

Research on the Returns of SSE STAR Market Chip Index Based on the ARMA Model

Yuqing Wu*

School of Statistics and Mathematics, Central University of Finance and Economics, Beijing,
100098, China

*Corresponding author: 2019310842@email.cufe.edu.cn

Abstract. Since 2021, China's semiconductor industry and chip industry have been in the hot stage. While the scale of the two industries has continued to expand, there are also problems of bubble economy and excessive competition. In the first half of 2022, the stock market has entered an adjustment cycle. In order to predict the returns trend of the adjustment cycle and help investors make reasonable decisions. This paper uses AIC and BIC criteria to determine the order of the model, test the significance of the model and its parameters, then build the ARMA(2,1) model to predict the closing price of the SSE STAR Market Chip Index from January 2nd, 2020 to February 21st, 2022. According to the results, this paper finds out that both fitting effect and forecasting effect are good. The index will maintain a downward trend in the next 30 periods. Through the above empirical analysis, this paper puts forward two suggestions: firstly, in the face of industry differentiation, people need to invest rationally; secondly, semiconductor companies and chip companies should attach importance to mergers and acquisitions.

Keywords: ARMA model; Time series; Chip index

1. Introduction

With the advent of the digital economy era, the diversification of consumer demand has led the integrated circuit industry chain to expand into digital and intelligent application fields. The industry chain covers semiconductor material processing, chip manufacturing, network communications and other popular industries. In recent years, it is increasingly valued by investors because it is not only the frontier industry of the information technology industry, but also an emerging industry that highlights national strength, promotes economic development and ensures national security [1]. The semiconductor industry is the hottest in this industry chain, because semiconductor chips involve various downstream terminals such as automobiles, communications, medical care, and consumer electronics. Almost every electronic device in our daily life is inseparable from the application of semiconductors. With continuous popularity of new energy vehicles, automotive chips have become an important growth opportunity for semiconductor production. The demand of chips outstripping supply and the profit margin of the industry remaining huge both make semiconductor investment the focus of the market [2].

However, in the first half of 2022, the semiconductor industry has entered a downward cycle. The enthusiasm of consumers' purchase has widened the gap between industries, showing a situation of uneven popularity. For example, new energy vehicles and energy storage related industries are still improving steadily, while the chips involved in smart home appliances are affected by the decline in the real estate industry and the demand volume fluctuates violently. In general, the market growth rate and dividends are decreasing. External reasons include repeated epidemics, inflation, wars, etc., and internal reasons include high valuations and blind competition in relevant industries, forming a bubble economy. The growth of returns has slowed down and there generates a future trend of industry reshuffle.

This paper believes that observing an industry-representative index can reflect the long-term fluctuations of the market, predict the income trend of the adjustment cycle and help some investment institutions to slow down, which is of great significance for making investment decisions.

With the rapid development of the securities market, scholars have conducted extensive research on the forecasting of individual stocks, industry indices, and funds. Among them, there are many

applications of time series models for financial time series forecasting. These statistical models include Auto Regressive Moving Average model (ARMA) [3], Autoregressive Integrated Moving Average Model (ARIMA) [4], Generalized Auto Regressive Conditional Heteroskedasticity model (GARCH), ARMA-GARCH [5], etc., which are all based on the relationship between current and historical value.

In terms of stock price, Yang and Zhang modeled the opening price of Volkswagen public stock with small prediction error, indicating the ARIMA model could provide stock investment decision-making for relevant institutions [6]. Sun and Zhou used the BIC principle to determine the ARIMA model's order and predict the closing price of China Merchants Bank in the next 40 periods [7]. In terms of indexes, Li et al., established ARIMA model to forecast the Shanghai Composite Index closing price of the last two months in 2016, the fitting effect of which was very significant [8]. K Murali et al., established the ARIMA(2,1,2) model to predict the closing BSE Index in the next 12 months, expounding the excellent forecasting effect of the ARIMA model [9]. Paulo Rotela Junior et al., used AR(1) model, ARIMA(0,2,1) model and exponential smoothing model to fit the Brazilian stock market Index respectively. The MAPE of AR(1) is 0.052%, which is the lowest when compared to other options [10].

