

# ***Analysis and Prediction of the China Securities Bank Index Based on the ARIMA Model***

**Yuxin Wang<sup>1,a,\*</sup>**

<sup>1</sup>*School of Financial Technology, Shanghai Lixin University of Accounting and Finance, Shanghai, China*

*a. 201360212@stu.lixin.edu.cn*

*\*corresponding author*

**Abstract:** Banking is crucial for a country's financial system. Also, the performance of banks can indicate the confidence of investors in the financial market and the overall situation of the country's economy to a certain extent. As the CSBI index reflects the overall performance of Chinese banking stocks, predicting the relevant price movements can predict how the banking industry will move over time to come and help investors make investment decisions. Under this circumstance, the paper selects the CSBI closing price from January 2th, 2019 to August 11th, 2023 and build the ARIMA (0,1,0) model to make predictions based on these data. The conclusion shows that the ARIMA model has good forecasting abilities in the short term, but its timeliness is relatively poor in the long term. Finally, on the basis of these analyses, the paper summarizes the development trend of the industry in the short term, lists the reasons for the inaccuracy of the long-term forecast from different aspects and provides suggestions for investors.

**Keywords:** Banking industry, CSBI, ARIMA, prediction

## **1. Introduction**

Bank stocks play a crucial role in a country's financial system. As a provider of financial services, banks' earnings and performance are directly affected by changes in the country's economy. The China Securities Bank Index (CSBI) is made up of banking stocks in the CSI Broad Market Index constituent stocks, which are weighted by average daily turnover and average daily total market capitalization. It reflects the overall performance of stocks in the banking industry. Therefore, the forecasting of its future price can indicate the direction of China's economy and short-term policies. It also provides valuable insights for investors' decision-making direction. In terms of the method, the Autoregressive Integrated Moving Average Model (ARIMA) is generally appropriate for the analysis and application of time series. In addition, it is capable of accurately predicting the future price data of stocks.

In the past, it was a hot topic to study the changes in different indexes of the financial market, both at home and abroad. Scholars often undertook experiments through time series analysis in order to provide data references for their future investment decision-making. The practicality of the ARIMA model for short-term data prediction was confirmed by some scholars who selected specific stocks in the financial market for predictive analysis. Liu et al. established an ARIMA model and used Python tools to predict the historical closing price data of Southwest Securities stocks. Then the practicality

of the model was confirmed [1]. Huang established an ARIMA model in R to analyze and predict Ping An's stock price. The ideal results were also obtained successfully [2]. Khan used the ARIMA model to predict the Maruti Suzuki stock price, which also confirmed that the short-term prediction of the model had better effect [3]. Shakir et al. used three different custom ARIMA models to analyze Netflix's five years of stock history price data and predict the future trend. Finally, the most accurate stock prediction model was obtained and the desired results were got [4]. Some scholars also made prediction analysis on the bank stock price and related indicators. Kong predicted and judged the up-and-down trend of ICBC stocks price in the next trading day by constructing a variety of models and algorithms. After comparison, there was a conclusion that the ARIMA model was suitable for price prediction and machine learning was suitable for judging the accuracy of predictions [5]. Yin introduced the ARIMA (2,1,2) model and the Box-Jenkins method to analyze and predict the P/E ratio of the BCM stocks. Finally, there was a conclusion that the model had high prediction accuracy [6]. However, some of these above scholars' predictions didn't contain data to show their conclusion of the evaluation of long-term predictions. They just made simple descriptions. Weng constructed the ARIMA (6,1,6) model to analyze and predict the stock price of CCB. Then through a variety of indicators, the data fit was judged. On the basis of data analysis, the conclusion that the long-term prediction effect error was large was also drew [7]. Many scholars also conducted predictive analysis of other indexes in the stock market. Li used the ARIMA model to predict the Shanghai Stock Exchange Index (SSE Index). Based on the experimental results, the model was believed that it had good prediction ability in a short term but the timeliness of long-term prediction was not good enough [8]. Zou constructed the ARIMA (3,1,3) model to conduct an empirical analysis of the closing price of the SSE Index and came to a similar conclusion [9]. Wang et al. established the ARIMA (0,1,6) model to predict the annual trend of the SSE 50 Index in 2017. It could be concluded that the index had a slight recovery and maintained a trend of volatility. Also, the model's error gradually increased with the extension of the forecast time [10]. Li used the combination of ARIMA (3,1,3) model and artificial neural network model to predict the CSI 300 Index. Finally, a prediction curve with high accuracy was obtained [11].

