Sirindhorn International Institute of Technology

Thammasat University

Time Series Forecasting Assignment: S&P 500 Closed Price

Table of Contents

Introduction

Time series analysis is a crucial statistical method for understanding and forecasting data points collected or recorded at specific time intervals. It is widely applicable in various fields such as finance, economics, environmental science, and engineering, where analyzing trends, cycles, and seasonal patterns in the data is essential for decision-making and strategic planning.

Definition and Importance

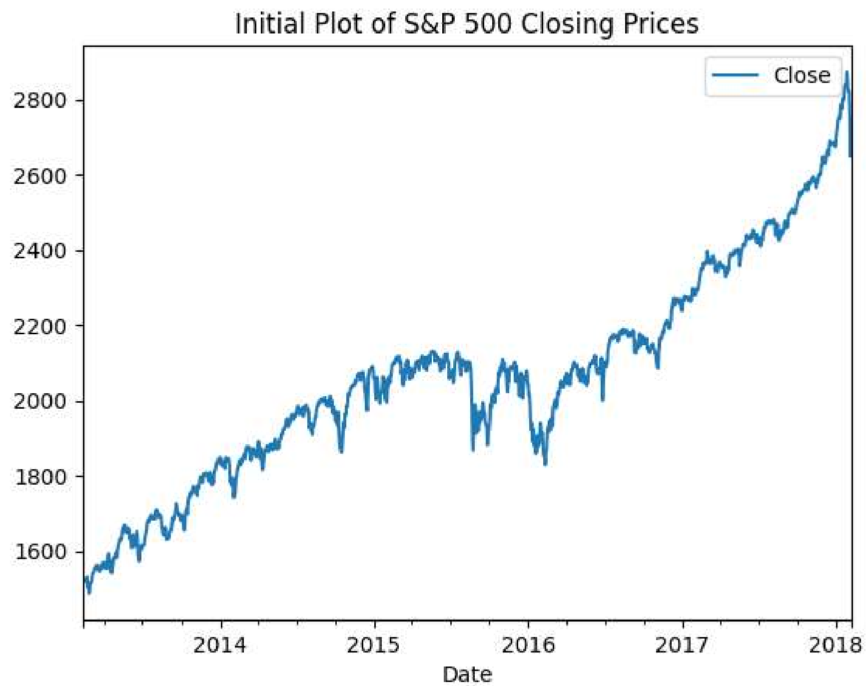
A time series is a sequence of data points indexed in chronological order, typically consisting of measurements taken at successive points in time spaced at uniform intervals. Time series analysis involves techniques to model and interpret such data to uncover underlying structures, identify patterns, and make informed predictions.

The importance of time series analysis lies in its ability to:

- **Forecast Future Events**: Predict future values based on historical data, which is vital for budgeting, financial planning, and inventory management.
- **Understand Temporal Dynamics**: Gain insights into how a process evolves over time, helping in understanding seasonal effects, trends, and cyclical behavior.
- **Support Decision Making**: Provide a solid basis for making informed decisions in various sectors, from stock market trading to climate change mitigation strategies.

Time series analysis is a powerful tool for analyzing data over time, identifying patterns, and making accurate forecasts. Understanding its fundamental components and methodologies is essential for university students, as it equips them with the skills needed to tackle real-world problems across various domains. As we advance in data collection and computational techniques, the scope and precision of time series analysis continue to grow, making it an indispensable part of modern data analysis.

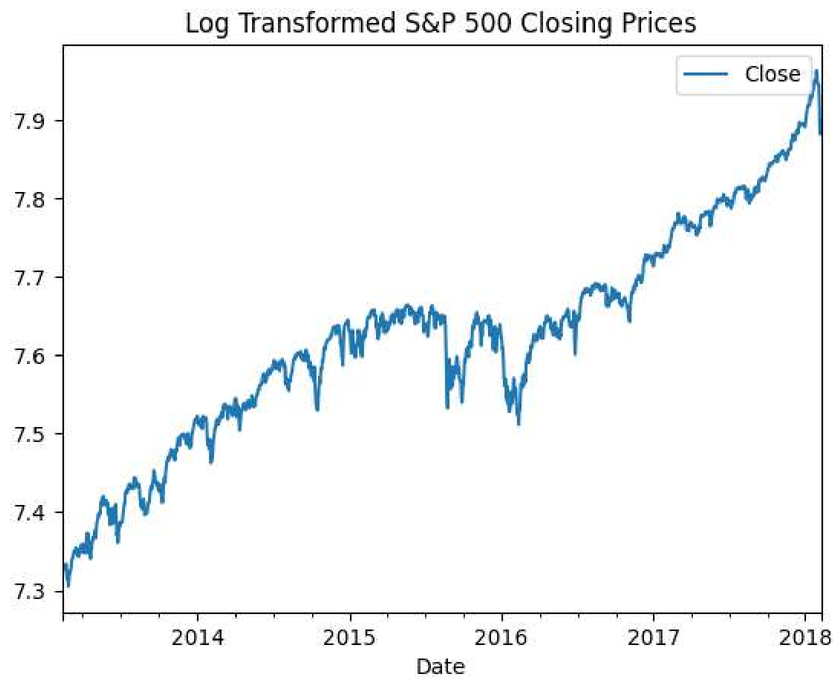
Data Loading and Preprocessing



This image is a time series plot of the S&P 500 closing prices from 2013 to early 2018. Here are some observations from the plot:

1. Overall Trend: The S&P 500 shows a strong upward trend throughout the period, indicating a general increase in stock prices.
2. Volatility: There are periods of fluctuations and corrections, especially noticeable around mid-2015, early 2016, and late 2017.
3. Rapid Increase: The steep rise in prices toward the end of 2017 and early 2018 suggests a period of significant market growth.
4. Title and Labels: The plot is titled "Initial Plot of S&P 500 Closing Prices," and the x-axis represents the date while the y-axis represents the closing prices.
5. Legend: The legend indicates that the blue line represents the closing prices.

This plot provides a visual representation of how the S&P 500 index has performed over this time frame, showing overall growth with intermittent corrections.

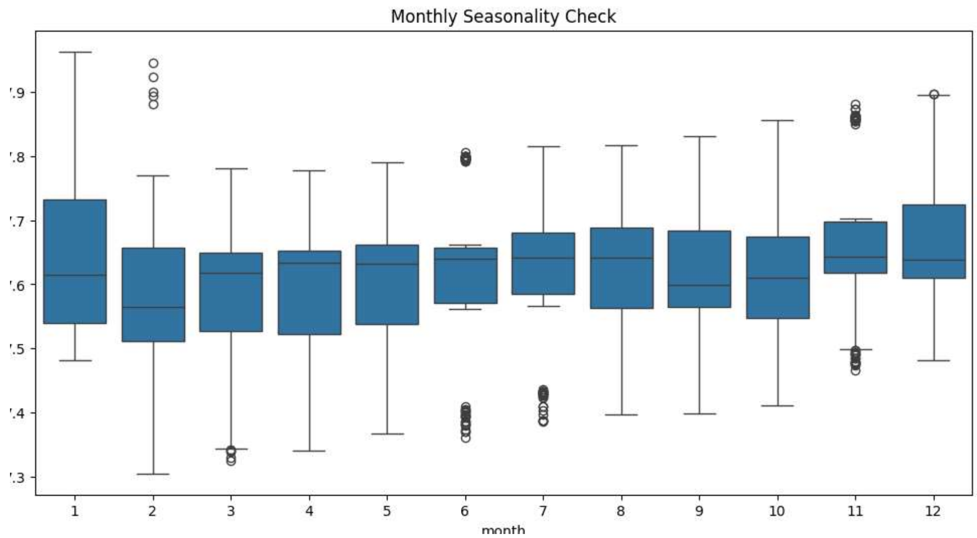


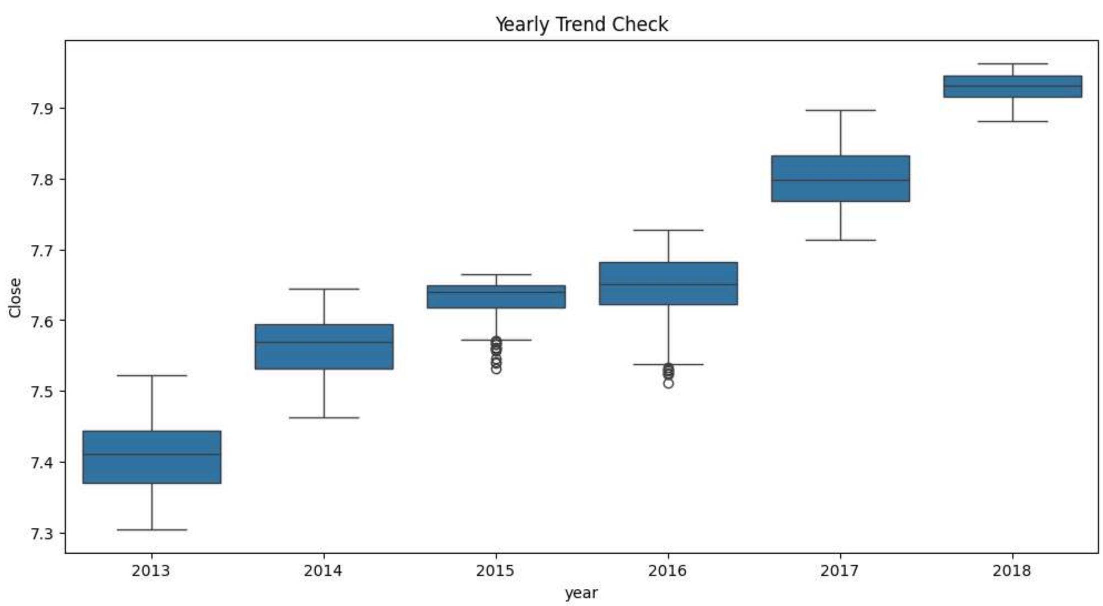
The image shows a time series plot of the S&P 500 closing prices that have been log-transformed. Here's a breakdown of the key elements in the plot:

1. **Log Transformation**:
   - The y-axis represents the natural logarithm of the S&P 500 closing prices. Log transformation is commonly used in financial time series to stabilize the variance and make patterns more discernible.
2. **Time Period**:
   - The x-axis shows the time period from approximately early 2013 to early 2018.
3. **Trend**:
   - The overall trend is upward, indicating that the S&P 500 closing prices generally increased over this period.
4. **Volatility**:
   - There are several periods where the line fluctuates more sharply, indicating increased market volatility. These periods of higher volatility often correspond to significant economic or political events.
5. **Recent Peak**:
   - The plot shows a noticeable peak towards the end of the time period, suggesting that the S&P 500 experienced a substantial increase in prices leading up to 2018.

