

David Sobek
Madison Ahern
Vibha Naiknavare

Project Report

Predicting Amazon Stock Price: A Time Series Analysis

Introduction:

In this project, we aim to delve into comprehensive exploration with the goal of predicting Amazon's stock price through time series analysis. As a cornerstone of the global marketplace, Amazon's stock performance not only reflects the company's dynamic but also serves as a crucial indicator for broader economic trends and investor sentiment. The analysis encompasses crucial data like daily open, high, low, close, and volume data from Amazon stock, particularly using split-adjusted data, we can gain insights into how fluctuations in Amazon's stock price may affect the broader economic landscape. Our exploration begins with an analysis of the correlation between Amazon's stock performance and the S&P 500, highlighting the importance of closing prices in this context. Moving past correlation, we then look closely at the different ways of predicting Amazon's stock price over the next 24 months using four distinct models. The insights we gain from this analysis help us understand how Amazon is related to the broader market and will provide useful advice for investors navigating the changing landscape of stock markets, looking at Amazon's performance for guidance.

Data Description:

The data we will be using for this project is Amazon stock data which was acquired from a dataset posted on Kaggle. We had access to raw Amazon stock data, as well as Amazon data which was adjusted for stock splits. Our team decided to conduct our analysis using the split-adjusted data to provide a more accurate model for our stock price forecast. The split-adjusted dataset contains daily Amazon stock data ranging from May 15, 1997, through November 7th, 2023. Our data contains several measures, including open and closing prices compared to the previous day, high and low prices compared to the previous day, the total volume of stock, the raw closing price of the stock, and the percentage change in the stock's

price. Additionally, we used a dataset of S&P 500 market prices, also found on Kaggle, to use as part of a correlation analysis to ensure that we can use our Amazon data as a metric to predict US economic health, however, we did not forecast this data or use it in any other part of our analysis.

Questions:

When considering our overall project goal of using forecasted prices of Amazon stock to predict the health of the United States economy, there are several important questions that can be asked when conducting this analysis. Firstly, is there a correlation between the price of Amazon Stock and overall economic health? Understanding this correlation can show us how changes in Amazon's stock price relate to more general market trends. Next, we look to see if there are any historical trends that could have affected Amazon's stock price? These trends could be positive, such as a price increase due to a new Amazon product launch, or negative such as during a national financial crisis like the housing market crash in 2007. Finally, we ask how do four separate predictive modeling techniques perform when forecasting Amazon stock data over the next 24 months? By evaluating several different models, we will be able to identify what predictive analytic techniques are more accurate when evaluating Amazon stock price movements.

Data exploration:

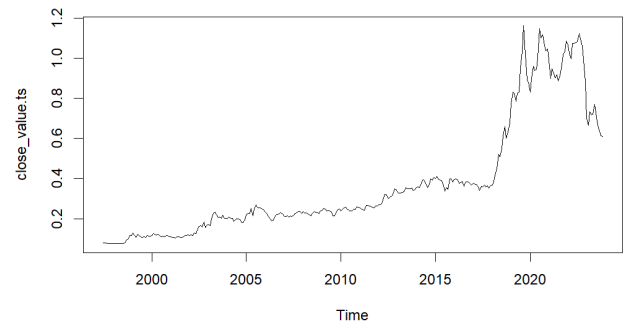
When we first read in our data, there were thousands of data points within our split-adjusted Amazon stock dataset, as we had daily stock evaluations for the last 26 years. Despite our extensive dataset, there were very few missing values. There were 19 missing values that all appeared during the first month of Amazon's IPO or Initial Public Offering in 1997. Due to the low amount of missing values compared to the large quantity of data that remained, our group decided to omit these values from our dataset and create a new dataset called "Amazon Clean" which had these null values removed.

When looking at our initial data summary, we can see that the average of our open, high, low, and close variables are all relatively similar. Stock open, low, and close, all have an average value of around 33.5, and the average high stock value was slightly higher at 34.2. The maximum and minimum values of these variables were all relatively similar as well, and there

appeared to be no significant outliers within our data that would dramatically affect the results of our time series analysis.

