Prediction of Apple stock price based on ARIMA model

Zihan Lin, Xihan Tang, Zhiting Tian, Shuhan Zhao

May 2023

1 Introduction

Apple Inc. is an American multinational technology company headquartered in Cupertino, California. It designs, develops, and sells a range of consumer electronics, computer software, and online services. Apple is now one of the most valuable publicly traded companies in the world, and its stock price has experienced significant growth over the years. The research will focus on the prediction of Apple stock price in June 2023 based on ARIMA model. And the result will indicate the overall trend of Apple stock price.

2 Background

The stock market is inherently uncertain and highly erratic. As a result, investors constantly take chances in an effort to profit. Supply and demand, market trends, the global economy, and historical price are all factors that can influence stock price and may lead to an increase or decrease in the number of buyers. The stock price of Apple has been volatile over the years, with fluctuations influenced by a variety of factors such as product releases, changes in market conditions, and global events. However, Apple has generally been a strong performer in the stock market, with a track record of consistent growth and healthy financials. Autoregressive Integrated Moving Average (ARIMA) is a popular time series forecasting model used in statistics and econometrics. ARIMA model is particularly useful in forecasting time series data with trends and seasonality. According to our research, it has been used widely in finance, economics, social science, and engineering. It can also be an effective tool in stock price prediction.

3 Method

A time series is simply an observation made in a chronological order of time. Time series forecasting models use mathematical techniques based on historical data to predict the future. It is based on the assumption that the future is an extension of the past; that's why we can use historical data to predict the future trend. In this article, ARIMA model is used to predict stock price. An ARIMA model typically present by ARIMA(p,d,q), which

- p: AR(p) – autoregression model
- d: I – times of difference
- q: MA(q) – moving average model

Autoregression model refers to that the variable of interest is predicted based on the combination of historical data of the variable of interest. Equation(1) can represent a AR(p) model. The equation replace predictor variable to the historical data of variable of interest.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t \tag{1}$$

Differential sequence is the time series consisting of changes between successive observations of the original series. It can be presented by equation(2)

$$y'_t = y_t - y_{t-1} \tag{2}$$

Moving average model uses error of historic prediction to construct the model. It can be defined by equation (3).

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_p \varepsilon_{t-p} \tag{3}$$

To start with, we use Box-Jenkins method, a suitable model for the purpose of estimating and predicting univariate time series, to analyze our data through the whole process. To construct the ARIMA model only on the basis that the time series is stationary. A time series is said to be stationary if there is no systematic change in the mean (no trend), no systematic change in the variance, and periodic changes are strictly eliminated. Eventually, we want to get a series of data for which the mean is a constant independent of time; the variance is a constant independent of time t, a property called homogeneity of variance; and the covariance is a constant that depends only on the interval k and is independent of time t. If the original series is not stationary, we need to apply difference to the series to stabilize it.

We also need to know whether our data is predictable, which is non-white noise. For a pure random series, also known as a white noise series, there is no correlation between the values of the series, and it is undergoing completely disordered random fluctuations. It also means that the past behaviors having no effect on future development, so there is no information to extract and no need to analyze it further.

After that, we start to calculate the model parameters. Here we may get several ARIMA(p, d, q) models. In order to determine which model is the most suitable, information criteria AIC, BIC, AICc can be used to determine. And we want the model with the least AIC, BIC, and AICc values, being the most suitable p and q values.

Finally, we need validate the model. Basically, by using plot of residuals and diagnostic statistic, we can assess the adequacy of future values. After all, we can compare and make prediction to the data.

4 Result

In this article, Apple stock price for each day is conducted based on real data, and the accuracy of model will be measured.

