Proceedings of Business and Economic Studies, 2022, Volume 5, Issue 5
http://ojs.bbwpublisher.com/index.php/PBES
ISSN Online: 2209-265X
ISSN Print: 2209-2641

# ARIMA and Facebook Prophet Model in Google Stock Price Prediction

**Beijia Jin[1], Shuning Gao[1], Zheng Tao[2]***

[1]College of Business & Public Management, Wenzhou-Kean University, Wenzhou 325060, Zhejiang Province, China

[2]Department of Statistics and Data Science, National University of Singapore, 10 Kent Ridge Crescent 119077, Singapore

****Corresponding author:*** Zheng Tao, e0729756@u.nus.edu



**Abstract:** We use the Autoregressive Integrated Moving Average (ARIMA) model and Facebook Prophet model to predict the closing stock price of Google during the COVID-19 pandemic as well as compare the accuracy of these two models' predictions. We first examine the stationary of the dataset and use ARIMA(0,1,1) to make predictions about the stock price during the pandemic, then we train the Prophet model using the stock price before January 1, 2021, and predict the stock price after January 1, 2021, to present. We also make a comparison of the prediction graphs of the two models. The empirical results show that the ARIMA model has a better performance in predicting Google's stock price during the pandemic.

**Keywords:** ARIMA model; Facebook Prophet model; Stock price prediction; Financial market; Time series

***Online publication:*** October 21, 2022

## 1. Introduction

Since the full-scale outbreak of COVID-19 in the middle of 2020, the financial markets have been severely hit. The world economy was subsequently affected by the spread of the pandemic, which resulted in financial markets slumping, stock prices plummeting, and worldwide companies being affected. Therefore, this time-specific series of changes has its research significance and can provide valuable references for future generations of the financial markets. We selected the closing price of GOOGL from 2007 to 2022 from Yahoo Finance for our research. We chose Alphabet Inc. as our research subject because the parent entity already has a market cap of 1.79 trillion U.S. dollars, which is over Amazon, thus making it the most valuable internet company. Moreover, the firm had over 257.6 billion in revenue in 2021, creating the highest annual net income at 76 million dollars [1]. Therefore, the study of Alphabet Inc. has significant implications for the financial markets. In this research, we use two models, which are the Autoregressive Integrated Moving Average (ARIMA) model and the Facebook Prophet model, to forecast the stock price during the pandemic and compare them with the actual situation in order to discover the advantages and disadvantages of the two models.

In the literature review section, we present the current research achievements of ARIMA and Facebook Prophet as well as the reason we choose these two models to predict stock price during the pandemic. In the method section, we first confirm the stationarity of the data, and then introduce the ARIMA and Facebook Prophet methods, respectively, while explaining their design principles. In the experimental analysis section, we conduct the experiments and analyze the prediction results. This includes the source of the data and its interpretations, the application of the algorithms, and the analysis of the predictions separately with graphs. We then we present a comparative analysis of the two models using errors. In the conclusion section, we highlight the value of this research and its implications in real life.

## 2. Literature review

There are some literatures that emphasize on the use of the ARIMA and Facebook Prophet model. ARIMA, as an abbreviation of "Auto Regressive Integrated Moving Average," is a class of models that interprets a given time series based on its past values, *i.e.*, its lags and lagged forecast errors, to use the equation to predict future values. Stevenson highlights the significance of the predicting method in the stock market, revealing that ARIMA models are instrumental in predicting broad market trends, but the forecasting results differ significantly from reality while using other models [2]. Jadevicius and Huston argue that ARIMA can be used to evaluate a wide range of market price changes, which can help governments and central banks to forecast the increase in national housing prices and enable investors to predict how to invest [3]. ARIMA is an effective model to forecast what we are focusing on in short-term, which is in line with the pandemic period.