In the semiconductor-related industries field, the existing literatures mainly include research on the valuation of specific semiconductor and chip companies and modeling the prosperity index of the semiconductor industry. There is no time series analysis of the comprehensive index of the semiconductor industry and chip industry. Therefore, the innovation of this paper is to analyze the SSE STAR Market Chip Index to explore the returns when the relevant markets are down.

This paper uses ARMA(2,1) model to predict the closing price of the SSE STAR Market Chip Index in order to explore the future trends of the semiconductor industry and the chip industry during the adjustment period from a comprehensive perspective. The research steps are as follows: firstly, use the AIC and BIC criteria to determine the order of the ARMA model; secondly, perform a significance test on the model and its parameters to obtain the predicted value of the closing price in the next 30 periods; finally, compare with the actual value to analyze fitting and forecasting effect. This paper believes that the ARMA model has excellent properties in predicting the short-term yield of the chip index. In the short term, the demand side of the semiconductor industry and the chip industry will continue to tighten, and the lack of enthusiasm for consumers' purchase will not be eased soon. In the face of industry differentiation, investors need to return to rationality and semiconductor-related companies should pay attention to mergers and acquisitions.

The following framework of this paper is: data and methodology, results and conclusion.

2. Data and Methodology

2.1 Data Interpretation

The data studied in this paper comes from Wind. The selected research object is the closing price of the SSE STAR Market Chip Index (Index code: 000685) from January 2nd, 2020 to February 21st, 2022, with a total of 516 pieces of data, denoted as y_t . The training set is the closing price from January 2nd, 2020 to December 31st, 2021, and the test set is from January 2nd, 2022 to February 21st, 2022, during which the semiconductor industry is undergoing the adjustment cycle. This index selects securities related to semiconductor materials and equipment, chip design, chip manufacturing, chip packaging and testing from STAR Market companies as index samples, which can reflect the overall performance of the representative securities of companies in the chip industry.

This paper then calculates the linear returns and compounded returns of the series and obtains the following data characteristics. As shown in table 1, the maximum and minimum values of the linear returns are 12.480% and -18.549% respectively, showing a right-skewed distribution while the maximum and minimum values of the compounded returns are 11.761% and -20.516% respectively, showing a left-biased distribution. Besides, both returns show steep spikes.

Table 1. Characteristics of returns

Type	Mean	Std	Max	Min	Kurt	Skew
Linear returns	0.11%	0.029	12.48%	-18.55%	5.502	0.096
Compounded returns	0.07%	0.029	11.76%	-20.52%	6.714	-0.258

2.2 Method Introduction

2.2.1 Stationarity Test

Before building a time series model, it is necessary to confirm that the series is stationary. There are two main stationarity test methods and both are introduced below [11].

Graph test method: Stationary time series have constant mean and variance, which means that the time series diagram should show that the series always fluctuates around a constant value, without obvious trend and periodicity. In addition, the autocorrelation plot of stationary time series usually decays to 0 in the short term.

DF test/ADF test: The DF test was proposed by statisticians Dickey and Fuller, which is only suitable for testing AR(1) model, while the ADF test can test the AR(p) model. Both of them are based on the relationship between stationary series and the unit roots, that is, the stationarity of the series is judged by checking whether there are unit roots on or outside the unit circle.

2.2.2 White noise test

If the time series are not completely disordered random fluctuations, it is meaningful to continue modeling. As a result, it needs to pass the white noise test next. According to Bartlett's theorem, the autocorrelation coefficient of the non-zero delayed period of a white noise sequence with n observation periods $\{x_t, t = 1, 2, \dots, n\}$ approximately obeys the normal distribution with mean value of zero and variance of $1/n$, which can be written as $\hat{\rho}_k \sim N\left(0, \frac{1}{n}\right), \forall k \neq 0$.

From this, it is easy to obtain the hypotheses for testing the white noise sequence, which are $H_0: \rho_1 = \rho_2 = \dots = \rho_m = 0, \forall m \geq 1$ and $H_1: \text{at least one } \rho_k \neq 0, \forall m \geq 1, k \leq m$.

