Stock Price Prediction: Comparison between ARIMA Model and LSTM

Use Boeing (BA) and Zoom Video Communications (ZM) during COVID-19 Period

Xizhu Lin

Abstract

This study investigates the impact of the COVID-19 pandemic on Boeing and Zoom stock prices using historical daily data from January 1, 2020, to May 12, 2023. Employing ARIMA and LSTM models, the analysis focuses on forecasting stock prices amid significant market volatility. The dataset undergoes comprehensive pre-processing, and model parameters are selected based on statistical tests and criteria like AIC and BIC. Performance evaluation metrics, including MSE and RMSE, reveal distinctive outcomes for Boeing and Zoom, shedding light on the strengths and limitations of ARIMA and LSTM models in capturing stock price dynamics. The findings contribute to a nuanced understanding of stock market behavior during crises and the effectiveness of different forecasting models.

Data Introduction. In this study, I utilize historical daily stock price data for two prominent companies, Boeing (BA) and Zoom Video Communications (ZM), obtained from Yahoo Finance using the yfinance Python library. The dataset spans from January 1, 2020, to May 12, 2023, during the COVID-19 period. No missing data is found during this period.

Data Format. I choose the closing prices of these two companies to serve as a crucial indicator of the market sentiment and are widely used in financial analyses.

Date	Close
2020-01-02	333.320007
2020-01-03	332.760010
2020-01-06	333.739990
2020-01-07	337.279999
2020-01-08	331.369995

Table 1: Sample Data Format.

The dataset is structured as a time series, with each row representing a specific trading day. The 'Date' column denotes the respective date, and the data type is datetime64, while the 'Close' column indicates the closing price of the stock on that particular day, and the data type is float64.

Data Pre-processing

(1) First-order differencing: Apply first-order differencing to the closing prices to stabilize the mean and make the time series stationary,

(2) Log transformation of prices: Take the logarithm of the stock prices to help mitigate the impact of extreme values and normalize the distribution.

(3) Log returns calculation: Apply first-order differencing to the log transformation of closing prices. Log returns are often preferred in financial analysis as they provide a percentage change in the stock price, making them more interpretable.
(4) Use the Min-Max method to scale price data: Apply the MinMaxScaler to normalize the original stock prices between the range of 0 and 1.
(5) Build a lookback window with 10 timestamps: To prepare the training data for input into LSTM models, build a function that employs a lookback window approach to create sequences of input features and their corresponding target values.

Companies and Period Chosen Reason. Boeing's (BA) stock price experienced a sharp decline at the onset of the COVID-19 pandemic. Subsequently, as the epidemic came under control, the stock rebounded but exhibited significant fluctuations. In contrast, Zoom Video Communications' (ZM) stock price witnessed a sharp rise at the onset of the pandemic, soaring from around \$70 to more than \$500. However, as the epidemic was brought under control, the stock swiftly retreated to approximately \$70.

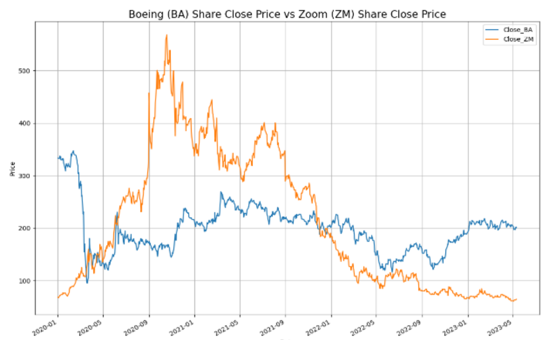
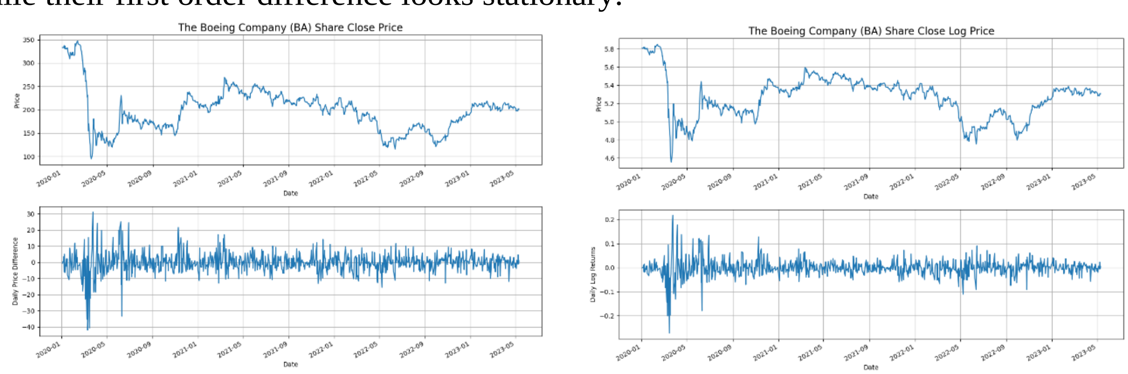


