****

**VIRGINIA COMMONWEALTH UNIVERSITY**

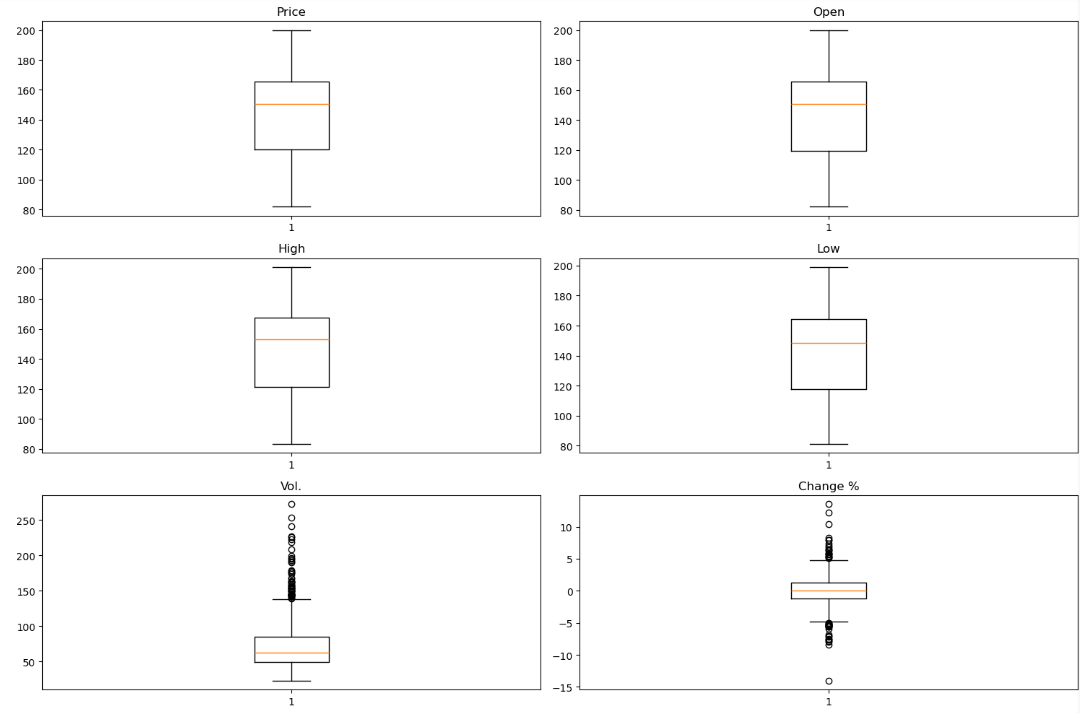
**Statistical analysis and modelling (SCMA 632)**

**A6a and A6b**

**Rohan Arpit Dungdung**

**V01106712**

**Date of Submission: 25-07-2024**

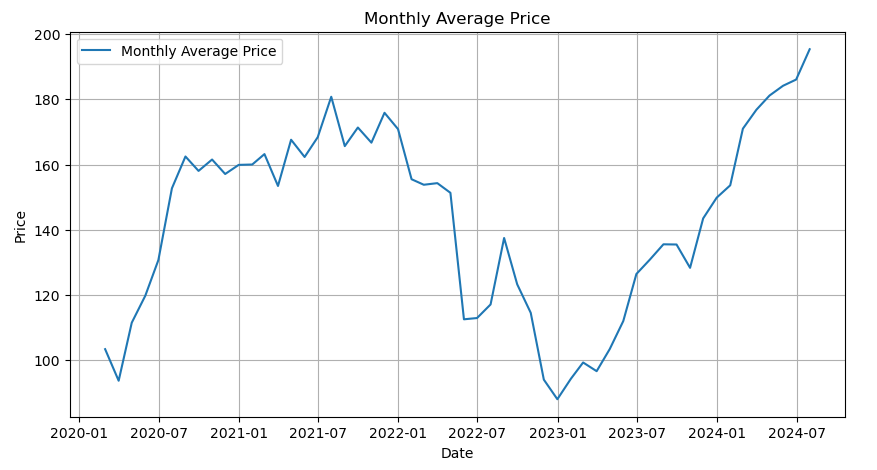


Interpretation

The above variables display a relatively normal distribution with no significant outliers. The median represented by the orange line is centered within the interquartile range (IQR), indicating a balanced distribution of values. The whiskers extend to the minimum and maximum values within 1.5 times the IQR, suggesting the data points are mostly concentrated around the median with a few data points at the extremes.