In order to test this joint hypothesis, the Q statistic in the case of large samples and the LB statistic in the case of small samples can be constructed, both of which approximately obey the chi-square distribution with m degrees of freedom, namely:

$$Q = n \sum_{k=1}^m \hat{\rho}_k^2 \sim \chi^2(m) \quad (1)$$

$$LB = n(n+2) \sum_{k=1}^m \left(\frac{\hat{\rho}_k^2}{n-k} \right) \sim \chi^2(m) \quad (2)$$

If the statistic is greater than the $1 - \alpha$ quantile of the chi-square distribution with m degrees of freedom, the sequence will be considered as a non-white noise sequence.

2.2.3 Definition of the ARMA(p,q) model

The ARMA model was jointly proposed by Box and Jenkins, also known as the Auto Regressive Moving Average model, and has the following structure:

$$\begin{cases} x_t = \phi_0 + \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \text{Var}(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(x_t \varepsilon_s) = 0, \forall s < t \end{cases} \quad (3)$$

This model is also called a centralized ARMA(p,q) model when $\phi_0 = 0$. It can be abbreviated as the following formula:

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (4)$$

After introducing the backshift operator, the ARMA (p, q) model can be written as $\Phi(B)x_t = \Theta(B)\varepsilon_t$.

The regression coefficient polynomial of p orders and the moving average coefficient polynomial of q orders are written respectively as $\Phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ and $\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$.

2.2.4 Significance test of the model

After using AIC and BIC criteria to determine the order of the model, the validity of the model needs to be checked and it is necessary to observe whether the extracted information is already sufficient. The fitted residual sequence should be a white noise sequence.

2.2.5 Significance test of parameters

In order to simplify the model, it is also necessary to perform a significance test on the parameters and then retain the parameters that have a significant impact on the dependent variable. The null and alternative hypotheses are $H_0: \beta_j = 0, \forall 1 \leq j \leq m$ and $H_1: \beta_j \neq 0, \forall 1 \leq j \leq m$.

Due to the assumption of normal distribution, the least squares estimate $\hat{\beta}_j$ of the j^{th} position parameter follows a distribution with mean value of 0 and variance value of $a_{jj}\sigma_\varepsilon^2$, where a_{jj} is the j^{th} element on the diagonal of matrix $(X'X)^{-1}$. Then t-test statistic can be constructed as the following formula:

$$T = \sqrt{n-m} \frac{\hat{\beta}_j}{\sqrt{a_{jj}Q(\tilde{\beta})}} \sim \chi^2(n-m) \quad (5)$$

If the absolute value of the statistic is greater than the $1 - \frac{\alpha}{2}$ quantile of the t distribution with $n - m$ degrees of freedom, the null hypothesis will be rejected. That is to say, the parameter is considered to be significantly non-zero and the model should retain this parameter.

2.2.6 Forecast

Next, Box and Jenkins use linear minimum variance forecast to estimate the value of a future period and obtain a 95% confidence interval. In the case of the ARMA(p,q) model. The predicted value of each period can be written as:

$$\begin{aligned} x_t(l) &= E(\phi_0 + \phi_1 x_{t+l-1} + \dots + \phi_p x_{t+l-p} + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \dots - \theta_q \varepsilon_{t+l-q} | x_t, x_{t-1}, \dots) \\ &= \begin{cases} \phi_0 + \phi_1 \hat{x}_t(l-1) + \dots + \phi_p \hat{x}_t(l-p) - \sum_{i=1}^q \theta_i \varepsilon_{t+l-i}, & l \leq q \\ \phi_0 + \phi_1 \hat{x}_t(l-1) + \dots + \phi_p \hat{x}_t(l-p), & l > q \end{cases} \end{aligned} \quad (6)$$

In this equation, $\hat{x}_t(k) = \begin{cases} \hat{x}_t(k), & k \geq 1 \\ x_{t+k}, & k \leq 0 \end{cases}$.

The variance of the predicted value is:

$$Var[e_t(l)] = (G_0^2 + G_1^2 + \dots + G_{l-1}^2) \sigma_\varepsilon^2 \quad (7)$$

Finally, it is common to calculate the average error to check the accuracy of the model and make appropriate decisions according to the future trend of the time series.

