influential indicators.*

Keywords: *Multiple linear regression, ARIMA time series model, SPSS, Least square method*

1. Research background

Quantitative investment refers to a trading method that sends out buying and selling orders through quantitative methods and computer programming to obtain stable income. Investors explore the market operation rules through data analysis and predict the market trend, so as to make decisions and transactions. With the development of big data technology, quantitative investment plays an increasingly important role in the global financial trading market. However, due to the huge amount of market information and the influence of many other factors on the price of products, it is a challenging task to extract effective indicators from massive market information and formulate trading strategies.

This paper mainly studies the data information given by the "Digital Economy" section every five minutes from July 14, 2021 to January 28, 2022. It filters and removes the given attachments, and deletes the unnecessary time data to obtain new data. According to the macro market indicators, domestic market indicators, technical indicators, international market indicators, and 46 subcategories of exchange rate categories, it adopts the clustering method of taking the average value to express the indicators of the month, Using the method of multiple linear regression to construct a multiple linear regression model to get the correlation coefficient, and then analyze to get the more influential indicators.

2. Introduction

The development of securities market plays an important role in the development of national economy. According to the statistics of the National Development and Reform Commission, from 2010 to 2020, the asset scale of China's securities companies presented a volatile situation and continued to expand. In 2019, the total assets of the national securities industry were 7.26 trillion yuan, and the net assets were 2.02 trillion yuan. By the end of 2020, the total assets of the securities industry were 8.90 trillion yuan, and the net assets were 2.31 trillion yuan, up 22.50% and 14.10% year on year respectively.

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

By 2022, the size of our securities market will not decrease, but will increase.

The research on the stock market, especially the stock market, has always been a hot issue in economics. A large number of literatures have studied the stock market. However, from the perspective of research methods, very few of them are purely mathematical modeling. And because of the great current situation of the stock market, most of the current published literature does not have a strong reference.

This paper constructs a multiple linear regression prediction model to predict the trading volume of the "digital economy" sector index every five minutes based on the "digital economy" sector index every five minutes from July 14, 2021 to December 31, 2021 of a stock as the training set, and the "digital economy" sector index every five minutes from January 4, 2022 to January 28, 2022 as the test set. And the actual situation is fitted, and it is found that the model constructed in this paper is in line with the actual situation and has a high practical utilization. And in the stock plate, there is a strong promotion.

3. Establishment of "Digital Plate" Analysis Model of Index Influence

3.1 Data classification

This paper uses 46 indicators ① macro market indicators (purchasing manager index, total retail sales of consumer goods, consumer price index, GDP, RMB deposit interest rate, RMB loan interest rate); ② Domestic stock market indicators (trading volume of Shanghai Composite Index, trading amount of Shanghai Composite Index, total value of stock market, circulation value of stock market (market value of stock circulation), Shanghai Shenzhen 300 Index, Shanghai Composite Index, China Securities 500 Index, GEM Index, Shanghai 50 Index, Shanghai A-share Index, Shenzhen Component Index, Shenzhen Composite Index Kechuang 50 Index); ③ Technical index Bid (VMA, VMACD, ARBR, OBV, BBI, DMA, MA, EXPMA, MTM, MACD, BIAS KDJ、RSI、BOLL）； ④ International stock market indicators (Dow Jones Industrial Index, Nasdaq Composite Index, Standard&Poor's 500 Index, American Stock Exchange Index, USD/RMB exchange rate, Hong Kong Hang Seng Index, Tokyo Nikkei 225 Index, London Financial Times 100 Index, France Paris CAC40 Index, Netherlands AEX Index, Russia RTS Index, Italy MIB Index, Euro/USD exchange rate) Construct multiple linear regression equation for 46 independent variables with trading volume and closing price as dependent variables, and then conduct correlation analysis through ADF test to obtain several more relevant indicators.

