Stock Price Forecasting Based on ARIMA Model an Example of Cheung Kong Hutchison Industrial Co.

Qijun Xue *

Department of Finance, Soochow University, Soochow, China

* Corresponding Author Email: 2010408016@stu.suda.edu.cn

Abstract. Previous studies in stock prediction mainly researched the ARIMA model, however, scant studies take Hong Kong stock price prediction as a data source for this model. The present study takes Cheung Kong Hutchison Industrial Co. as an example. With conducting descriptive analysis and the ARIMA model, this study goes through four steps: 1) Stationarity Processing: the time series data are pre-processed with differential operations to be stationary, and the series are tested for stationarity by the unit root (ADF) test. 2) Order determination: the order can be observed by drawing autocorrelation and partial autocorrelation plots. The ARIMA model's p and q values can then be calculated through SPSS. 3) Validity tests: A Q-value Table is set to determine whether the residual term conforms to the white noise distribution. Moreover, a Ljung-Box test is performed to determine whether the residual series correlated. 4) Prediction: The stock prices in the next few trading days are predicted and analyzed according to the model. The main finding of this research suggests that the price of the stock will decline at a relatively gradual pace from 48.55 to 48.51 over the next five trading days. Moreover, the ARIMA model predicts preciously this time with an error of no more than 3%. The findings can inspire researchers with the application of the ARIMA model, which can also offer suggestions for traders to obtain returns in the HK stock market.

Keywords: ARIMA model; Hongkong stock; Time series prediction.

1. Introduction

Time series refers to a series of values of a quantity obtained at successive times, often with equal intervals between them [1]. According to the random characteristics of time series, econometrics divides it into two types: stationary and non-stationary [2]. The representative forecasting methods mainly include the moving average method, time series decomposition method, exponential smoothing method, seasonal multiplier method, ARCH method, and Box-Jenkins method [3].

Financial time series is a special high-noise dynamic time series, showing complex features of discrete and non-stationary [4]. Analyzing, predicting, and controlling such data is a requirement in the context of the era of data science, which is also the critical research object of technologies such as robot advisors in financial technology. One of the most important models used in the academic world to fit non-stationary time series is ARIMA. The full name of the ARIMA model, firstly proposed by Box and Jenkins, is the differential autoregressive moving average model [5]. The advantages and features of numerous time series prediction models are combined in this model.

2. Research Questions

What is the general profile of the chosen stock price's time series?
To what extent will this stock price fluctuate in the next seven days?
What is the performance of the ARIMA model's prediction?

3. Research Methods

3.1. Data Selection

This paper selects Cheung Kong Hutchison Holdings (CKH HOLDINGS) as the research object since the stock was listed in Hong Kong earlier in 2015, which has a certain degree of representativeness. Specifically, this paper analyzes the daily closing prices of CKH HOLDINGS (00001.HK) for each trading day from February 8, 2019, to February 7, 2023. The paper gets data from the China Stock Market & Accounting Research Database (CSMAR), with a total of 986 sets of data.

3.2. Model Selection

This paper chooses the ARIMA model to forecast the stock's price in the following months. It was born based on the ARMA model. Both these two models include the AR model and the MA model [6]. The improvement of the ARIMA model is its Universality. The ARMA model can only be used to analyze Stationary Time Series, while the ARIMA model can simultaneously investigate the non-Stationary Time Series. Because time series are typically non-stationary Stationary Time Series, the ARIMA model receives more application scenarios in practice.

When the time series is stationary, the ARIMA (p,d,q) model form can be written as follows. In this model, AR stands for autoregressive; p is the order of autoregressive; MA is for moving average; q is the number of moving average components; and d is the number of differences (order).

$$\Delta^d y_i = \phi_1 y_{i-1} + \phi_2 y_{i-2} + ... + \phi_p y_{i-p} + \varepsilon_i - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - ... - \theta_q\varepsilon_{t-q} \quad (1)$$

Where $\Delta^d y_i$ denotes the Stationary Time Series formed by the non-Stationary Time Series y_t after d-difference processing, ε_t represents random error term, ϕ_i (i=1, 2, ..., p) and θ_j (j=1, 2, ..., q) are the pending coefficients of the model, with p and q values are the orders of the model.