3. Results

3.1 Preprocessing of Time Series

3.1.1 Stationarity test

First, this paper observes the stationarity of the index closing price time series by graph test method. Figure 1 shows that the time series diagram of y_t has no significant non-stationary characteristics, that is, there is no obvious trend and periodicity. However, figure 2 shows that the autocorrelation coefficient of y_t falls within twice the standard deviation after 30 lags and the decay rate is slow, indicating that the autocorrelation coefficient has a constant correlation, which is the characteristics of a non-stationary series. So, it is difficult to judge the stationarity of this series based on the above.

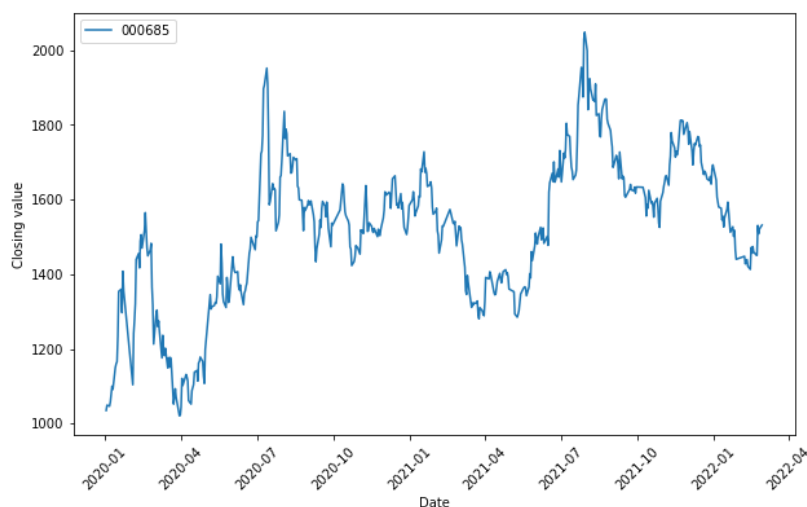


Figure 1. The time series diagram of the chip index

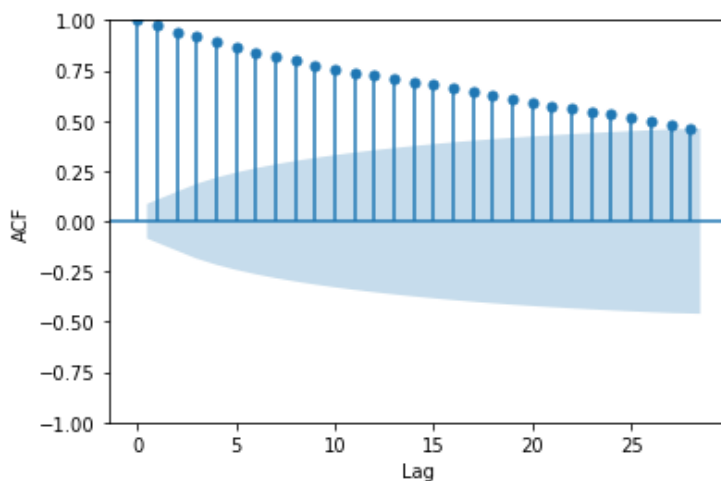


Figure 2. The autocorrelation coefficient of the chip index

It is highly subjective to judge the stationarity of the series only by the graph test, so this paper is going to carry out the ADF test next. As shown from table 2 that the t statistic of the ADF test is -2.881 and the p value is significant, which means that null hypothesis is rejected at the 5% confidence level, that is, the series is considered stationary. The test formula obtained according to the coefficient output results in table 2 is:

$$\Delta \hat{Y}_t = -0.277Y_{t-1} + 43.444 \quad (8)$$

Table 2. Results of ADF test

	Variable	Coefficient	Std. Error	t-Statistic	Prob.
Parameters	Y(-1)	-0.277	0.010	-2.881	0.004
	C	43.444	14.744	2.947	0.003
ADF value	t-Statistic	Prob.	1% level	5% level	10% level
	-2.881	0.048	-3.444	-2.868	-2.570

3.1.2 White noise test

The Q statistic and LB statistic of lag 6 and their p-values were calculated respectively (table 3). The results significantly rejected the null hypothesis that the sequence was a white noise sequence and considered the sequence to be a non-white noise sequence.