The stock market is the most active and main component of the stock market. The operating situation of the stock market is a thumbnail of a country's macroeconomic development. With the continuous development of China's market economy, the increasingly perfect financial market, and the soaring enthusiasm for investment in society, stocks, as a way and means of financing, play an increasingly important role in promoting investment and financing, improving the turnover rate and efficiency of funds, promoting the rational allocation of scarce resources, stabilizing the capital chain of the real economy, and thus improving the welfare of the whole society. In recent years, the emergence of Shanghai Shenzhen Stock Connect, Shanghai Hong Kong Stock Connect and other connectivity mechanisms has made the development of the capital market more perfect, expanded the choices of the majority of shareholders, attracted more potential investors, and provided greater development space for enterprises. The status of stocks as a social hot topic has only increased, but also has to provide more support and impetus for the operation and development of the real economy. This also makes the impact of stock trading on the market and economy more significant, and the stock market has been paid more attention in the process of macroeconomic policy formulation. In this context, it is of theoretical and practical significance to study the relationship between stock trading volume and macroeconomic variables, explore the influencing factors of stock trading volume, so as to make more accurate analysis and prediction in the future, grasp the market situation more accurately, and better serve the economy and society.

3.2 Variable setting and theoretical analysis

The total number of shares traded is the representative of the stock trading volume as the explained variable Y. The total number of shares traded is the number of shares actually traded in each year, and it is also a thermometer to measure the changes in the stock market. Therefore, it is of great significance to measure the overall financial market.

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

CPI as the first explanatory variable X3. The consumer price index is a macroeconomic indicator that reflects the change of household consumption price level, and is also one of the main indicators to measure the inflation rate. Moderate inflation will promote the economy: on the one hand, it will stimulate and activate market transactions, increase enterprise sales and profits, and promote investor confidence; On the other hand, the rise of nominal wage will cause people to have a monetary illusion. When disposable income increases, investment expenditure will increase, and the total number of shares traded will increase. However, excessive inflation will have a negative impact: the production cost of the enterprise will rise significantly, and then the output will be reduced, so the profit may decline, which will reduce the expected rate of return of the stock. In this case, the demand for stock investment will decrease, and the total number of shares traded will also decrease. Therefore, there is an uncertain positive correlation between the total number of shares traded and the consumer price index.

Stock market: the total market price at the end of the period is used as the second explanatory variable X2. Stock market: The total market price at the end of the period is a relative indicator reflecting the changes of the stock market at different time points. The Shanghai Composite Index, which is widely representative and highly convincing, was chosen as the representative of the stock price index. Generally speaking, the stock market: the total market price at the end of the period rose, indicating that the overall development situation of the enterprise was good, and the performance supported its issuance of more stocks. The stock market was favored by investors, and the demand for stocks increased, so the total number of shares traded increased. Therefore, theoretically, there is a positive correlation between the total number of shares traded and the total market price of the stock market at the end of the period.

EPMI index is the third explanatory variable X3. Most of the research on PMI index focuses on its forecasting function, especially on the macro-economy [1.2]. However, the research on the relationship between the stock market (several security indexes) as another barometer of macro-economy and PMI is rare. Norbert J. Ore (1999) pointed out that in the period of no obvious inflation, the trend of PMI index change is obviously consistent with the trend of Dow Jones Index and Standard&Poor's Index, The consistency of change trend can even reach more than 80%; However, in the period of obvious inflation, the prediction effect is not satisfactory [3]. Cai Jinhe and Yu Ying (2011) also pointed out that the manufacturing PMI index tends to synchronize with the Shanghai Composite Index, sometimes showing a leading relationship, and this relationship has been particularly evident since 2007 [4]. In view of the fact that domestic research on the relationship between the PMI index system and the capital market is still lacking, This paper expects to explore the relationship between China's manufacturing PMI index and Shanghai and Shenzhen stock price composite index through empirical analysis.

