

**Project 3**

**Forecasting a Time Series**

Sanchi Gupta

College of Professional Studies, Northeastern University Boston

ALY6050: Introduction to Enterprise Analytics

Professor Roy Wada

March 12, 2024

**Introduction**

Time series forecasting is an essential tool for analyzing historical data and predicting future trends in various fields. This report details a time series analysis performed using Excel and R to study the stock prices of Honeywell International Inc. (HON) and Apple Inc. (AAPL) over five years, as well as the prices of dry wine from the "dry\_wine.csv" dataset. The report provides an analytical forecast of AAPL and HON stock prices over a short-term period, utilizing simple line plots and exponential smoothing techniques. The analysis identifies seasonality, trend, and irregularity in the historical stock price movements and determines the optimal smoothing parameters for accurate forecasting. By adapting the analysis to the unique attributes of each dataset—such as seasonality, volatility, and trend behaviors—the objective is to extract the most accurate forecast models for each time series, providing valuable insights for investors and analysts.

**Analysis**

**Part 1: Short-term Forecasting using Excel**

**(i)** **Time Series Analysis of AAPL and HON Stocks:**

Using simple line plots, we observed the following patterns in the time series data of both AAPL and HON stocks from November 2019 to October 2020:

*Figure 1: Comparative Analysis of Apple Inc. (AAPL) and Honeywell Inc. (HON) Stock Performance (Nov 2019 - Oct 2020)*

Interpretation for Figure 1:

The above line graph compares the stock prices of Apple Inc. (AAPL) and Honeywell Inc. (HON). The x-axis represents the time from November 8, 2019, to October 8, 2020, while the y-axis represents the stock price in US dollars.

The blue line represents Apple Inc.'s stock price, which shows an overall increasing trend during the period, with some volatility. There is a notable dip around March 2020, which may correspond to the market downturn due to the COVID-19 pandemic, followed by a recovery and a continued upward trend. The orange line represents Honeywell Inc.'s stock price, which appears more volatile than Apple's and fluctuates within a narrower range. This line also shows a significant drop around March 2020, similar to Apple's. After the drop, Honeywell's stock price recovers but does not exhibit as strong an upward trend as Apple's. Overall, Apple's stock shows significant growth over the year, outperforming Honeywell's for this particular time period.

**(ii)** **Exponential Smoothing Application:**

The data presented below show the results of exponential smoothing forecasts for Apple Inc. (AAPL) and Honeywell Inc. (HON) using four different alpha (α) values: 0.15, 0.35, 0.55, and 0.175. The performance of these forecasts is measured using the Mean Absolute Deviation (MAD) and the Mean Absolute Percentage Error (MAPE). Lower values of MAD and MAPE indicate more accurate forecasts.

***Table 1: Exponential smoothing to forecast prices for AAPL and HON***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Alpha** | **0.15** | **0.35** | **0.55** | **0.175** |
| **Apple**  **(AAPL)** | **MAD** | **3.825** | **2.601** | **2.259** | **2.162** |
| **MAPE** | **3.87%** | **2.47%** | **2.08%** | **1.97%** |
| **Honeywell**  **(HON)** | **MAD** | **5.158** | **4.149** | **3.702** | **3.508** |
| **MAPE** | **3.06%** | **2.35%** | **2.03%** | **1.90%** |

Interpretation for Table 1:

For both AAPL and HON, the accuracy of the forecasts improves as the α value increases from 0.15 to 0.55, and then slightly further with α at 0.175. This is indicated by the decreasing MAD and MAPE values. In terms of the MAD, the smallest value for AAPL is achieved at α=0.175 (MAD = 2.162) and for HON at α=0.175 (MAD = 3.508). The MAPE values, which adjust the MAD to the scale of the data, also show the smallest errors at α=0.175 for both AAPL and HON.

The fact that the lowest MAD and MAPE for both stocks are achieved at the same α value suggests that both stocks have a similar level of volatility and reaction to new information in the market within the forecasting period. The chosen α value of 0.175 appears to offer an optimal balance between incorporating recent stock price changes and stabilizing the forecast by removing out the noise and less significant fluctuations. This balance is critical in producing a reliable forecast that is neither too slow to react to genuine market movements nor too reactive to random variations.