Table 3. Results of white noise test

Lag	Q statistic	Prob.	LB statistic	Prob.
1	458.106	0.000	460.940	0.000
2	889.439	0.000	895.837	0.000
3	1297.228	0.000	1307.848	0.000
4	1680.543	0.000	1695.934	0.000
5	2042.239	0.000	2062.893	0.000
6	2385.370	0.000	2411.744	0.000

3.2 Analysis of the ARMA Model

3.2.1 Model order determination

Next, this paper draws the partial autocorrelation coefficient diagram of the sequence and then combines figure 2 with figure 3. The autocorrelation diagram shows the characteristics of tailing and the partial autocorrelation diagram shows the characteristics of second-order truncation. So, this paper considers using AR(2), ARMA(2,1) and ARMA(2,2) to fit the data respectively.

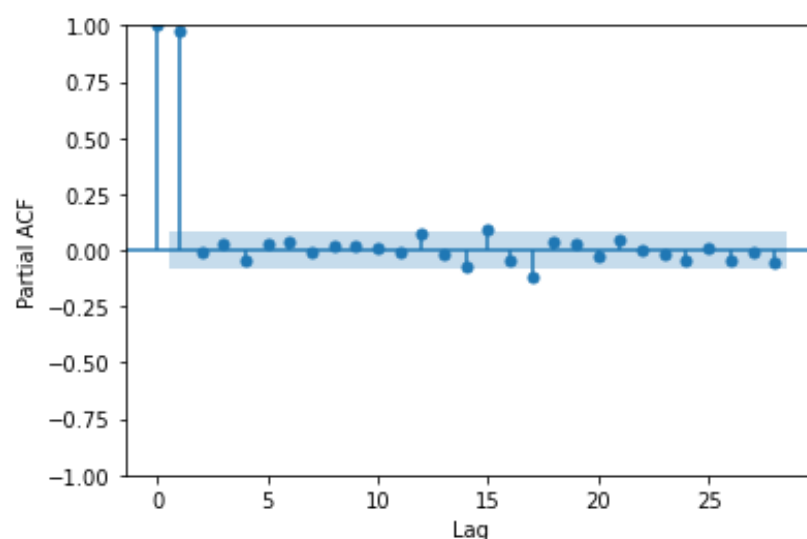


Figure 3. The partial autocorrelation coefficient of the chip index

3.2.2 Significance test of the model

This paper then compares the AIC and BIC values of AR(2), ARMA(2,1), and ARMA(2,2) and determine the order of model. Table 4 shows that the AIC value of ARMA (2,1) model is 5060.544, which is the smallest among the three models and the BIC value is very similar to that of AR(2) model. Therefore, ARMA(2,1) model is finally used to fit the time series.

Table 4. The AIC and BIC value of each model

Model	AR(2)	ARMA(2,1)	ARMA(2,2)
AIC	5063.503	5060.544	5062.472
BIC	5080.248	5080.475	5087.589

As shown in figure 4(c), the white noise test results of the residual sequence show that the p-values of each order of delay are all higher than the dotted line position, that is, greater than 0.05. Figure 4(d) shows the residual series are densely distributed around the diagonal and are considered to be approximately normally distributed. In summary, the ARMA(2,1) model is significantly effective.