The modeling of ARIMA model can be divided into four steps as follows:

1) Stationarity Processing. The time series data were pre-processed with differential operations to be stationary, and the series were tested for stationarity by the unit root (ADF) test.

2) Order determination. The order can be observed by drawing autocorrelation and partial autocorrelation plots. Then ARIMA model's p and q values were finally determined through SPSS.

3) Validity tests. A Q-value Table was set to determine whether the residual term conforms to the normal distribution. Moreover, a white noise test was performed to determine whether the residual series was white noise.

4) Prediction. The stock prices in the next few months were predicted and analyzed according to the model.

4. Results and Discussions

4.1. Descriptive analysis

In terms of the descriptive analysis, this study imported the data into SPSS and noted a few indicators. The following key indicators are selected from them:

Table 1. Descriptive Indicators of the stock's data

Indicators	Sample Size	Standard Deviation	Variance	25th Percentile	Median	75th percentile
Values	59.144	10.836	117.414	51.675	56	68.413
Indicators	Kurtosis	Skewness	IQR	Mean 95% CI(LL)	Mean 95% CI(UL)	coefficient of variation (CV)
Values	-0.508	0.678	16.738	58.468	59.821	0.18321

This paper uses SPSS to plot a line plot of the stock's price from February 8, 2019 to February 7, 2023. The graph gives a general indication of the stock's share price movement trend during that period. According to Table 1, the sample size is 59.144, with the mean of 58.468 (Mean 95% CI (LL)) and the Standard Deviation of 10.836.

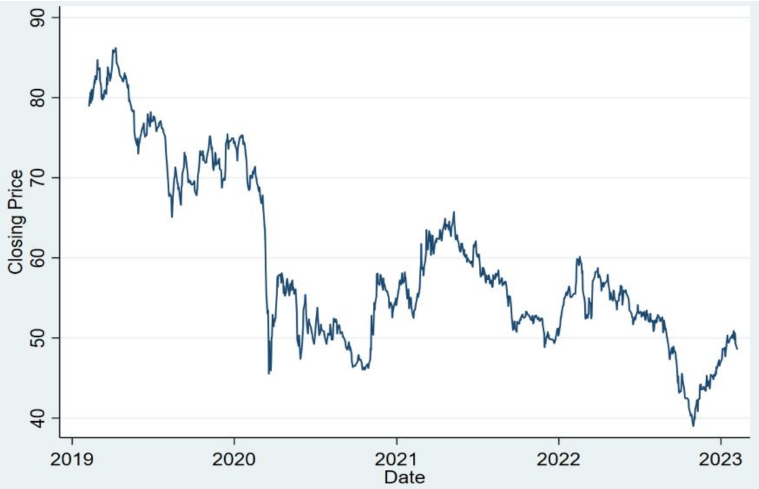

Fig 1. CKH Holdings' Stock Price Time Series Line Chart

According to Figure 1, this stock follows a downward trend in general. Specifically, the whole process can be divided into three stages.

Stage 1: Sharp decline period (2019-2020). In this period, the stock price dropped rapidly from a peak of ￥86.20 to a trough of ￥45.55, meaning the stock shrank by 47.16% in just one year.

Stage 2: Stable fluctuation period (2020-2022 2nd quarter). In this period, the stock price fluctuated steadily in the range of ￥50 to ￥60 most of the time.

Stage 3: New Stable fluctuation period (2022 2nd quarter-2023 1st quarter). In this period, the stock still fluctuated steadily but gradually transformed into a new range of ￥40 to ￥50.

A stock's price will not fluctuate much in a short amount of time. Therefore, using the following trend analysis, it is possible to summarily forecast that the price will remain between $30 to $60.

4.2. ARIMA Model Prediction

4.2.1 Stationarity Processing

Due to the high volatility of the stock market, the stock time series is usually non-stationary [7]. Therefore, the original data needs to be differentiated until it has passed the stationary test.