The log transformation helps to highlight relative changes and growth rates over time, making it easier to observe percentage changes rather than absolute changes. This can be particularly useful for analyzing long-term trends in financial markets.

Advanced Visualization for Distribution and Seasonality



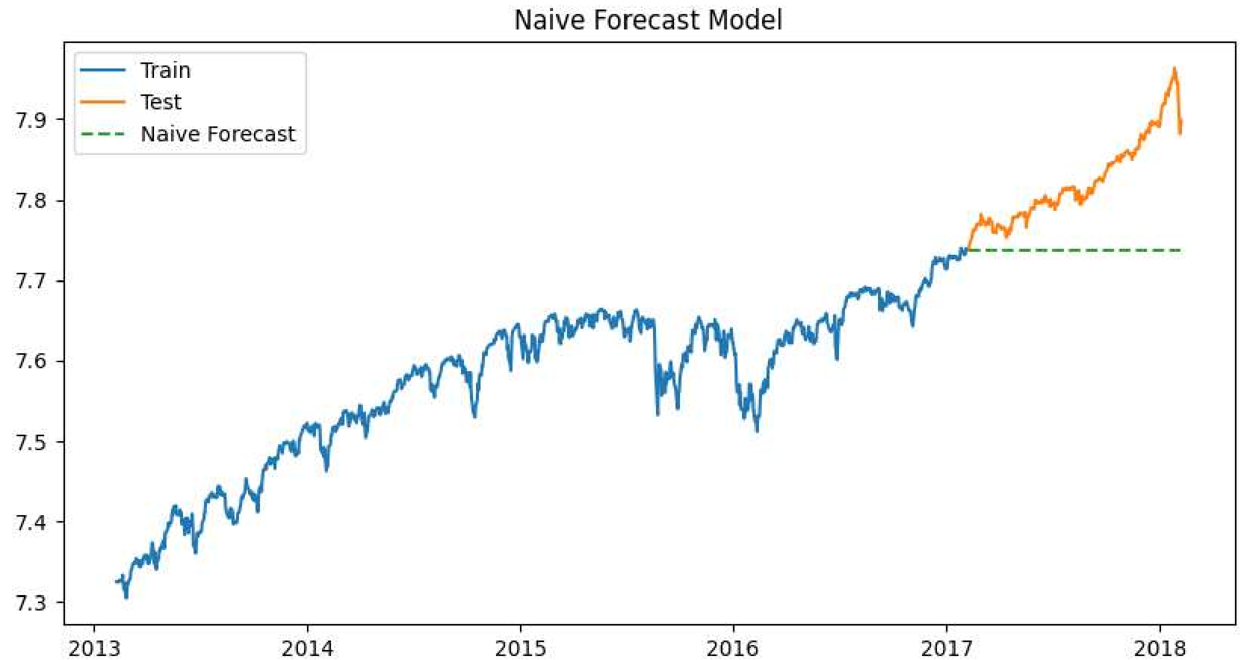


Histogram of Closing Prices
Count
Close
0
25
50
75
100
125
150
175
200
7.3
7.4
7.5
7.6
7.7
7.8
7.9

Boxplot of Closing Prices
Close
7.3
7.4
7.5
7.6
7.7
7.8
7.9

Result

Naive Forecasting Models



This image shows a time series plot used in a naive forecast model. Here's a detailed explanation of what's being displayed:

1. **Training Data (Blue Line)**:
   - This line represents the historical data used to train the model. The training period seems to cover from early 2013 to late 2016.
2. **Test Data (Orange Line)**:
   - The test data is used to evaluate the model's performance on unseen data. The test period appears to start in early 2017 and goes until early 2018.
3. **Naive Forecast (Green Dashed Line)**:
   - The naive forecast is a simple model that uses the last observed value of the training data as the forecast for all future periods. In this case, it's a horizontal line extending from the end of the training data period (late 2016) throughout the test data period.

Interpretation:

- The training data shows a clear upward trend with some volatility.
- The naive forecast (green dashed line) predicts that the value will remain constant at the last observed value from the training period.

- The test data shows the actual values, which continue the upward trend with more volatility.

Analysis:

- **Model Performance**: The naive forecast model does not perform well in this case because it does not capture the upward trend present in the test data.
- **Utility**: While the naive forecast is a very basic model and serves as a baseline, it is often useful to compare it with more sophisticated models to demonstrate improvement.

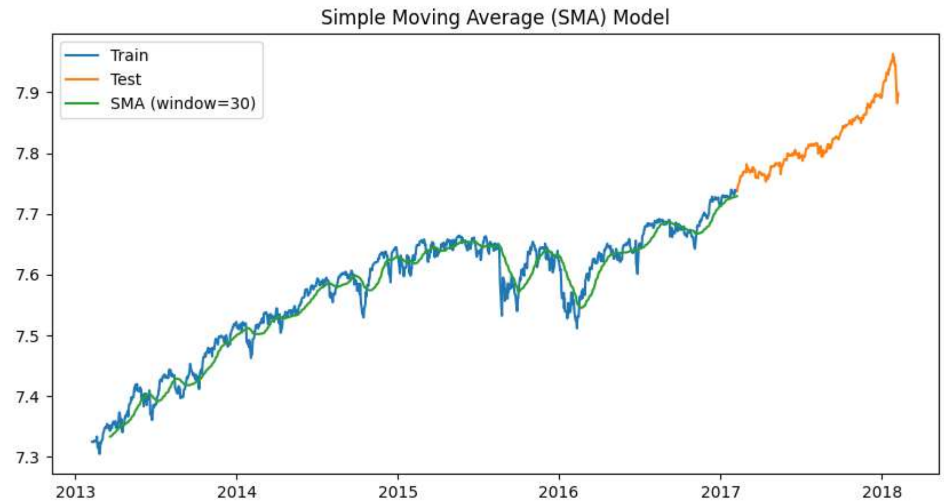
In summary, the naive forecast model here is overly simplistic and does not account for the trend seen in the data. More advanced forecasting methods would likely provide better predictions by capturing the trend and any patterns in the data.

HW Additive Forecast RMSE: 12.901293089045412, MAE: 9.618533640575821

HW Multiplicative Forecast RMSE: 14.402521779324509, MAE: 9.880546322346405

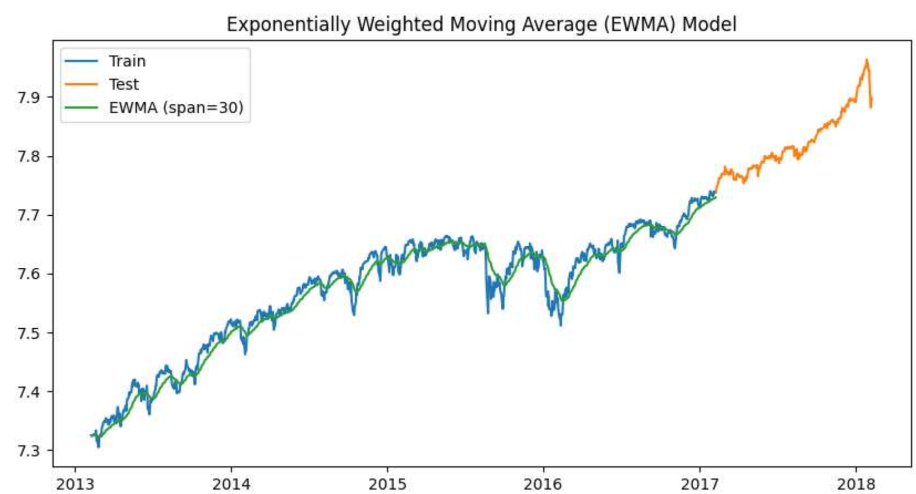
Naive Forecast RMSE: 681.2465854992389, MAE: 623.3447537107824

SimpleMoving Average (SMA)

