Prediction on the Price of Technology Stock and COVID-19 Based on Multiple Linear Regression

Xitao Cen [1, *]

[1]College of Art and Science, New York University, NY, the USA

*Corresponding author: xc1475@nyu.edu

Abstract: The outbreak of COVID-19 hit the world economy. Represented as an ownership share in the company, stock is one of the most common financial products and economic indicator. However, the stock market is known as being volatile, which makes it hard to predict. Many factors may affect the stock market, e.g., interest rate, trade wars, political scandals, and widespread pandemic such as COVID-19. While stock prediction is extremely challenging, the passage uses multiple linear regression model to predict the price of three technology stock: Apple, Amazon, and Google. The author builds the linear model based on the 5 factors of stock price, including VR (Volume Variation Index), WR (Williams Overbought/Oversold index), ATR (Average True Range), TRIX (Triple Exponential Average) and Log Return of the Close. From the website, yahoo finance, the author obtained the raw data of these three technology companies in recent five years. Compared the final linear regression model with the real adjusted closing price of the stock, there is a clear difference between them. Additionally, the passage finds that TRIX, ATR and Log Return of the Close have larger influence on the predicted closing price as they have larger coefficient. According to the regression model, one can predict the price of the stock in the future. Based on the predicted stock price, it is feasible to investigate the future influence of COVID-19 on the technology price. These results shed light on guiding further exploration of impact of COVID-19 on the stock market.

Keywords: Stock price prediction, COVID-19, technology industry, Volume Variation Index, Williams Overbought/Oversold index, Average True Range, Triple Exponential Average.

1. Introduction

Since the end of 2019, the COVID-19 widely spread around the world. The global economy went into a recession after the outbreak of coronavirus. The rapid spread of the COVID-19 largely reduces the human interaction and movement. The COVID-19 also had a strong impact on the stock market, which is one of the most important indicators of the world economy. On 11th March 2020, World Health Organization (WHO) officially declared COVID-19 a pandemic [1]. It is not the first time for human to experience the pandemic and people have record of past pandemic's influence on the stock market. The pandemic severely hit the economy around the world and impact people's daily life [2]. through different ways, like reducing trading and killing labors which will decrease the production of the market. When the industries are in recession, the stock price of those industries will also decrease. In 2003, when the SARS infected people worldwide, the S&P 500, which represents the average stock price of 500 companies in American, lost 12.8% of its value [3]. Chandra once did the similar research as this paper. He used the Bayesian neural networks to predict the stock price and finds the influence of COVID-19 on the prediction of the stock price [4].

There are also other people predicting the stock price based on the linear regression model. Chaudhary once predicted the adjusted closing price of Netflix with the similar linear regression model [5]. He followed the similar steps as the study did and finally predicted the stock price well after testing the equation. Cakra and Trisedya also predicted the stock price based on the linear regression model and find the influence of sentiment on the stock price [6]. Other models are also used to predict the stock price, e.g., LSTM and ARIMA. Charles used the autoregressive integrated moving average models to predict the stock price and this model has a strong potential for short-term prediction [7]. Selvin and Soman also predicted the stock price, but they used a different model called

LSTM (Long Short-Term Memory). The model has the learning problem and fails to capture the dynamical changes precisely [8].

This passage predicts of the technology stock price to find the future influence of COVID-19 on the stock market. The stock price change of the stock, as an indicator of the economy, could help to see the future economic fluctuation. To predict the stock price, the author uses linear regression model to run the data of three technology stocks. Finding the correlation of each factor, the research predicts the price of those stocks based on the linear equation of the adjusted closing price. The rest part of the paper is organized as follows. The Sec. 2 will discuss the methodology which is about the data and models. The Sec. 3 will conclude on the result and discuss the limitation of the models based on the result of the models. In the last section, the passage will discuss the future development and the significance of this research.

2. Methodology

The data is collected from the website "yahoo finance" and decides to choose the top three technology stocks on the S&P 500. The data was chosen from April 27th, 2017, to April 27th, 2020, and the total number of data is 1259 because the stock market was not open during the weekends and holidays. The independent variable is the ACP (Adjusted Closing Price), and the dependent variable are the stock statistic, including VR (Volume Variation Index), WR (Williams Overbought/Oversold index), ATR (Average True Range), TRIX (Triple Exponential Average) and LRC (Log Return of the Close). The linear function of the model can be given as:

$$ACP = constant + w_1 VR + w_2 WR + w_3 ATR + w_4 TRIX + w_5 LRC \quad (1)$$

All the data could be shown on the graph with stock price on y-axis and date on x-axis which will show the change of stock price. The linear regression model is used to predict the future price of three technology companies based on the past record. The first step is to see the correlation between the dependent and independent variables. The larger coefficient between two variables represents the higher correlation between them. Based on the Table. 1, Table. 2, and Table. 3, the independent variable, the adjusted closing price, is highly related to ATR (Average True Range), TRIX (Triple Exponential Average) and Log Return of the and has lower correlation with VR (Volume Variation Index), WR (Williams Overbought/Oversold index).