GDP index as the fourth explanatory variable X4. Under the condition that the system is not established, the changes in the institutional arrangements of the government's functional departments will lead to the relaxation and contraction of liquidity, the expansion and reduction of the scale of securities issuance and the scale of fiscal expenditure, etc. These situations will lead to changes in investment, consumption and exports at the macroeconomic level, which will lead to fluctuations. The stock market practice of countries around the world shows that when it is in the rising stage, if the monetary policy is relaxed and the pressure on the financing of liquid securities is not large, and the government fiscal expenditure is expanded, the stock market will generally have a prosperous scene; When in the declining stage, if the above institutional arrangements still exist, the stock market will not immediately move from prosperity to depression. On the contrary, if the institutional arrangements reduce the government's fiscal expenditure and liquidity and further accelerate the scale of securities financing, even if it is in the rising stage, the extent of the stock market's rise in the short term will be greatly limited, and there will be inflection points in the stock market operation over time, so that it will show a downward trend in the long term bear market) If the stock market has been in the declining process, liquidity will be tightened The institutional arrangement of shrinking fiscal expenditure and expanding the scale of securities financing will lead to an accelerated decline in the stock market. That is to say, there is a clear correlation between volatility and the rise and fall of the stock market [5].
Exchange rate index as the fourth explanatory variable X5 [6-7]. After the reform of the RMB exchange rate formation mechanism, there is a one-way relationship between the exchange rate and the stock price. After the exchange rate reform, a managed floating exchange rate system will be implemented, with reference to a basket of world currencies and no longer a single dollar peg. Since then, the exchange rate of RMB against the US dollar has continued to grow by a small unilateral margin, and it did not retreat until the beginning of 2014, because the withdrawal of the quantitative easing policy of the United States had a certain impact on the value of the RMB. The biggest contribution of this reform is that the exchange rate can largely depend on the power of market supply and demand and the expected role, and is no longer solely controlled by policies. In addition, the gradual implementation of free convertibility

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

under capital account allows foreign capital to flow into the stock market and play an influential role. At the same time, the reform of non tradable shares ended the status quo that state-owned listed companies are dominant. A fair and reasonable pricing mechanism has been established in the stock market. Enterprises no longer try to raise share prices through speculation, but focus more on the company's operation, cash flow and management, and highlight the true value of shares. This is not only conducive to the development of enterprises, but also to fair market competition. More importantly, it makes China's stock market more mature and stable, which is conducive to financial reform and economic development. The effectiveness of the stock market and foreign exchange market has been improved, making the linkage between them clear and obvious, which can better play its role. In the process of rapid economic development in China, a large amount of hot money flows into the stock market and other financial markets, and arbitrage is carried out using asset price differentials and exchange rate price differentials. Due to the rapid development of China, all countries in the world have expectations for the appreciation of the RMB, and the wind of undervaluation of the RMB has been released economically and politically. The more such expectations exist, the more hot money will flow into speculative markets such as the housing market and the stock market, and the greater volatility of the stock market will be boosted. The empirical study also shows that there is a correlation between the exchange rate of RMB against USD and the stock price, as shown in Figure 1.

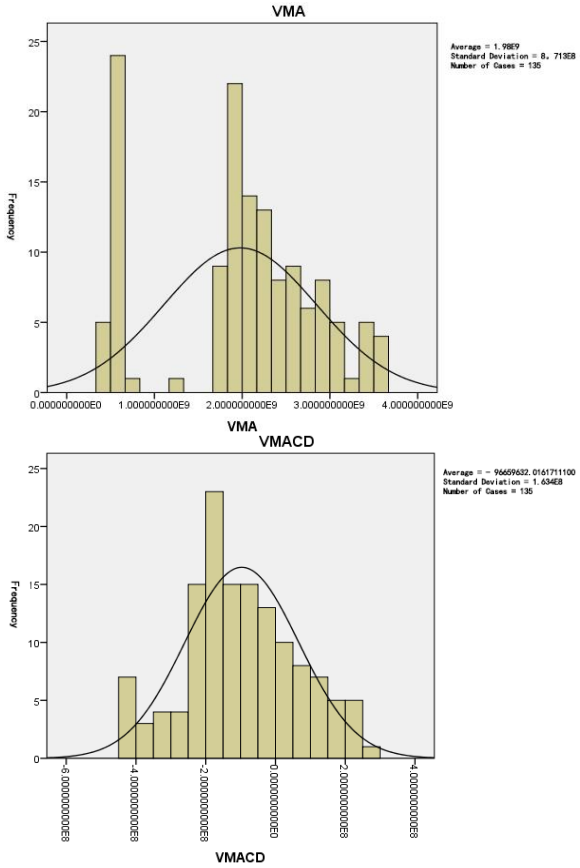

Figure 1: Normal Distribution of Technical Indicators

3.2.1 Establishment of multiple linear regression model

This paper uses multiple linear regression to explain the impact of changes in multiple values on a numerical variable, as follows: CPI, stock market: total market price at the end of the period, EPMI, GDP, exchange rate and technical indicators as independent variables on the opening price, highest price, lowest price and closing price of the "digital economy". The mathematical expression of the multiple linear regression model is:

$$y_t = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + u \tag{1}$$

It explains that variables are not related to each other, but there is no multicollinearity.Suppose there are M observed values y1, y2 , y3 , y4 , x0i , x1i i=1,2…9； n=1,2,3,4 we can get the multiple regression model as follows:

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

$$y_n = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_{1i} x_{1i} + u_n \tag{2}$$

It is expressed as:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 1 & x_1 & x_2 & \cdots & x_{1i} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \cdots \\ b_{1i} \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \tag{3}$$

As shown in Table 1, the least squares principle is still used to calculate the corresponding coefficient matrix B, that is, the optimal coefficient is Minimum According to the first order condition,

$$\sum u_n^2 - \sum \left(y_n - \left(\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_{1i} x_{1i} \right)^2 \right)$$

we have the optimal parameter

$$\begin{cases} \dfrac{\partial \sum \left(y_n - \left(\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_{1i} x_{1i} \right)^2 \right)}{\partial b_0} = 0 \\ \dfrac{\partial \sum \left(y_n - \left(\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_{1i} x_{1i} \right)^2 \right)}{\partial b_1} = 0 \\ \cdots \\ \dfrac{\partial \sum \left(y_n - \left(\hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \cdots + \hat{b}_{1i} x_{1i} \right)^2 \right)}{\partial b_{1i}} = 0 \end{cases} \tag{4}$$

Table 1: According to SPSS analysis, the indicators with strong correlation are obtained

coefficient						
Model			Non standardized coefficient	Standardization coefficient	t	Significance
		B	Standard error	Beta		
1	Constant	1553.937	4.469		347.719	0.000
	MACD	2.863	0.267	0.712	10.722	0.000
2	Constant	4691.260	449.980		10.425	0.000
	MACD	3.096	0.226	0.769	13.687	0.000
3	exchange rate	488.737	70.096	-0.392	6.972	0.000
	Constant	5008.221	432.066		11.591	0.000
	MACD	1.128	0.553	0.280	2.041	0.044
	exchange rate	538.554	67.327	-0.432	7.999	0.000
	DMA	0.701	0.181	0.536	3.861	0.000
4	Constant	5284.807	346.198		15.265	0.000
	MACD	-1.431	0.544	-0.356	-2.629	0.010
	exchange rate	581.671	53.947	-0.466	10.782	0.000
	DMA	1.410	0.170	1.079	8.310	0.000
	MTM	0.560	0.070	0.416	8.004	0.000

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

According to the above formula, it can be concluded that $\hat{B}=\left(x^{'}x\right)^{-1}x^{'}x$ standardized regression coefficient or partial regression coefficient in multiple linear regression model. For an explanatory variable, the significance of the pre regression coefficient is that when the relationship with other analytic variables remains unchanged, each unit of this variable changes, the size of the average partial regression coefficient between the analyzed variables will change, and the multivariate linear regression model will be sampled for numerical calculation and data analysis, as shown in Figure 2.

The relevant multiple linear regression model is obtained: $y_1=5284.807-1.431x_{10}-581.671x_5+1.41x_6+0.56x_9$

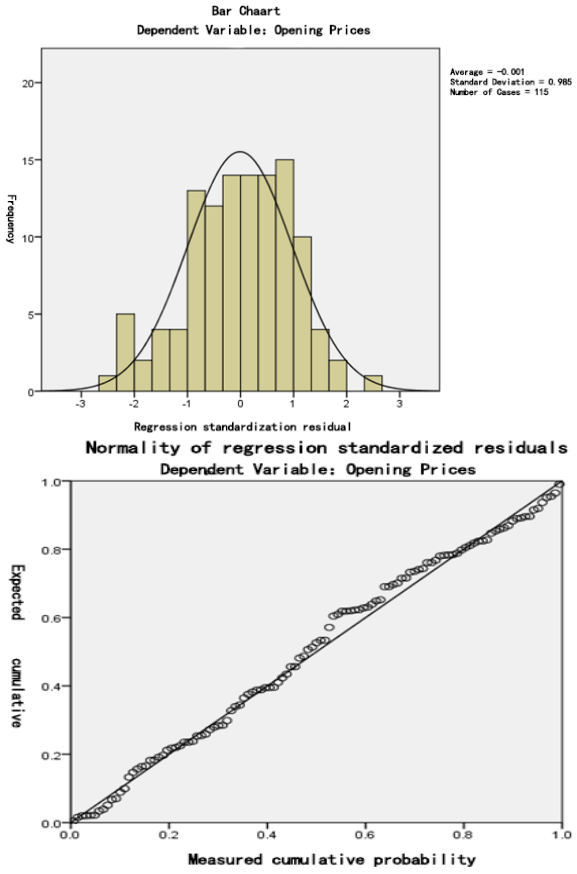

Figure 2: Histogram

Similarly, we can see that:

$$y_2=5309.316-1.43x_{10}-585.66x_5+1.41x_6+0.558x_9$$
$$y_3=5297.459-1.431x_{10}-583.57x_5+1.411x_6+0.558x_9$$
$$y_4=5296.176-1.432x_{10}-583.496x_5+1.410x_6+0.559x_9$$

$$\begin{cases} y_1=27256.384-0.078X_1-243.655X_2-0.002X_3-39.691X_4 \\ y_2=27198.856+0.077X_1-242.954X_2-0.002X_3-39.635X_4 \\ y_3=27118.447-0.077X_1-242.224X_2-0.002X_3-39.646X_4 \\ y_4=27092.370-0.077X_1-241.966X_2-0.002X_3-39.466X_4 \end{cases} \quad (5)$$

The selected indicators are:CPI (X1), stock market: total market price at the end of the period (X2), EPMI (X3), GDP (X4), MACD (X10), exchange rate (X5), DMA (X6), MTM (X9).

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

4. Establishment of multiple linear regression prediction model

4.1 Data processing

Based on the obtained indicators, this paper constructs a multiple linear regression prediction model to predict the trading volume of the "digital economy" sector index every five minutes from July 14, 2021 to December 31, 2021 as the training set, and the "digital economy" sector index every five minutes from January 4, 2022 to January 28, 2022 as the test set.

4.2 Establishment of multiple linear regression model

Multivariate linear regression is selected to explain the impact of changes in multiple values on a numerical variable, as follows: MACD (xa), exchange rate (xb), DMA (xc), MTM (xd) indicators are independent variables on the trading volume of the "digital economy".

The mathematical expression of multivariate linear regression model is:

$$y_t = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_k x_k + u \tag{6}$$

It explains that variables are not related to each other, but there is no multicollinearity.

Suppose there are M observation values $\left(y_m, x_a, x_b, x_c, x_d\right)$, we can get the multiple regression model as follows:

$$y_m = b_0 + b_1 x_a + b_2 x_b + b_3 x_c + b_4 x_d + u \tag{7}$$

The least squares principle is still used to calculate the corresponding coefficient matrix B, that is, the optimal coefficient is $\sum u_m^2 - \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)$, Minimum According to the first order condition, the optimal parameter $\hat{\beta}$ satisfies.

$$\begin{cases} \dfrac{\partial \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)}{\partial b_0} = 0 \\ \dfrac{\partial \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)}{\partial b_1} = 0 \\ \dfrac{\partial \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)}{\partial b_2} = 0 \\ \dfrac{\partial \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)}{\partial b_3} = 0 \\ \dfrac{\partial \sum\left(y_m - \left(\hat{b}_0 + \hat{b}_1 x_a + \hat{b}_2 x_b + \hat{b}_3 x_c + \hat{b}_4 x_d\right)^2\right)}{\partial b_4} = 0 \end{cases} \tag{8}$$

According to the above formula, it can be concluded that $\hat{B} = \left(x'x\right)^{-1} x'x$ standardized regression

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

coefficient or partial regression coefficient in multiple linear regression model. For an explanatory variable, the significance of the pre regression coefficient is that when the relationship with other analytic variables remains unchanged, each unit of this variable changes, the size of the average partial regression coefficient between the analyzed variables will change, and the multivariate linear regression model will be sampled for numerical calculation and data analysis, as shown in Table 2 and Figure 3.