Data Visualization:

- **Initial Time series:** The time series plot of Amazon's closing prices, spanning from June 1997 to November 2023, provides a visual narrative of the historical performance of the stock. The vertical axis captures the fluctuating closing prices, revealing patterns, trends, and potential outliers. This graphical representation serves as a powerful tool for understanding the dynamics of stocks, offering insights into growth trends and notable turning points. The lowest closing price, 17.880 was in May 1998, and the highest, 134.752 in August 2022. The latest observations in 2023 suggest a decreasing trend in closing prices.

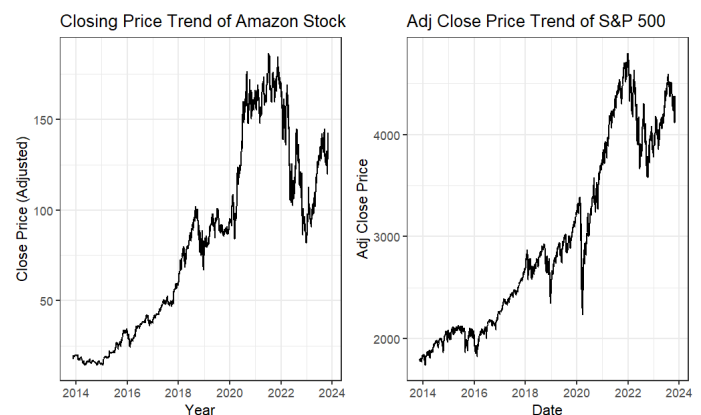


- **Line Graph of closing price over time for Amazon and S&P data:** The side-by-side graph for Amazon showcases the adjusted closing prices over time, while the S&P 500 plot illustrates the trend in its adjusted closing prices from the year 2017.

Notably, from 2020 to 2022, Amazon experienced a significant increase, in contrast to a low period for the S&P 500 in 2020, which later increased by 2022.

When interpreting these visualizations, we noticed that the spike or dip in prices, respectively, occurred alongside the

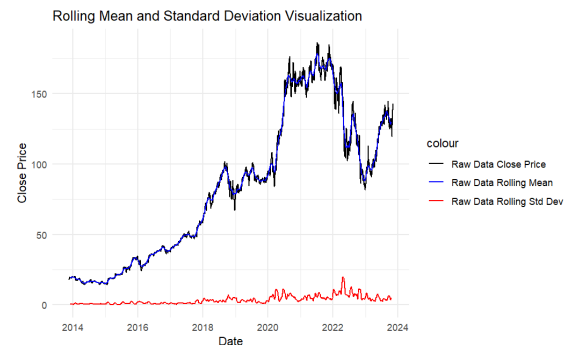
COVID-19 pandemic. As we know, e-commerce rose in prosperity at this time, while other facets of business and the economy as a whole plummeted (as seen in the S&P 500 data). These years are noteworthy due to these historical events, and will be discussed



further in our analysis. In 2022, both the closing prices indicate growth. Despite subsequent fluctuations, this dual presentation enhances data interpretability, serving as a valuable tool for stakeholders and decision-makers. Importantly, it is observed that both Amazon and the S&P 500 share similar trends over the analyzed period.

- **Rolling Statistic line graph Amazon:**

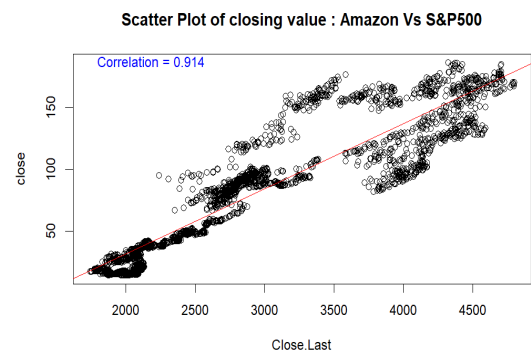
The graph illustrates Amazon stock data, depicting raw close prices (black line), a 30-day rolling mean (blue line) for trend identification, and standard deviation (red line) for volatility insights. The rolling mean smoothens fluctuations, revealing long-term trends, while standard deviation highlights periods of increased price variability.



