Time Series Forecasting of Google's Stock Prices

Statistical Learning
Professor: Gianluca MASTRANTONIO

Author: Ghassane BEN EL AATTAR

Politecnico di Torino

Contents

1 Introduction

This report presents a comprehensive analysis of the closing stock prices of Google (GOOG) over a specific time period, from 2013 to 2017.

The stock market is a complex system influenced by a myriad of factors, both internal corporate events and external global events. Analyzing stock prices can provide insights into the company's performance, market trends, and potential future movements. The primary objective of this analysis is to understand the historical behavior of Google's stock prices and to forecast its future movements.

Forecasting stock prices is a pivotal task in finance, aiding investors in making informed decisions. Specifically, closing prices are of significant interest as they represent the consensus of value for a stock at the end of a trading day, encapsulating the day's volatility and investor sentiment.

By delving deep into the patterns, seasonality, and other characteristics of Google's closing prices, this report aims to provide a robust foundation for predictive modeling and strategic investment planning.

2 Time Series

A time series is a sequence of data points, typically consisting of successive measurements made over a time interval.

Mathematically, a time series $\{X_t\}$ can be represented as:

$$X_t = f(t, S_t, \epsilon_t) \tag{1}$$

where t is the time index, S_t is the systematic structure of the series, and ϵ_t is the random error term. The systematic structure can further be decomposed into trend, seasonality, and cyclic components.

The essence of time series analysis lies in decomposing the series to understand its underlying patterns and then modeling these patterns for forecasting.

In the context of our task, the time series represents the closing stock prices of Google. Analyzing this time series can unveil patterns in stock price movements, the influence of external events, and the overall trajectory of the stock's value.

By modeling these patterns, we aim to forecast future stock prices, providing valuable insights for investment strategies and decision-making processes.

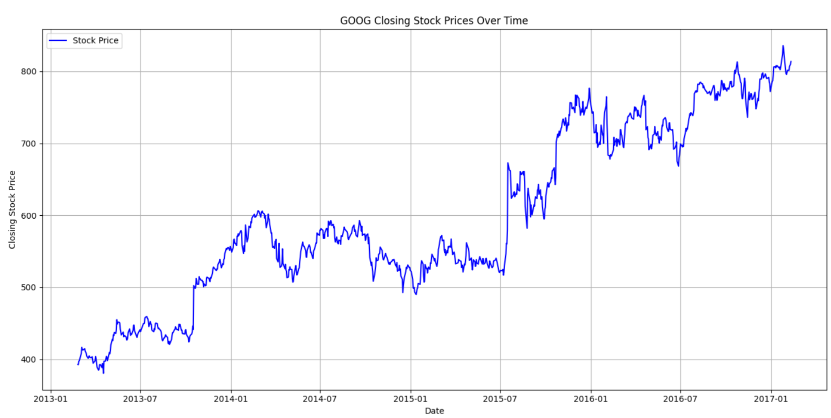

Figure 1: Google Stock Price Over Time

2.1 Time Series Decomposition

Time series decomposition is a foundational method in time series analysis, aiming to deconstruct a time series into its constituent components. Most time series data, including stock prices, can be described using three systematic components:

- **Trend**: Represents the underlying trajectory of the data over a long period. In the context of Google's closing stock prices, the trend might capture the company's overall growth trajectory, influenced by factors such as increasing market share, product innovations, or global expansion.

- **Seasonality**: Captures regular and predictable changes that recur at fixed intervals. For stock prices, seasonality might manifest in patterns related to quarterly earnings reports, annual sales events, or other regular corporate activities.

- **Cycles**: Unlike seasonality, cycles represent medium-term changes caused by circumstances or events that don't have a fixed repetition period. In stock prices, cyclical components might capture broader economic cycles, industry-wide shifts, or market sentiments that don't adhere to a fixed calendar schedule.

In addition to these systematic components, there's a non-systematic component called **noise**. This captures random fluctuations in the data that can't be attributed to the trend, seasonality, or cycles. Noise might result from unforeseen events, such as sudden market shocks, unexpected corporate news, or other unpredictable influences.