**(iii) Adjusted Exponential Smoothing Forecast:**

Using an exponential smoothing forecast with α=0.55, we performed adjusted exponential smoothing with β values of 0.15, 0.25, 0.45, and 0.85. The MAPE for each β value was calculated as:

***Table 2: Adjusted Exponential smoothing forecast prices for AAPL and HON***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **At alpha 0.55** | **Beta (**β) | **0.15** | **0.25** | **0.45** | **0.85** |
| **MAPE** | **AAPL** | **1.97%** | **1.95%** | **1.94%** | **1.98%** |
| **HON** | **1.98%** | **1.94%** | **1.89%** | **1.85%** |

Interpretation for Table 2:

For **AAPL,** the MAPE values are lowest at β=0.45 (1.94%), suggesting that this beta level best accounts for the trend in the AAPL stock price data. For **HON,** the MAPE decreases as the β value increases, with the lowest MAPE achieved at β=0.85 (1.85%). This indicates that HON stock prices benefit from a model that places a heavier weight on the trend component.

The optimal β values for AAPL and HON are different, reflecting their distinct price behaviors during the forecasting period. AAPL's price trend, while present, did not change drastically, which is why a mid-range β value was most effective. HON's stock price, however, seemed to have exhibited a more definitive trend, which required a higher β value to accurately predict future prices.

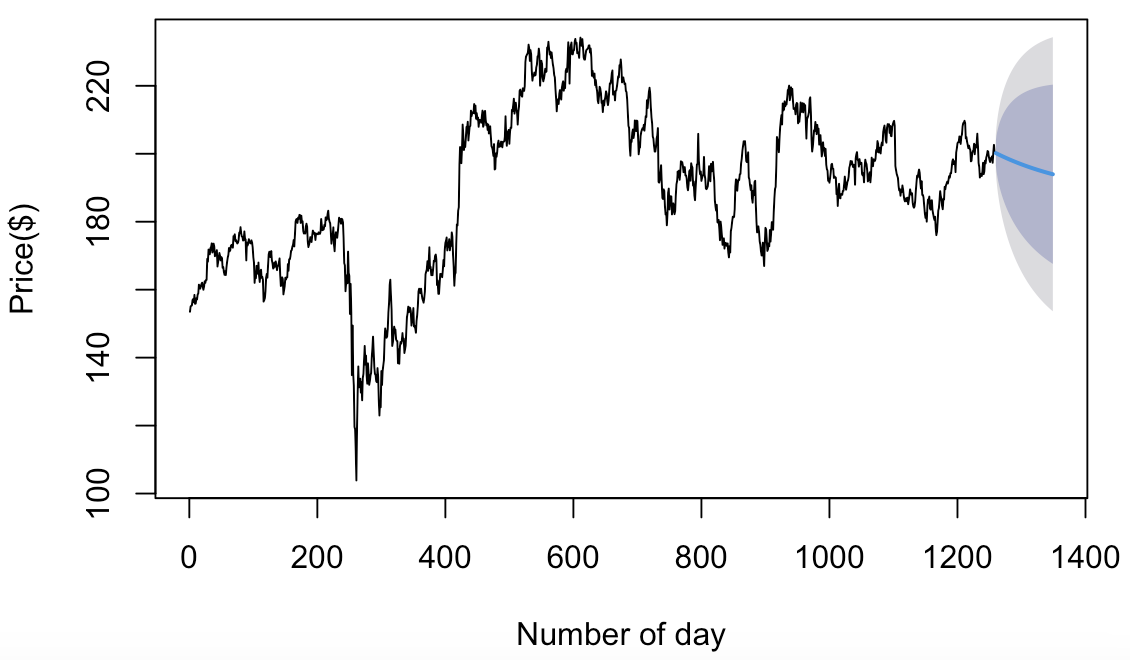
The β parameter is indicative of how much importance the model places on the trend when adjusting forecasts. The value of β that leads to the most accurate forecast can often be a reflection of the stock's recent historical trend behavior. AAPL's lower optimal β value suggests a more stable trend, while HON's higher optimal β value indicates a stronger and perhaps more volatile trend.