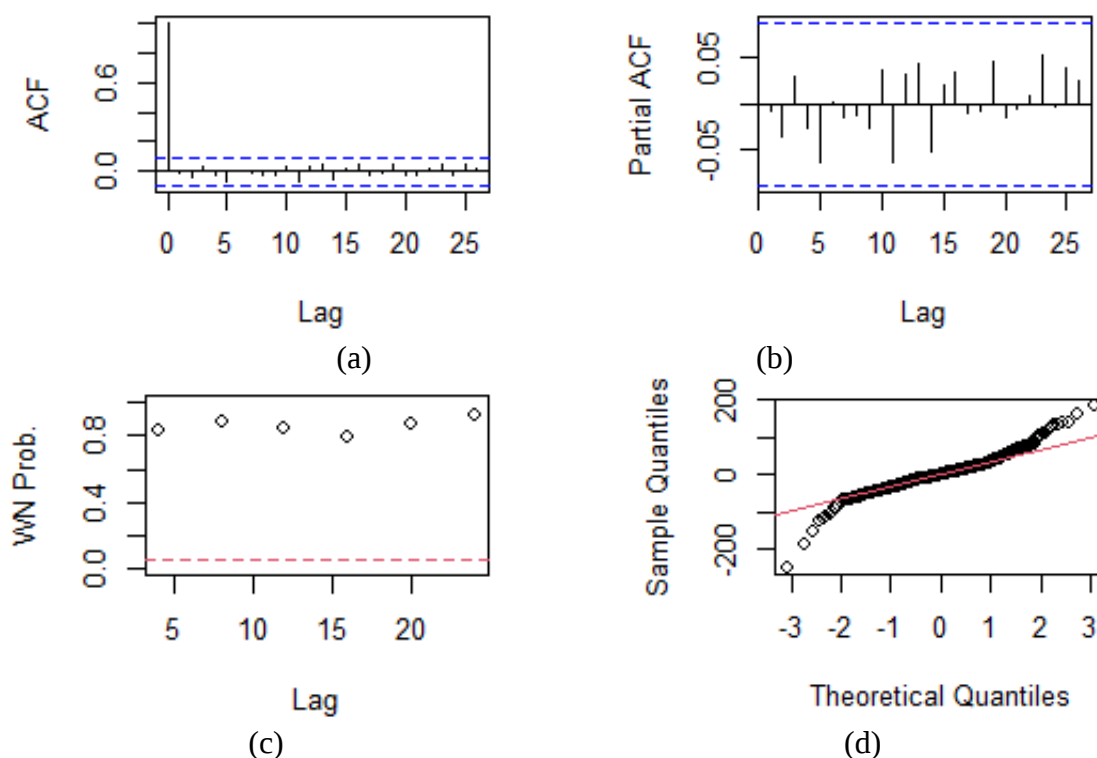


Figure 4. Significance Test of the Model

3.2.3 Significance test of parameters

This paper then estimates the coefficients of the ARMA(2,1) model. As shown in table 5, the estimated coefficient of AR(1) is 0.009, the estimated coefficient of AR(2) is 0.953, and the estimated coefficient of MA(1) is 0.999. Except for AR(1), the remaining p-values are significant and the fitting effect is good. So, the final expression of the model is:

$$y_t = 1517.593 + 0.009\Delta y_{t-1} + 0.953\Delta y_{t-2} + \varepsilon_t - 0.999\varepsilon_{t-1} \quad (9)$$

Table 5. Results of Significance Test of Parameters of ARMA(2,1)

Variable	Coefficient	Std. Error	z	Prob.
C	1517.593	81.429	18.637	0.000
AR(1)	0.009	0.017	0.542	0.588
AR(2)	0.953	0.018	53.444	0.000
MA(1)	0.999	0.082	12.214	0.000
Sigma ²	1889.165	166.821	11.325	0.000

3.3 Forecast

3.3.1 Forecast of closing price

The predicted values are calculated according to the final expression of the model, as shown in table 6:

Table 6. 30-period forecast

l	$x_{486}(l)$	l	$x_{486}(l)$	l	$x_{486}(l)$
1	1694.469	11	1663.081	21	1637.275
2	1686.434	12	1656.810	22	1632.371
3	1687.690	13	1657.508	23	1632.693
4	1680.045	14	1651.539	24	1628.023
5	1681.171	15	1652.149	25	1628.287
6	1673.897	16	1646.467	26	1623.839
7	1674.903	17	1646.996	27	1625.050
8	1667.981	18	1641.586	28	1619.814
9	1668.876	19	1642.040	29	1619.975
10	1662.288	20	1636.890	30	1615.940

Finally, the variance of the predicted values is calculated to obtain the 80% and 95% confidence intervals. It can be seen from table 7 that the error values are all at a low level and fitted values are very close to the real values. As is shown in figure 5, the 30-period forecast values completely fall within the 80% confidence interval and almost fall within the 95% confidence interval. The model predicts that the index will continue to decline sharply in the future, which is consistent with the actual trend, but the decline in the predicted values is smaller than the actual values. In general, the prediction effect of the model is good.