In the task of analyzing Google's closing stock prices, decomposition allows us to isolate these components, providing a clearer understanding of the underlying patterns and behaviors in the data. By understanding each component separately, we can more effectively model and forecast future stock prices. For instance, recognizing a consistent upward trend can inform long-term investment strategies, while understanding seasonality can guide short-term trading decisions. Similarly, identifying cyclical downturns might prompt more cautious investment approaches, even if the overall trend is positive.

$$Y_t = T_t + S_t + C_t + R_t \tag{2}$$

where Y_t is the observed series, T_t is the trend component, S_t is the seasonal component, C_t is the cyclical component, and R_t is the residual or noise component.

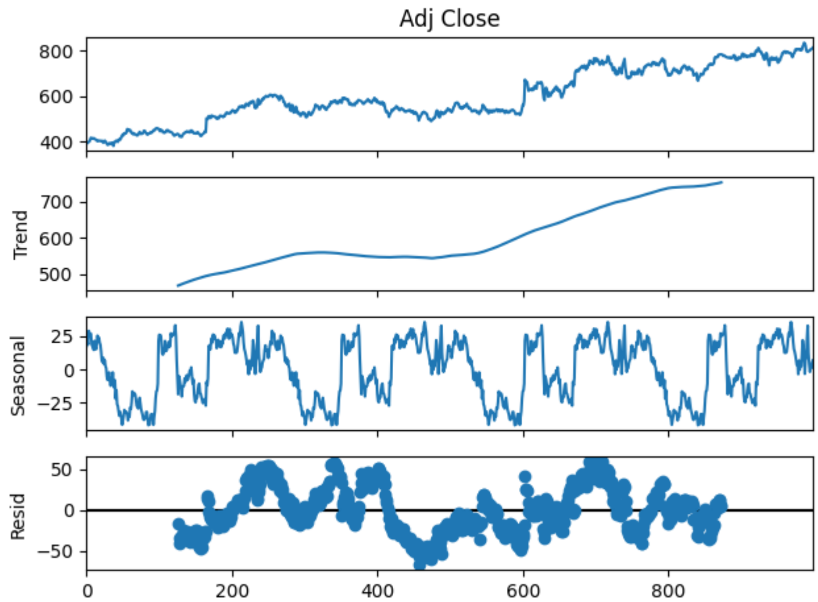

Figure 2: Time Series Decomposition of Google Closing Stock Prices.

3 Trend and Seasonality Analysis

3.1 Trend Analysis of Google's Stock Prices

Upon analyzing the trend plot for Google's closing stock prices, several key observations emerge:

1. There is a **distinct upward trend** over the years, signifying a consistent increase in the value of Google's stock. This is indicative of the company's robust performance and positive market reception.

2. Some intervals showcase a **flattened trend**, suggesting periods of stagnation or minimal growth. Notably, these phases are transient and are succeeded by renewed upward trajectories.

3. Occasional **dips in the trend line** are evident, pointing to temporary declines in stock prices. Crucially, these downturns are short-lived and are invariably followed by a return to the prevailing upward trend.

In essence, the trend plot underscores a predominant theme: Google's stock prices have largely witnessed an upward trajectory over the years, punctuated by brief spells of slower growth or minor declines.

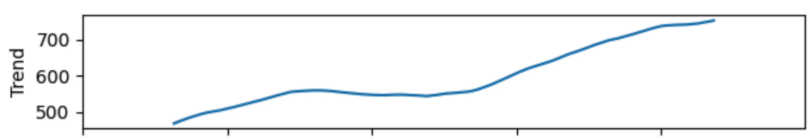

Figure 3: Trend of Google's Stock Prices

3.2 Seasonality Analysis

As described before, seasonality is a characteristic of a time series in which the data experiences regular and predictable changes that recur every calendar year. These fluctuations are often tied to the time of year, such as increased sales of winter clothing during winter months.

Let's now analyse seasonality for Google's stock prices, from the plots below.