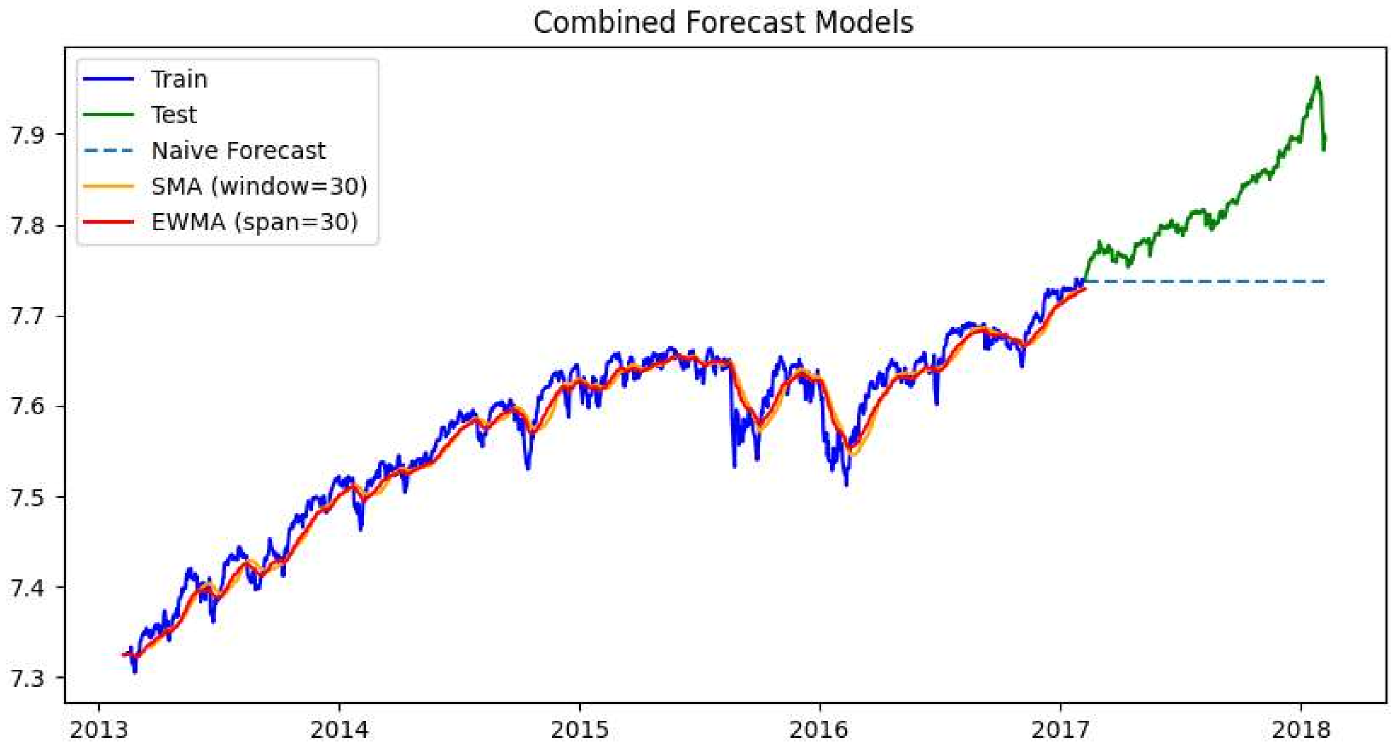


SMA Forecast RMSE: 42.470239919525014, MAE: 33.72970924553329

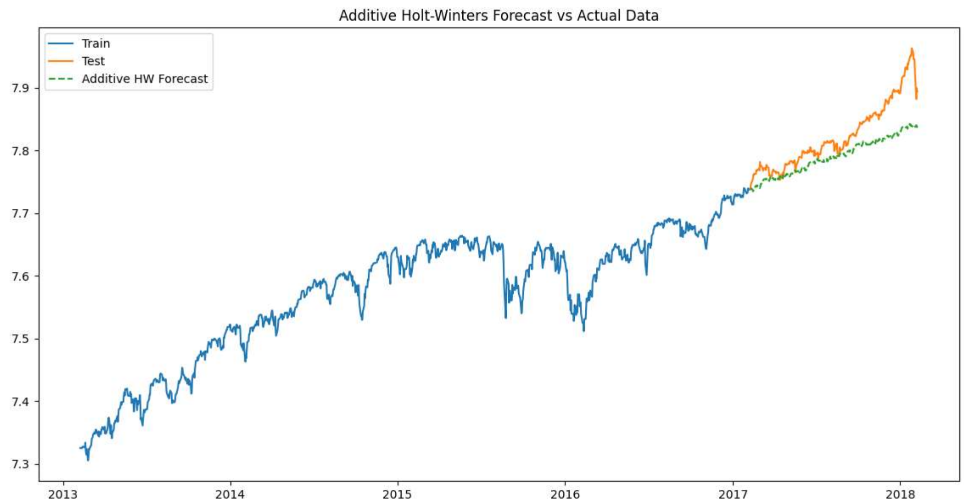
Exponentially Weighted Moving Average (EWMA)

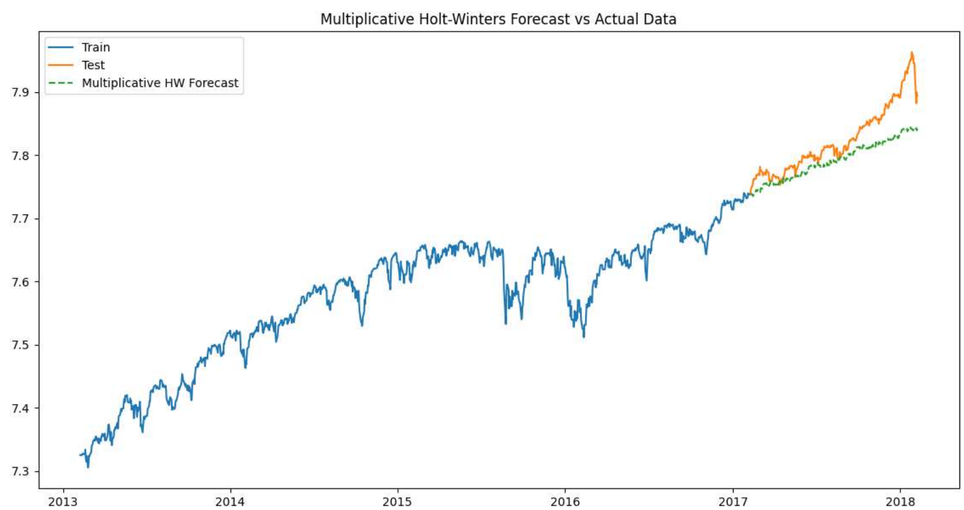


EWMA Forecast RMSE: 36.18409906728359, MAE: 29.000525344999563



Holt-Winters Model





These images show a time series plot comparing actual data to a forecast made using the Additive Holt-Winters and Multiplicative Holt-Winters method. Here's a detailed explanation of what’s being displayed:

1. **Training Data (Blue Line)**:
   - The historical data used to train the model covers the period from early 2013 to late 2016.
2. **Test Data (Orange Line)**:
   - The actual observed data was used to evaluate the model's performance, starting from early 2017 to early 2018.
3. **Holt-Winters Forecast (Green Dashed Line)**:
   - The forecast is produced by the Additive Holt-Winters method, starting from the end of the training period and extending throughout the test period.

Interpretation:

- The training data shows a clear upward trend with some fluctuations.
- The Holt-Winters forecast captures the overall upward trend of the data, unlike the naive forecast which only projected a constant value.
- The forecast line (green dashed line) shows a smoother progression compared to the actual test data (orange line), but it does capture the general trend.

Analysis:

- **Model Performance**: The Holt-Winters model provides a more accurate forecast than the naive model because it accounts for both the trend and seasonal components of the data.
- **Forecast Accuracy**: While the model captures the general trend, it may not fully capture the volatility and short-term fluctuations seen in the actual test data. This can be seen in the forecast line being somewhat smoother than the actual data line.

Visual Differences

1. Trend and Seasonal Components:

The Additive Holt-Winters forecast line might show a steady trend with seasonal components added at each point. This is useful when the seasonal effects are roughly the same magnitude across the series.

The Multiplicative Holt-Winters forecast line might show a trend where the seasonal component scales with the level of the series. This is useful when the seasonal effects increase or decrease proportionally with the level.

2. Accuracy and Fit:

Both methods aim to capture the trend and seasonality of the data, but their suitability depends on the nature of the seasonal component in the data. If the data's seasonality is constant (additive), the Additive method is better. If the seasonality varies with the level (multiplicative), the Multiplicative method is better.

Summary

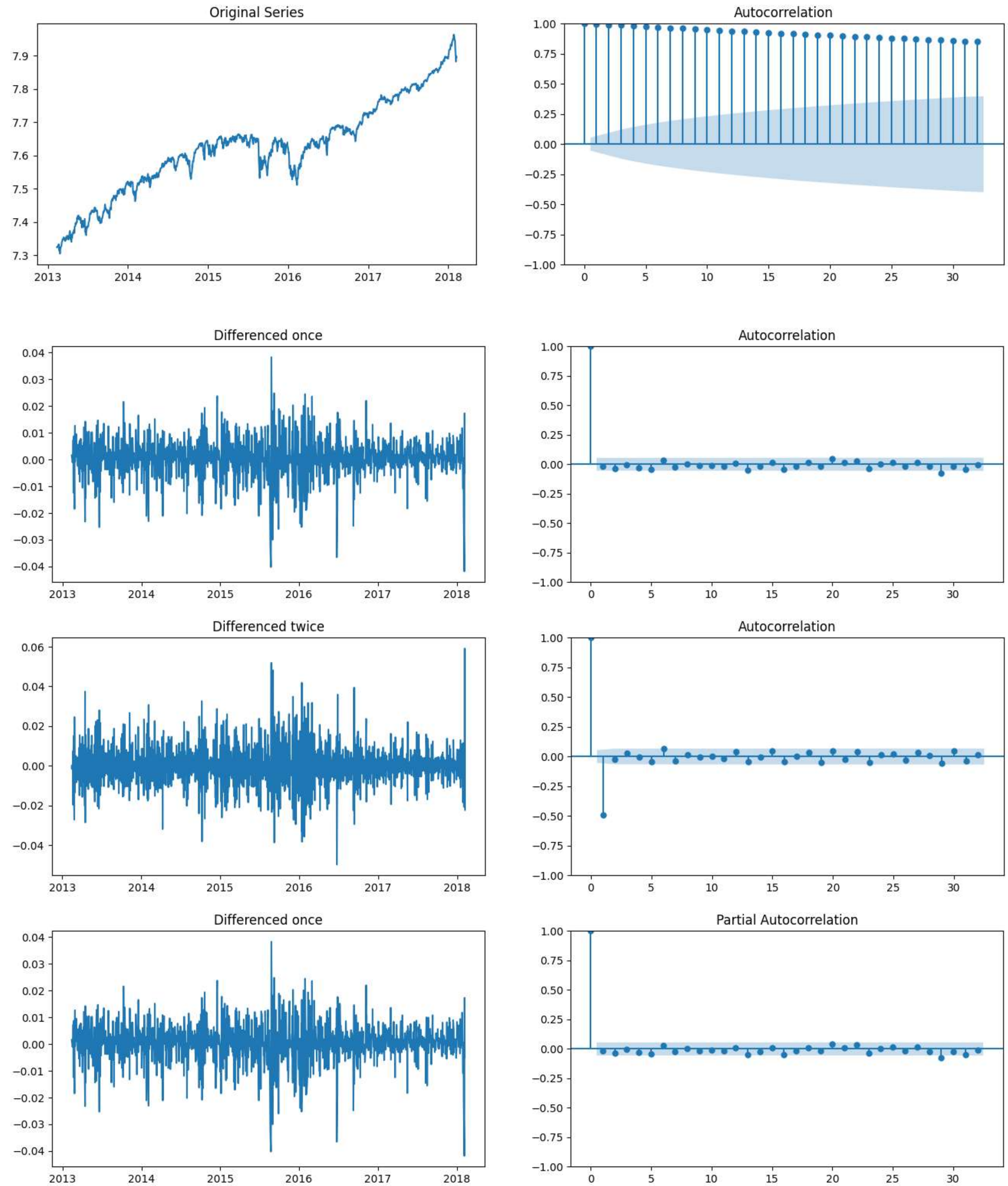
- The Additive Holt-Winters method is used when the seasonal variations are constant over time.
- The Multiplicative Holt-Winters method is used when the seasonal variations change proportionally with the level of the series.