Figure 1: Time series data of Boeing (BA) and Zoom (ZM) stock closing price

These distinctive trajectories provide a dynamic context for evaluating the forecasting capabilities of ARIMA and LSTM models amid substantial market volatility, to test which model has a better ability to capture the dramatic changes in stock prices.

Training Set and Test Set Split. To compare the performance of different models on different stocks, I use the same period for Boeing (BA) and Zoom (ZM). Because I use the time series data to do the price prediction, I split the first 80% of the total dataset as the training dataset and the last 20% of the total dataset as the test dataset, to make sure that the units are comparable.

Stationarity. I observe that stock price and log stock price time series appear non-stationary, while their first order difference looks stationary.



(a) BA Price and Price Difference

(b) BA Log Price and Log returns

(c) ZM Price and Price Difference

(d) ZM Log Price and Log returns

Figure 2: Time series data of price, price difference, log price, and log returns

A formal statistical test to detect unit roots is the Augmented Dicky-Fuller test. The details of the ADF test for all eight-time series are in the Model ARIMA Parameter Selection part.

Model Choice.

ARIMA. ARIMA(p,d,q) is a classical time series forecasting model that combines autoregressive (AR) and moving average (MA) components. The model is particularly effective in capturing linear trends and seasonality in time series data. The three main components of ARIMA are:

● AutoRegressive (AR): The AutoRegressive (AR) component, denoted by p, represents the correlation between the current value and its past values.

● Integrated (I): The Integrated (I) component, denoted by d, involves differencing the time series data to achieve stationarity.

● Moving Average (MA): The Moving Average (MA) component, denoted by q, incorporates the weighted sum of past error terms.

ARIMA is well-suited for time series data with clear patterns and trends, making it a valuable tool for financial forecasting.

LSTM (Long Short-Term Memory). Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed for sequential data, known for its ability to capture long-term dependencies and intricate patterns. Unlike traditional neural networks, LSTM has a memory cell that allows it to store and access information over extended sequences. Key features of LSTM include:

● Memory Cell: Enables the model to retain and utilize information over prolonged periods.

● Gates: Control the flow of information, facilitating the selective update or forgetting of specific information.

● Non-linearity: Allows the model to capture complex, non-linear relationships in the data.

LSTM is particularly adept at handling time series data with irregular patterns, making it suitable for capturing the dynamic nature of stock prices, especially during periods of volatility.

Model Evaluation.

Evaluation Criterion: MSE, RMSE

In Model Evaluation, I choose Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) as criteria to evaluate model performance. Comparing the MSE and RMSE on the test dataset, the model with lower MSE and RMSE means better performance.

Mean Squared Error (MSE) is defined as the average of the squared differences between the predicted value ($\hat{y}_i$) and the actual value (y_i). Mathematically, it can be expressed as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

Root Mean Squared Error (RMSE), calculated as the square root of the MSE:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}$$

Prediction: Evaluate Training and Out-of-Sample Performance

ARIMA. Use arima_model.predict for in-sample forecasting and arima_model.get_forecast for out-of-sample forecasting. Apply an inverse transformation to revert predicted values to the original scale and calculate MSE and RMSE using the respective original prices.