In short, there is a growing number of data analysis and forecasts on composite indexes and individual stocks in the stock market. However, less attention is paid to indicators which reflect the development of the entire banking industry. At the same time, more scholars emphasize the applicability of ARIMA model for short-term predictions, while the analysis of the reasons for the limitations of long-term predictions is insufficient. The purpose of this paper is to solve the two problems mentioned above using the ARIMA model. It aims to analyze and predict the CSBI's price over the next 15 trading days based on the historical data. Additionally, a summary is provided of the changes in China's banking industry or even the national economy. Moreover, the paper compares the fit between the predicted and actual data to analyze both short-term and long-term accuracy, then further summarizes the reasons why the prediction is not accurate enough in the long-term.

## 2. Methods

### 2.1. Data Source

The prices of CSBI fluctuate at every point in the trading period, so the selection of the daily closing price as the research object is more meaningful in predicting the overall fluctuation of the index than its opening, the highest and the lowest prices. The paper extracts the closing price data for 1120 trading days from 2/1/2019 to 11/8/2023. The data is gathered from Orient Wealth Online.

## 2.2. Variable Selection

This paper mainly analyzes and predicts the change of the CSBI closing price from the dimension of time, so the variables are the day of trading and the CSBI closing price. Sometimes the day of trading can be shown in the format of date. Table 1 shows the variable attributes.

Table 1: Basic data attributes.

Name	Type	Width	Decimal
Day	Number	8	0
Corresponding date	Date	10	0
Closing price	Number	11	4

## 2.3. Research Protocol

The research uses Excel software to list the basic outline of the selected data and SPSS Statistics 27.0 software to build an ARIMA model for analysis. Also, the research process is systematic.

Firstly, the paper carries out an initial stationary test. The sequence diagram, autocorrelation coefficient function (ACF) and partial autocorrelation coefficient function (PACF) analysis plots of the historical closing price data are viewed and analyzed. After the previous analysis, the non-stationary data is smoothed by difference method and the stationarity test is carried out again. Then, according to the ACF and PACF plots, the model parameters  $p$  (autoregressive order),  $d$  (the number of differences required to change the original time series into a stationary time series) and  $q$  (moving average order) are determined to build the optimal model. After that, the applicability of the model is tested by performing a white noise test on the residual sequence of the model. Finally, the future value of the indicator is predicted and analyzed using the model which passes the test.

## 3. Results and Discussion

### 3.1. Stationary Test

Firstly, the original time series is analysed and shown in Figure 1. It can be preliminarily determined that the data is distributed in a non-stationary time series.

