Time Series Analysis
Course

Final Project -
Apple Inc. Stock Analysis

Winter 23/24

<u>Submitted by:</u>

Saleem Kheralden - Technion
Bashar Massad - Technion

Introduction:

In the realm of financial markets, understanding the patterns and dynamic of stock prices is of paramount importance in order to blossom in this realm.

The stock price of a company reflects not only its intrinsic value but also the collective perception of its future prospects. In this context, the analysis of stock price time series data becomes a crucial endeavor, offering insights into market trends, investor behavior, and potential forecasting opportunities.

For our time series analysis project, we have chosen to delve into the world of Apple Inc. [5], one of the most iconic and influential companies in the technology field. the dataset we have selected spans over a long period of time, capturing the rise and the fluctuations in Apple's stock price over time.

In this part of the project, we will conduct a comprehensive preliminary analysis of the Apple stock price dataset. We will present the patterns, trends and seasonality inherent in the data. We will be doing this through graph and visualization.

The dataset's shape is (10468, 7), where the columns are Date, Open, High, Low, Close, Adj Close, Volume.

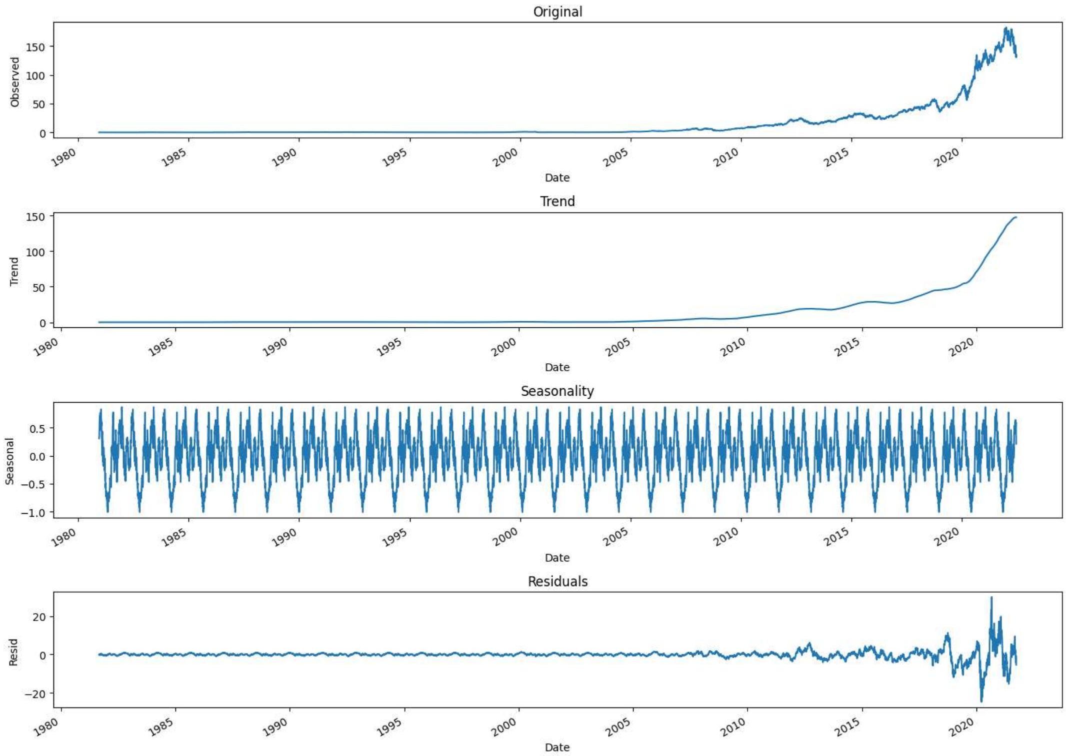
The dataset has points that span over almost 40 years, starting from 12/12/1980 until 17/06/2022.

	Date	Open	High	Low	Close	Adj Close	Volume
10463	2022-06-13	132.869995	135.199997	131.440002	131.880005	131.880005	122207100
10464	2022-06-14	133.130005	133.889999	131.479996	132.759995	132.759995	84784300
10465	2022-06-15	134.289993	137.339996	132.160004	135.429993	135.429993	91533000
10466	2022-06-16	132.080002	132.389999	129.039993	130.059998	130.059998	108123900
10467	2022-06-17	130.070007	133.080002	129.809998	131.559998	131.559998	134118500

The first part of the project our goal was to find whether the time series was stationary, seasonality’s and patterns in the data that will help us in the research.

Then we move to the second goal which is to find the best model that fits the data.

In this project we will be focusing on the ‘Close’ column.



Original:

The first plot shows the original time series data for the AAPL closing stock prices. It demonstrates a long-term upward trend with some fluctuations over time. There is a notable acceleration in price increase in the latter part of the series, indicating periods of rapid growth.

Trend:

The second plot isolates the trend component from the time series. It smoothes out the short-term fluctuations and highlights the long-term movement. The upward trend is gradual at first and then becomes more pronounced after 2000, reflecting the company's growth and the market's increasing valuation of Apple over time.