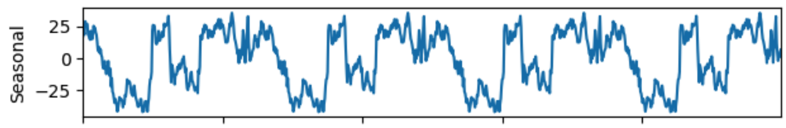

Figure 4: Seasonality Component of Google's Stock Prices

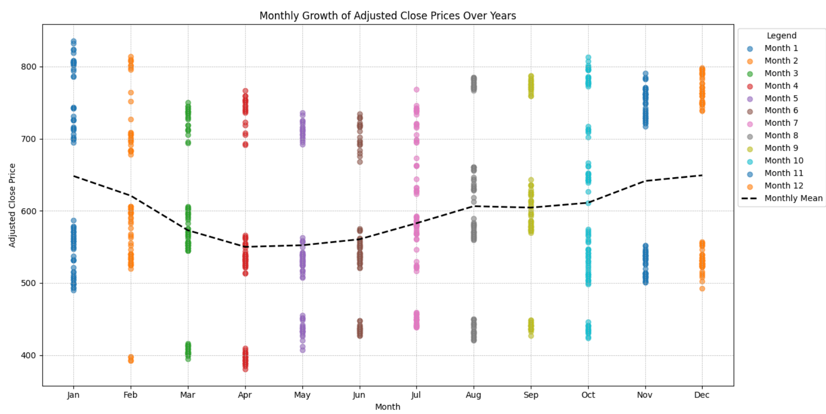

Figure 5: Monthly Growth of Google's Stock Prices of the Years

Analyzing the seasonality plot for Google's closing stock prices yields several noteworthy insights:

- A **recurring pattern** in the data is discernible, hinting at some degree of seasonality. Nevertheless, the amplitude of these seasonal oscillations is modest when juxtaposed with the overarching trend. This observation underscores the trend as the primary driving force in this time series.

- The seasonality lacks a **uniform rhythm**. This inconsistency can be attributed to the intricate nature of stock prices, which are swayed by a plethora of factors, many transcending seasonal influences.

- Certain intervals exhibit a **heightened seasonal component**. Such pronounced seasonality could align with specific events or annual periods that wield a marked influence on Google's stock valuation, such as product launches, annual sales events, or industry-wide phenomena.

To encapsulate, the seasonality plot intimates the presence of seasonal patterns in Google's stock prices. However, these patterns exhibit variability and their magnitude is overshadowed by the dominant trend. This infers that, while seasonality might contribute to stock price fluctuations, a myriad of other determinants exert a more pronounced influence.

Let's now further confirm our assumptions by analysing the polar seasonal plot generated from the dataset, which is a way of visualizing the average monthly closing stock prices for each year. It can potentially help identify any seasonal patterns in the data.

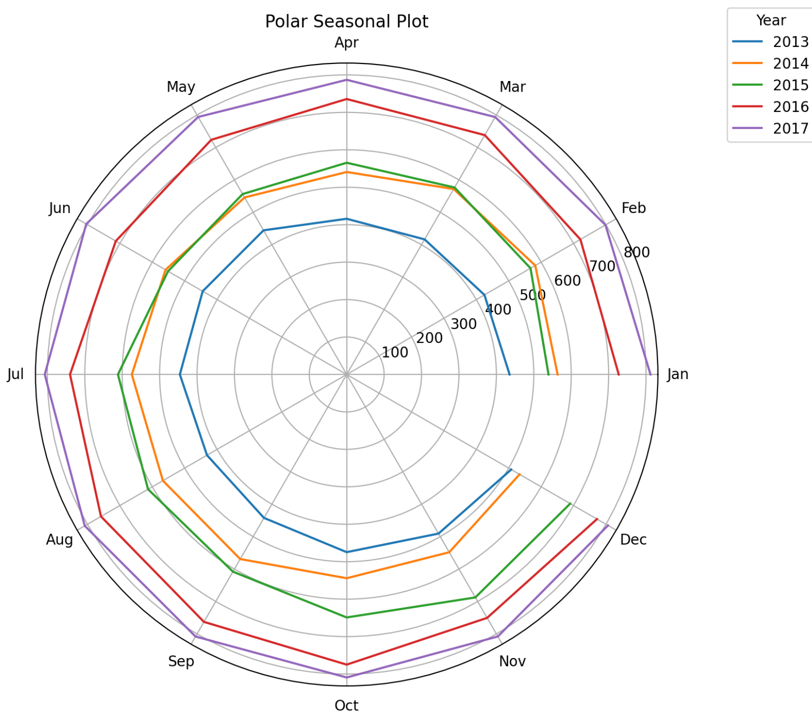

Figure 6: Polar Seasonal Plot of Google's Stock Prices Over the Years

As it's possible to see, the polar seasonal plot offers a unique perspective on the cyclical behavior of Google's closing stock prices:

- Each trajectory in the plot corresponds to a distinct year, with individual points on the trajectory denoting the average closing stock price for each month within that year.