LSTM. Use lstm_model.predict to get the predicted values for training data and test data separately and apply the Min-Max scaler inverse transformation to the predicted values to revert them to the original scale and calculate MSE and RMSE using the respective original prices.

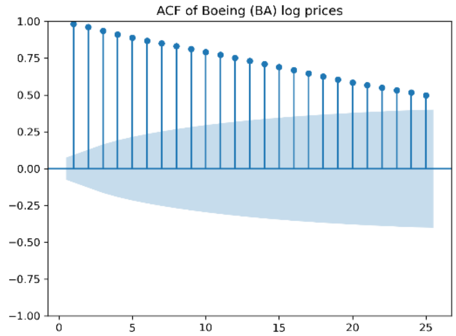
Model Parameters Selection.

ARIMA. Find the parameters of (p,d,q)

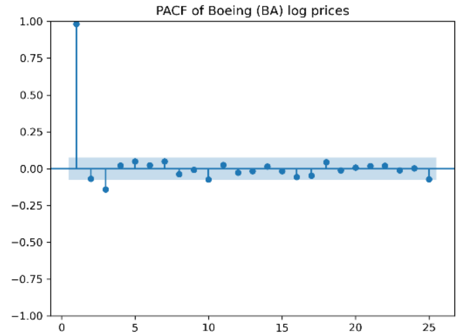
First, I need to determine d for. The idea is to take d^{th} order difference to the time series to get a stationary sequence. To test stationary, I use the Augmented Dicky-Fuller (ADF) test to find the parameter d. The null hypothesis is non-stationarity (the existence of unit root). The alternative hypothesis is that the time series is stationary.

After the ADF test for price, log price, the first order difference of price, and the first order difference of log price (log returns) for both Boeing and Zoom, I found that for both Boeing and Zoom, the first order difference of price, and the log returns time series rejects the null at 99% confidence level, the price and log price time series of Boeing rejects the null at 95% confidence level, but the price and log price time series of Zoom fails to reject the null at any conventional confidence level. In conclusion, I will use the first-order difference of log price (log returns) for both Boeing and Zoom to train the ARIMA model, I will use the training log price time series data with d=1.

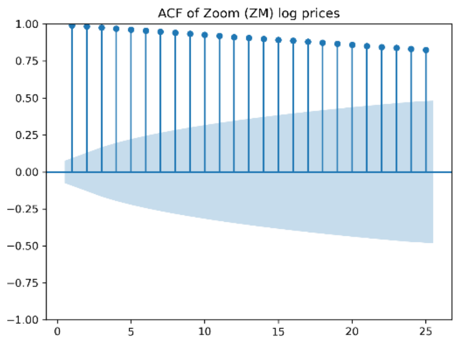
Then, I need to find the value of parameters p and q in the ARIMA model. I use two methods separately. First, I choose proper p and q from ACF and PACF plots and use the p-value of predictors to test whether they are significant. For Boeing and Zoom, I plot ACF (for choosing q) and PACF (for choosing p) of log price together. For Boeing, the proper p=2 and q=0. For Zoom, the proper p=0 and q=0. Then I use the Boeing training log prices for ARIMA (2,1,0) and the Zoom training log prices for ARIMA (0,1,0). Also, I use the p-value test to check whether all predictors in these two models are significant. At a 95% confidence level, all predictors in these two models are significant.