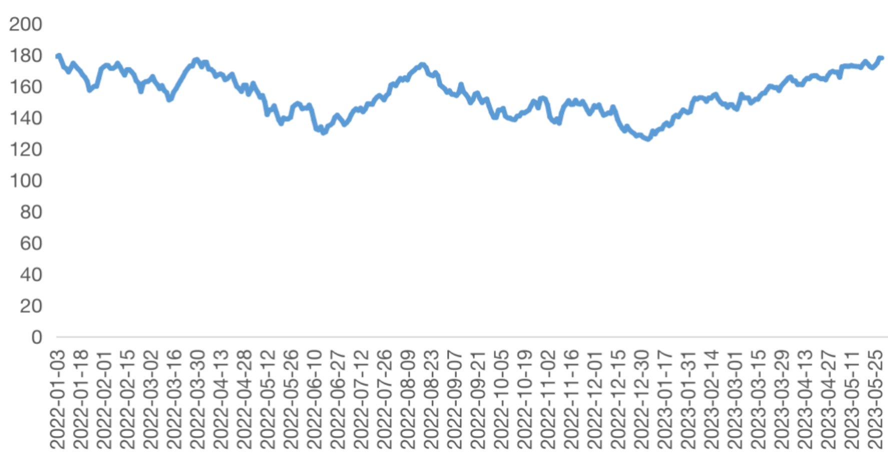

Figure 1: Average Apple stock price per day

Based on the Box–Jenkins approach, our study will be carried out in three parts: identification, estimation, and verification. The model shown in figure 1 is the average price of the highest and lowest price

of Apple stock for each day, from 2022.1.3 to 2023.5.31. However, based on the 359 data, we will first use the former 200 data to predict the later 159 data to determine the accuracy of our model and then forecast the future stock price.

4.1 Identification

In this step, we are going to identify the conditions of the data for the model, including the initial processing of difference to make data stationary and identifying possible values of p and q. We start with checking whether the data is stationary and determining the degree of difference for the sequence. To test whether or not it is stationary, we can use the unit root test is to check whether there is a unit root in the sequence, because the existence of a unit root is a non-stationary time series. The Dickey-Fuller test and Augmented Dickey Fuller test (ADF) can test whether an autoregressive model has a unit root.

H_0**: the series has a unit root, and it is nonstationary.**

H_a**: the series has no unit root, and it is stationary.**

Table 1: ADF test results

Differential times	P-value
0	0.7205
1	$<$0.01
2	$<$0.01

When the series is after one difference, the resulting p-value has already smaller than 0.01, which is smaller than the significance level 0.05. As a result, we reject the null hypothesis and accept the alternative hypothesis, which shows that the data is stationary after one difference. Figure 2 and 3 shows the model after one difference and two difference. The figure after one and two difference shows that there's not much of a difference, so we use data after one difference to build our model.

Additionally, to further assure whether the data is stationary, we use the graph of ACF and PACF shown in figure 4 and 5, which indicates that the data fluctuates around an average value of 0 and ACF is fairly approaching to zero, which proves the stationary of the sequence.

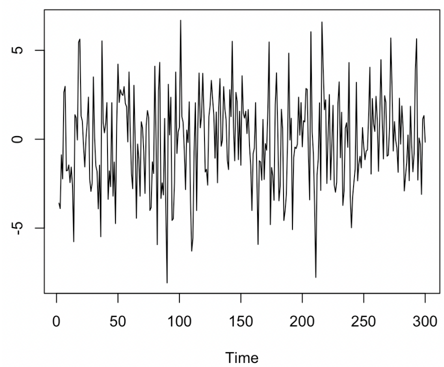

Figure 2: Data after one difference

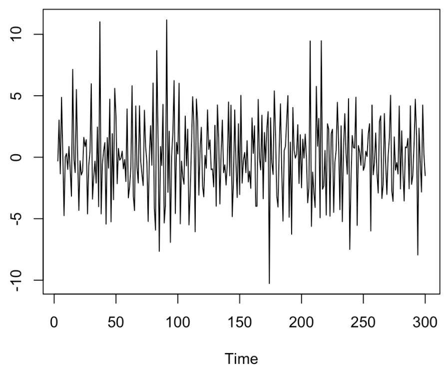

Figure 3: Data after two difference

Moreover, we need to figure whether our sequence is predictable. To test the white noise of the series, we use Ljung-Box test. It is based on a series of lag orders to determine whether the correlation or randomness of the whole sequence exists. The result of the test is in Table 2, which showed that the sequence is non-white noise and the sequence can be predicted.

H_0**: The data are random, and the series is white noise sequence.**

H_a**: The data are not random, and the series is no white noise sequence.**

Table 2: Ljung-Box test result

X-squared	335.33
df	1
P-value	<2.2e-16

Since the calculated p-value is smaller than the threshold significance level $\alpha = 0.05$, the null hypothesis H_0 can be rejected. Therefore, the data in the series is not independent, and the series can be predicted.