The Vol. boxplot shows a larger number of outliers compared to the other variables. This indicates that there are several trading days with significantly higher trading volumes than usual. The upper whisker and the outliers above it suggest that trading volume varies more widely than the other metrics, potentially due to market events or news impacting trading activity.

The Change % boxplot also has several outliers, both on the higher and lower ends. These represent days with unusually high gains or losses in the stock price percentage. The median is close to zero, which makes sense since the stock price typically oscillates with some days of gains and some of losses. However, the outliers suggest that there are days with significant percentage changes, which could be due to major market events, earnings reports, or other significant announcements affecting the stock price.



Interpretation

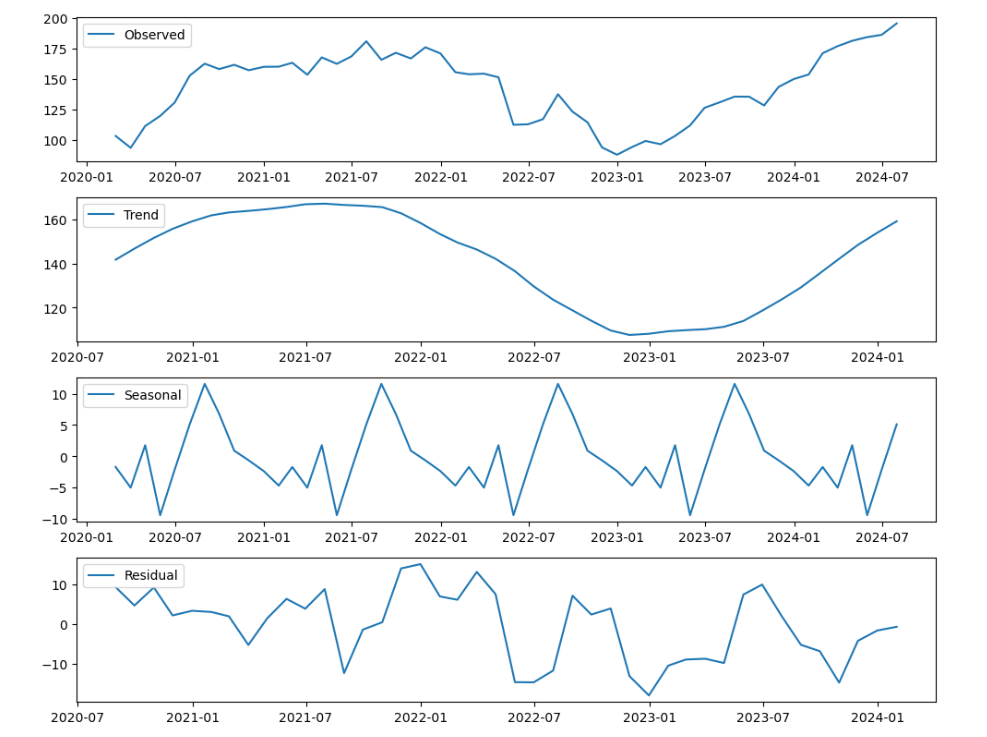
The graph shows a significant upward trend in the stock's average price starting in early 2020, which continues into early 2021. This increase suggests strong positive market sentiment or performance, possibly driven by company growth, market trends, or broader economic factors.

The price continues to rise, reaching a peak in mid-2021, followed by a series of fluctuations. These ups and downs could indicate volatility due to market reactions to various events such as earnings reports, macroeconomic data, or industry-specific news.

There is a noticeable decline starting around early 2022, which continues until mid-2022. This downward trend could result from negative news, market corrections, or other factors affecting investor sentiment or company performance.

From mid-2022 onwards, the stock price begins to recover, with a significant increase observed from early 2023 through mid-2024. This recovery could be attributed to improved company performance, positive market conditions, or other favorable factors.

As of mid-2024, the stock price is at a new high, indicating strong market performance. The latest data point shows a sharp rise, which might be due to recent positive developments, such as favorable financial results, strategic moves by the company, or broader market trends.