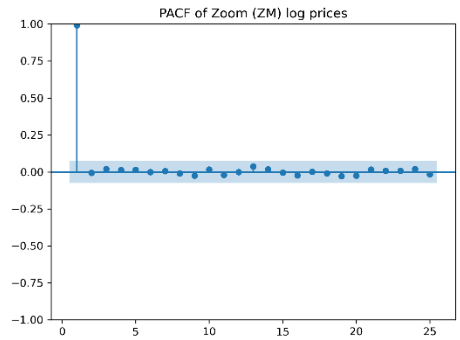
(a) BA Log price ACF



(b) BA Log price PACF



(c) ZM Log price ACF



(d) ZM Log price PACF

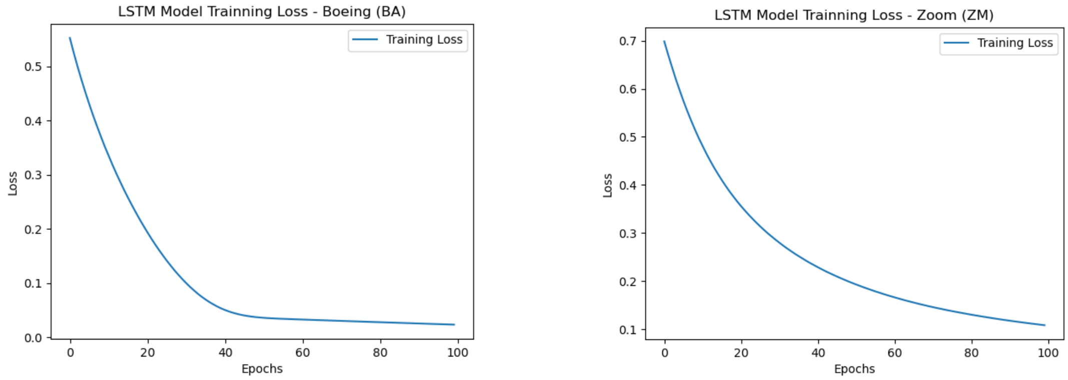
Figure 3: ACF and PACF of Log Price of BA and ZM (25 Lags).

Alternatively, I use both AIC and BIC criteria to select the ARIMA (p,d,q) model within the class of $p \leq 10$ and $q \leq 10$. For Boeing, the minimum AIC or BIC suggests model ARIMA (0,1,2) and ARIMA (4,1,6). For Zoom, the minimum AIC and BIC suggest model ARIMA (0,1,0).

For the Boeing training log price, I train ARIMA (2,1,0), ARIMA (0,1,2), and ARIMA (4,1,6) separately. For the Zoom training log price, because the parameter values of p and q found from ACF and PACF align with p and q in the ARIMA model with the minimum AIC and BIC, I train ARIMA (0,1,0). Also, I use the p-value test to check whether all predictors in these two models are significant. At a 95% confidence level, all predictors in these two models are significant.

LSTM: Show evidence of convergence for the LSTM model

For the LSTM model, I set the mean squared error as the loss function to find the proper number of epochs. I use different numbers of epochs to fit the LSTM model and plot the value of the loss function during the training to check the. When setting the number of epochs as 50, the value of the loss function decreases as the number of epochs increases for both LSTM models, but the values are not stable and not flat even when the number of epochs is 50. When setting the number of epochs as 200, the value of the loss function of both LSTM models becomes stable and flat when the number of epochs is around 100. In general, I use the number of epochs equal to 100 for both LSTM models.



(a) BA LSTM Training Loss (b) ZM LSTM Training Loss

Figure 4: LSTM Training Loss for BA and ZM.

Results Explanation

Model Performance Comparison

Company	Model	Training MSE	Training RMSE	Test MSE	Test RMSE
Boeing (BA)	ARIMA (2,1,0)	210.72	14.52	1475.72	38.42
	ARIMA (0,1,2)	210.63	14.51	1491.44	38.62
	ARIMA (4,1,6)	209.78	14.48	1556.52	39.45
Zoom (ZM)	ARIMA (0,1,0)	153.99	12.41	87.11	9.33
Boeing (BA)	LSTM	257.11	16.03	85.99	9.27
Zoom (ZM)	LSTM	323.68	17.99	62.43	7.90

Table 2: Model Performance (MSE, RMSE) Comparison.