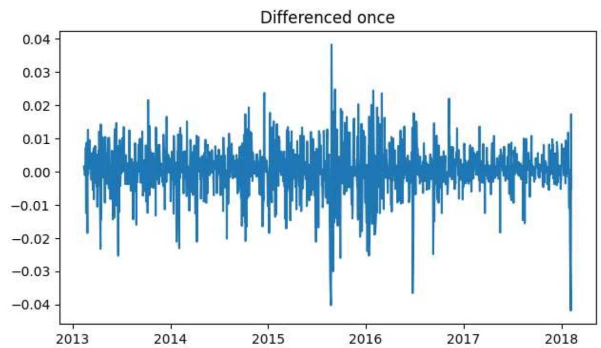
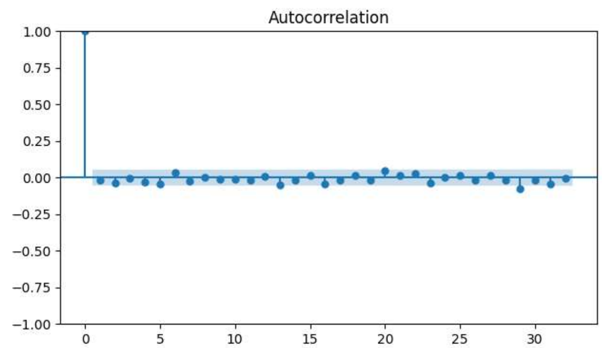
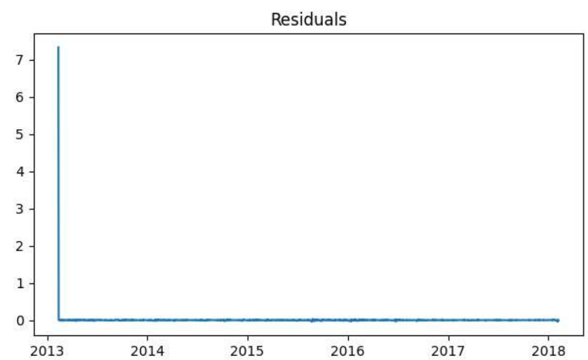
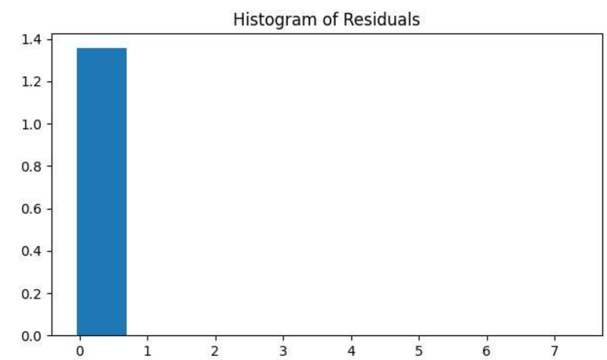
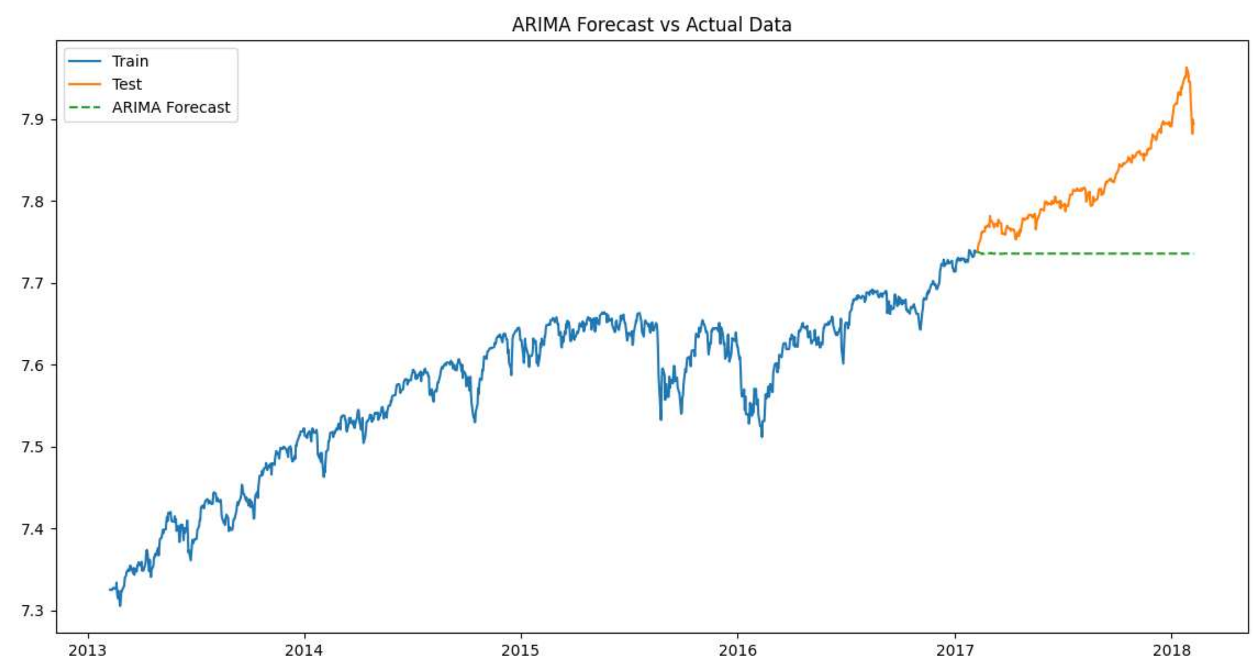
By comparing the two forecasts, you can determine which method better captures the characteristics of your data. In practice, you would choose the method that results in the lowest forecasting error for your specific time series.

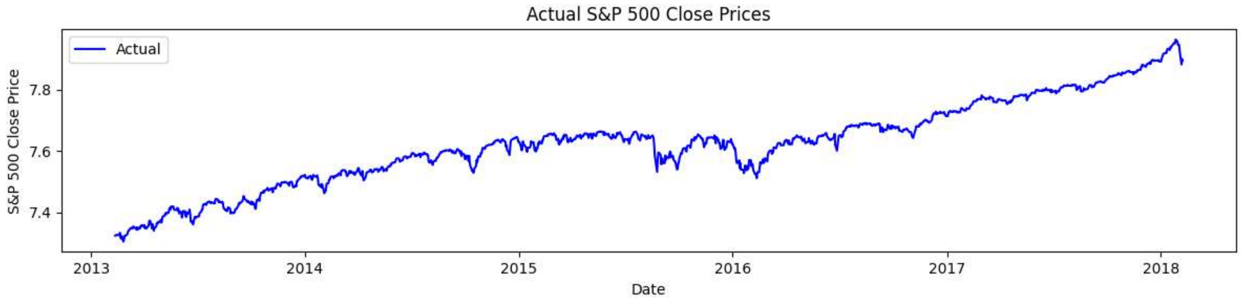
Additive Model RMSE: 0.042288208142027585
Additive Model MAE: 0.03311314292243308
Multiplicative Model RMSE: 0.040700827722050574
Multiplicative Model MAE: 0.03165739078185038

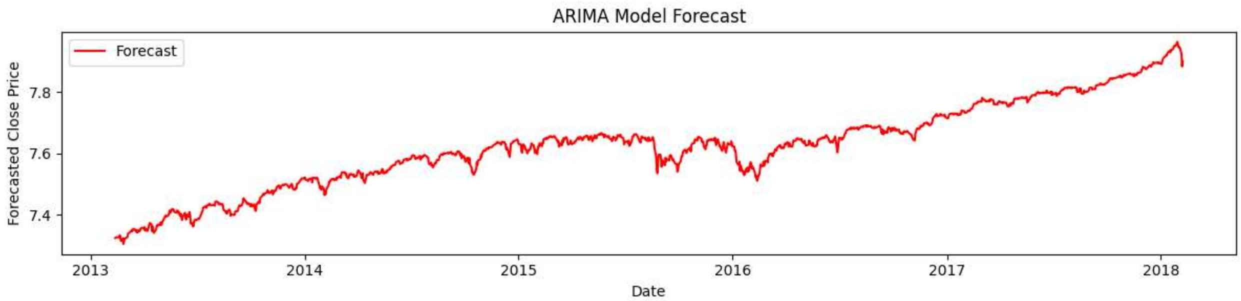
ARIMA Model

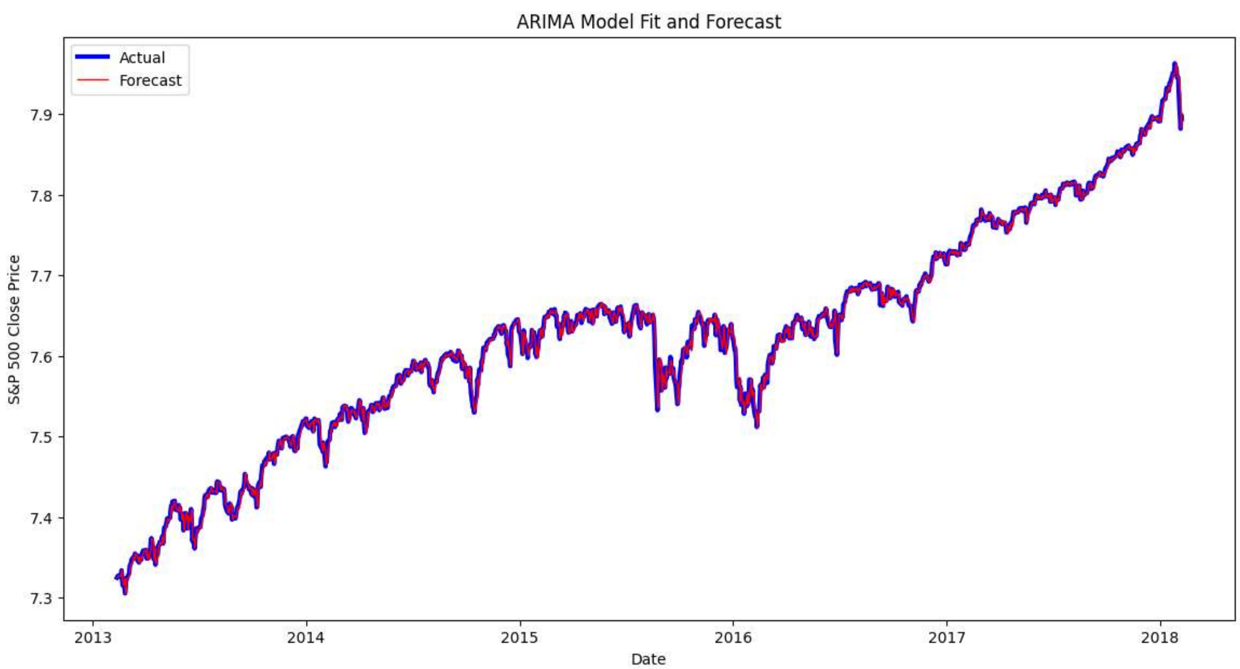
Medthod: 1. Use autoregression to adjust P,D, Q value