- The circular configuration of the plot symbolizes the cyclical progression of a year. This design ensures that the commencement and culmination of the year (i.e., January and December) are contiguous, illustrating the seamless transition from one year to its successor.

- The radial magnitude from the circle's epicenter to a specific point on a trajectory is indicative of the stock price. Consequently, an augmented radial distance signifies an elevated stock price.

- Despite the plot's illustrative nature, discerning definitive patterns is challenging due to the superimposition of trajectories. A preliminary observation suggests an absence of consistent peaks or troughs during identical months across different years. This insinuates that Google's stock price might not be profoundly influenced by a robust seasonal component on a monthly scale.

In essence, while the polar seasonal plot provides a cyclical overview of Google's stock prices, it intimates a potential lack of strong monthly seasonality, reinforcing the earlier observations from the seasonality analysis.

4 Stationarity

Stationarity is a pivotal concept in time series analysis. A time series is said to be stationary if its statistical properties, such as mean, variance, and autocorrelation, remain constant over time. Mathematically, for a stationary time series, the joint probability distribution of $X_t, X_{t+1}, \ldots, X_{t+n}$ remains unchanged for any lag n. In simpler terms, a stationary series does not exhibit trends, seasonality, or cycles that might change over time.

- **Relevance to Forecasting**: Stationarity is crucial because many forecasting methods, including ARIMA and its variants, assume that the time series is stationary or can be transformed to become stationary. Stationary time series are inherently more predictable as their future statistical properties are not different from their past.

- **Application to Google's Stock Prices**: In the context of our task with Google's closing stock prices, ensuring stationarity is essential for accurate forecasting. If the series is non-stationary, we might need to apply transformations, such as differencing, to stabilize its statistical properties before modeling. Recognizing and addressing non-stationarity can significantly enhance the reliability of our forecasts.

4.1 Augmented Dickey-Fuller (ADF) test

The ADF test is a widely-used statistical procedure to determine the stationarity of a time series. The test's null hypothesis posits that the series is non-stationary, implying the presence of a unit root. A significant test result (typically a p-value below 0.05) allows us to reject this null hypothesis, indicating stationarity.

For our Google's stock prices dataset, the ADF test yields:

$$\text{ADF Statistic} : -0.9840162878560302$$
$$\text{p-value} : 0.7590492618329442$$

Given the obtained p-value of 0.759, which exceeds the 0.05 threshold, we cannot reject the null hypothesis. This outcome suggests that Google's closing stock prices exhibit non-stationarity, necessitating potential transformations or different modeling approaches for accurate forecasting.

4.2 Autocorrelation Function (ACF) Plot Analysis

Autocorrelation, in the context of time series analysis, refers to the correlation of a series with its own past values. It measures the linear predictability of the series at time t using its value at time $t - k$, where k is the lag. The Autocorrelation Function (ACF) provides a systematic way to visualize and quantify this correlation across various lags.

The ACF plot offers a graphical representation of a time series' autocorrelation across a range of lags. For a stationary time series, the autocorrelation typically decays to zero swiftly, indicating that past values have diminishing influence on future values. In contrast, a non-stationary series often manifests a more gradual decline in autocorrelation, suggesting that past values retain their influence over a longer span.

Upon examining the ACF plot below for Google's stock prices, we discern a slow attenuation of autocorrelation as the lag increases. This behavior further substantiates the non-stationary nature of the series.