Facebook Prophet is an open-source procedure for forecasting time-series data, which works best with time series that have strong seasonal effects and several seasons of historical data. Ignacio Medina *et al*. first introduced this web-based algorithm – Prophet. There are two main options, which are "train" and "predict." Under the "train" part, Prophet builds a prediction rule based on genes by finding the optimal predictor, then the rule is applied to give the prediction of a new dataset with high accuracy [4]. Wen-Xiang Fang *et al*. used Prophet in the finance markets to forecast the stock value of Microsoft Corporation (MSF), and the empirical results showed that Prophet has certain advantages in predicting future price trends [5]. Sumedh Kaninde *et al*. also used Prophet to predict the stock price and concluded that this model can be used to predict stock prices for a long period of time with reasonable accuracy [6]. Due to its training characteristic and effectiveness in predicting future time-series data, including stock prices, we use it to predict the closing stock price of GOOGL during the COVID-19 pandemic to determine whether the model can constantly predict with high accuracy under the contingency. In addition, we also compare the prediction of ARIMA and that of Facebook Prophet.

## 3. Methodology

In order to ascertain reliable predictions of the closing stock price during the COVID-19 pandemic, two algorithms are used. Before performing the algorithm, we need to check the stationary of our data by applying the augmented Dickey-Fuller (ADF) test. Thereafter, we will use the ARIMA model and Facebook Prophet model to predict the stock price during the pandemic.

### 3.1. ARIMA

ARIMA is obtained from combining differencing with autoregressive model and moving average model, in which the integration refers to the reverse of differencing. It was first popularized by Box and Jenkins [7], and it belongs to a class of models that interprets a given time series based on its past values, *i.e.*, its lags and lagged forecast errors, to use the equation to predict future values [8].

The ARIMA model can be written as follows:

$$y'_t = c + \emptyset_1 y'_{t-1} + \cdots + \emptyset_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon_t \tag{1}$$

where $y'_t$ is the differenced time series data, and the functions on the right contain the lagged values of $y'_t$, which can be used as the predictors for prediction. Any "non-seasonal" time series that exhibits a pattern and is not random white noise can be modeled with an ARIMA model. An ARIMA(p,d,q) model is characterized by three terms: p, d, and q [9].

(1) p: The order of the autoregressive term
(2) q: The order of the moving average term
(3) d: The number of differences required to make the reversed time series stationary

We will use the backshift notation when building the complicated models, the equation can be written as follows [8]:

$$\left(1 - \emptyset_1 B - \cdots - \emptyset_p B^p\right)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q)\varepsilon_t \tag{2}$$

### 3.2. Facebook Prophet

Facebook Prophet is a time-series data forecasting algorithm, which was developed by Facebook's data science team. It is based on an additive model, in which the nonlinear trend fits to the seasonality of the year, week, and day, along with the holiday effect [10].

The Prophet model can be expressed as follows:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon(t) \tag{3}$$

where $y(t)$ refers to the additive regression model, $g(t)$ refers to trends, $s(t)$ refers to seasonality, which shows periodic changes, $h(t)$ refers to holiday, which reflects predictable annual abnormal days on irregular schedules, $\varepsilon(t)$ refers to error, which is typically modeled as normally distributed noise, showing information that is not included in the model [12]. Due to the seasonality and holiday components, Prophet is most effective for time series with strong seasonal effects and historical data of several seasons (however, we do not expect seasonal or holiday effects for stock market data).

.

## 4. Empirical analysis

The data set comprises 15 years of stock prices from 2007–2022 from Google Inc., whose ticker symbol is "GOOGL." We use the R package "Quantmod" to obtain publicly available data from Yahoo Finance. The data selected include the company's close price on Google. Moreover, we have a total of 3,946 data with no missing values. The structure of the data is shown in **Table 1**. The average closing price of Google Inc. is 41.065, and the minimum value is 6.443.