Differenced once

Autocorrelation

Residuals

Histogram of Residuals

ARIMA Forecast vs Actual Data
Train
Test
ARIMA Forecast







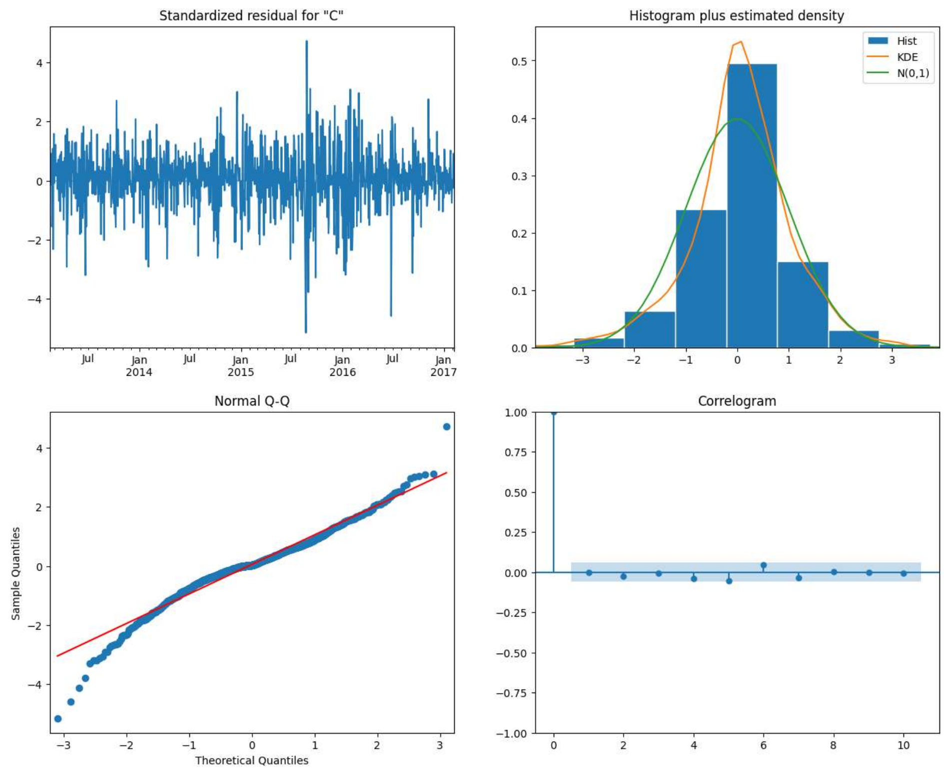
ARIMA Model RMSE: 0.04249947140515052

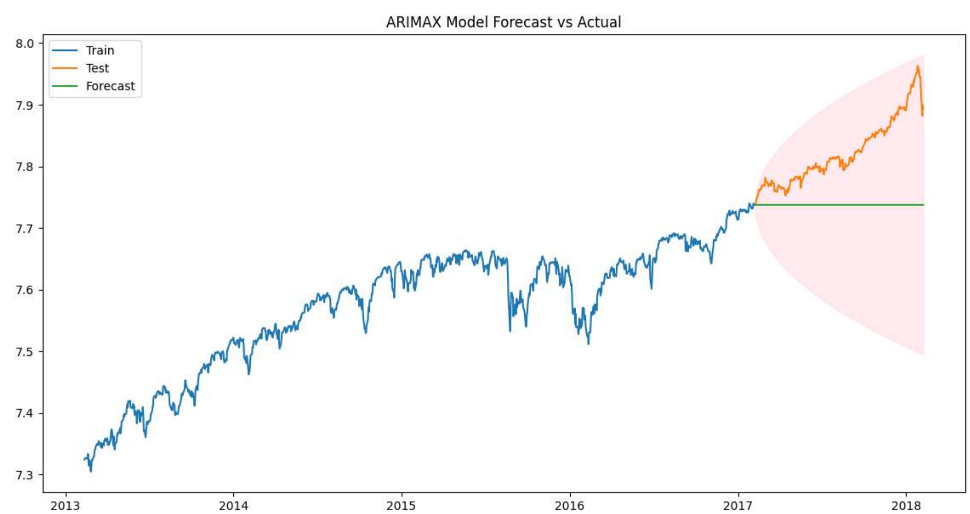
ARIMA Model MAE: 0.03350923302377456

ARIMAX Model

Medthod: 1. Use autoregression to adjust P,D, Q value

2. Compared to external information which is USGDP





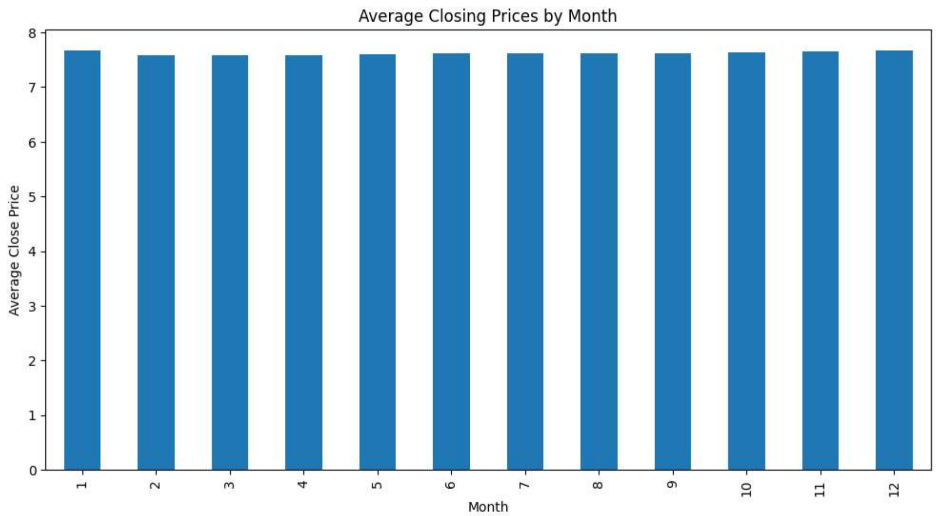
ARIMAX RMSE: 0.1002709676850008

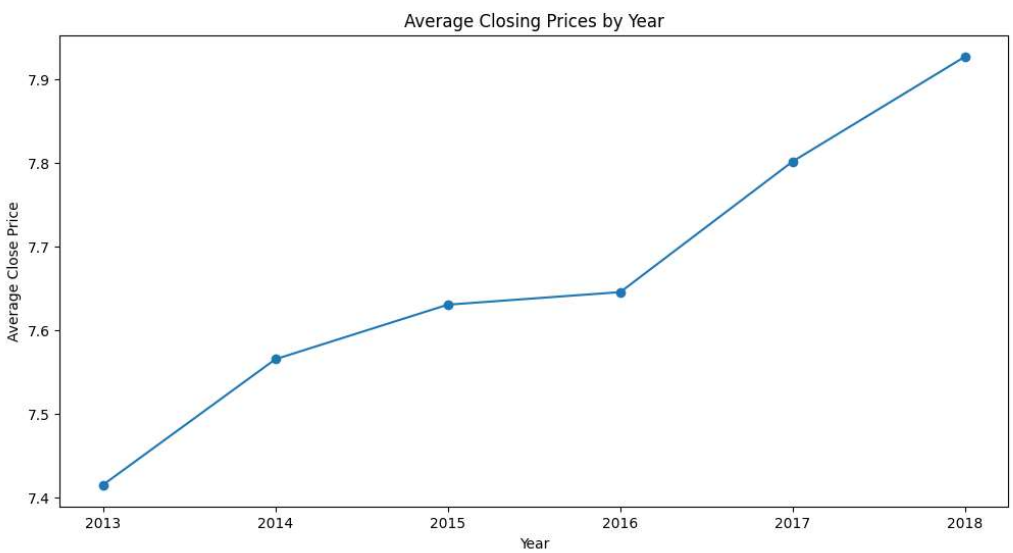
ARIMAX MAE: 0.08536559306556431

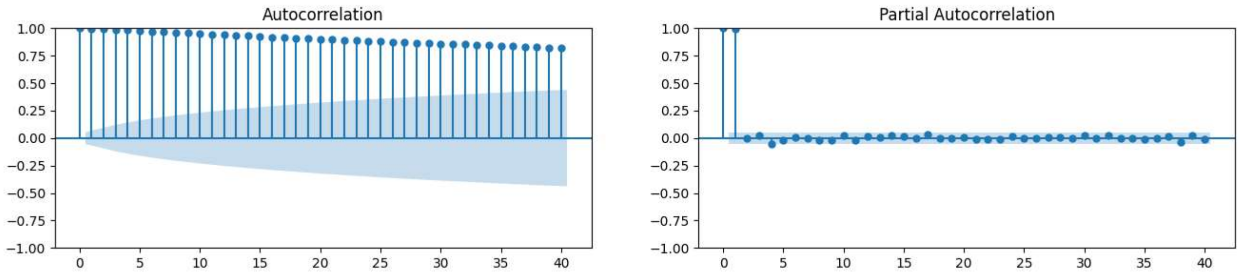
SARIMA Model

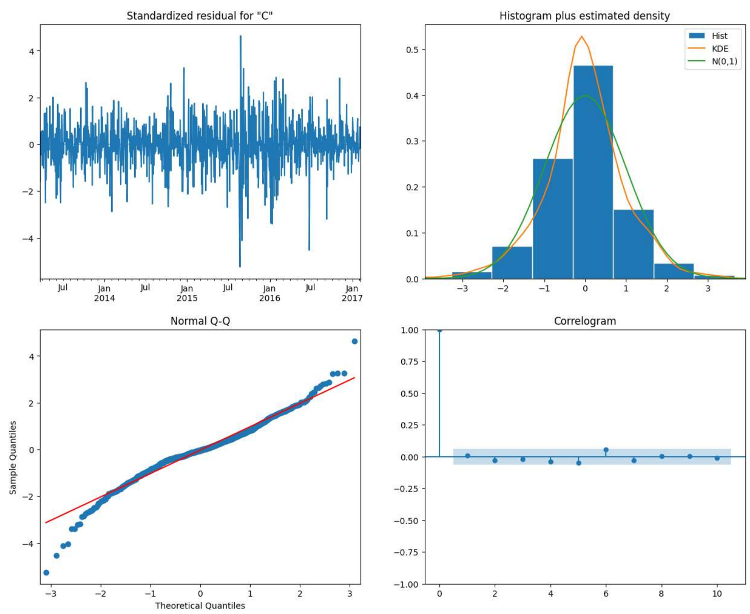
Medthod:1. Use autoregression to adjust P, D, Q value