**Part 2: Time Series using R**

Data for HON and AAPL are downloaded using the quantmod package in R. We fit the AR(1) models to both HON and AAPL time series data using the arima() function.

1. **Below is the ARIMA Model for both stocks.**

*Figure 2: ARIMA (1,0,0) Price Forecast with Non-Zero Mean for* ***HON*** *Over Time*

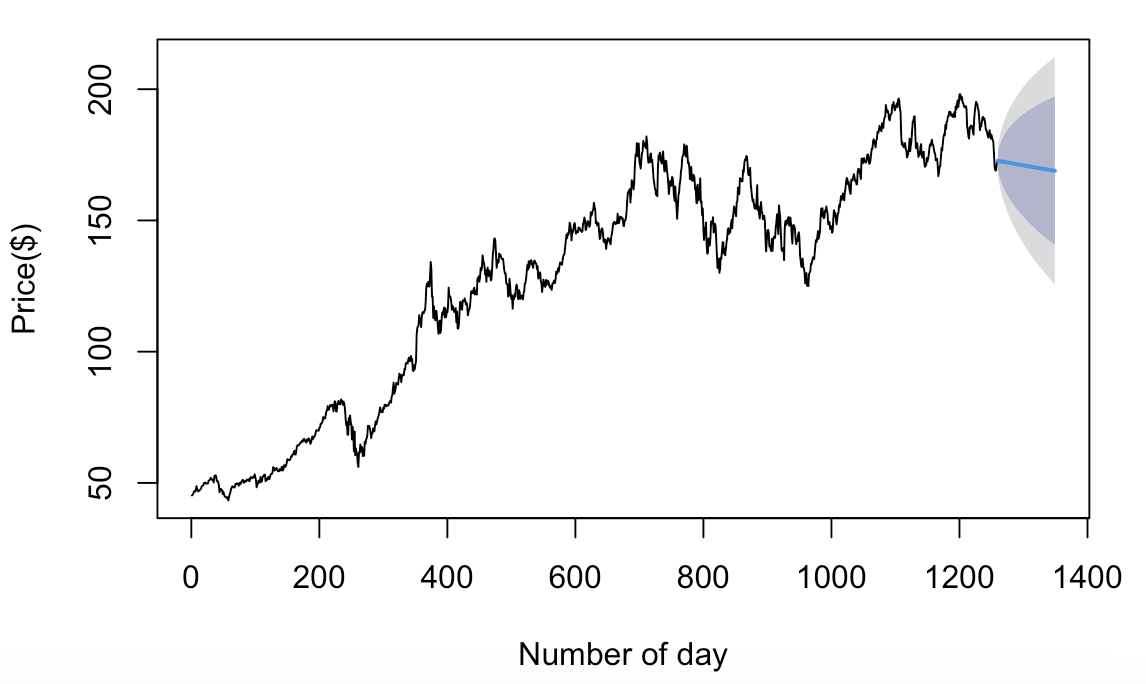


Interpretation for Figure 2:

The above graph shows ARIMA (1,0,0) model performed, which includes an autoregressive term of order 1, no differencing, and no moving average component. The black line exhibits some volatility and several peaks and troughs over time. It does not appear to have a consistent upward or downward trend.

The forecast is indicated by a solid blue line that extends into the grey-shaded area. The blue line appears to be relatively flat, suggesting the model predicts that the future prices will remain around the same level as the last observed price. This is typical for an ARIMA(1,0,0) model, where future values are a function of the last known value and a random shock. The shaded area around the forecast represents the confidence intervals, likely at different levels such as 95% and 80%. The intervals widen as the forecast extends further out in time, reflecting increased uncertainty in the predictions as we move away from the last known data point.