Table 2. ADF Test Table

Difference Order	t	p	Critical Value		
			1%	5%	10%
0	-1.99	0.291	-3.437	-2.864	-2.568
1	-16.019	0	-3.437	-2.864	-2.568

In Table 2, On the one hand, the result of the ADF test of the original data shows that p=0.291>0.1, meaning the original hypothesis cannot be rejected. Hence, the series is non-stationary, which is required to take the first difference and then the ADF test.

On the other hand, after first order difference, the result of the test shows that p=2536.916>0.1, indicating that the initial hypothesis cannot be ruled out and that the series is non-stationary. The second difference is required to make the series stationary. The ADF test result of the data after the second difference shows that p=-2.864<0.1. Therefore, the second difference was chosen (i.e., d=2).

4.2.2 Order Determination

This paper uses SPSS to plot ACF and PACF to determine the autoregressive order p and the moving average order q. From Figure 2, the ACF plot is truncated at 1st order lag (ACF is 0 after 1st order lag), and the PACF plot is not truncated, then the ARIMA model can be simplified to MA(q). The most significant order in the ACF plot is chosen as the q value (i.e., 1st-order lag). Therefore, the final autoregressive order p-value is 0, while the moving average order q-value is 1.

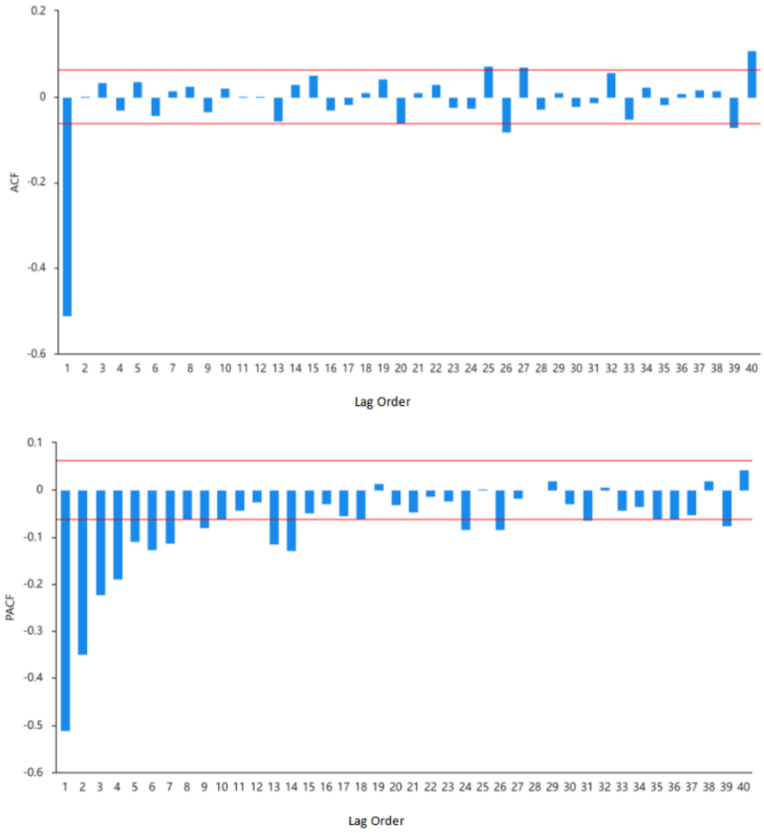

Fig 2. ACF and PACF Plot

4.2.3 Validity Tests

The ARIMA model requires the model residuals to be white noise, which can be tested by the Q statistic test. Normally, Q_t is used to test whether the first t orders of autocorrelation coefficients of the residuals meet the white noise. If its p-value is larger than 0.05, the original hypothesis is satisfied, which means the residuals are white noise. Among previous literature, it is more common practice to take t as 6 [8-10]. From the Q statistic results (Table 3), the residuals of the model are white noise, and the model meets the requirements.