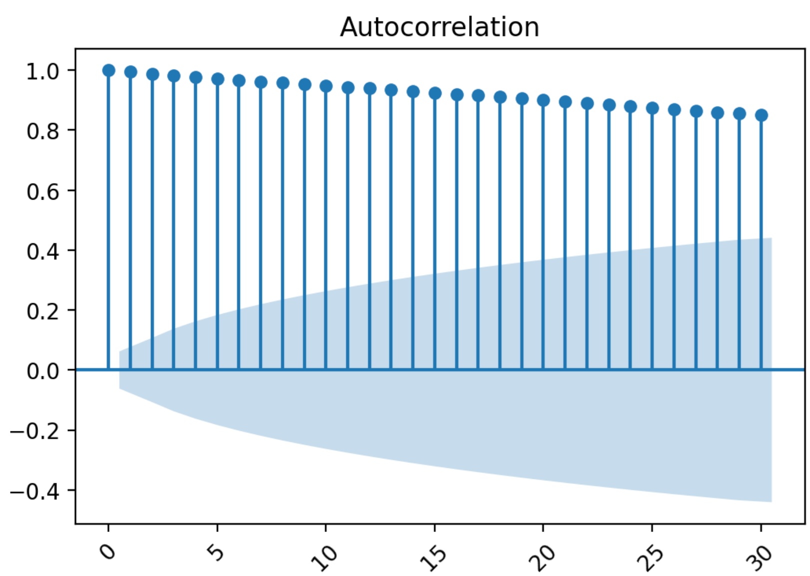

Figure 7: ACF Plot

Both diagnostic tools, the ADF test and the ACF plot, converge on the inference that Google's closing stock prices are non-stationary. This suggests that the series' statistical properties evolve over time. Such a characteristic is anticipated for stock prices, given their susceptibility to a plethora of external and internal factors, often leading to discernible trends and other non-stationary dynamics.

5 Time Series Forecasting Models

5.1 ARIMA (AutoRegressive Integrated Moving Average)

ARIMA is a time series forecasting method that combines autoregressive (AR) and moving average (MA) models, along with differencing to make the series stationary (Integrated).

5.1.1 Components of ARIMA

- **AR (AutoRegressive)**: It models the relationship between an observation and a number of lagged observations (previous time steps). Mathematically, the AR component is represented as:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t$$

  where ϕ are the parameters of the AR model, p is the order of the AR model, and ε_t is white noise.

- **I (Integrated)**: It represents the number of differences needed to make the series stationary. Differencing is the transformation of the series to the difference between consecutive observations.

- **MA (Moving Average)**: It models the relationship between an observation and a residual error from a moving average model applied to lagged observations. The MA component is given by:

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

  where θ are the parameters of the MA model, q is the order of the MA model.

Given the observed non-stationarity in Google's stock prices, the integrated component of ARIMA can be instrumental in transforming the series to a stationary one, thereby making it amenable to forecasting. The AR component can capture the momentum and mean reversion effects often observed in stock prices, while the MA component can account for any shock effects in the market.

If there's a discernible seasonal pattern in the stock prices, perhaps due to quarterly financial reports or yearly market trends, the SARIMA model becomes particularly relevant. The seasonal components in SARIMA can capture these periodic fluctuations, offering more accurate forecasts during such periods.

5.2 SARIMA (Seasonal AutoRegressive Integrated Moving Average)

SARIMA extends ARIMA by adding a seasonal component. It is particularly useful for series with clear seasonal patterns.

5.2.1 Components of SARIMA

In addition to the ARIMA components, SARIMA introduces:

- **Seasonal AR**: Represents the autoregressive term for the seasonal component.
- **Seasonal I**: Indicates the differencing required to remove seasonality.
- **Seasonal MA**: Represents the moving average term for the seasonal component.

The seasonal part of the model is represented as $SARIMA(p, d, q)(P, D, Q)_s$, where s is the number of time steps in each season (e.g., 12 for monthly data with yearly seasonality).

5.3 ETS (Error, Trend, Seasonality)

ETS is another versatile forecasting method that decomposes a time series into error, trend, and seasonality components.

5.3.1 Components of ETS

- **Error**: Can be additive or multiplicative. It represents the residuals or errors in the model.
- **Trend**: Can be none, additive, or multiplicative. It captures the underlying trend of the series.

- **Seasonality**: Can be none, additive, or multiplicative. It captures the seasonal fluctuations in the series.

The ETS model is often represented as $ETS(Error, Trend, Seasonality)$, with each component specified by its type (e.g., ETS(A, A, M) for additive error, additive trend, and multiplicative seasonality).

This model's decomposition approach can be particularly useful in isolating the error, trend, and seasonality components of Google's stock prices. By understanding the trend, we can gauge the long-term trajectory of the stock. If there's a consistent upward or downward trend, it can inform long-term investment strategies. The seasonality component, on the other hand, can be used to identify specific times of the year when the stock price is likely to rise or fall, guiding short-term trading decisions.

6 Forecasting

As seasonality could not be noticed in the data, we'll focus the rest of the analyss on ARIMA and ETS, comparing their behaviour and performances in terms of Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).

The following plot shows the results generated by applying these two models to out data.