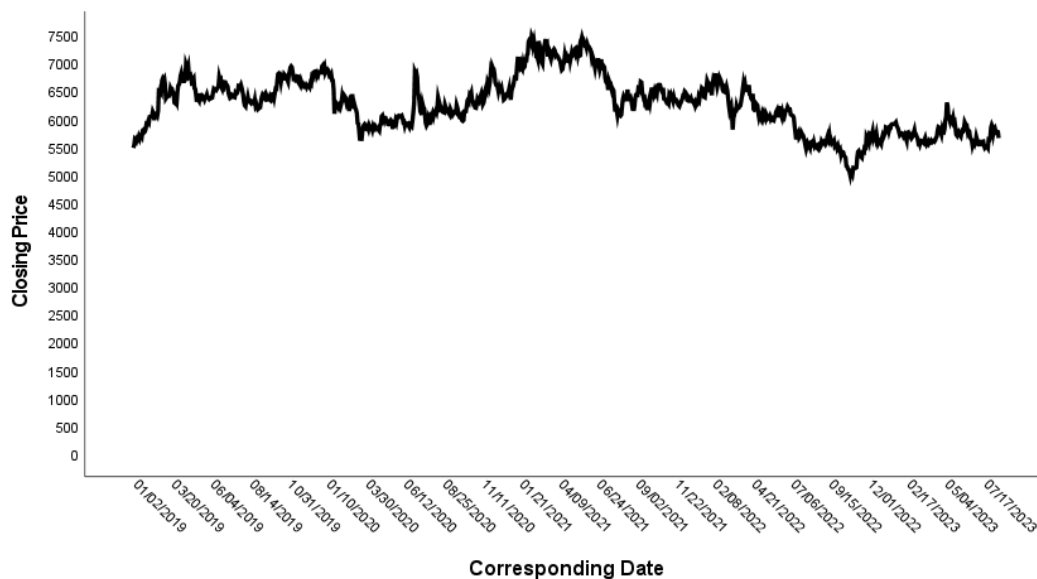


Figure 1: Original time series plot of CSBI closing price.

Also, the stationary test is further carried out to verify this kind distribution in Figure 2 and 3. It further proves the characteristics of the original data.

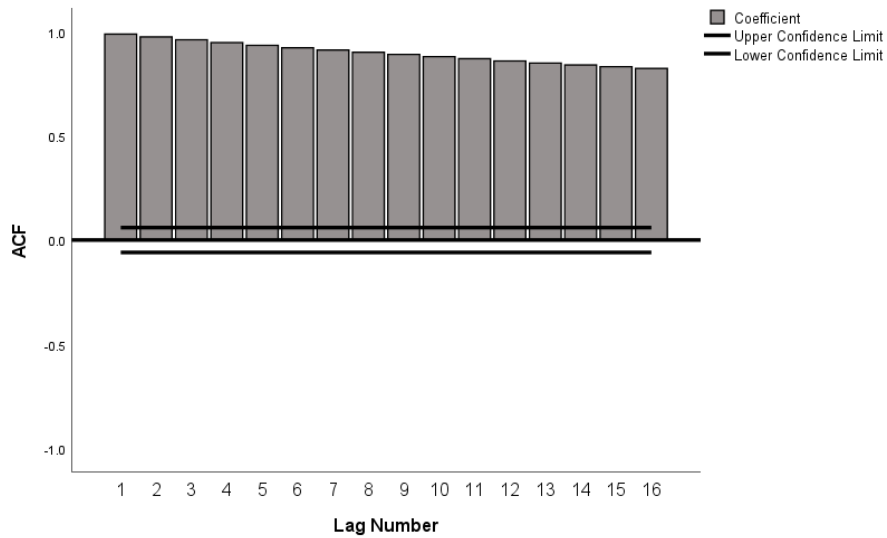


Figure 2: Original time series ACF plot of CSBI closing price.

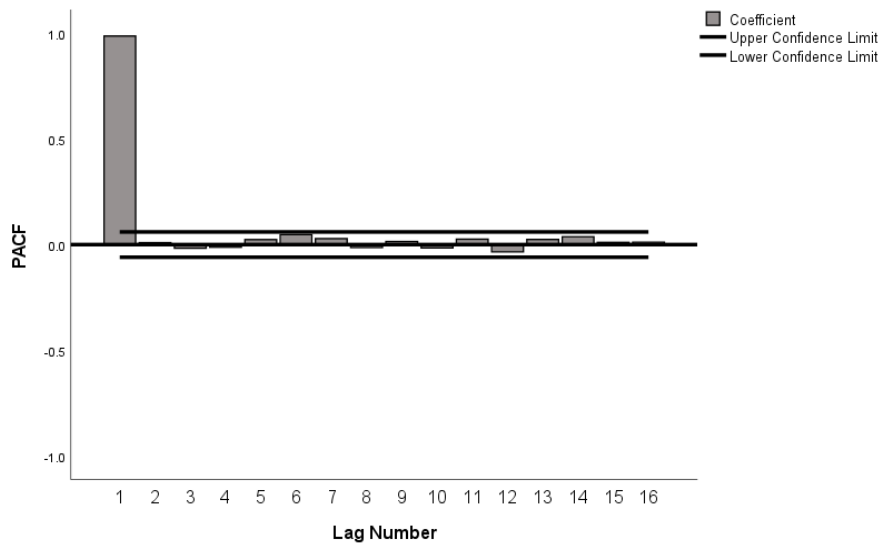


Figure 3: Original time series PACF plot of CSBI closing price.