Table.1. Apple's correlation between each factor

	adj close	volume	change	vr	wr	log-ret	trix	atr
adj close	1.00	-0.21	0.02	-0.03	0.01	0.02	0.13	0.85
volume	-0.21	1.00	-0.08	-0.05	-0.25	-0.10	-0.19	0.08
change	0.02	-0.08	1.00	0.14	0.43	1.00	0.03	-0.01
vr	-0.03	-0.05	0.14	1.00	0.49	0.14	0.68	-0.19
wr	0.01	-0.25	0.43	0.49	1.00	0.44	0.43	-0.15
log-ret	0.02	-0.10	1.00	0.14	0.44	1.00	0.03	-0.02
trix	0.13	-0.19	0.03	0.68	0.43	0.03	1.00	-0.11
atr	0.85	0.08	-0.01	-0.19	-0.15	-0.02	-0.11	1.00

Table.2. Amazon's correlation between each factor

	adj close	**volume**	**change**	**vr**	**wr**	**log-ret**	**trix**	**atr**
adj close	1.00	-0.21	0.02	-0.03	0.01	0.02	0.13	0.85
volume	-0.21	1.00	-0.08	-0.05	-0.25	-0.10	-0.19	0.08
change	0.02	-0.08	1.00	0.14	0.43	1.00	0.03	-0.01
vr	-0.03	-0.05	0.14	1.00	0.49	0.14	0.68	-0.19
wr	0.01	-0.25	0.43	0.49	1.00	0.44	0.43	-0.15
log-ret	0.02	-0.10	1.00	0.14	0.44	1.00	0.03	-0.02
trix	0.13	-0.19	0.03	0.68	0.43	0.03	1.00	-0.11
atr	0.85	0.08	-0.01	-0.19	-0.15	-0.02	-0.11	1.00

Table.3. Google's correlation between each factor

	adj close	volume	change	vr	wr	trix	atr	log-ret
adj close	1.00	-0.22	0.02	0.02	0.05	0.14	0.75	0.02
volume	-0.22	1.00	-0.09	-0.04	-0.24	-0.24	0.11	-0.10
change	0.02	-0.09	1.00	0.12	0.42	-0.01	0.00	1.00
vr	0.02	-0.04	0.12	1.00	0.38	0.53	-0.07	0.12
wr	0.05	-0.24	0.42	0.38	1.00	0.35	-0.14	0.42
trix	0.14	-0.24	-0.01	0.53	0.35	1.00	-0.21	-0.01
atr	0.75	0.11	0.00	-0.07	-0.14	-0.21	1.00	-0.01
log-ret	0.02	-0.10	1.00	0.12	0.42	-0.01	-0.01	1.00

One runs the Ordinary Least Square regression and tests the fitness of the OLS model based on R square and MSE (Mean Square Error). R square represents the proportion of the variance for a dependent variable which is explained by an independent variable in the model [9]. The formula of R^2 can be given as:

$$R^2 = 1 - \frac{SSR}{SST} \tag{2}$$

MSE measures how close a fitted line is to data points which can test whether the model fits well [10]. The formula of MSE can be given as:

$$MSE = \frac{1}{n}\Sigma(actual - forcast)^2 \tag{3}$$

For R^2, the result should be larger than 0.7 which represents a strong relationship. For MSE, the model fits better when the value of MSE is smaller. The next part will discuss the result based on the methodology the research chooses.

3. Results & Discussion

The R^2 for APPLE, Amazon and Google are 0.78, 0.62 and 0.67, separately. The value of R^2 shows that the relationship between the independent and dependent variable is strong for Apple which is larger than 0.7 and is moderate for Amazon and Google which is between 0.5 and 0.7 [11]. The MSE for Apple, Amazon and Google are 1324.91, 317370.94 and 315667.12 separately. The MSE is high due to large amount of data for all three companies. The high value of MSE represents that the OLS model may not fit well for the three companies. For Apple, the fitted equation can be given as:

$$ACP = 28.80 - 0.05VR + 0.09WR + 33.52ATR + 32.85TRIX + 29.40LRC \tag{4}$$

For Amazon, the fitted equation can be given as:

$$ACP = 990.53 - 0.12VR + 0.70WR + 306.41ATR + 478.58TRIX + 29.40LRC \tag{5}$$

For Google, the fitted equation can be given as:

$$ACP = 750.82 - 1.60VR + 2.06WR + 907.50ATR + 126.78TRIX + 29.80LRC \tag{6}$$

From the equations above, the coefficient of ATR is largest for all equations which shows it has the largest impact on the adjusted closing price among the five dependents. Instead, the coefficient of VR is smallest for all equations which shows it has the smallest impact on the adjusted closing price among the five dependents.