After verification of the conditions of the series, we notice from the ACF and PACF correlation that our model is not pure AR or pure MA. We therefore tested several models to identify the most suitable

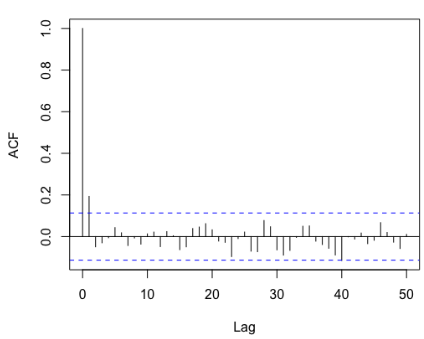

Figure 4: ACF graph

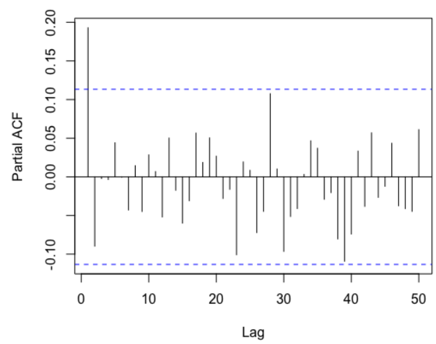

Figure 5: PACF graph

one for sales. In figure 4 and 5, graph of ACF and PACF are shown.

4.2 Estimation

The estimation of coefficients in ARIMA is done by the usage of ACF and PACF graph to provide parameters p, q, and d, using a fast maximum likelihood estimation algorithm. The execution of the procedure used Hyndman-Khandakar algorithm that combined with multiple root of unity tests. The best model is as simple as possible and minimizes the value of AIC, AICs, BIC, and maximum likelihood estimation. The chosen model is that of ARIMA(0,1,1).

Table 3: ADF test results

Model	AIC	AICc	BIC
ARIMA(0,1,0)	1432.83	1432.85	1436.54
ARIMA(0,1,1)	1421.66	1421.7	1429.06
ARIMA(1,1,0)	1423.35	1423.39	1430.75
ARIMA(1,1,1)	1423.28	1423.36	1434.38
ARIMA(0,1,2)	1423.2	1423.28	1434.3
ARIMA(1,1,2)	1425.07	1425.21	1439.88
ARIMA(0,1,1) with drift	1423.39	1423.47	1434.49
ARIMA(1,1,0) with drift	1425.08	1425.16	1436.18
ARIMA(2,1,2) with drift	1428.7	1428.99	1450.91

For other models, the AIC values were estimated, and Table 3 summarizes AIC value for each model. From the result, we can estimate

ARIMA(0,1,1) is the best model.

Table 4: Coefficients of model

	Coefficient	S.E.	T value	p Value
MA(1)	0.2209	0.0584	3.782534	0.0001554
Constant	-0.001	0.158	-0.005	0.996

Based on the the coefficients of the model, we first perform the parameter significance test. We use coefficient test to determine whether the estimates of the model parameters are significant. In Table 4, for coefficient of MA(1), T-statistic is larger than 1.96 and p-value is smaller than 5% shows that the parameter of the model is prominent. For the constant, since P-value is larger than 5% significance, we can remove this constant from the fitted model. The developed model is given by equation(4).

$$y_t = y_{t-1} - 0.2209\varepsilon_{t-1} \tag{4}$$

4.3 Validation

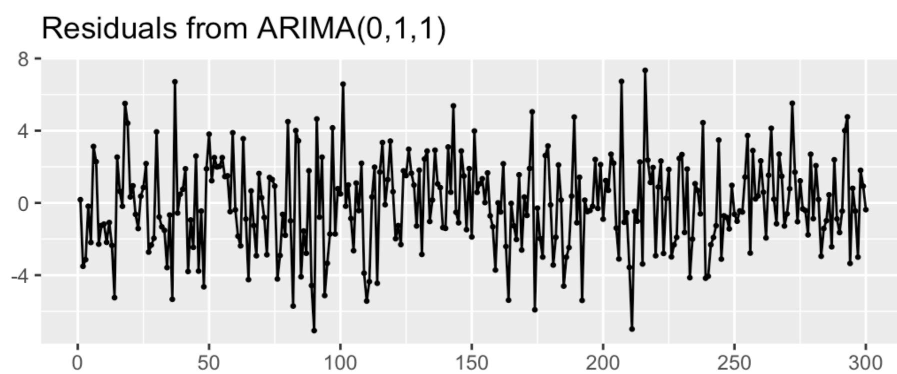

Figure 6: Average Apple stock price per day

To check the validity of the model, we should consider model's residual. The model residue should be stationary, follows the white noise process, and is normally distribution. Figure 6 shows the residual plot of ARIMA(0,1,1). Specifically, we examine the ACF plot for model residues, as in figure 7. The figure shows that the ACF of residuals are within the standard error at each different lag, which data has no correlation.