Figure 2 and 3 demonstrate that the ACF plot tends to tail. The coefficients show a slow rate of periodic decrease and most of the data does not fall within the confidence interval. The PACF plot shows a 1st order truncation and the coefficients fluctuate in a small range above and below the zero line. In summary, the original time series is a nonstationary time series.

### 3.2. Model Recognition

In this case, the difference method should be used to perform differential processing of appropriate order on the original data in order to form a stationary time series. Therefore, the paper carries out the 1st order differential processing to smooth the original sequence and then observes the stationarity of the new one. Figure 4 shows the result of the process. It indicates that the data after processing tends to be stable.

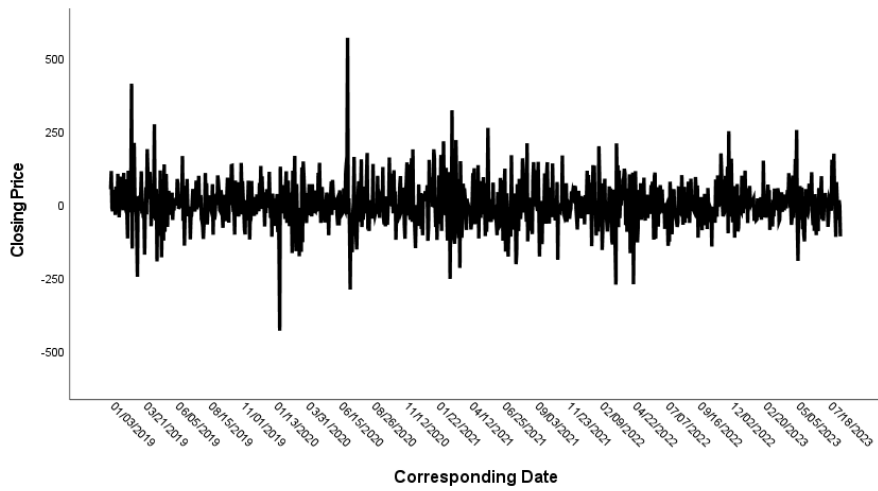


Figure 4: Time series plot of CSBI closing price after 1st order differential processing.

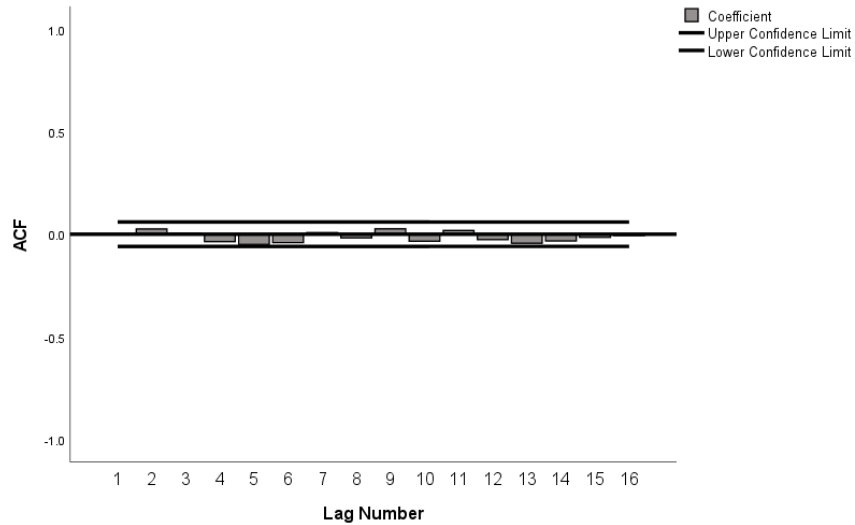


Figure 5: ACF plot of CSBI closing price after 1st order differential processing.