Table 7. Results of the prediction

Error type	ME	RMSE	MAE	MPE	MAPE
Value	1.620	43.445	30.600	0.033	2.046

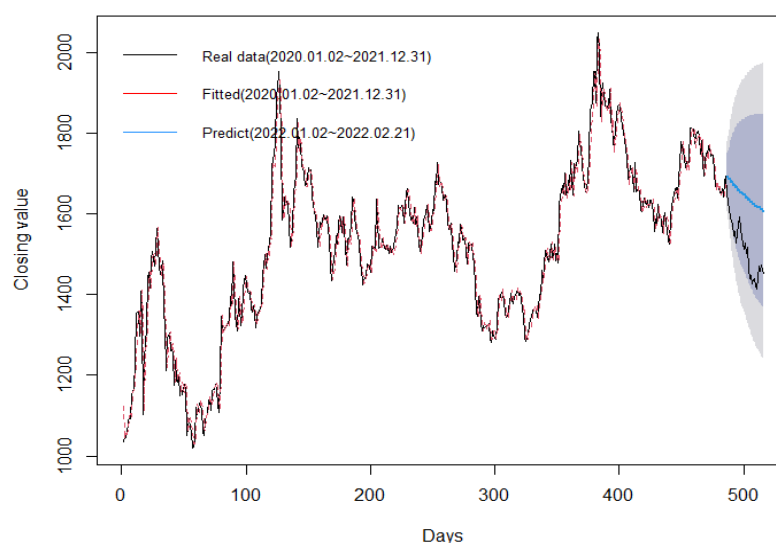


Figure 5. Plot of fitted values and forecast values

3.3.2 Forecast of returns

The linear returns are calculated based on the predicted values of the closing price. As is shown in figure 6, the number of periods of ups and downs in the next 30 periods of forecasted returns is similar, but the decline is more significant than the increase, so the overall trend is a sharp decline. As is

shown in figure 7, the number of periods of decline of real returns is 19 and it also shows a downward trend.