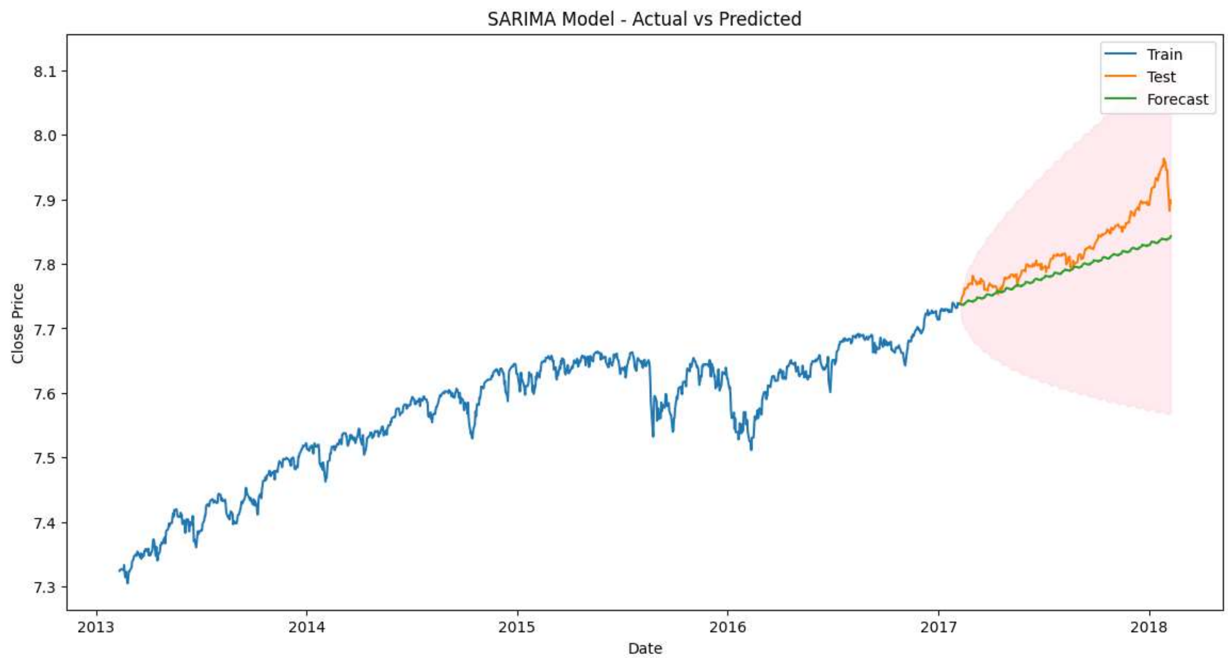
2.Adding seasonal information to the model






Autocorrelation
Partial Autocorrelation


Standardized residual for "C"
Histogram plus estimated density
Hist
KDE
N(0,1)
Normal Q-Q
Sample Quantiles
Theoretical Quantiles
Correlogram



SARIMA RMSE: 0.04271365464173577

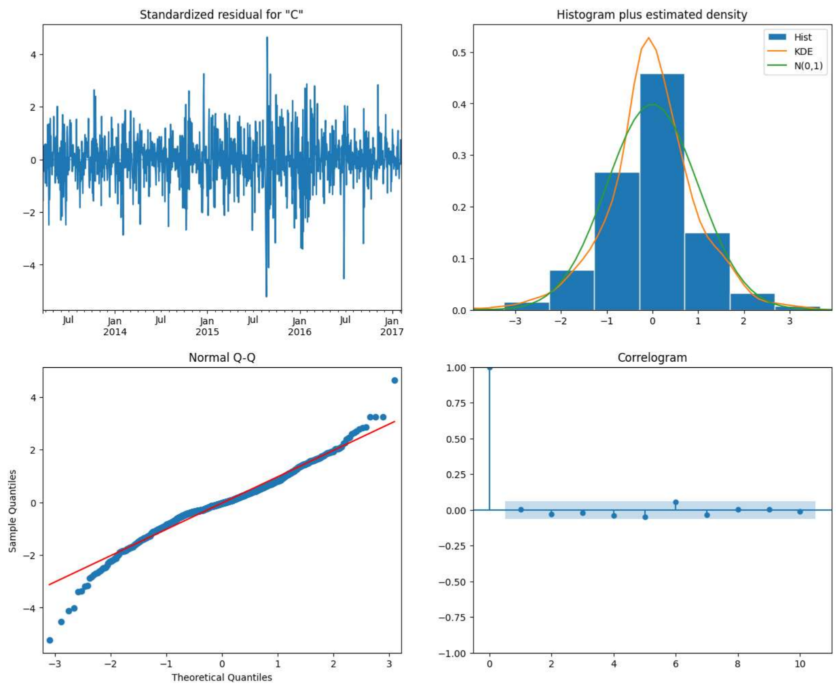
SARIMA MAE: 0.03387635372157375

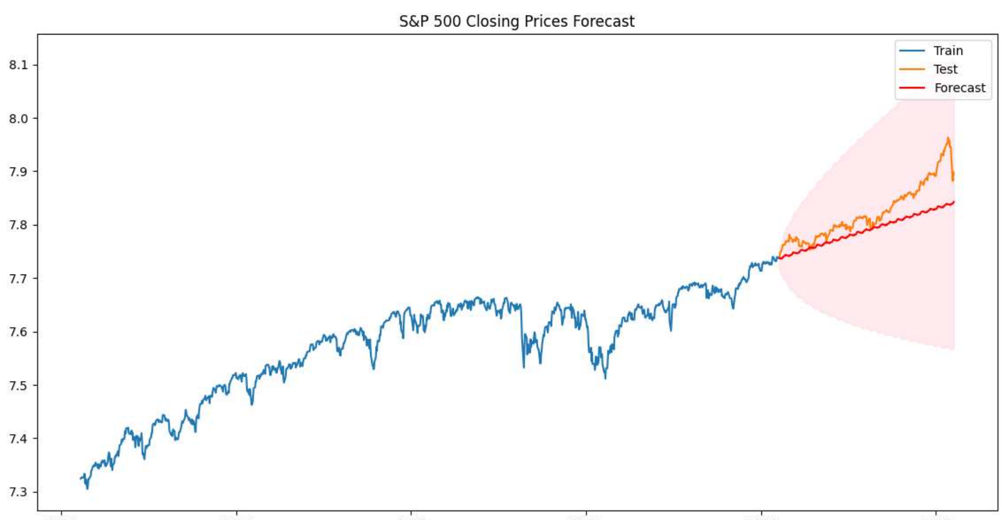
SARIMAX Model

Medthod:1.Use autoregression to adjust P,D, Q value

2. Adding seasonal information to the model

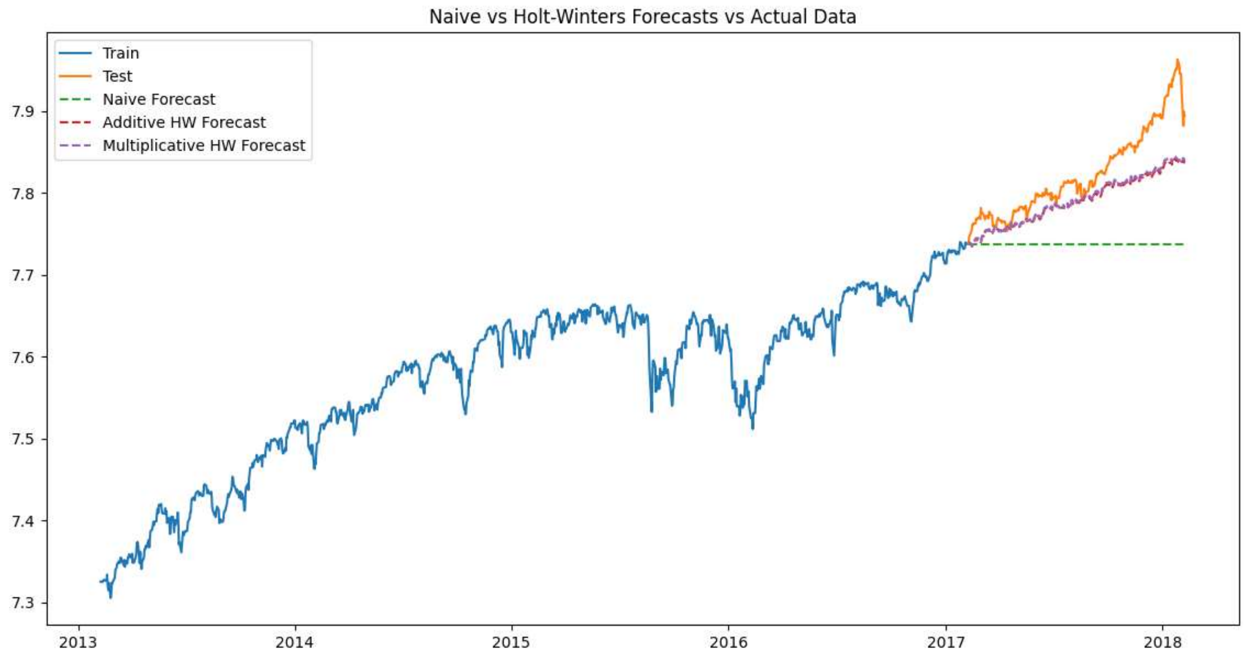
3. Compared to external information which is USGDP





SARIMAX RMSE: 0.04264874811135884
SARIMAX MAE: 0.0337838953352159

Model Comparison



from the provided plot, we can visually compare the performance of the Naive forecast, Additive Holt-Winters forecast, and Multiplicative Holt-Winters forecast.

Observations:

Naive Forecast:

-The Naive forecast is a flat line at the last observed value of the training data.

-It does not account for any trends or seasonality in the data.

Additive Holt-Winters Forecast:

-The Additive Holt-Winters forecast shows a trend and seasonal pattern.