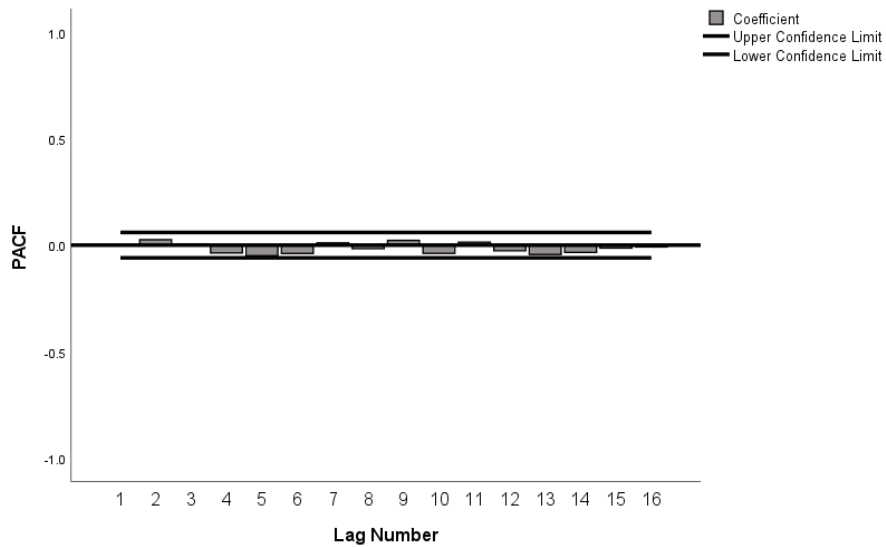


Figure 6: PACF plot of CSBI closing price after 1st order differential processing.

By plotting first order difference sequence ACF and PACF plots in Figure 5 and 6, it can be observed that the coefficients are distributed within the confidence interval. It indicates that after the first differential process, a stationary sequence for the Arima model is obtained. The model parameter is ARIMA (0,1,0) with  $p = 0, d = 1, q = 0$ .

### 3.3. Model Testing

To ensure the applicability of the model to the CSBI closing price, the model is constructed using SPSS 27.0 software. Some of the relevant model parameters and features are shown in the tables below.

Table 2: Model statistics.

Model	Number of Predictors	Model Fit statistics		Ljung-Box Q (18)			Number of Outliers
		Stationary R-squared	R-squared	Statistics	DF	Sig.	
Closing Price-Model	0	0.148	0.979	16.643	18	0.548	6

Table 2 shows that the R-squared is 0.979, close to 1, indicating that the model fits well and can explain 97.9% of the phenomenon. This is a preliminary demonstration that the model fits well.

At the same time, according to the BIC criterion, the smaller the BIC value, the better the model fits the data. The BIC value for this model is shown in Table 3. Compared to models with other parameters, the BIC value 8.578 in this model is the smallest. It indicates that the data fits well in this case. In summary, it can be concluded that this model is the optimal one.

Table 3: Model fit.

Fit Statistic	Mean	SE	Minimum	Maximum
Stationary R-squared	0.148	0.000	0.148	0.148
R-squared	0.979	0.000	0.979	0.979
RMSE	71.295	0.000	71.295	71.295
MAPE	0.848	0.000	0.848	0.848
MaxAPE	4.430	0.000	4.430	4.430
MAE	53.155	0.000	53.155	53.155
MaxAE	273.127	0.000	273.127	273.127
Normalized BIC	8.578	0.000	8.578	8.578

At the same time, according to the BIC criterion, the smaller the BIC value, the better the model fits the data. The BIC value for this model is shown in Table 3. Compared to models with other parameters, the BIC value 8.578 in this model is the smallest. It indicates that the data fits well in this case. In summary, it can be concluded that this model is the optimal one.

### 3.4. Residual White-noise Test

After selecting the final model, the residual terms also need to be tested. The paper chooses to test the characteristics of these terms in 2 different ways. The first one is performing a stationary test on the residuals.