Based on linear equation of the three companies, the author built the model and tested the accuracy between the actual price and predicted price. In the first step, the research chooses ten percent of the data as tested data, which is 126 entries. Using the equation of each stock, the tested data is used to find the predicted adjusted closing price in last 126 dates which is the most recent data. Then, the scattered plot is drawn for the predicted price and actual price of each stock. Based on the Figs. 2-3, one can vaguely see the linear equation which shows that the linear regression may not be the best predicted method for the stock price. The dot spreads without any rules for Figs. 2-3. One can approximately see a linear line for Figs. 1.

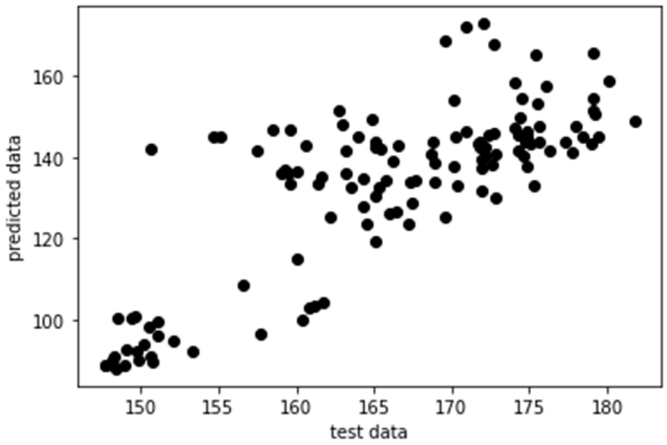

Figure 1. Scatter plot of APPLE's predicted and tested data.

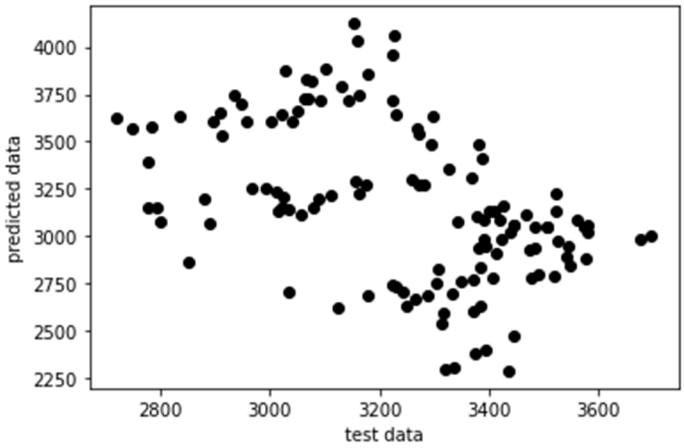

Figure 2. Scatter plot of AMAZON's predicted and tested data.

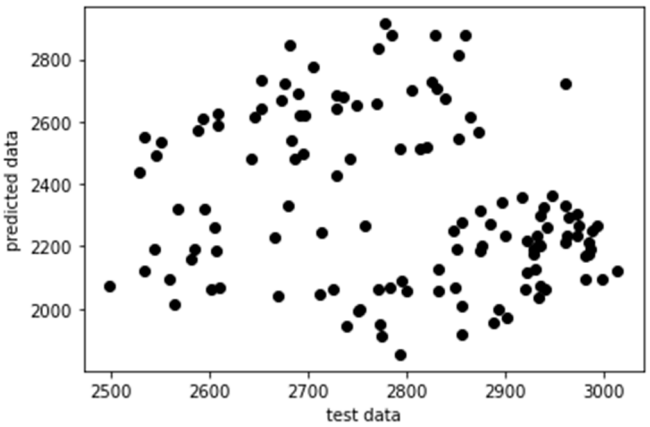

Figure 3. Scatter plot of GOOGLE's predicted and tested data.

Besides, based on the Fig. 4-6, one can find that the blue line, which is the predicted line, seems not to fit the actual price perfectly because there is a large difference between the actual and predicted price. Seen from Fig. 4, the predicted price for Apple has the similar trend as the actual price but has a price different overall. As illustrated in Fig. 5, the predicted price for Amazon has an increasing trend. However, the trend of the actual price is going downward. As presented in Fig. 6, the predicted price for Google is similar to the actual price from January to March in 2022. However, for other time, the predicted price is different from the actual price. After building the model and get the equation, one can use this equation to predict the future stock price and see the trend of the stock price based on the graph. The COVID-19 had a large influence on the stock price in recent two years as the

pandemic blocked humans' activities, including trading and interactions. With the prediction model, one could see the future influence of COVID-19 on the stock price as the virus keeps mutation and spreading around the world which may not be controlled nowadays.