Model Interpretation & Performance Difference between Boeing and Zoom

Boeing

For Boeing, the MSE and RMSE values for all training sets closely align, as do those for the test set. However, the test set exhibits notably higher MSE and RMSE values, indicating a challenge in accurately predicting prices. Visualizing the predictions reveals that, in the training set, the ARIMA model approximates real values closely, but in the test set, the predicted prices form a straight line. This suggests that ARIMA struggles to capture Boeing's significant price fluctuations. Because the MSE of the test set is much higher than the training set, the ARIMA models of Boeing are at risk of overfitting.

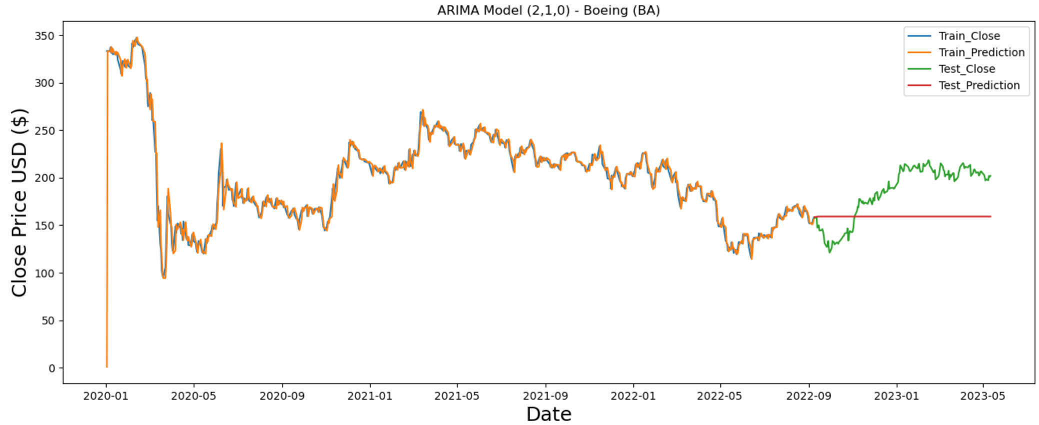


Figure 5: ARIMA Model for BA

On the other hand, the LSTM model exhibits similar MSE and RMSE values to ARIMA in the training set, but significantly smaller values in the test set. The test MSE is 85.99, and the RMSE is 9.27. Visual inspection of predicted prices illustrates that, whether in the training or test set, the LSTM model adeptly captures price trends, demonstrating superior predictive performance compared to ARIMA.

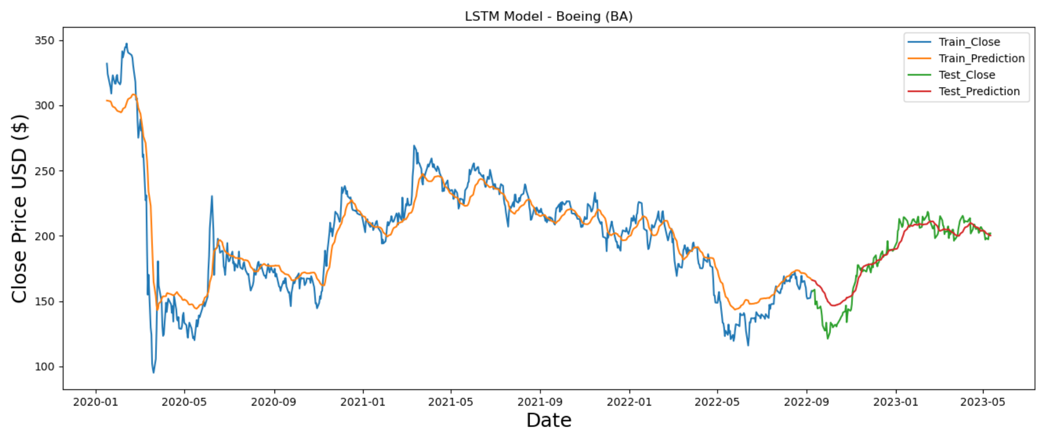


Figure 6: LSTM Model for BA

Zoom

For Zoom, ARIMA performs well in both the training and test sets, with a test set MSE of 87.11 and an RMSE value of 9.33. Visual predictions indicate a good fit in the training set, but the test set's predicted prices form a straight line.