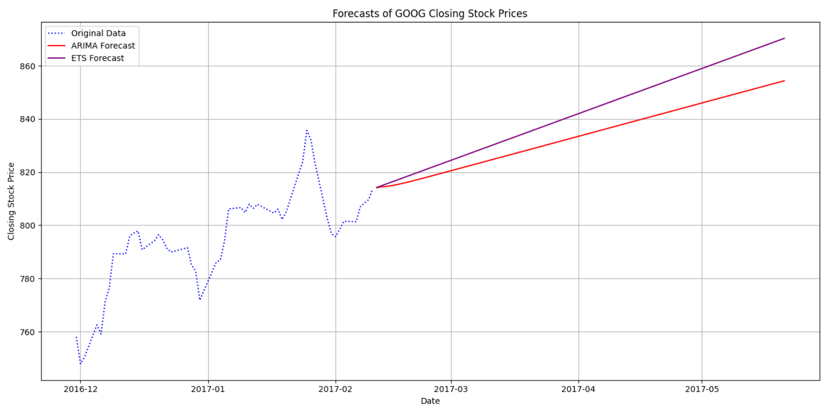

Figure 8: Forecasts of Google's Stock Prices

6.1 Forecasting Insights

Both the ARIMA and ETS models project an upward trajectory for Google's closing stock prices in the subsequent 100 days. This forecast aligns with the historical data, which predominantly showcases an ascending trend. While the forecasts from both models bear resemblance, indicating a consistent understanding of the underlying trend, they also exhibit subtle differences. These variations can be attributed to the distinct assumptions and structures inherent to each model. For instance, while ARIMA necessitates the time series to be stationary post differencing, ETS can accommodate non-stationary time series.

6.2 Analysis of Forecasting Errors

To evaluate the accuracy and reliability of our forecasting models, we computed two common error metrics: the Mean Absolute Error (MAE) and the

Root Mean Squared Error (RMSE). These metrics provide insights into the average magnitude of the errors between predicted and observed values.

6.3 Error Metrics Defined

- **Mean Absolute Error (MAE)**: Represents the average of the absolute differences between the forecasted and actual values. It provides a direct measure of the average magnitude of errors without considering their direction. Mathematically, it is defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

  where y_i is the actual value, $\hat{y}_i$ is the forecasted value, and n is the number of data points.

- **Root Mean Squared Error (RMSE)**: Measures the square root of the average of squared differences between forecasted and actual values. It gives more weight to larger errors than smaller ones, making it sensitive to outliers. It is defined as:

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

6.4 Error Values

- **ARIMA Model**:

$$\begin{aligned} \text{MAE} &: 46.815972247167764 \\ \text{RMSE} &: 49.580851847849104 \end{aligned}$$

- **ETS Model**:

$$\begin{aligned} \text{MAE} &: 55.366687566163584 \\ \text{RMSE} &: 58.07181916269611 \end{aligned}$$

6.5 Interpretation

The ARIMA(1,1,3) model, characterized by one autoregressive term, one differencing term, and three moving average terms, exhibits a slightly lower MAE and RMSE compared to the ETS model, suggesting that, on average, its forecasts are closer to the actual values. However, both models have relatively close error metrics, indicating that their performance is comparable.

It's essential to consider these error values in the context of the dataset's scale and the potential financial implications of stock price forecasting. While a difference of a few units in MAE or RMSE might seem small, it can translate to significant financial discrepancies in the stock market realm.

7 Conclusion

Our comprehensive analysis of Google's closing stock prices aimed to understand the underlying patterns and to forecast future values. The time series decomposition revealed a dominant trend component, with no confirmable seasonality. This observation influenced our choice of models, leading us to the ARIMA and ETS models.

The ARIMA(1,1,3) model, characterized by one autoregressive term, one differencing term, and three moving average terms, captured the inherent patterns in the data effectively. On the other hand, the ETS model, with an additive trend component, provided a different perspective, leveraging error, trend, and (if present) seasonality components.

Comparing the forecasted results, both models predicted an overall upward trend for Google's closing stock prices in the subsequent 100 days. The error metrics further quantified the models' performance, with the ARIMA model slightly outperforming the ETS model in terms of MAE and RMSE.

However, it's crucial to approach these forecasts with caution. While the models are grounded in historical data and robust statistical methods, stock prices are inherently volatile and influenced by myriad factors, both predictable and unforeseeable. The models, despite their sophistication, cannot account for sudden market changes, global events, or internal corporate decisions that might significantly sway stock prices.