We also check the white noise of residue to determine whether two data are related. By using Ljung-Box test, there is a 97.1% probability of not rejecting the null hypothesis, which shows that the residue follows the white noise rule.

The residue histogram shows whether the distribution of residues is approximately normal. From figure 8, we can conclude that the residues is approximately normal distribution.

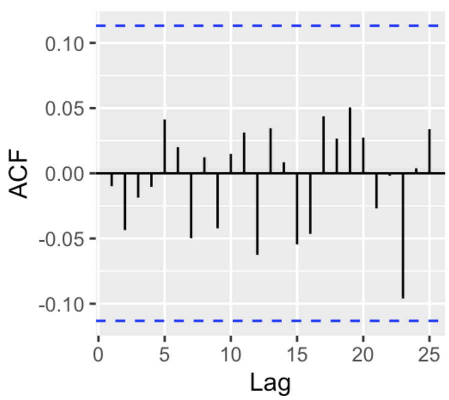

Figure 7: ACF graph of residue

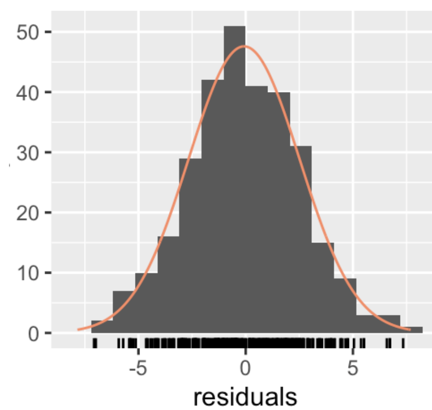

Figure 8: Residuals histogram

Through these process, the selected model of ARIMA(0,1,1) is appropriate for the prediction of Apple's stock price.

4.4 Accuracy

The accuracy of the model is resolved by comparing the real data and the predicted data in the same period. Figure 9 shows this comparison and reflects that the model is highly accurate and can predict the future stock price. Red line shows the real stock price, whereas the blue line is the prediction of our model with a confidence interval of 95%.

4.5 Forecast

After we check the accuracy of the model, we start to make the prediction for one future month. Figure 10 and Table 5 shows the final prediction using the ARIMA(0,1,1) model for a coming month with a confidence level of 95%.

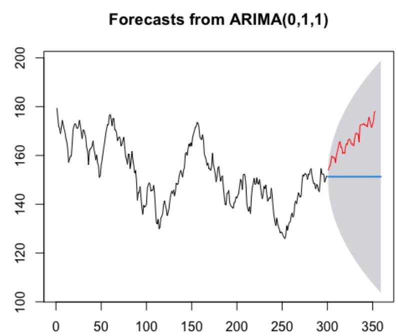

Figure 9: Real and predicted value

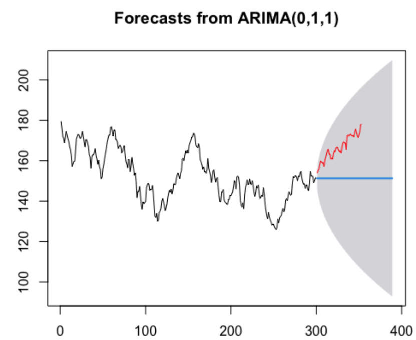

Figure 10: Future prediction

Table 5: Forecast stock price of June 2023

Model	360	361	362	363	364	365	366	367	368	369
Point estimate	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29
Lo 95	103.33	102.93	102.53	102.14	101.75	101.36	100.98	100.60	100.22	99.84
Hi 95	199.24	199.64	200.04	200.44	200.83	201.21	201.60	201.98	202.36	202.73
Model	370	371	372	373	374	375	376	377	378	379
Point estimate	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29
Lo 95	99.47	99.10	98.73	98.74	98.00	97.64	97.28	96.93	96.58	96.22
Hi 95	203.11	203.48	203.85	204.21	204.57	204.93	205.29	205.65	206.00	206.35
Model	380	381	382	383	384	385	386	387	388	389
Point estimate	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29	151.29
Lo 95	95.88	95.53	95.19	94.84	94.50	94.16	93.83	93.49	93.16	92.83
Hi 95	206.70	207.05	207.39	207.74	208.08	208.41	208.75	209.08	209.42	209.75

5 Discussion

Due to its simplicity and operability, ARIMA models are now increasingly used for price forecasting. This study uses time series ARIMA model to forecast the future stock price of Apple. The results show that the model has high predictive power and is helpful to predict the potential of Apple stock.