**Table 1.** Summary statistics

| Variable | N | Mean | Std. Dev. | Min | Pctl. 25 | Pctl. 50 | Pctl. 75 |
|---|---|---|---|---|---|---|---|
| GOOGL.Close | 3945 | 41.065 | 34.853 | 6.442 | 14.515 | 28.065 | 55.84 |

### 4.1. Data analysis and processing

**Figure 1** clearly shows the trend of the stock over a period of time. The volatility of the stock market can be seen clearly. Though there are peaks and troughs in the short term, we can see a clear upward trend in the long term, but it is not linear. At the same time, we can see a significant downward trend in the share price after the COVID-19 period. We want to ascertain whether a simple linear regression can capture the trend of the data and help us forecast the stock price. We assume that a simple linear regression does not capture the trend of the data.

We use date as the independent variable and stock price as the dependent variable. We obtain a goodness-of-fit of 74.4% through the linear regression model. By comparing the p-value tables, the results are found to be statistically significant since the p-value is less than 0.001, thus the original hypothesis is rejected. This means a relationship exists between the date and the stock price. The coefficient of the date indicates that with each changing day, the closing stock price increases by \$0.02.

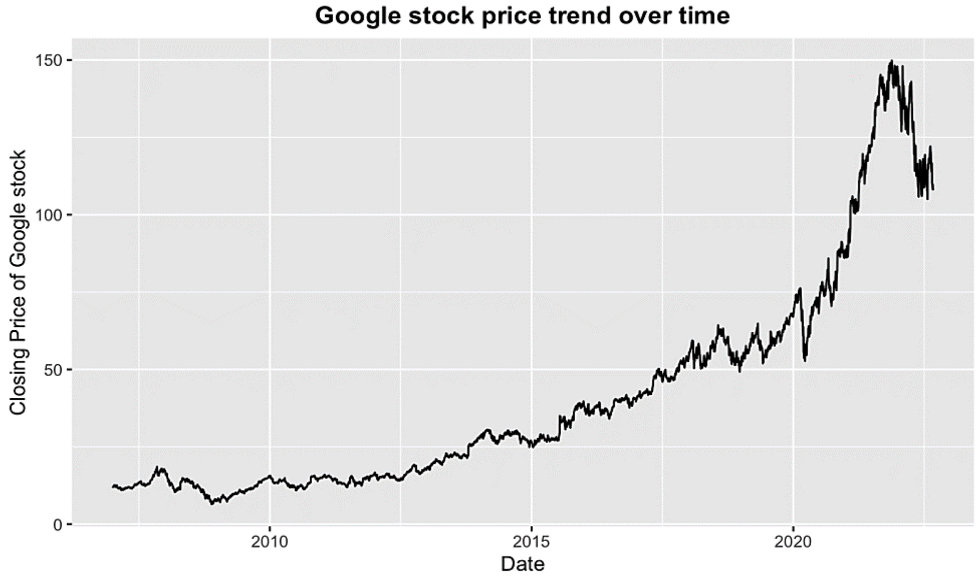


**Figure 1.** Google stock price trend over time

Based on **Figure 2**, we can clearly see that the time series is not stationary, *i.e.*, its variance is not constant across time. In addition, a clear positive trend indicates that there is no recovery to a constant value, thus suggesting that the time series is not stationary on average. In general, these two properties are undesirable for better and easier time-series modeling. In order to confirm that it is non-stationary, we compared the original time-series data with the transformed data. We applied log transformation and Box-Cox transformation to do so. Moreover, we used the popular log transformation method to determine the mean stationarity.