Table 3. Residues Test

Term	Value	*p* -Value
Q1	1.137	0.286
Q2	4.076	0.13
Q3	9.728	0.021*
Q4	10.362	0.035*
Q5	11.759	0.038*
Q6	12.131	0.059

The model was also set based on the assumption that the residuals were not correlated with each other [11]. The Ljung-Box test was designed to examine whether the residuals were correlated with each other. With the application of SPSS, results showed that the p-value was 0.221, much larger than 0.05. Hence, it can be concluded that the residuals were not correlated with each other. The model was close to perfect because all of the valid information from the residual terms had been extracted.

4.2.4 Predicting Stock Prices and Comparing it with its Actual Price

According to test results above, it can be seen that the ARIMA model fits well. Therefore, the established ARIMA (0,2,1) model can be used to forecast stock prices. Since the ARIMA model is suitable for short-term forecasting and has high accuracy, it is chosen to forecast stock prices for the next seven trading days. The forecasting results are shown in Table 4.

Table 4. Price Prediction of Next 7 Trading Days

Date	Pridicted Price	Actual Price	Error Rate
Feb. 8	48.545	48.90	0.73%
Feb. 9	48.541	49.35	1.64%
Feb. 10	48.536	49.55	2.05%
Feb. 13	48.531	49.00	1.00%
Feb. 14	48.527	49.05	1.07%
Feb. 15	48.523	47.80	1.51%
Feb. 16	48.518	47.90	1.29%

The stock prediction value is closer to the actual closing price, with the error is no more than 3%, indicating high model prediction accuracy. It also verifies that the ARIMA (0, 2, 1) model constructed in this paper is accurate, which can reflect the future trend of stock prices.

5. Conclusion

The ARIMA model is used to predict the price of a Hong Kong Stock, with selecting the closing price of CK Hutchison (0001. HK) from February 8, 2019 to February 7, 2023 as the research data. First, this paper imports the data into SPSS for descriptive analysis. According to the changing law of the closing price, the whole period is divided into three stages. Secondly, this paper carries out the processing steps of the stationarity test, difference processing, determination of auto(partial) correlation coefficient, and residual error test so that the data can finally be adapted to the application conditions of the model. Then, this paper obtains the predicted closing price of the following five trading days and compares the result with the actual closing price to obtain the forecasting effect of the ARIMA model. The final results show that the ARIMA model can be applied to the stock price sequence analysis efficiently, which has a better short-term forecasting effect on the stock price. It can provide investors with a reasonable investment direction and offer suggestions for obtaining returns in the stock market.

References

[1] Wang Yan. Applied Time Series Analysis [M]. Beijing: Renmin University of China Press, 2015.

[2] Wu Yuxia, Wen Xin. Short-term Stock Price Forecast Based on ARIMA Model[J]. Statistics and Decision Making, 2016(23):83-86.

[3] Alfarra A N K, Hagag A. Forecasting of the American Digital Economy Using ARIMA Model[C]. 2021 International Conference on Electronic Engineering IEEE,2021: 1-5.

[4] Qian Long. Research and application of financial time series long-term and short-term forecasting model based on ARIMA and recurrent neural network[D]. Xihua University, 2022.

[5] Box G E P, Jenkins G M. Time Series Analysis, revised edition[J]. San Francisco, 1976.

[6] Wang Ying. Analysis and Forecast of Stock Price Based on ARMA Model[J]. Productivity Research,2021, No.350(09):124-127.

[7] Wang Yan. Research on stock price prediction based on ARFIMA-GARCH-LSTM hybrid model[D]. Shanghai Normal University, 2022.

[8] Liu Song, Zhang Shuai. An Empirical Study on Stock Price Prediction Using ARIMA Model[J]. Economic Research Guide, 2021(25):76-78.

[9] Xiong Zheng, Che Wengang. Application of ARIMA-GARCH-M Model in Short-term Stock Forecasting[J]. Journal of Shaanxi University of Technology (Natural Science Edition),2022,38(04):69-74.

[10] Lapedes A, Farber R. Nonlinear signal processing using neural networks: Prediction and system modelling[R]. 1987.

[11] Huang Shimin. Analysis and Forecast of Stock Price Based on ARIMA Model——Taking China Merchants Bank as an Example[J]. SME Management and Technology,2022(11):184-187.