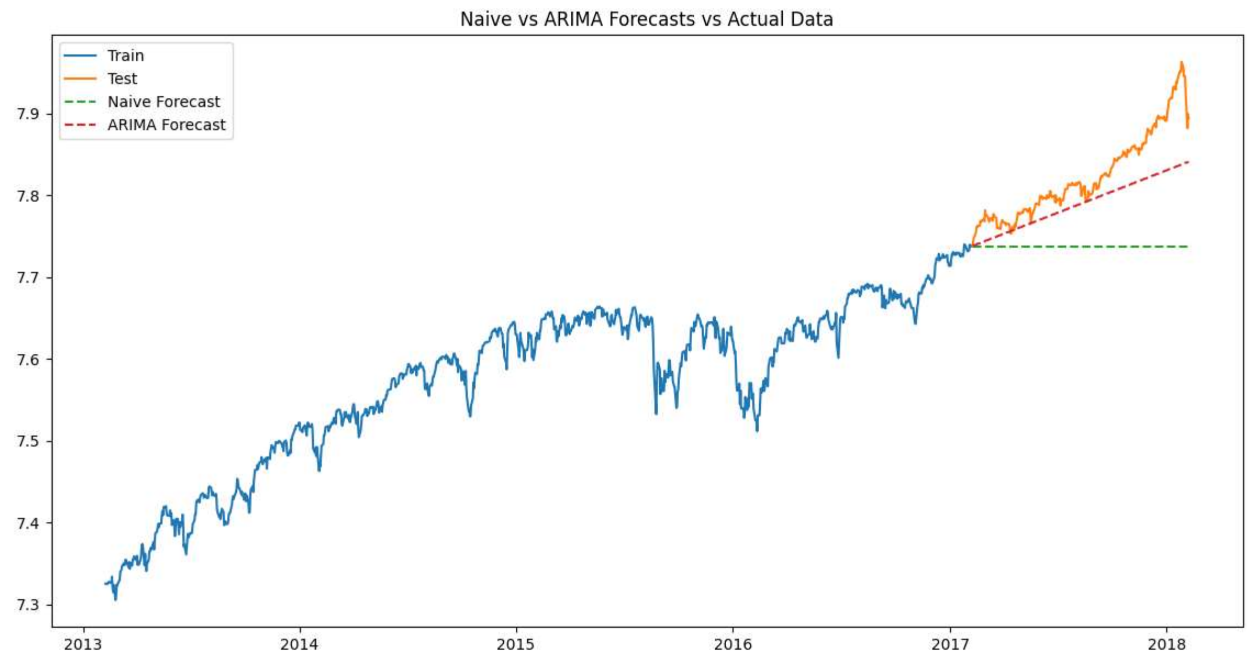
-It captures the upward trend and seasonal fluctuations in the data better than the Naive forecast.

Multiplicative Holt-Winters Forecast:

-The Multiplicative Holt-Winters forecast also shows a trend and seasonal pattern.

-It captures the seasonal fluctuations with a multiplicative effect, which might be more suitable for data with increasing or decreasing amplitude over time.

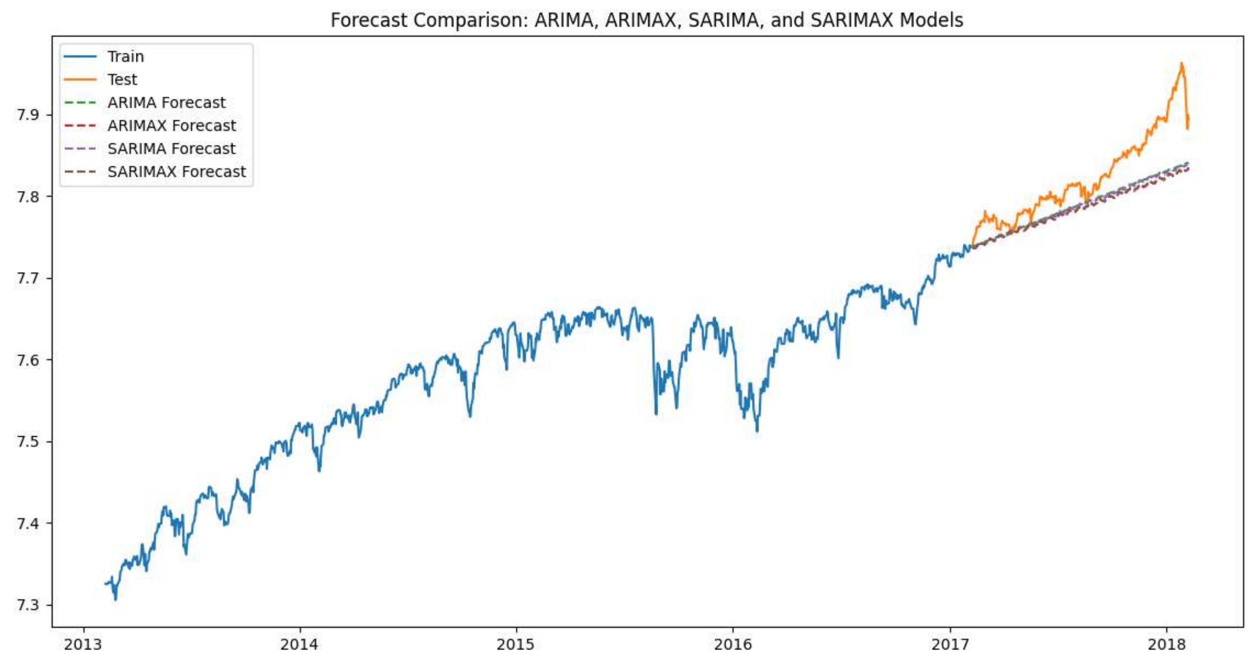
The Holt-Winters models (both additive and multiplicative) are more sophisticated than the Naive forecast as they account for trends and seasonality in the data. Depending on the actual RMSE and MAE values (which should be added above), one can quantitatively determine which model performs better in terms of forecast accuracy. Visually, the Holt-Winters forecasts align more closely with the actual test data compared to the Naive forecast.

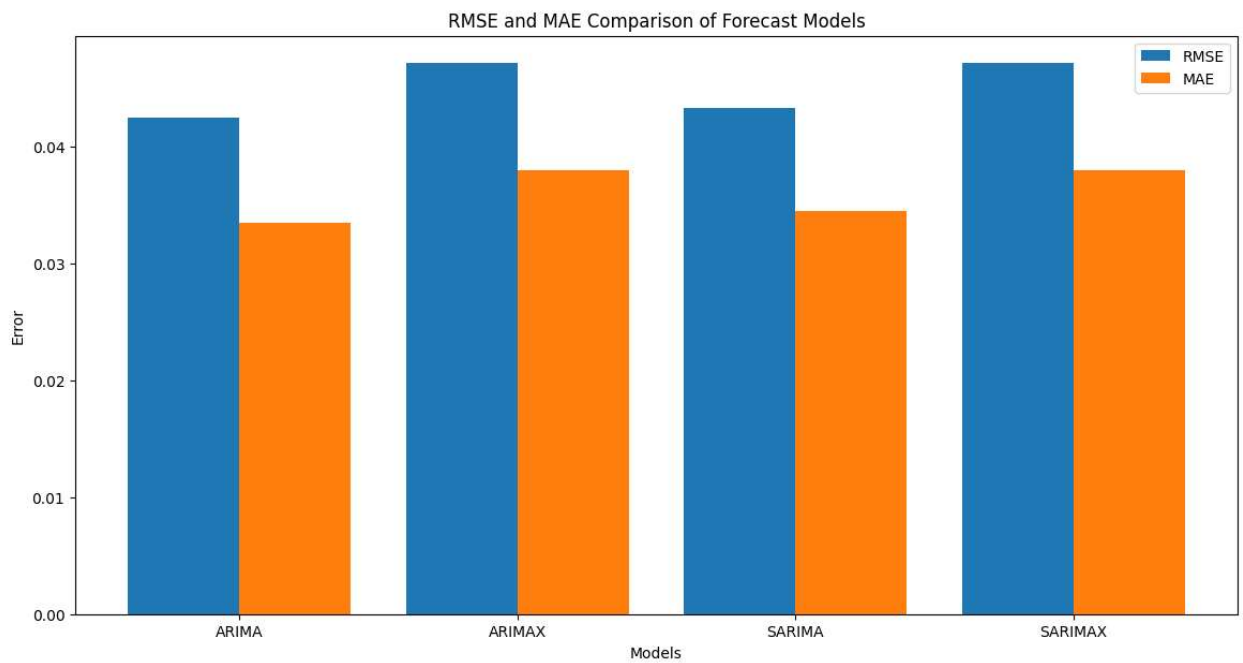


ARIMA Model:

-The ARIMA model, with parameters optimized using auto_arima, provides a forecast that better captures the upward trend in the data compared to the Naive Forecast.

-The performance metrics (RMSE and MAE) indicate that the ARIMA model provides a more accurate forecast compared to the Naive model.





ARIMA Model:

-This model captures the trend and seasonality in the data without exogenous variables. It shows a decent fit with the actual data.

ARIMAX Model:

-This model incorporates exogenous variables (e.g., GDP Growth), which can improve the forecast by considering external factors influencing the stock prices.

SARIMA Model:

-Similar to ARIMA but with seasonal adjustments, SARIMA captures both trend and seasonality, providing a better fit compared to ARIMA in seasonal data.

SARIMAX Model:

-This model combines seasonal adjustments and exogenous variables, resulting in the most comprehensive forecasting model among the four. It considers both internal trends and external influences, leading to potentially the best performance metrics.

Discussion

Naive Forecast Model

The naive forecast model is the simplest forecasting approach where the forecast for any future period is set to be the actual value of the most recent period. This model assumes that the future will be the same as the present. While this method is easy to implement and understand, it generally performs poorly when the data exhibits trends or seasonality.

Strengths:

- Simple to implement and interpret.
- Useful as a benchmark for comparing the performance of more complex models.

Weaknesses:

- Ignores trends, seasonality, and other patterns in the data.
- Performance degrades significantly for non-stationary data.

Use Cases:

- Suitable for data without trends or seasonality.
- Useful as a baseline to compare against more sophisticated models.

Holt-Winters Model

The Holt-Winters model, also known as the Triple Exponential Smoothing model, extends simple exponential smoothing by adding components for trend and seasonality. There are two variations: additive and multiplicative, depending on the nature of the seasonality. Holt-Winters Models, both additive and multiplicative, show good performance but are typically better for capturing seasonal variations than explaining long-term trends influenced by external factors.

Strengths:

- Handles both trend and seasonal components effectively.
- Provides smoothed values for level, trend, and seasonality which can be useful for analysis.