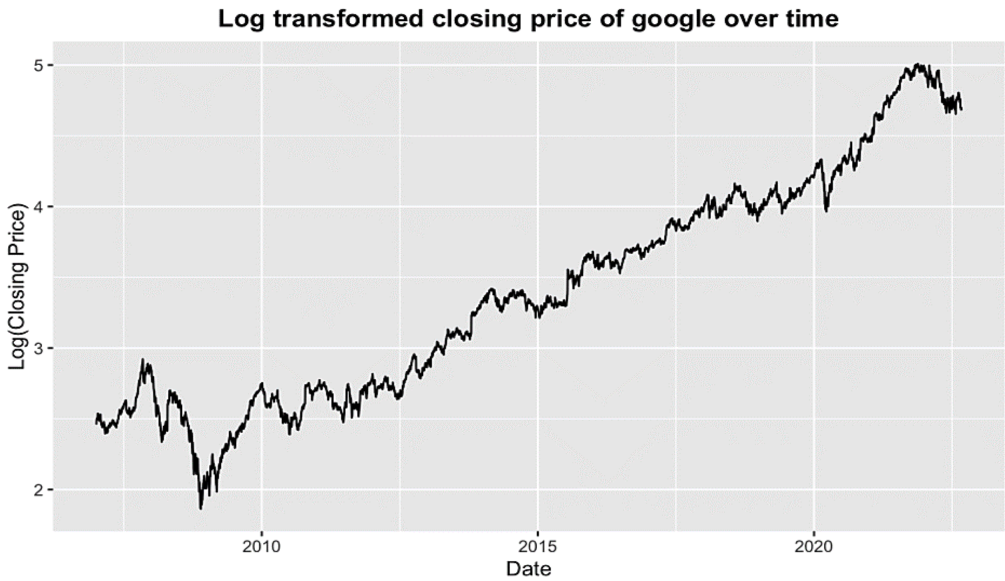


**Figure 2.** Log-transformed closing price of Google over time

The ADF test is a standard statistical test used to determine whether a given time series is stationary. It is one of the most commonly used statistical tests when analyzing the stationarity of a series, which is largely dependent on the presence of a unit root. The time series is non-stationary if the unit root is present.

The null hypothesis of the ADF test is as follows: the covariance of the first lag of Y is equal to 1. When the test statistic is below the critical value, we reject the null hypothesis and infer that the time series is stationary. The ADF test was performed on the original closing price, and the closing price was Box-Cox transformed. Three sets of p-values were derived from the ADF test when the lag order is 15 (**Table 2**).

**Table 2.** ADF test on three variables

| Variable | Dickey-Fuller | p-value |
|---|---|---|
| GOOGL.Close | -1.68 | 0.715 |
| GOOGL.log_close | -3.15 | 0.0967 |
| GOOGL.bc_close | -3.33 | 0.647 |

Although the transformed p-values dropped sharply, we cannot confidently reject the null hypothesis at 95% significance level. In order to address the non-smoothness of the mean, we calculated the first-order lag of the closing price. However, we also have outliers that deviate more than most points. By running an ADF test on the first-order lagged closing price to see if we have achieved the mean stationary property, the p-value is apparently equal to 0.01 when the Dickey-Fuller is equal to -15.484, below the critical value of 0.05. This indicates that the series is now stationary.

### 4.2. ARIMA model

We will use Akaike information criterion (AIC) and Bayesian information criterion (BIC) to evaluate the performance of the model. According to the BIC value, the order of the best model is c(0,1,1). We then bring ARIMA(0,1,1) into the model.

In **Figure 3**, the dashed line represents the model's prediction, and the solid line represents the actual value. As we can observe from **Figure 3**, the dashed and solid lines have a very high degree of overlap, and ARIMA(0, 1, 1) has an excellent performance in predicting stock prices. It not only predicts the overall upward trend of stock prices well, but also the upward and downward fluctuations of stock prices in the short term, except that the value of the predicted fluctuations is higher than the actual stock prices. Moreover, this model perfectly predicts the overall downward trend of stock prices after the COVID-19 pandemic.