Predictive Task and Proposed Models(Forecast)

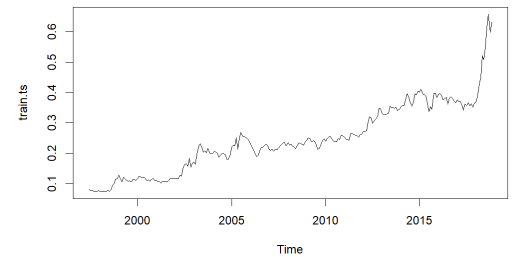
- **Correlation Analysis:**

Our exploration aimed to understand how Amazon's stock performance connects with the S&P 500. To establish a quantitative measure, we conducted a correlation analysis. The correlation coefficient was calculated using the 'cor()' function, revealing a substantial positive correlation of 0.914 between Amazon's closing price and the corresponding closing price of the S&P 500. The objective of this correlation analysis was to discern the patterns and trends that might indicate similar movements between Amazon and the broader market. A scatter plot was generated to visually explore the relationship between the two closing prices. The plot illustrates a clear positive correlation and a linear trend between the two variables. Overall, the positive correlation indicates a strong association with the closing price. Moving ahead, we can start building predictive models for Amazon's stock anticipating future movements and trends



- **Models: Preparing training and validation time series data**

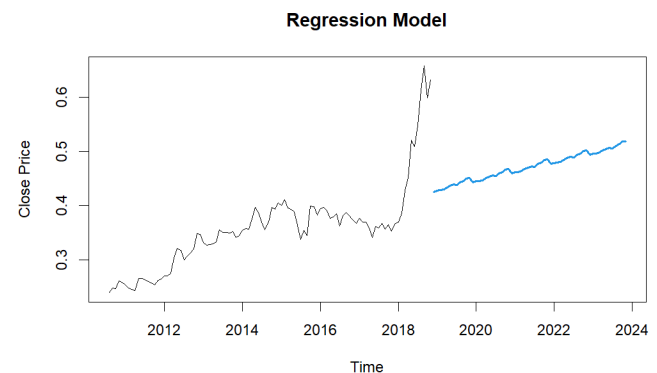
In our analysis, we constructed a time series dataset 'close_value.ts' representing Amazon's monthly stock closing prices from June 1997 to November 2023. To ensure thorough model evaluation, we split this dataset into a training set, spanning the majority of the timeline, and a validation set containing the most recent 60 observations. The training set enables our model to grasp historical patterns, while the validation set provides a comprehensive context for assessing predictive performance.



The plotted training time series graph captures the evolution of Amazon's stock price over time, establishing a strong foundation for upcoming time series modeling work.

1. Linear Regression model:

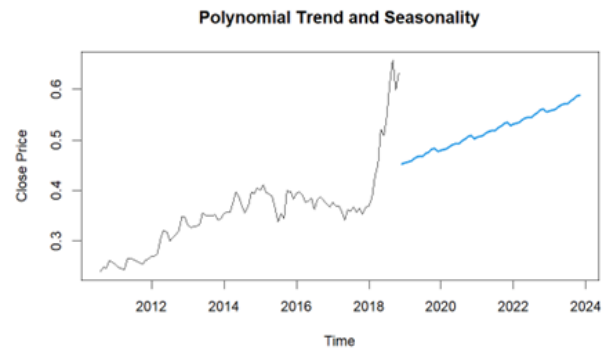
In this time series linear regression model, we assessed the relationship between Amazon's stock prices and time, considering both trend and seasonal components. The coefficients provide valuable insights into the influence of these factors. The intercept represents the estimated closing price when all predictors are zero, which is 0.05894. The trend coefficient of 0.001418 suggests that, on average, the closing price increases by 0.001418 units for each unit increase in time. The seasonal coefficients (season 2 to season 12) capture the impact of each season on the closing prices relative to a reference season. The residuals exhibit a spread from -0.064599 to 0.230281, emphasizing the variability captured by the model. The R-squared values of 0.8822 suggest that the model explains 88.22% of the variance in closing prices. The F-statistics further supports the model's overall significance.



The forecasted values from the model offer valuable insights into Amazon stock prices from January 2018 to June 2023. The projections capture the upward trend, with notable increases over the years. This demonstrates the model's ability to anticipate stock movements over an extended period, providing valuable insights for investors.