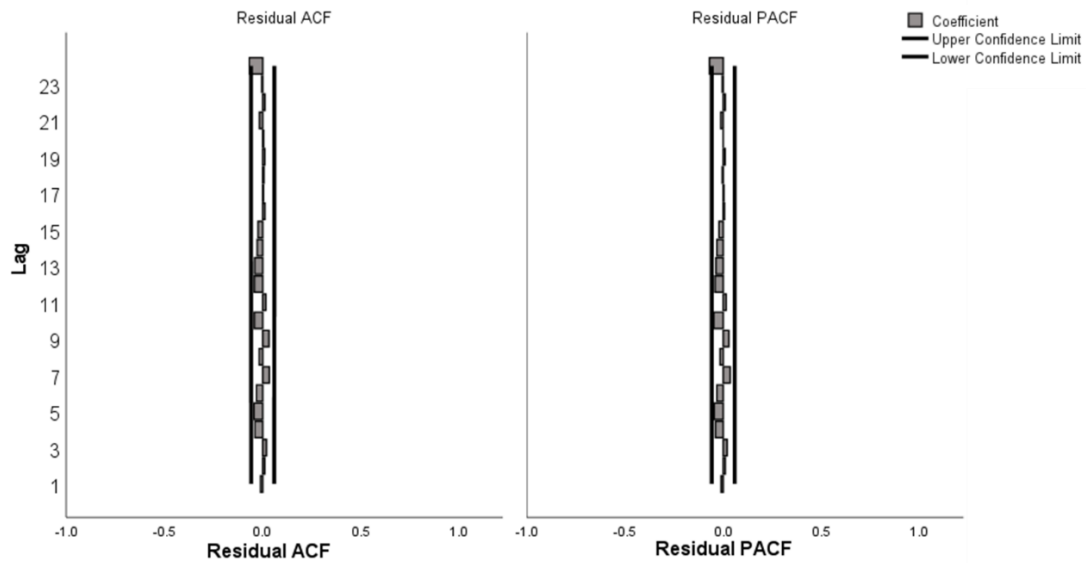


Figure 7: Residual ACF and PACF plots.

It can be seen in Figure 7 that the values of ACF and PACF for each lagging order are within the confidence interval and there is no significant change in trend. However, only through this analysis cannot accurately show the distribution features of the residual terms. Further proofs are needed from other dimensions and results.

Table 4: Residual ACF graph statistics.

Lag	Autocorrelation	SE	Ljung-Box Statistic		
			Value	DF	Sig.
1	-0.011	0.030	0.132	1	0.716
2	0.011	0.030	0.267	2	0.875
3	0.019	0.030	0.684	3	0.877
4	-0.040	0.030	2.464	4	0.651
5	-0.044	0.030	4.674	5	0.457
6	-0.032	0.030	5.828	6	0.443
7	0.034	0.030	7.116	7	0.417
8	-0.018	0.030	7.467	8	0.487
9	0.032	0.030	8.616	9	0.473
10	-0.043	0.030	10.667	10	0.384
11	0.016	0.030	10.961	11	0.447
12	-0.042	0.030	12.978	12	0.371
13	-0.041	0.030	14.873	13	0.315
14	-0.028	0.030	15.783	14	0.327
15	-0.024	0.030	16.419	15	0.355
16	0.012	0.030	16.578	16	0.413

At the same time, as shown by the Ljung-Box test, Table 4 shows that the residual terms have P values that are greater than 0.05. According to this result, the residual terms are part of the white noise sequence. Additionally, it indicates that the useful information in the residual terms has been extracted and the model is essentially flawless.

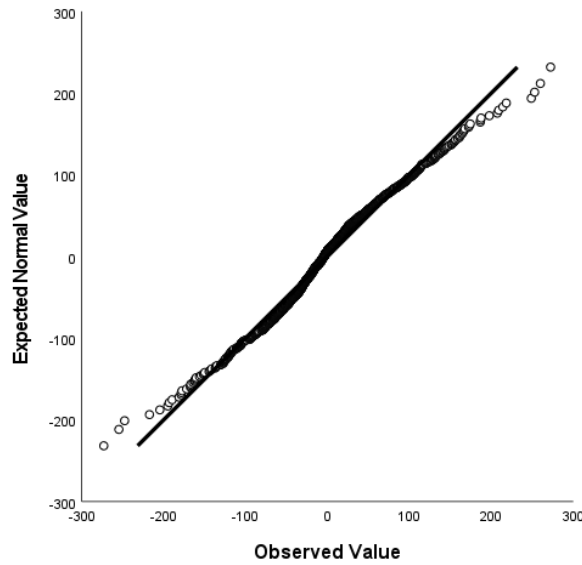


Figure 8: Q-Q plot of residual from the final model.