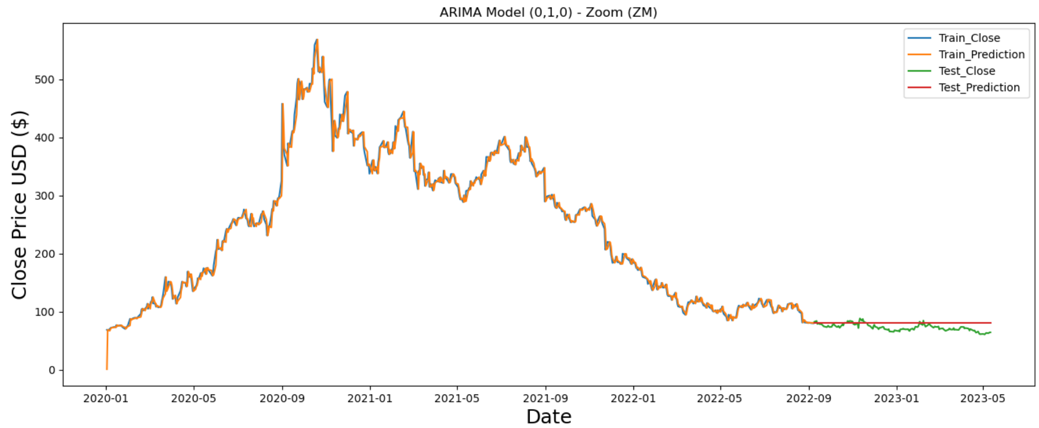


Figure 7: ARIMA Model for ZM

Contrarily, the LSTM model doesn't exhibit significantly smaller values in the test set, with an MSE of 62.43 and an RMSE of 7.90. Despite the test set's price fluctuation nuances, the LSTM model captures the price trend effectively, showcasing superior predictive performance compared to ARIMA.

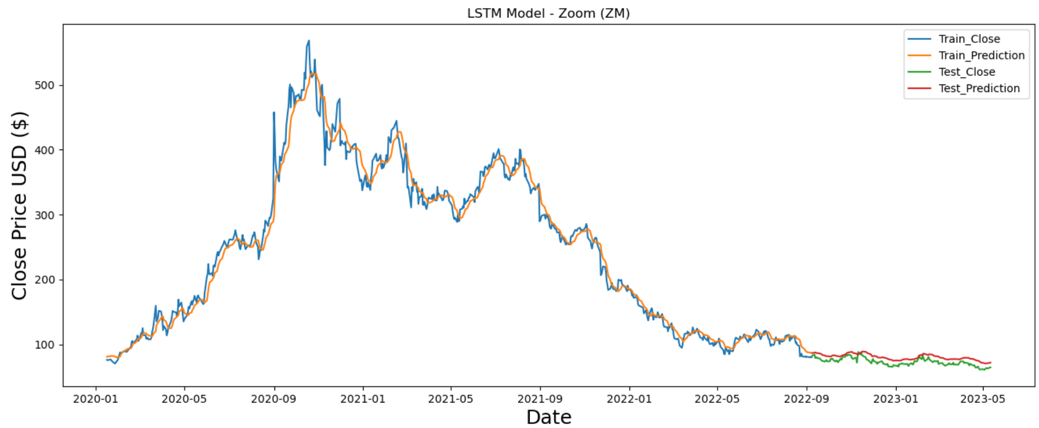


Figure 8: LSTM Model for ZM

Despite the overall significant stock price changes for Zoom, the reason why the LSTM model isn't notably superior to the ARIMA model might be attributed to minimal changes in Zoom's price range during the test set. During such periods, even if ARIMA fails to capture price fluctuations, it can still yield better prediction results.