Interpretation

The above graphs shows Amazon stock's monthly average price from early 2020 to mid-2024. It shows the overall fluctuations in the stock price over time, including rises and falls.

The trend component shows the underlying direction of the data over time, smoothing out short-term fluctuations. The trend line reveals:

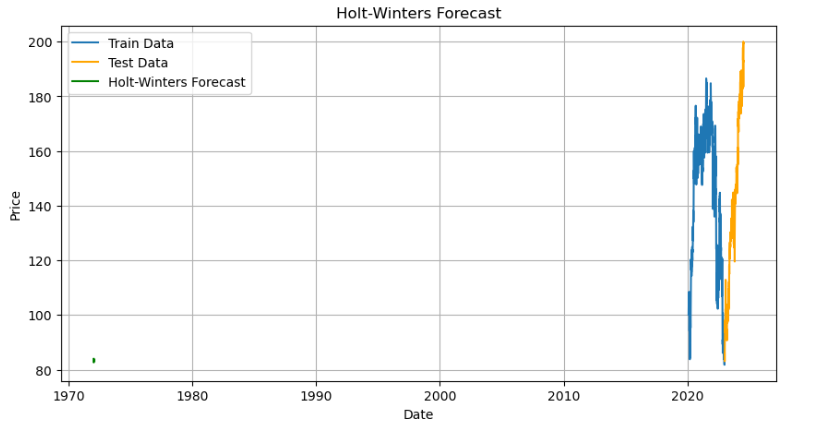
An upward movement from 2020 to mid-2021, indicating growth in the stock price.

A downward trend from mid-2021 to early 2023, reflecting a general decline in prices during this period.

A recovery starting from early 2023, with prices rising steadily towards mid-2024.

The seasonal component captures the repetitive, cyclical patterns that occur at regular intervals within the data. In this case, the seasonal fluctuations show variations within a range of approximately ±10 units. This indicates that the stock price exhibits some seasonal behavior, potentially due to regular financial cycles, market conditions, or company-specific factors like product releases or earnings announcements.

The residual component represents the irregular fluctuations that are not explained by the trend or seasonal components. These are the random or irregular movements in the stock price, capturing anomalies or unexpected events. The residuals fluctuate around zero, indicating that the decomposition has effectively captured the primary patterns, leaving mostly noise.

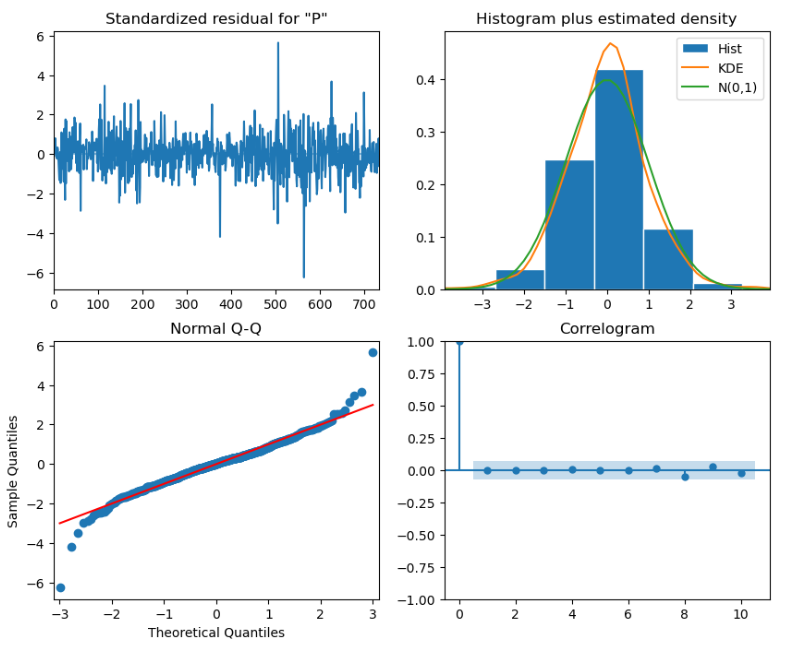


Interpretation

This above graph represents the historical data used to train the Holt-Winters model. It covers a significant period, capturing the variations in the stock price over time. The training data is typically selected up to a certain cutoff date to fit the model and predict future values.