Furthermore, by plotting the QQ plot, Figure 8 shows that most of the residual values are near the straight line. This indicates that the residual terms conform to the characteristics of a normal distribution. Therefore, combined with the stationary test results, it is determined that the residual terms after the optimal model fitting are a white noise sequence.

### 3.5. Model Prediction

The CSBI closing price is predicted using the optimal model. The results are shown in Table 5 and Figure 9.

Table 5: Comparison of CSBI closing price forecasts for the next 15 trading days (08/14/2023-09/10/2023).

Date	Actual Closing Price (yuan)	Predicted Closing Price (yuan)	95% UCL	95% LCL	Relative Error
08/14/2023	5572.90	5646.31	5786.19	5506.42	-1.30%
08/15/2023	5634.39	5645.95	5843.78	5448.12	-0.20%
08/16/2023	5625.07	5645.59	5887.89	5403.30	-0.36%
08/17/2023	5578.96	5645.24	5925.01	5365.46	-1.17%
08/18/2023	5604.33	5644.88	5957.68	5332.08	-0.72%
08/21/2023	5515.26	5644.53	5987.18	5301.87	-2.29%
08/22/2023	5575.90	5644.17	6014.28	5274.06	-1.21%
08/23/2023	5560.51	5643.81	6039.47	5248.15	-1.48%
08/24/2023	5536.68	5643.46	6063.12	5223.80	-1.89%
08/25/2023	5583.72	5643.10	6085.46	5200.74	-1.05%
08/28/2023	5669.67	5642.75	6106.70	5178.79	0.48%
08/29/2023	5593.05	5642.39	6126.97	5157.81	-0.87%
08/30/2023	5527.76	5642.03	6146.40	5137.66	-2.03%
08/31/2023	5509.30	5641.68	6165.09	5118.27	-2.35%
09/01/2023	5561.97	5641.32	6183.10	5099.54	-1.41%



Table 5 shows that the forecast values' relative errors of the first 5 trading days are relatively small. The absolute values of three of them are within 1%, which are more accurate. Due to a situation of more than 2%, the relative error's absolute values of the forecast value have become relatively large in the last 3 predicted trading days. In addition, the average absolute value of the relative error in this period is greater than that in the first 5 days. It can be seen that as the prediction interval of the model increases, the predictive error will gradually increase. The accuracy of short-term prediction is fairly high.

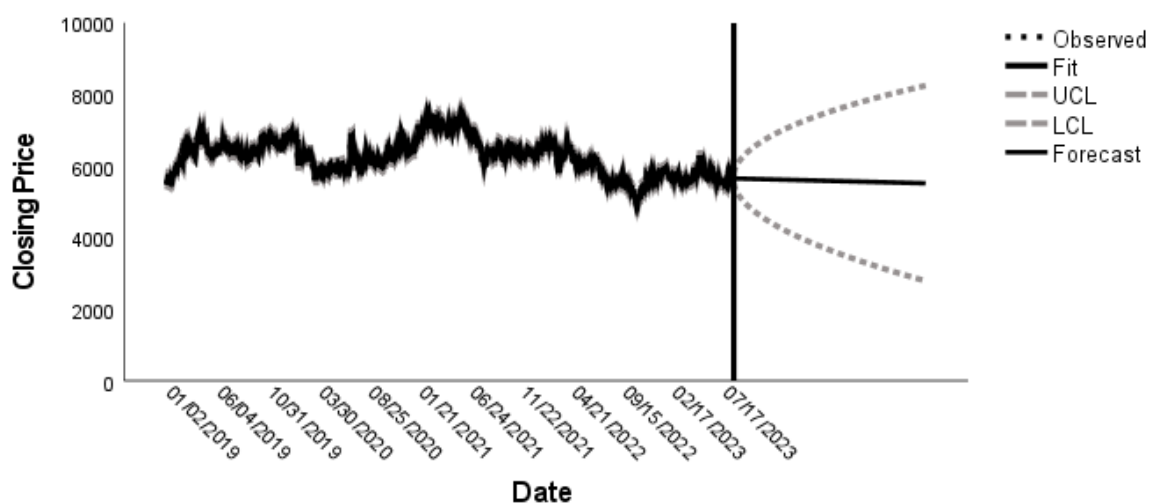


Figure 9: Prediction plot of CSBI closing price.