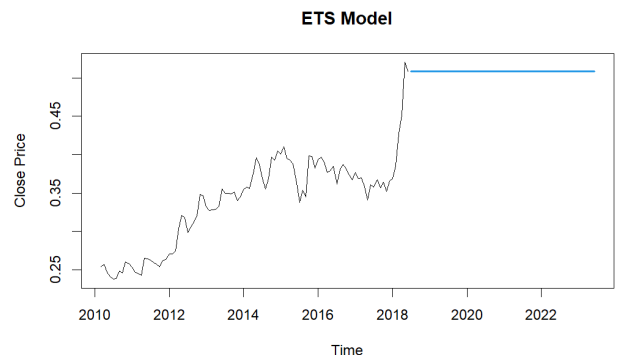
2. Polynomial and Seasonal Time Series Linear Model:

In the polynomial and seasonal time series model, we explored the relationship between Amazon's stock prices and time, incorporating a quadratic trend and seasonal components. The model indicates an average price increase with time, and the quadratic trend introduces a curving effect. Residuals, depicting the difference between actual and predicted values, show low variability. The model demonstrates strong performance with a low residual standard error, high Multiple R-squared, and adjusted R-squared. Compared to the linear model, the polynomial model slightly outperforms, as evidenced by lower residual standard error and higher R-squared values. Further evaluation using metrics like MAPE and RMSE will provide a more comprehensive accuracy assessment.



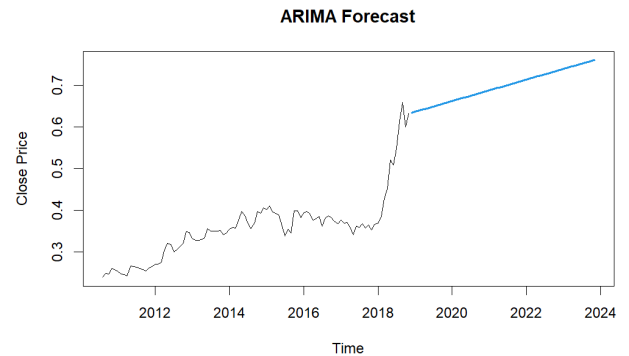
3. ETS:

For our ETS (Error, Trend, Seasonal) analysis, we selected an ETS A,N,N model. We set our smoothing parameter at 0.9999, which emphasizes the most recent data point of training data. Our AIC, AICc, and BIC all have very low values, each of which is around -700. Low values in these metrics indicate a better model fit, and therefore our ETS model fits this set of data quite well. When looking at our ETS accuracy measures, we were very pleased to see that all of our error metrics were very low. Our Mean average error was 0.0021, our RMSE was 0.0141, and our MAPE was 3.69%. These low error metrics indicate that the ETS model is a good predictor for the closing price of Amazon stock, and should provide accurate predictions of stock price for future months. However, when we apply this model to our validation data, we see that the prediction for the next 24 months shows a stagnant closing stock price of 0.6318 for every month, which is incorrect due to the continuous rise and fall of stock price. Because of this, we have not decided to incorporate the ETS model in our final predictive analysis.

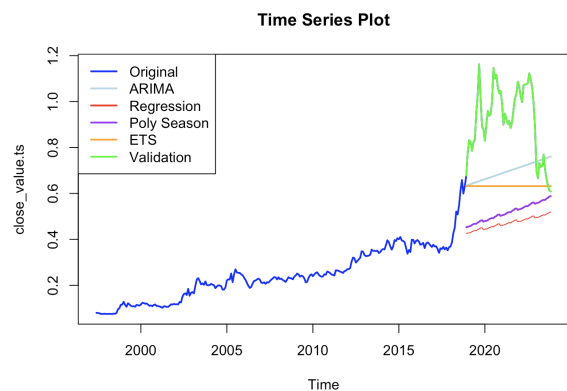


4. ARIMA:

In assessing the ARIMA(1,1,1) model for predicting Amazon's stock prices, we used Autocorrelation Function (ACF) analysis. The results indicate that the model isn't a random walk. The chosen ARIMA model includes an autoregressive term ($ar1 = 0.2523$), differencing, and a moving average term ($ma1 = -1.0000$). These terms, backed by their significant coefficients ($ar1\ se = 0.0545$, $ma1\ se = 0.0101$), reject the idea that Amazon's stock price follows a random walk. Moreover, the application of an auto ARIMA model to the training data further supports the non-random nature of the time series. The result forecasts, generated using the fitted Arima model for the validation period, contribute to an understanding of future stock price movements. The plotted ARIMA forecasts visually showcase the model's ability to capture underlying patterns, revealing an increasing trend.