The test data comprises the actual stock prices following the training period. This section is used to evaluate the performance of the model by comparing the forecasted prices against the actual observed values.

The green line represents the forecasted prices generated by the Holt-Winters model for a period following the training data. This forecast aims to predict future stock prices based on the patterns identified in the training data, including trends and seasonality.



Interpretation

The top-left plot displays the standardized residuals over time. Residuals are the differences between the observed values and the values predicted by the model.

Ideally, residuals should be randomly distributed around zero, with no discernible pattern. In this plot, the residuals appear to fluctuate randomly around zero, suggesting that the model has captured the main patterns in the data.

However, a few spikes indicate occasional larger errors, which may point to anomalies or periods where the model underperformed.

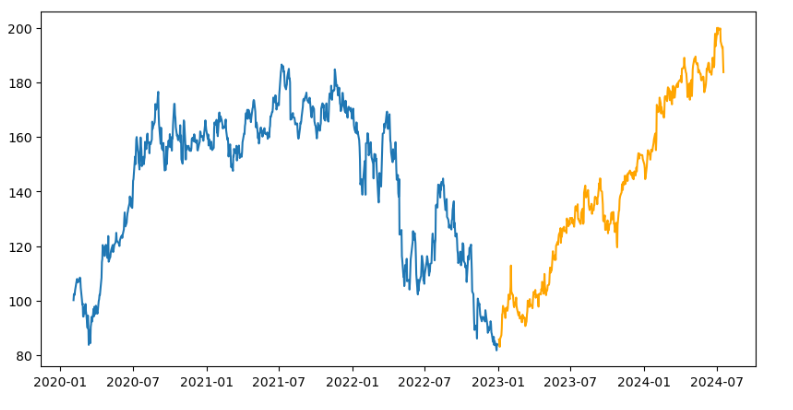
The top-right plot shows the histogram of residuals with an overlaid Kernel Density Estimate (KDE) and a normal distribution (N(0,1)) for comparison.

The residuals should ideally follow a normal distribution if the model is appropriate. The histogram and KDE appear to be roughly bell-shaped, suggesting that the residuals are approximately normally distributed. However, there are slight deviations, especially in the tails, indicating some departures from normality.

The bottom-left plot compares the quantiles of the residuals with the theoretical quantiles of a normal distribution. If the residuals are normally distributed, the points should lie along the red line. In this plot, the residuals mostly follow the red line, indicating that they are approximately normally distributed. However, deviations at the extremes suggest some residuals are larger than what would be expected under a normal distribution, indicating potential outliers or periods of poor model fit.

The bottom-right plot shows the autocorrelation function (ACF) of the residuals. It measures the correlation between residuals at different lags.

For a well-fitted model, residuals should not be autocorrelated, meaning the autocorrelation coefficients (blue dots) should fall within the confidence intervals (shaded area).

In this plot, most residual autocorrelations are within the confidence bounds, suggesting that there is no significant autocorrelation left in the residuals. This indicates that the model has effectively captured the temporal dependencies in the data.

Interpretation

This above graph represents the historical stock prices used to train the forecasting model. It covers from early 2020 to the end of 2022.