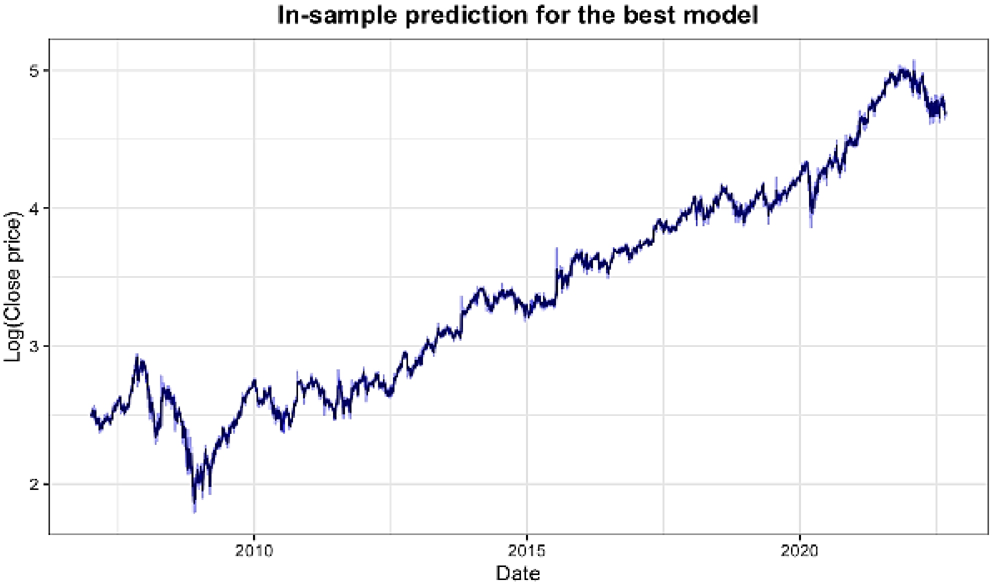


**Figure 3.** In-sample prediction for the best model

The root mean square error (RMSE) for the ARIMA model is 0.01842, which means that the error between the predicted value and true value is 0.01842461, signifying a relatively small error. This indicates that this model completes the prediction with a relatively high degree of accuracy. We will predict five time periods from the best model.

### 4.3. Facebook Prophet model

We consider the data before January 1, 2021, as the training dataset and the data from January 1, 2021, to present as the test dataset.

In **Figure 4**, the x-axis is time in years, while the y-axis is the closing price of Google, which is the target variable. The black points represent the real data, whereas the blue points represent the predicted values; the light blue area represents the 95% confidence interval. Using the dataset before January 1, 2021, the Prophet model predicts the stock closing prices with relatively high accuracy. However, when predicting short-term fluctuations, the value of the fluctuations is in fact lower than the real data. Using the dataset from January 1, 2021, to present, the model predicts a continuous upward trend in the closing stock price, which is also not consistent with the actual situation, in which the stock price shows a significant downward tendency.