Figure 9 shows that the forecast trend of the price for CSBI is relatively stable in the coming period. Also, it is slightly downward but it doesn't have huge fluctuations.

#### 4. Conclusion

As for the generally stable trend of the price for CSBI, it can be inferred that China's banking industry is operating in a stable manner. Also, at present, the economic policy orientation promoted by the state is to maintain the stable operation of the banking industry and ensure the overall requirement of 'economic stability and development security'. At the same time, the country's attitude towards the banking industry is to seek progress while the credit risk is manageable and market risk and liquidity risk are generally stable.

The presentation of prediction results and relative error analysis demonstrate that the ARIMA (0,1,0) model is a reliable prediction method in the economic field. From the comparison of the predicted and the actual value of the price in the next 15 days, it can be seen that the model has a good prediction effect in the first 5 trading days. Nevertheless, the prediction error after that is somewhat large. It indicates that the timeliness of this model should be taken into account when using it for prediction. It is better suited for short-term forecasting.

There are numerous reasons for larger errors in long-term forecasting results resulting from this consequence. In addition to the internal volatility, the stock market is also impacted by various external factors. Firstly, the government may introduce new policies to promote economic development. For example, at the end of August this year, the government introduced policies such as halving the stamp duty on securities transactions and tightening the pace of IPOs in stages. The implementation of these policies will bring new vitality to the market and boost the growth of China's overall economy. Therefore, the CSBI index prices may increase significantly. Secondly, changes in the social environment and the international situation are also uncertain factors. If the COVID-19

epidemic is repeated and the new virus is infested, the national economy may experience a sluggish development again in the future. As a major part of the national economy, the banking industry will inevitably be affected to a certain extent. Also, changes in the international political situation and fluctuations in the international financial market may have an impact on the domestic stock market. Finally, investor psychology will be affected by changes in the investor base. If more investors become more aggressive, the prices of the relatively unpopular types of banking stocks and indexes will fall.

In summary, the development of China's banking industry is relatively stable because of the steady trend of the CSBI index. The paper recommends that conservative investors choose CSBI as their investment choice. However, investors should make decisions based on both machine forecasts and economic facts when investing in the future.

## References

- [1] Liu, S., and Zhang, S. (2021) *An Empirical Study on Stock Price Forecast by ARIMA Model*. *Economic Research Guide*, 25, 76-78.
- [2] Huang, L.X. (2020) *Stock Price Analysis and Prediction based on ARIMA Model: A Case Study of Ping An Insurance Company of China*. *Science and Technology Economic Market*, 10, 62-63.
- [3] Khan, K. (2022) *Stock Price Forecasting of Maruti Suzuki using ARIMA Model*. *Parikalpana: KIIT Journal of Management*, 2, 90-98.
- [4] Shakir, K., Hela, A. (2020) *ARIMA Model for Accurate Time Series Stocks Forecasting*. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 7, 524-528.
- [5] Kong, Y.P. (2023) *ICBC Stock Forecast Analysis based on Time Series and Machine Algorithm*. *China Management Informatization*, 6, 146-148.
- [6] Yin, Y. (2015) *Stock P/E Ratio Analysis and Prediction based on ARIMA Model*. *CO-Operative Economy and Science*, 2, 36-37.
- [7] Weng, Z.X. (2023) *Stock Price Analysis and Prediction based on ARIMA Model—Taking China Construction Bank as an Example*. *Modern Information Technology*, 14, 137-141.
- [8] Zou, Z.Y. (2021) *Analysis and Prediction of Shanghai Composite Index based on ARIMA Model*. *Hebei Enterprise*, 3, 36-37.
- [9] Li, X.X. (2017) *Prediction of Shanghai Composite Index Trend based on ARIMA Model*. *Liaoning Economic Journal*, 7, 94-96.
- [10] Wang, H.X. and Lin, J.X. (2017) *Analysis and Prediction of Shanghai Stock Index 50 Index based on ARIMA Model*. *Times Finance*, 24, 143-144+147.
- [11] Li, B. (2019) *The Prediction of CSI 300 Stock Index—A Hybrid Method based on ARIMA Model and Artificial Neural Network Model*. Jinan University.