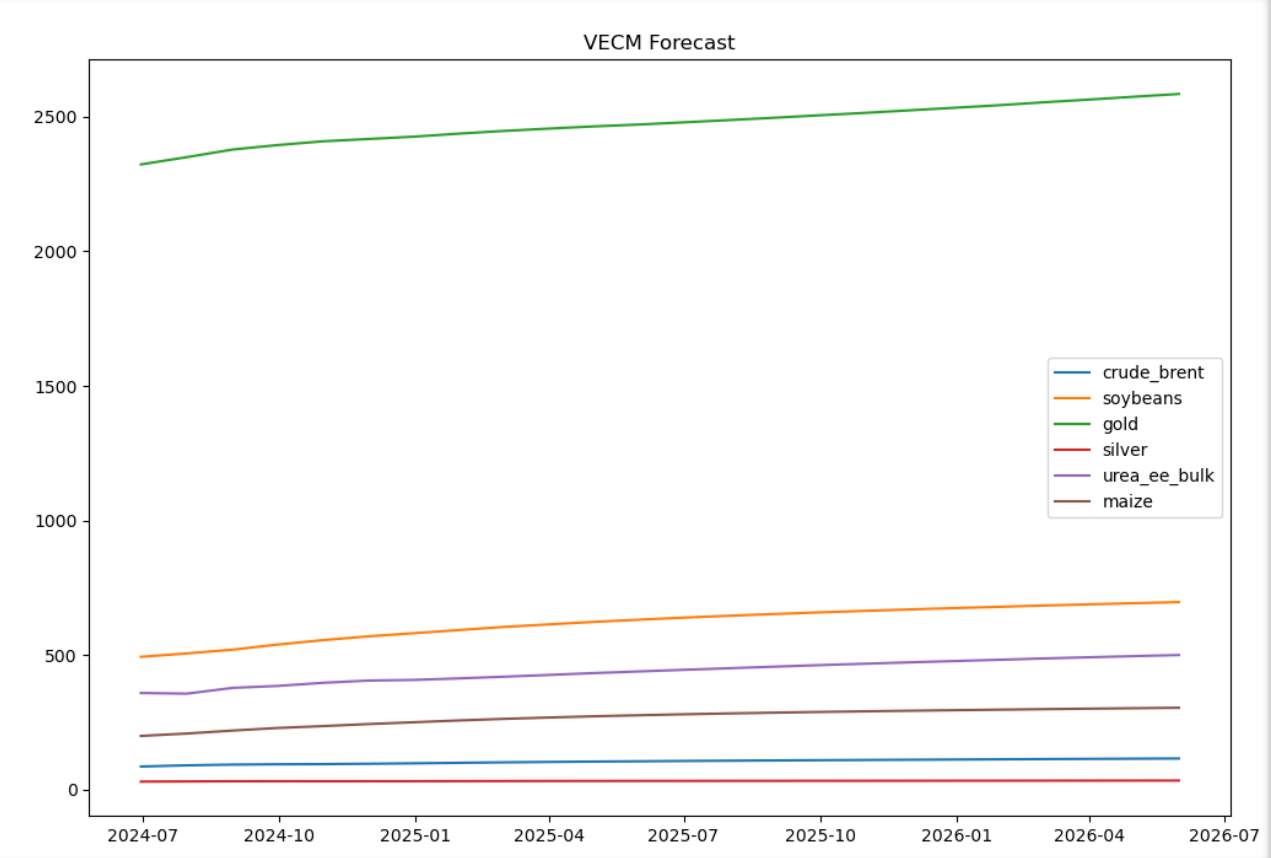
The stock price shows various trends and patterns:

An initial rise from early 2020, peaking around mid-2021.

A period of decline, with significant drops observed throughout 2022, reaching a low point at the end of 2022.

The orange segment represents the stock prices after the training period, typically used to validate the model's forecasts. This data covers from early 2023 to mid-2024.

The stock price in this period shows a strong recovery, with a consistent upward trend starting in early 2023. This trend continues through mid-2024, with the price reaching new highs.



Interpretation

All commodities in the forecast appear to have an upward trend over the forecast period. This suggests an expected increase in the prices of these commodities.

Gold (Green Line): Gold prices are significantly higher than other commodities, showing a continuous upward trend. This suggests a strong increase in gold prices over the next two years.

Soybeans (Orange Line): Soybeans also show a noticeable upward trend, although less steep than gold. The consistent increase indicates a steady rise in soybean prices.

Crude Brent Oil (Blue Line): The blue line representing crude Brent oil shows a relatively flat but slightly upward trend, indicating a mild increase in oil prices.

Silver (Red Line): Similar to crude Brent oil, silver shows a slight increase over time, though the growth rate appears less steep than that of gold.

Urea (Purple Line): Urea (ee bulk) shows a gradual upward trend, indicating a slow but steady increase in prices.

Maize (Brown Line): The forecast for maize prices also shows a steady but modest increase.

There is a clear disparity in the price levels of different commodities, with gold being the highest, followed by soybeans, and then the other commodities. This indicates differing factors influencing each commodity's price, such as supply-demand dynamics, geopolitical influences, and market conditions.

The upward trends across all commodities suggest inflationary pressures or increased demand. For investors or stakeholders, this forecast could indicate potential investment opportunities in commodities expected to rise in value. Policymakers and businesses may use this forecast to plan for potential cost increases in raw materials and adjust their strategies accordingly.