Figure 4. Predicted and actual model of APPLE.

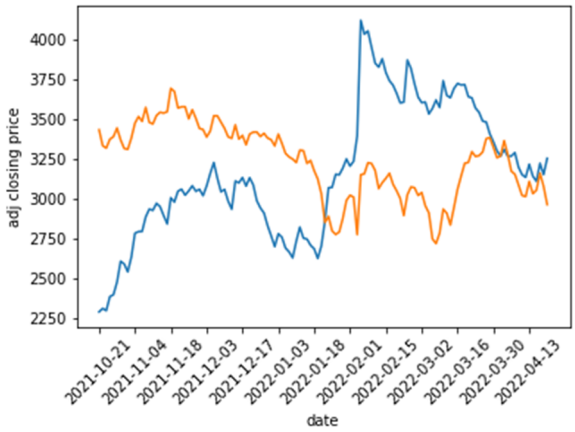

Figure 5. Predicted and actual model of AMAZON.

Figure 6. Predicted and actual model of GOOGLE.

The prediction model also has several limitations. To be specific, the used model is linear regression model which is the simplest one and may not be the most suitable one. Based on the result, the predicted value seems to have a large difference from the actual value which decreases the accuracy of the model. Moreover, this research only focused on the three technology stocks which may only represent the top of the technology stock. However, for other stocks, this model may have a different result which needs to be tested. For the dataset, the 5-years data may be not enough to

conclude the whole stock because people don't know how the stock works before 5 years and the stock price may suddenly increase or decrease due to some special situations.

4. Conclusions

In summary, this study investigates the impact of COVID-19 on the stock price based on the prediction of the stock price, which can be calculated from the linear regression model. The research firstly introduces methodology, including how to collect data and which model is used for the prediction of stock price. Subsequently, it discusses the result of the model and makes a conclusion based on the figure and the equation. Finally, the study mentioned the limitation of model and the procedure. In the future, one can build other models and find which one is most suitable for prediction of stock price among all the models. Moreover, stocks of other industries can be chosen, and one may try to find the ways to conclude the whole stock market. Based on the result of the prediction, one can find the effect of COVID-19 on the stock market better with the most suitable model. Overall, these results offer a guideline for the prediction of stock price though models and helps to investigate the impacts of the COVID-19 on it by comparing the prediction price and the current price of stock.

References

[1] Cox J, Greenwald D L, Ludvigson S C. What explains the COVID-19 stock market?[R]. National Bureau of Economic Research, 2020.

[2] Jabeen S, Farhan M, Zaka M A, et al. COVID and World Stock Markets: A Comprehensive Discussion[J]. Frontiers in Psychology, 2022: 4837.

[3] Dure E. The history of pandemics and stock markets [EB/OL]. Investopedia, Investopedia, 2021-10-03. Information on: https://www.investopedia.com/the-history-of-pandemics-and-stock-markets-5093256.

[4] Chandra R, He Y. Bayesian neural networks for stock price forecasting before and during COVID-19 pandemic[J]. Plos one, 2021, 16(7): e0253217.

[5] Team T AI. Stock price prediction model for netflix[EB/OL]. Towards AI, Information on: https://towardsai.net/p/machine-learning/netflix-stock-prediction-model-a-comparative-study-of-linear-regression-k-nearest-neighbor-knn-4527ff17939b.

[6] Cakra Y E, Trisedya B D. Stock price prediction using linear regression based on sentiment analysis[C]. 2015 international conference on advanced computer science and information systems (ICACSIS). IEEE, 2015: 147-154.

[7] Ariyo A A, Adewumi A O, Ayo C K. Stock price prediction using the ARIMA model[C]. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation. IEEE, 2014: 106-112.

[8] Selvin S, Vinayakumar R, Gopalakrishnan E A, et al. Stock price prediction using LSTM, RNN and CNN-sliding window model[C]. 2017 international conference on advances in computing, communications and informatics (icacci). IEEE, 2017: 1643-1647.

[9] Fernando J. What is R-squared? [EB/OL]. Investopedia, Investopedia, 2022-02-08. Information on: https://www.investopedia.com/terms/r/r-squared.asp.

[10] What are mean squared error and root mean squared error? [EB/OL]. Technical Information Library, -1-11-30. (-1-11-30). Information on: https://www.vernier.com/til/1014.

[11] [11] Alhyari S. What is the acceptable R-squared in the information system research [EB/OL]. Information on: https://www.researchgate.net/post/What-is-the-acceptable-R-squared-in-the-information-system-research-Can-you-provide-some-references.