The limitations of the ARIMA model are that it has strict requirements for large sample sizes and stationary time series, and can only be used to predict linear relationships. This requires complex condition checks before calculations, and many data that do not survive rigorous requirements will have to be predicted using more complex models. Furthermore, while prediction accuracy has shown high efficacy, improving the accuracy of the ARIMA model to near-perfect predictions has remained the primary goal of countless studies over the past few decades. The data we have so far is close to the truth and only gives us a rough idea of where Apple's stock price will go. For the future development of the technology, there is still room to rise.

As mentioned earlier, the technique presupposes that future stock prices follow a trend consistent with previous data, which excludes consideration of urgent factors not reflected in past prices, such as the release of new policies. Therefore, in the actual use of these data should be combined with changes in the external environment to consider, in order to achieve a more accurate forecast.

6 Conclusion

Although short-term stock values may fluctuate with large contingencies around this average value due to the turbulent nature of stock prices, the level of the series over the long term will remain unchanged. Technical analysis provides compelling evidence of a trend in Apple average stock price. For the days in June 2023, the Apple stock price is around 151.3. It also can be seen that ARIMA model has a good prediction effect on stock price.

References

[1] Stock Price - Apple. (n.d.). https://investor.apple.com/stock-price/default.aspx

[2] Wikipedia contributors. (2023). Apple Inc. Wikipedia. https://en.wikipedia.org/wiki/Apple_Inc.

[3] Workneh, D. (2021b, December 14). Stock price prediction using ARIMA Model - Dereje Workneh - Medium. Medium. https://medium.com/@derejeabera/stock-price-prediction-using-arima-model-251ddb4ee52a

[4] Wikipedia contributors. (2023a). Autoregressive integrated moving average. Wikipedia. https://en.wikipedia.org/wiki/Autoregressive_integrated_moving_average

[5] ARAVIND G., ADARSH K. (2021, Aug). Stock Price Prediction using ARIMA Model. IRJET. https://www.irjet.net/archives/V8/i8/IRJET-V8I829.pdf

[6] Hyndman, R. J., Athanasopoulos, G. (2018). Forecasting: Principles and Practice. (2nd ed.) OTexts. https://otexts.org/fpp2/

[7] Adhikari, R. (2013, February 26). An Introductory Study on Time Series Modeling and Forecasting. arXiv.org. https://arxiv.org/abs/1302.6613

[8] Fattah, J., Ezzine, L., Aman, Z., Moussami, H. E., Lachhab, A. (2018). Forecasting of demand using ARIMA model. International Journal of Engineering Business Management, 10, 184797901880867. https://doi.org/10.1177/1847979018808673

[9] Prabhakaran, S. (2022). Augmented Dickey Fuller Test (ADF Test) – Must Read Guide. Machine Learning Plus. https://www.machinelearningplus.com/time-series/augmented-dickey-fuller-test/

[10] Zach. (2020). Ljung-Box Test: Definition + Example. Statology. https://www.statology.org/ljung-box-test/

[11] Ward, E. E. H. M. D. S. a. E. J. (2021, July 28). Applied Time Series Analysis for Fisheries and Environmental Data. https://atsa-es.github.io/atsa-labs/

[12] Zvornicanin, E., Zvornicanin, E. (2023). Choosing the best q and p from ACF and PACF plots in ARMA-type modeling — Baeldung on Computer Science. Baeldung on Computer Science. https://www.baeldung.com/cs/acf-pacf-plots-arma-modeling

[13] Fuqua School of Business. (n.d.). Introduction to ARIMA models. https://people.duke.edu/ rnau/411arim.htm

[14] Workneh, D. (2021, December 14). Stock price prediction using ARIMA Model - Dereje Workneh - Medium. Medium. https://medium.com/@derejeabera/stock-price-prediction-using-arima-model-251ddb4ee52a