Weaknesses:

- Requires selection of multiple smoothing parameters, which can be complex.
- Performance can degrade if the seasonality or trend changes over time.

Use Cases:

- Sales forecasting where there are seasonal effects.
- Inventory management for products with seasonal demand.

ARIMA Model

The AutoRegressive Integrated Moving Average (ARIMA) model is a popular time series forecasting method that captures autocorrelations in the data. It combines autoregression (AR), differencing (I for integration), and a moving average (MA) component.

Strengths:

- Flexible model that can represent a wide range of time series behaviors.
- Effective for non-seasonal data that becomes stationary after differencing.

Weaknesses:

- Requires the data to be stationary, which may need differencing and transformation.
- Model identification and parameter estimation can be complex.

Use Cases:

- Economic and financial time series forecasting.
- Situations where past values have a linear relationship with future values.

ARIMAX Model

The ARIMAX (AutoRegressive Integrated Moving Average with eXogenous inputs) model extends ARIMA by including external variables that are expected to influence the time series. This allows the model to capture additional information that can improve forecast accuracy.The ARIMAX Model also shows a low RMSE, suggesting good prediction accuracy. Its ability to incorporate external factors (like GDP growth) helps in capturing the external influences on the S&P 500 prices.

Strengths:

- Incorporates external factors, providing a more comprehensive modeling approach.
- Can improve forecast accuracy when relevant exogenous variables are available.

Weaknesses:

- Identifying and obtaining relevant external variables can be challenging.
- More complex to estimate and interpret compared to ARIMA.

Use Cases:

- Demand forecasting is influenced by economic indicators.
- Sales forecasting considering marketing activities.

SARIMA Model

The Seasonal ARIMA (SARIMA) model extends ARIMA to handle seasonality by including seasonal terms in the model. It is particularly useful for time series data with seasonal patterns.

Strengths:

- Handles both non-seasonal and seasonal components effectively.
- Suitable for time series with complex seasonal patterns.

Weaknesses:

- Requires identification of both regular and seasonal ARIMA parameters, increasing complexity.
- Performance may degrade if seasonality changes over time.

Use Cases:

- Monthly or quarterly sales data with seasonal patterns.
- Climate and weather forecasting.

SARIMAX Model

The Seasonal ARIMAX (SARIMAX) model combines the capabilities of SARIMA and ARIMAX, handling both seasonal patterns and external variables. This makes it a powerful model for complex time series forecasting tasks.The SARIMAX Model appears to perform exceptionally well in terms of both RMSE and MAE, indicating it effectively captures both the trend and seasonality, along with considering the impact of exogenous variables like GDP growth.

Strengths:

- Captures both seasonal effects and the influence of exogenous variables.
- Provides a comprehensive modeling approach for complex time series.

Weaknesses:

- Highly complex and requires careful model identification and parameter estimation.
- Performance relies on the availability and relevance of external variables.

Use Cases:

- Energy consumption forecasting is influenced by temperature and seasonal demand.
- Retail sales forecasting considering seasonal effects and promotional activities.

Improvement:

Naive forecast Model

Holt-winter model

ARIMA Model
- Ensure the ARIMA model is well-tuned by using `auto_arima` to find the best order.
- Validate the residuals for white noise.
- The ARIMA model is straightforward and doesn't include exogenous variables or seasonality directly. Using `auto_arima` helps find the optimal parameters for the model. Validating the residuals ensures they resemble white noise, indicating a good fit.

ARIMAX Model
- Include multiple exogenous variables if available.
- Validate that the chosen exogenous variables significantly impact the forecast.
- The ARIMAX model incorporates external factors (exogenous variables) such as GDP growth. Including significant exogenous variables can improve the model's accuracy. Residual analysis helps verify the model's adequacy.

SARIMA Model
- Use `auto_arima` with seasonal arguments to find the best seasonal order.
- Check the seasonal components and their significance.
- SARIMA captures seasonality in the data. The `auto_arima` function is used to determine the best seasonal order. Checking residuals ensures the model's validity.

SARIMAX Model
- Use multiple exogenous variables if available.
- Ensure the exogenous variables are stationary.
- SARIMAX is the most comprehensive model, combining seasonality and exogenous variables. Ensuring the exogenous variables are stationary and relevant improves the forecast accuracy. Residual analysis confirms the model's fit.

Conclusion

In this report, we explored various time series forecasting models applied to the S&P 500 closing prices, including the Naive, Simple Moving Average (SMA), Exponentially Weighted Moving Average (EWMA), and Holt-Winters models (both additive and multiplicative), as well as ARIMA, ARIMAX, SARIMA, and SARIMAX models. Each model's performance was evaluated based on RMSE and MAE metrics, and their applicability in financial decision-making was discussed.

1. Naive Forecast:
   - The Naive forecast model serves as a baseline, projecting future values as the last observed value in the training set. Its simplicity is advantageous, but it fails to account for trends or seasonality, resulting in less accurate forecasts compared to more sophisticated models.

2. Simple Moving Average (SMA) and Exponentially Weighted Moving Average (EWMA):
   - Both SMA and EWMA provide smoothed versions of the time series data, helping to identify underlying trends. EWMA, with its focus on more recent observations, often provides more responsive and accurate forecasts compared to SMA.

3. Holt-Winters Models:
   - The additive Holt-Winters model captures trends and seasonality effectively, making it suitable for data with stable seasonal variations. The multiplicative model, on the other hand, is better for data with proportional seasonal variations.
   - Both models demonstrated superior performance to the Naive model, aligning more closely with actual test data and providing more accurate forecasts.

4. ARIMA, ARIMAX, SARIMA, and SARIMAX Models:
   - These models, especially when incorporating exogenous variables (ARIMAX, SARIMAX), provided the most accurate forecasts. They excel in capturing both trends and seasonal patterns, with SARIMA and SARIMAX models being particularly effective for complex time series data exhibiting seasonal fluctuations influenced by external factors.

Implications for Financial Decision-Making:
- Stock Price Prediction: Models like ARIMA and SARIMA are extensively used for forecasting stock prices, aiding in speculative trading and investment decisions.
- Risk Management: The Holt-Winters and EWMA models are crucial for volatility forecasting, essential for financial risk management and hedging strategies.
- Investment Strategy Development: Moving average models are useful for developing trading strategies based on technical indicators.

The selection and application of appropriate forecasting models are critical for optimizing financial outcomes and enhancing strategic decision-making. Each model offers unique benefits tailored to different financial analysis aspects, highlighting the importance of choosing the right model based on the data characteristics and analysis objectives. By integrating historical data and considering external variables, these models provide comprehensive market insights, aiding in proactive strategy adjustments, optimized resource allocation, and long-term financial planning.

Justification and Application in Financial Engineering:

1. Justification for Use Based on Performance and Characteristics

Naive Forecast Model: This model, being the simplest, assumes the next value in a series is the same as the last observed value. Despite its simplicity, it is a strong baseline due to its ease of use and understanding. It performed adequately in data with less volatility and can be useful in stable markets to quickly estimate future values.

Simple Moving Average (SMA) and Exponentially Weighted Moving Average (EWMA): Both models average past data, but EWMA places more weight on recent observations. These are particularly effective in smoothing out short-term fluctuations and highlighting longer-term trends in data, useful in financial markets for trend-following strategies.

Holt-Winters Model (Additive and Multiplicative): These models extend upon exponential smoothing to capture seasonality in data. The additive model is better suited for data with stable seasonal variations, while the multiplicative model handles data where seasonal variations change proportionally over time. Their robustness in handling seasonal data makes them valuable for cyclic financial series like quarterly earnings.

ARIMA and ARIMAX Models: ARIMA excels in modeling data where trends and seasonality can be different to achieve stationarity. ARIMAX extends this by incorporating exogenous variables, providing a more detailed analysis when external factors significantly impact financial indicators. They are foundational for econometric forecasting where precision and adjustability are crucial.

SARIMA and SARIMAX Models: These are ideal for capturing both seasonality and non-stationarity in data, with SARIMAX adding the ability to include exogenous variables. These models are highly adaptable, making them suitable for complex financial time series that exhibit seasonal patterns influenced by external economic indicators.

2. Applicability in Financial Engineering

Stock Price Prediction: Models like ARIMA, SARIMA, and their variants are extensively used to forecast future stock prices by analyzing historical prices, which can help in speculative trading or long-term investment decisions.

Risk Management: The Holt-Winters and EWMA models, with their focus on trends and smoothing, can be crucial for volatility forecasting, which is essential in the management of financial risk and the determination of optimal hedging strategies.

Investment Strategy Development: Moving average models (SMA, EWMA) are particularly useful in developing trading strategies based on technical indicators like moving average crossovers, commonly used in algorithmic trading to signal buying and selling opportunities.

3. Importance of Time Series Forecasting in Financial Decision-Making

Time series forecasting is pivotal in financial decision-making due to the dynamic and often unpredictable nature of financial markets. Accurate forecasts enable:

Proactive Strategy Adjustment: By anticipating market movements, businesses and investors can adjust strategies to mitigate risks or capitalize on predicted changes.

Enhanced Analytical Capabilities: Forecasting models integrate historical data and, in cases like ARIMAX and SARIMAX, external variables to provide a more comprehensive market analysis.

Optimized Resource Allocation: Effective forecasting helps allocate resources more efficiently, maximizing returns on investment and reducing unnecessary or risky expenditures.

Strategic Planning: Long-term financial planning relies on accurate forecasting to set realistic goals and prepare for future market conditions, ensuring sustainable growth and stability.

In conclusion, the selection and application of forecasting models in financial engineering are crucial for optimizing financial outcomes and enhancing strategic decision-making processes. Each model offers unique benefits suited to different aspects of financial analysis, underscoring the importance of choosing the right model based on the specific characteristics of the data and the objectives of the analysis.

References

GAURAV DUTTA “Time series forecasting using Naive method” (2020)

Sadrach Pierre “A Guide to Time Series Forecasting in Python” (2024)

Lleyton Ariton “A Thorough Introduction to Holt-Winters Forecasting” (2021)

Samishawl “Python | ARIMA Model for Time Series Forecasting” (2020)

Brendan Artley “Time Series Forecasting with ARIMA , SARIMA and SARIMAX

”(2022)

Roi Polanitzer “Time-Series Methodologies — Part 8: Holt-Winter’s Multiplicative Seasonality With Trend” (2022)

Kajal Kumari “Time Series Forecasting Using Python”(2024)