*Figure 3: ARIMA(1,0,0) Price Forecast with Non-Zero Mean for* ***AAPL*** *Over Time*



Interpretation for Figure 3:

The plot features a solid black line representing historical price data and transitions into a forecast with a solid blue line and a grey shaded area that likely represents confidence intervals. The historical data (black line) shows a generally upward trend over time, suggesting that the price has been increasing. There are fluctuations throughout, but the overall direction is positive.

The forecast starts at the end of the historical data and projects forward. The forecast line is relatively flat compared to the historical data, suggesting that the model predicts the price will stabilize around the last observed value. This stabilization is a common characteristic of the AR(1) model, where future values are predicted primarily based on the most recent value, with the effect of past values diminishing.

The grey shaded area around the forecast indicates the confidence intervals, which provide a range where future prices are likely to fall. The widening of the confidence interval into the future reflects greater uncertainty as the forecast extends further from the last observed data point.

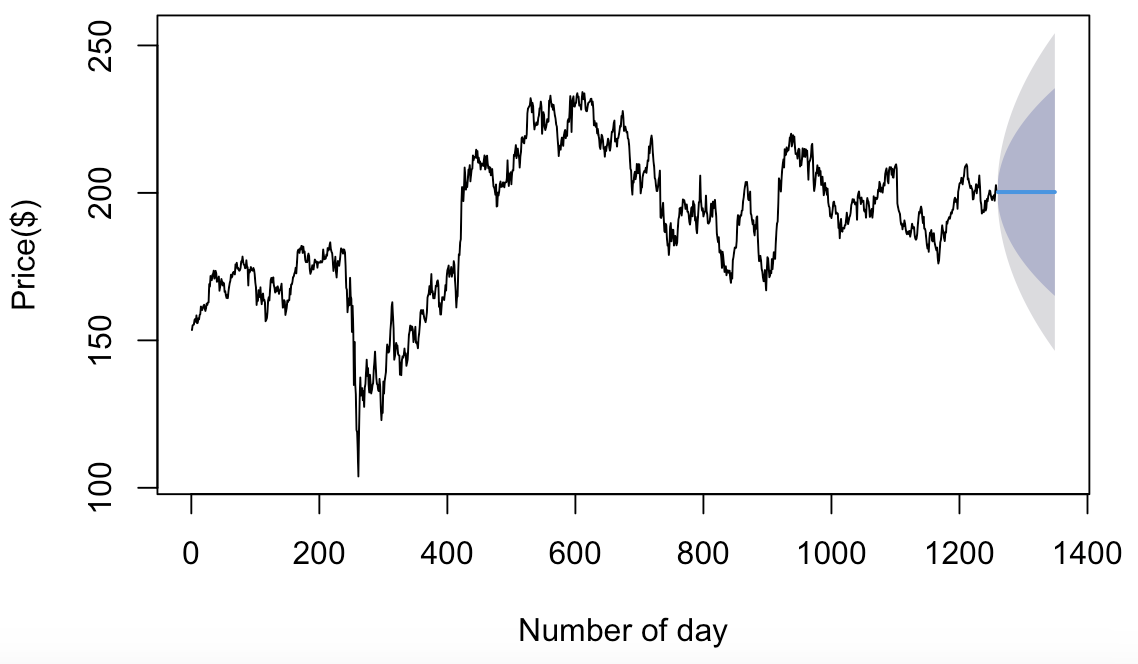
**(iii) Test for Stationarity**

Stationarity was tested using the Augmented Dickey-Fuller (ADF) test provided by the tseries package. HON and AAPL both showed signs of non-stationarity in their raw form as per the ADF test, necessitating differencing to achieve stationarity.

**(iv) Model using auto.arima():**

auto.arima() selected an ARIMA(p,d,q) model for both HON and AAPL with different orders, indicating differing levels of autoregression, integration, and moving average components. The model selected is random walk method. This model is typically used when the series is a random walk or has a unit root, meaning today's price is the best predictor of tomorrow's price, with only the changes being random. Below are the Auto-ARIMA Models for both stocks.