ARIMA – Pros & Cons

Pros	Cons
● High Interpretability. ● Suitable for stationary data with clear trends and seasonality. ● Fewer Hyperparameters. ● Computational Efficiency.	● Assumption of Linearity. ● Difficulty with Non-Stationary Data. ● Limited Memory.

LSTM – Pros & Cons

Pros	Cons
● Capture Long-Term Dependencies. ● Can deal with Non-Linear Relationships. ● Automatically Feature Learning.	● Complexity. ● Easy to Overfitting. ● Difficulty in Interpretability. ● Computational Intensity.

Financial Interpretation

During the initial phase of the COVID-19 pandemic, reduced outdoor activities had a significant impact on industries, notably affecting the aviation sector, as evident in Boeing's sharp decline in stock prices. The industry's susceptibility to negative sentiments stemming from the outbreak quickly translated into stock market repercussions. Simultaneously, the increased demand for online work, driven by limitations on outdoor movement, spurred the growth of platforms like Zoom. Despite the ongoing pandemic, optimism surrounding Zoom led to a surge in its stock price from $70 to around $550.

As the pandemic gradually comes under control, Boeing's stock prices have exhibited an upward trend, indicating a rebound despite market fluctuations. Over time, there is an expectation that Boeing's stock prices may return to their pre-pandemic levels. In contrast, Zoom's stock prices have gradually decreased from the peak of approximately $550 to around $70 as the pandemic's impact diminishes.

This scenario highlights three economic phenomena. Firstly, industry stock prices are invariably influenced by market demand. Secondly, emergencies, such as the COVID-19 pandemic, exert a profound impact on the affected industries' stock prices. Thirdly, over time, as emergencies subside, stock prices tend to revert to their original ranges. These phenomena underscore the short-term influence of market sentiment, leading to rapid overvaluation or undervaluation, but in the long run, stock prices tend to reflect their intrinsic value. While these explanations are speculative, further research and investigation are essential, and drawing definitive conclusions solely from the fluctuations in two stock prices may require a more comprehensive analysis.

Conclusion

In conclusion, this study employed historical daily stock price data from January 1, 2020, to May 12, 2023, to analyze the impacts of the COVID-19 pandemic on Boeing (BA) and Zoom Video Communications (ZM). The dataset underwent meticulous pre-processing, including first-order differencing, log transformation, log returns calculation, and Min-Max scaling. The chosen companies, Boeing and Zoom, were selected due to their contrasting stock price trajectories during the pandemic, providing a dynamic context for model evaluation.

The study compared the forecasting capabilities of ARIMA and LSTM models, considering their performance on both training and test sets. Model parameters were carefully selected, and evaluation criteria such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were employed to assess predictive accuracy. In General, when prices fluctuate greatly, using the LSTM model for price prediction is a better choice.

Extra Work.

Improve ARIMA performance: Try different d^{th} order for Boeing (BA) and using pm.arima.auto_arima to find the optimal parameters for (p,d,q) for Zoom (ZM)

Company	Model	Training MSE	Training RMSE	Test MSE	Test RMSE
Boeing (BA)	ARIMA (2,0,0)	74.28	8.62	382.84	19.57
Zoom (ZM)	ARIMA (4,2,1)	526.72	22.95	426.70	20.66

Use a stock with linear trends and seasonality in time series data, which is suitable for the ARIMA model, and test the performance of ARIMA and LSTM models, to see whether ARIMA can have better performance. Choose Apple as the example.

Company	Model	Training MSE	Training RMSE	Test MSE	Test RMSE
Apple (AAPL)	ARIMA (1,1,0)	15.27	3.90	151.68	12.32
	LSTM	163.43	13.02	126.18	11.23

As shown in the figures below, Apple, ARIMA, and LSTM for Apple perform similarly on the out-of-sample dataset prediction. It can be said that ARIMA does have better forecasting capabilities for this type of time series data. (I used all the same steps as the previous, the exact graphs and results can be found in my Extra Work pdf and ipynb file.)