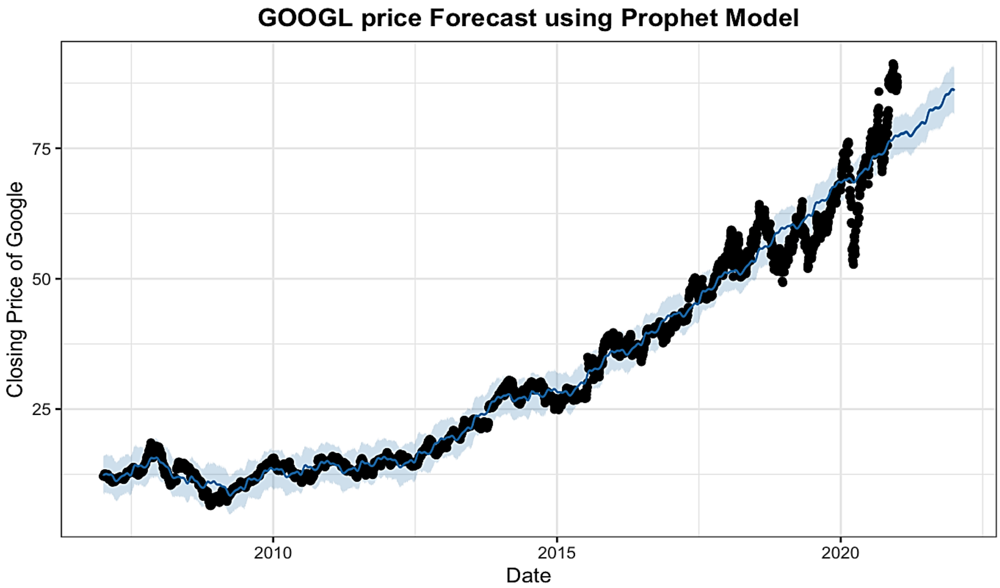


**Figure 4.** GOOGL price forecast using the Prophet model

### 4.4. Comparison between ARIMA and Facebook Prophet

There are some parameters that can be used to examine the performance of the Prophet model. We use RMSE for comparison, in which the lower the RMSE, the lesser the error. The RMSE of ARIMA and Prophet is 0.01842 and 47.56, respectively. Comparing the values, the error of the Prophet model is significantly larger than that of ARIMA. This indicates that ARIMA outperforms Prophet in terms of accuracy and overall performance when predicting Google's closing stock price. Moreover, considering the negative impact of COVID-19 on stock prices, ARIMA well predicts this situation, while Prophet derives the exact opposite trend, further validating its imperfection in predicting stock prices.

## 5. Conclusion

In this study, we used both the ARIMA model and the Facebook Prophet model to predict the trend of stock prices during the pandemic period. The RMSE shows that the error of the ARIMA model is less than that of the Facebook Prophet model, indicating that the ARIMA model is more accurate than the Facebook Prophet model. This suggests that the ARIMA model will play a significant role in forecasting the trend of stock prices during a particular time series. Therefore, the ARIMA model merits further investigation.

**Disclosure statement**

The authors declare no conflict of interest.

**Author contributions**

Z.T. conceived the idea of the study and performed the experiments. S.G. and B.J. analyzed the data and wrote the paper.

**References**

[1] Statista Research Department, 2022, Alphabet Inc. – Statistics and Facts, Statista, viewed September 15, 2022, https://www.statista.com/topics/6042/alphabet-inc/#dossierKeyfigures

[2] Stevenson S, 2007, A Comparison of the Forecasting Ability of ARIMA Models. Journal of Property Investment & Finance, 25(3): 223–240. https://www.emerald.com/insight/content/doi/10.1108/IJHM A-04-2014-0010/full/html

[3] Jadevicius A, Huston S, 2015, ARIMA Modelling of Lithuanian House Price Index. International Journal of Housing Markets and Analysis, 8(1): 135–147. https://www.emerald.com/insight/content/doi/10.1108/IJHMA-04-2014-0010/full/html

[4] Medina I, Montaner D, Tarraga J, et al., 2006, Prophet, a Web-Based Tool for Class Prediction Using Microarray Data. OUP Academic. Oxford University Press, https://academic.oup.com/bioinformatics/article/23/3/390/236076

[5] Fang W-X, Lan P-C, Lin W-R, et al., 2019, Combine Facebook Prophet and LSTM with BPNN Forecasting Financial Markets: The Morgan Taiwan Index. 2019 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS), http://doi.org/10. 1109/ISPACS48206.2019.8986377

[6] Kaninde S, Mahajan M, Janghale A, et al., 2022, Stock Price Prediction Using Facebook Prophet. International Conference on Automation, Computing and Communication 2022 (ICACC-2022), ITM Web Conf, 44: 03060. https://doi.org/10.1051/itmconf/20224403060

[7] Stellwagen E, Tashman L, 2013, Arima: The Models of Box and Jenkins. Foresight: The International Journal of Applied Forecasting, 2013(30): 28–33. https://ideas.repec.org/a/for/ijafaa/y2013i29p28-33.html

[8] Hyndman RJ, Athanasopoulos G, 2018, 8.5 Non-seasonal ARIMA models, in Forecasting: Principles and practice (2nd Edition), OTexts, Melbourne, Australia. https://otexts.com/fpp2/non-seasonal-arima.html

[9] Fattah J, Ezzine L, Aman Z, et al., 2018, Forecasting of Demand Using ARIMA Model. International Journal of Engineering Business Management, 2018: 10. https://doi.org/10.1177/1847979018808673

[10] Forecasting at Scale, n.d., Prophet, viewed September 18, 2022, https://facebook.github.io/prophet