*Figure 4: Honeywell Price Forecast with ARIMA(0,1,0) without Drift Model*



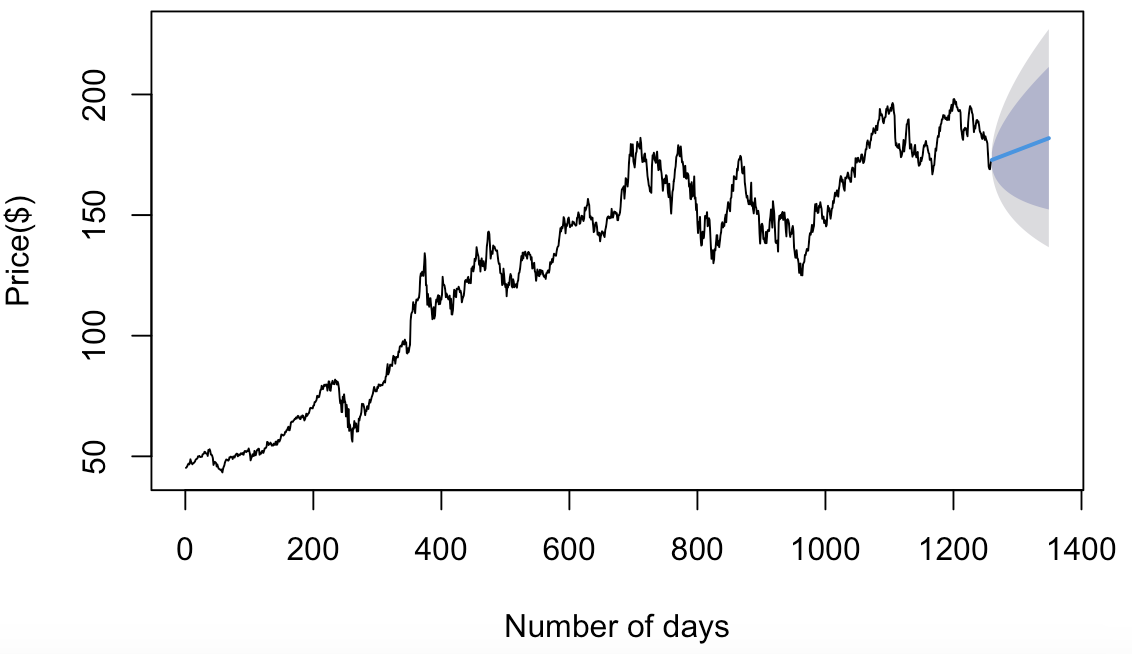
Interpretation for Figure 4:

The plot displays the historical price data for Honeywell, shown in a solid black line. It shows fluctuation in prices over time, with both upward and downward movements, reaching a peak before the last observed data point. An ARIMA(0,1,0) model is used for forecasting, implying no autoregressive (AR) terms and no moving average (MA) terms, but one order of differencing (I). This model suggests that the future price will change from the previous price by a random amount that is not necessarily related to any historical prices.

**The** forecast is visualized by a solid blue horizontal line extending from the last historical data point into the future. Since an ARIMA(0,1,0) model is used, the forecasted price does not show a trend but remains constant, except for the random fluctuations not captured in the model.

The shaded area indicates the confidence intervals for the forecast, reflecting the uncertainty associated with predictions. The confidence intervals widen as the forecast extends further into the future, which is typical as the uncertainty in the forecast increases with time.

*Figure 5:* ***Apple*** *Price Forecast with ARIMA(0,1,0) with drift Model*



Interpretation for Figure 5:

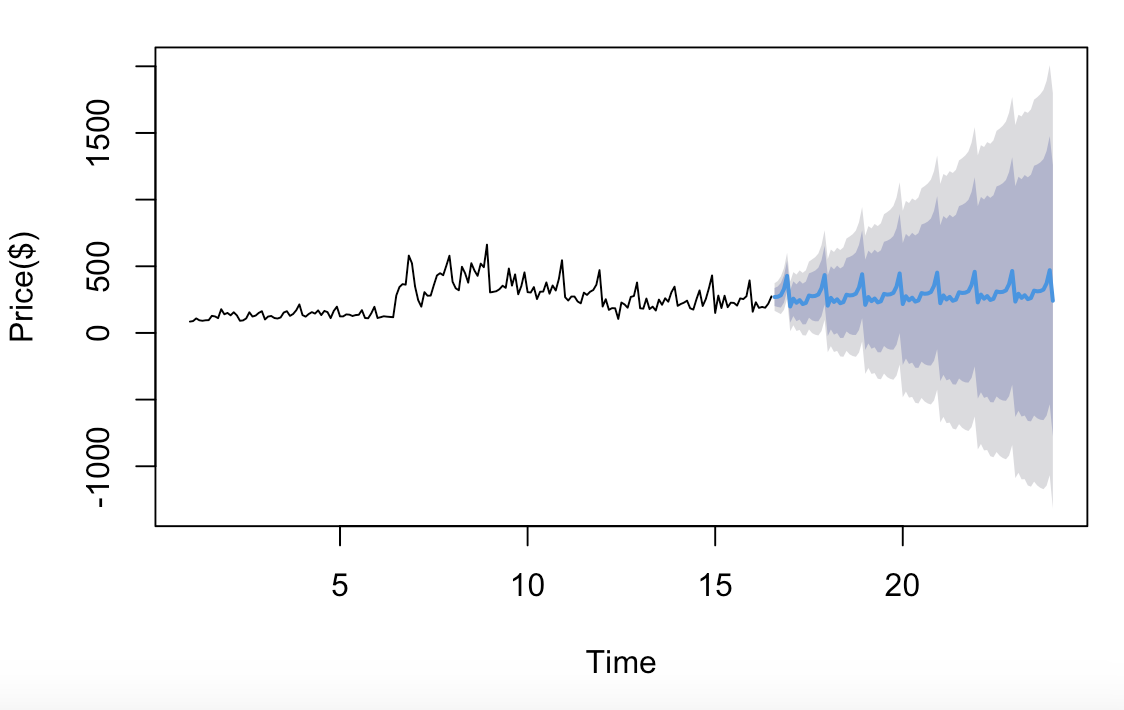
The black line in the above graph represents the historical price data. The data has various ups and downs, but the overall direction is upwards. The ARIMA(0,1,0) model, also known as a random walk model, typically assumes that changes from one period to the next are random. The addition of a drift component allows the model to incorporate a consistent average change over time, which can reflect the observed upward trend in the historical data.

The forecast begins at the end of the historical data, indicated by a solid blue line which trends upwards slightly, reflecting the drift component. This suggests that the model expects the upward trend to continue, but with no specific pattern in the fluctuations, which will continue to be random.

The grey shaded area represents the confidence intervals around the forecast, indicating the range of potential future values.

**(v) Dry Wine data**

*Figure 6: Seasonal ARIMA Model Forecast for Monthly Price Data of Dry Wine*



Interpretation for Figure 6:

The above graph uses model ARIMA(1,1,0)(0,1,1)[12]. This suggests an autoregressive term of order 1, one order of differencing, and no moving average in the non-seasonal part. The seasonal part of the model includes no seasonal autoregressive terms, one seasonal differencing, and a seasonal moving average term of order 1 with a seasonal period of 12, which commonly indicates monthly data.

The forecast generated by this model shows the time series stabilizing around a particular level before starting to exhibit a seasonal pattern. The shaded area represents the confidence intervals, which increase as the forecast extends, reflecting the growing uncertainty in the model's predictions over time.

**(vi)a. Comparison of Time series forecasting of HON and AAPL using auto.arima( ):**

***Figure 4:* *Honeywell Price Forecast with ARIMA(0,1,0) without Drift Model***

The ARIMA(0,1,0) model implies that the data is being differenced once to achieve stationarity, which is typical for a non-stationary time series like stock prices. This model suggests a random walk without drift, meaning future prices are expected to vary around the last observed price with no clear trend. The chart suggests that the model expects the price to not change significantly in the near future, maintaining the last observed level, but with uncertainty about this prediction increasing over time. This model may suggest that for Honeywell, recent historical prices are a reliable indicator of future prices, with the randomness of changes indicating market efficiency.