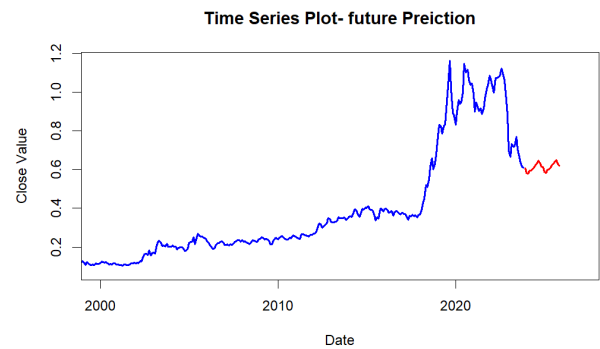


Accuracy: In the assessment of model accuracy for predicting Amazon's stock prices, the ARIMA model emerged as the most effective, showcasing superior performance across various metrics. The Mean Error (ME) of 0.22105, Root Mean Squared Error (RMSE) of 0.27420, Mean Absolute Error (MAE) of 0.24583, and other metrics such as Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Autocorrelation Function (ACF1), and Theil's U demonstrated favorable results. Notably, the MAPE value is 25.37826, lower than the other model indicating that ARIMA is the best predicting model.



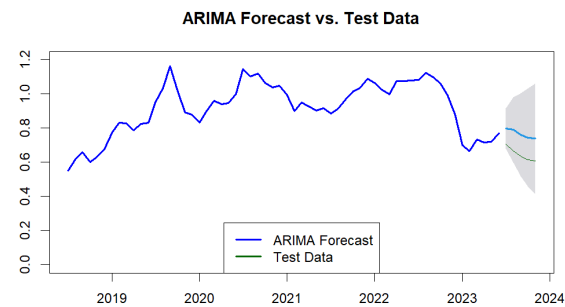
Next 2-year prediction forecast:

To predict Amazon's stock prices for the next two years, we used an ARIMA(1,1,1) model with seasonal differencing. The model breaks down into autoregressive (AR), differencing (I), and moving average (MA) components. We forecasted 24 months into the future, presenting the results with a 95% confidence interval. Notable points include a steady increase in the forecasted values from December 2023 to September 2024, suggesting a positive trend in the future. Of particular interest are the months from October 2024 to November 2025, where a slight decline is projected, potentially indicating a turning point or a period of stabilization. Additionally, its noteworthy that the confidence interval widens as we move further into the future, indicating increased uncertainty in the longer term predictions.



How accurate is the data?

We assessed the ARIMA model's predictive accuracy by comparing its output to the test data for the last 5 months. Our dataset was divided into training, validation, and test sets, covering June 1997 to June 2018, July 2018 to June 2023, and July 2023 to November 2023, respectively. The analysis involved contrasting actual data with the ARIMA point forecast, considering the lower (Lo95) and upper (Hi95) 95% confidence intervals from July to November 2023. The table showcases the variability in ARIMA predictions, emphasizing the closeness of actual values to the forecast for accuracy. The Lo95 and Hi95 intervals provide insights into prediction uncertainty, contributing to a comprehensive assessment of model reliability. The alignment of actual values within the confidence intervals contributes to the overall accuracy of the model, as seen in the table below.



	Actual	ARIMA	Lo95	Hi95
July 2023	0.7041667	0.7687901	0.6255975	0.9119827

August 2023	0.6635333	0.8049632	0.5694328	1.0404936
September 2023	0.6338667	0.8126209	0.5115884	1.1136535
October 2023	0.6140667	0.8111998	0.456561	1.1658384
November 2023	0.6083333	0.7796590	0.3785151	1.1808029

Conclusion:

Although our predictive models were very accurate when predicting Amazon Stock Prices, the stock market is a very volatile and unpredictable place. When looking at all of our models compared to the validation data, we can see that the actual Amazon stock prices for the time we predicted are much higher than our predictions. This is not an error on our part, but rather the result of the rise of E-Commerce over the past few years, especially during the COVID-19 Pandemic of 2020. Despite this, our models were accurate with a very low error rate, and we were able to conclude that the change in Amazon's stock price does indeed have an effect on the overall economic health of the United States. Finally, The analysis compares ARIMA predictions with actual data, showcasing accuracy through the alignment of actual values within the forecast confidence intervals from July to November 2023.