Table 2: Analysis according to SPSS

coefficient						
Model			Non standardized coefficient	Standardization coefficient	t	Significance
		B	Standard error	Beta		
1	Constant	-84885832990.0	10923572960.00		-7.771	0.000
	US: USD to RMB	13558935200.00	1701641112.000	0.676	7.968	0.000
	DMA	1486728.984	8175507.323	0.079	0.182	0.856
	MACD	-2893812.548	26786498.250	-0.050	-0.108	0.914
	MTM	4649444.302	2515769.165	0.233	1.848	0.068
a. Dependent variable: turnover						

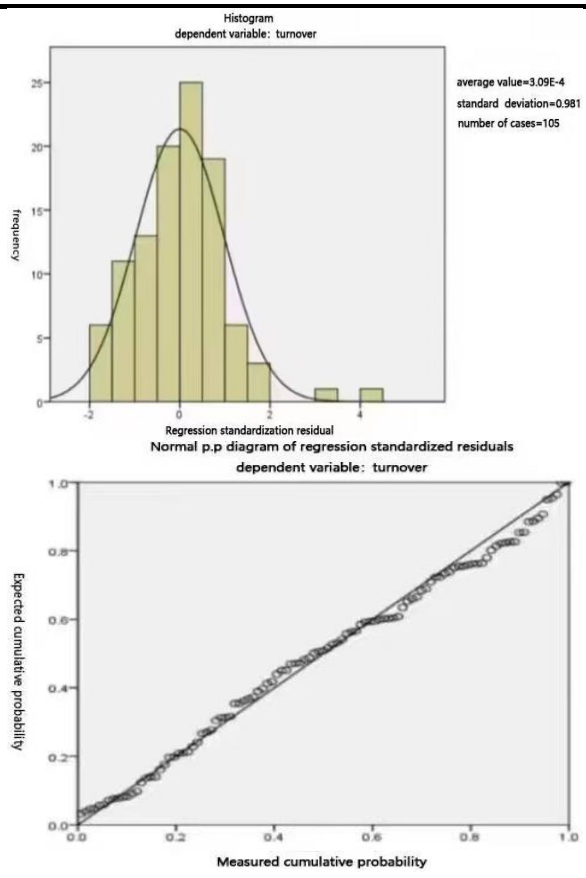

Figure 3: Normal Distribution of Technical Indicators

The multiple linear regression model is obtained:

$$y_m = -84885832985.678900 + 13558935201.517424x_a + 1486728.984202x_b + 2893812.548287x_c + 4649444.302215x_d \quad (9)$$

ISSN 2616-5902 Vol. 4, Issue 17: 54-62, DOI: 10.25236/AJBM.2022.041708

5. Conclusion

By using the selected technical indicators as the independent variables to construct the multiple linear regression equation, taking the trading volume as the dependent variable, and based on the ADF test, testing the correlation between the independent variables of the indicators and the influence of the dependent variables, seven more appropriate indicators are selected, namely, CPI, EPMI, GDP, exchange rate, MACD, DMA, MTM.

Based on the "digital economy" sector index every five minutes from July 14, 2021 to December 31, 2021 as the training set, and the "digital economy" sector index every five minutes from January 4, 2022 to January 28, 2022 as the test set, a multiple linear regression prediction model is constructed to predict the trading volume of the "digital economy" sector index every five minutes. And the model is fitted to the actual situation, which is in line with the actual situation and has a high practical utilization.

References

[1] Lu Jianjun. Research on the Development of PMI Index System in Hangzhou [D]. Hangzhou: Zhejiang University, 2005: 89

[2] Wang Yacan, Chen Qiong, Ru Yihong. On the Indicative Function of Purchasing Manager Index to Economic Operation [J]. Statistics and Decision Making, 2006 (22): 54-56

[3] Zhang Libin, Feng Yi, Liu Longfei, et al. The predictive role of manufacturing PMI on GDP trend [J]. Journal of Central South University for Nationalities: Natural Science Edition, 2012, 31 (3): 131-134

[4] Cai Jin, Yu Ying. Interpreting PMI: Leading the Market [M]. Beijing: Chemical Industry Press, 2011

[5] He Da'an, Su Zhihuang. Institutional arrangement, GDP fluctuation and stock market fluctuation [J] Academic Monthly, 2013, 45 (07): 90 - 99. DOI: 10.19862/j.cnki.xsyk.2013.07.11

[6] Zhou Yun. Empirical Study on the Linkage between RMB Exchange Rate and Stock Price [D]. Southwest University of Finance and Economics, 2014

[7] Zhang Wentong, Yan Jie. A Basic Course of SPSS Statistical Analysis [M]. Beijing: Higher Education Press, 2010