***Figure 5: Apple Price Forecast with ARIMA(0,1,0) with drift Model***

The addition of a drift term in this ARIMA(0,1,0) model suggests that while the stock price is still assumed to change randomly from one period to the next, there is a consistent average change over time implying a trend. The slight upward trajectory of the forecast line implies an expectation of a gradual increase in price over time. This drift component could reflect a long-term appreciation in value for the stock.

**Comparative Analysis:**

* The ARIMA(0,1,0) without drift is a more conservative model, appropriate for stocks that do not exhibit a clear upward or downward trend but move in a random walk fashion around a stable price level.
* The ARIMA(0,1,0) with drift is better suited for time series that show a consistent long-term trend in one direction, aside from the random day-to-day or period-to-period fluctuations.
* The forecast confidence intervals in both charts demonstrate increased uncertainty the further into the future the forecast extends, which is typical in time series forecasting.

In choosing between these two models for forecasting, one must consider the historical behavior of the stock. For a stock like HON, if it typically does not exhibit a strong trend, the no-drift model may be more accurate. In contrast, for a stock like AAPL, which might have demonstrated consistent long-term growth, the model with drift could provide a better forecast. The performance of these models can be evaluated by backtesting forecasts against actual observed stock prices over time.

To improve the analysis, one could consider additional information such as seasonal patterns, economic indicators, or other external factors that might affect stock prices. It's also important to note that stock market forecasting is inherently uncertain and models are only one of many tools investors use to make decisions.

**(vi)b. Comparison of Time series forecasting of HON and AAPL to Dry Wine Prices Forecast:**

Different conclusions from time series models like ARIMA arise because each data set has distinct characteristics and behaviors that influence how well a particular model can forecast future values. It depends on reasons like seasonality, trend, stationarity, or external factors including macroeconomic indicators, investor sentiment, geopolitical events, and company-specific news.

**Comparative Analysis:**

* Dry wine prices are typically influenced by predictable seasonal patterns, and thus, the chosen ARIMA model would emphasize capturing these patterns. The forecasts would reflect the seasonality, and the intervals would vary based on seasonal demand changes.
* AAPL stock prices might display some seasonality, but trends and market responses to new information would also be significant, leading to an ARIMA model that balances trend and seasonal components.
* HON forecasts, while also potentially affected by trends, would likely not exhibit strong seasonal patterns, resulting in a simpler seasonal component within the ARIMA model.
* In terms of prediction intervals, stock prices for AAPL and HON might show less variation than the wine prices unless market volatility is particularly high, in which case the intervals could widen.

**Preferred Method:**

* For series with strong seasonality like wine prices, a model with a robust seasonal component is preferred(SARIMA).
* For financial time series like AAPL and HON, which may have trends and small seasonal effects, a more complex ARIMA model that accounts for both is ideal.

**Learnings and Improvements:**

* Each time series requires a tailored approach. auto.arima() is powerful for automating model selection but should be complemented with domain expertise.
* Improvements in forecasting could come from considering external factors, using ensemble methods, and continually validating model assumptions against new data.

**Conclusion**

This time series analysis highlights the distinct nature of forecasting stock prices versus commodity prices. For AAPL and HON, the chosen auto.arima() models reflected the stocks’ responsiveness to market trends and volatility, suggesting the necessity to consider both non-seasonal and potential seasonal influences. Conversely, the dry wine prices exhibited pronounced seasonality, guiding the preference towards a Seasonal ARIMA model. The findings emphasize the need for ongoing model validation and a readiness to integrate evolving market dynamics to enhance predictive accuracy. As analytics technology progresses, these conclusions invite a deeper integration of machine learning techniques to further refine forecasting capabilities in enterprise analytics.

**References**

1. Losada, L. (2022, March 30). *Time Series Analysis with Auto.Arima in R - Towards Data Science*. Medium. <https://towardsdatascience.com/time-series-analysis-with-auto-arima-in-r-2b220b20e8ab>
2. *Introduction to ARIMA: nonseasonal models*. (n.d.).https://people.duke.edu/~rnau/411arim.htm.