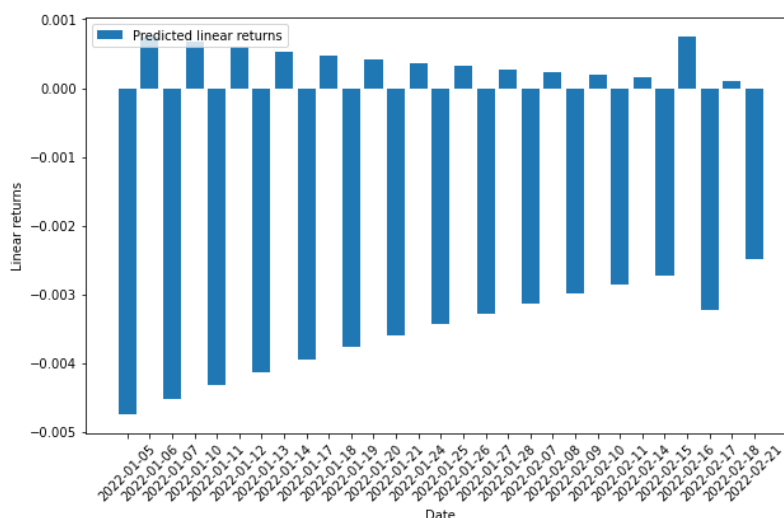


Figure 6. Linear returns of the forecasted value

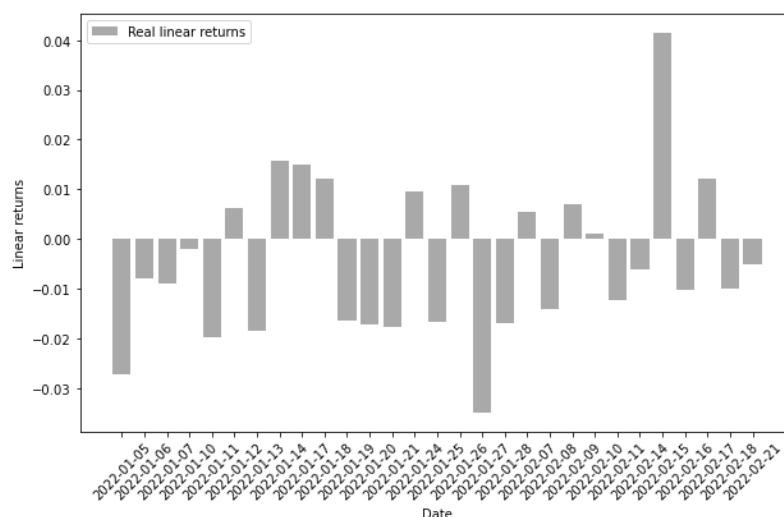


Figure 7. Linear returns of the real value

4. Conclusion

This paper found that although investors poured into the semiconductor industry and the chip industry in 2021, making relevant companies to rise together, 2022 will be an adjustment cycle for these industries and it will last for a longer time than people thought before. This paper uses the ARMA(2,1) model to predict the closing price of the SSE STAR Market Chip Index. According to the final expression of the best model, the chip index will still decline sharply in the next 30 periods, indicating that the demand side of the semiconductor industry and chip industry will continue to tighten, and the situation of consumers lacking of purchasing enthusiasm will not be eased up anytime soon.

Based on the empirical analysis results of the model, this paper proposes the following two suggestions for industry development and investment so that investors can better cope with the certain recession in the stock market brought about by the slowdown in the growth of the semiconductor industry and the chip industry.

Firstly, in the face of industry differentiation, investors need to return to rationality. Affected by varying degrees of demand weakening, downstream industries face a situation of different prosperity.

For example, chips in the general and consumer electronics fields are affected more seriously, therefore the price reduction is more obvious, while the prosperity of automotive chips and semiconductor equipment and materials is relatively strong. From above it can be seen that the price of this industry sector is being pulled back to a reasonable valuation point.

Secondly, semiconductor companies and chip companies should pay attention to mergers and acquisitions. In recent years, the capital market has been optimistic about these industries. Numbers of entrepreneurs have poured in making the homogenization competition between enterprises fierce, which will cause some companies facing operational difficulties to seek opportunities of mergers and acquisitions. It is expected that there will be a wave of mergers and acquisitions in the domestic semiconductor industry in the near future.

There are still some shortcomings in this paper. Although the predicted values are consistent with the future trend of the real values, the prediction error is positively correlated with the prediction step size. In addition, a single application of the ARMA model cannot determine the specific reasons that affect the returns. In the follow-up studies, this paper recommends to use the Granger Causality Test to explore the relationship between the rate of returns and other indicators of microeconomics and macroeconomics.

References

- [1] Zhu Jing. The current status of semiconductor parts industry and suggestions for China. *China Integrated Circuit*, 2022, 31(04): 10-17+36.
- [2] Xie Shenkun, Wang Zhizhe, Ren Yan et al. Challenges and opportunities of automotive chip shortage. *Electronic Product Reliability and Environmental Testing*, 2022, 40(03): 115-119.
- [3] Yao Yuqing. Prediction of stock return based on ARMA model. *Scientific Journal of Economics and Management Research*, 2021, 3(9): 183-191.
- [4] Jeffrey E Jarrett, Eric Kyper. ARIMA modeling with intervention to forecast and analyze Chinese stock prices. *International Journal of Engineering Business Management*, 2011, 3(3): 716-723.
- [5] Zhang Huilei. Empirical analysis of short-term and long-term volatility of SSE 50 ETF——based on the ARMA-GARCH model. Kunming, China, 2021.
- [6] Yang Yuyuan, Zhang Mei. Empirical analysis of stock price based on ARIMA model. *Science & Technology Information*, 2021, 19(29): 121-123+127.
- [7] Sun Yi, Zhou Longlong. Teaching of financial time series ARIMA model based on Python. *Modern Information Technology*, 2021, 5(10): 192-195.
- [8] Li Chenggang, Yang Bing, Li Min. Forecasting analysis of Shanghai stock index based on ARIMA model. *MATEC Web of Conferences*, 2017, 100: 02029.
- [9] K Murali, Sk Nafeez Umar, D Chandrakesavulu Naidu et al. Analysis and forecast of index monthly time series based on ARIMA model. *Journal of Innovation and Social Science Research*, 2021, 8(7): 20.
- [10] Paulo Rotela Junior, Fernando Luiz Riêra Salomon, Edson de Oliveira Pamplona. ARIMA: an applied time series forecasting model for the Bovespa stock index. *Applied Mathematics*, 2014, 5(21): 3383-3391.
- [11] Wang Yan, *Times series analysis with R*. Beijing: Renmin University of China Press, 2005.