Seasonality:

The third plot shows the seasonal component of the time series. It exhibits the regular and periodic fluctuations within the time series data that repeat over a specific interval. In this case, the fluctuations seem quite consistent, suggesting a stable seasonal pattern. The scale of seasonality is relatively small compared to the overall trend, which is typical for stock data that doesn't typically have strong seasonality like retail or agricultural sectors might.

Residuals:

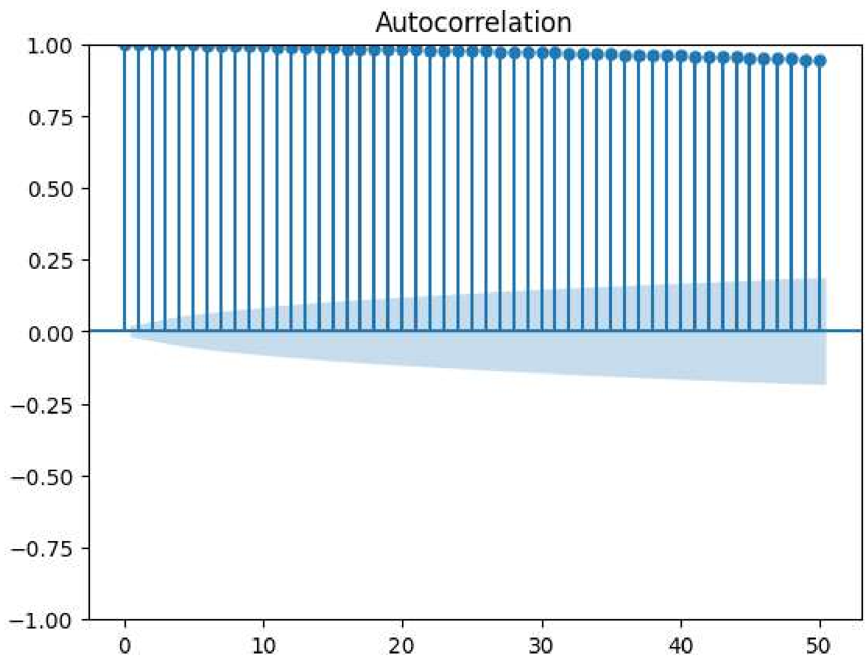
The final plot illustrates the residuals, which are the time series after the trend and seasonal components have been removed. This component represents the irregular or random noise in the data. Here the residuals seem fairly consistent in variance over time, with some bursts of increased variability. These could be due to external factors or events not accounted for by the seasonal and trend components, or they could indicate periods of market instability.

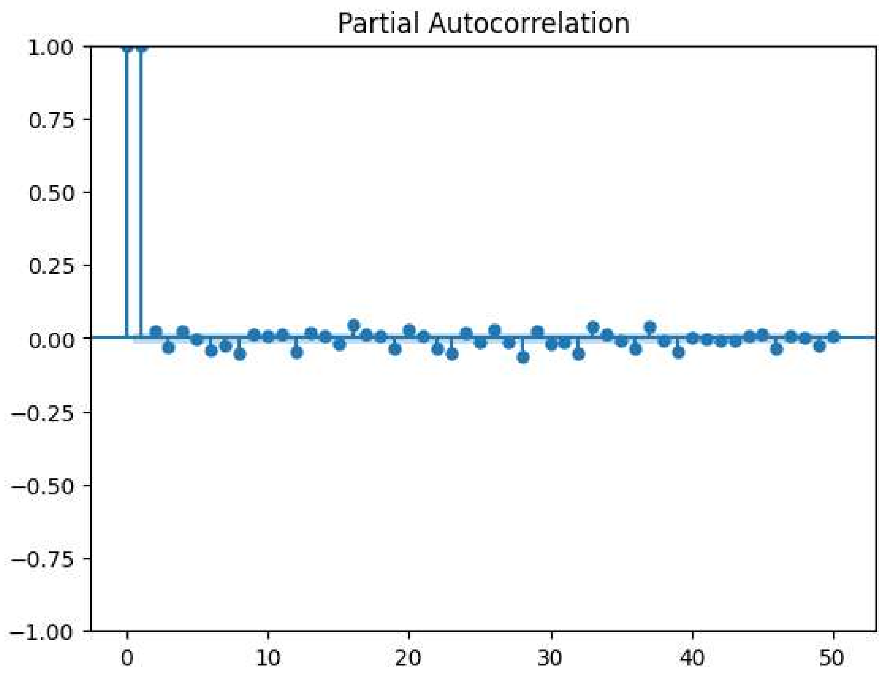
In summary, the plots indicate a strong long-term growth in the AAPL stock price with some periods of increased volatility. The stable seasonal component suggests regular patterns, possibly related to business cycles, product release schedules, or other recurring events. The residuals do not show any pronounced patterns, indicating that the trend and seasonal components have captured most of the systematic variation in the time series.

Methodology:

After exploring the data and trying to find patterns, we decided to delve into predicting the Close price of the apple stock. Following previous exploration, we suggest that the model which best fits the Close price could be AR(2), as a result, we're going to test the SARIMAX model on the data since there's seasonality. Also, after we conducted research about the relevancy of the years in our dataset, we concluded that the behavior of the stock Close price before the year 2007 (The First Release of the iPhone) is not relevant since the Close prices was significantly low compared to today's stock price, So we chose to consider the data from a year before (2006) and on.

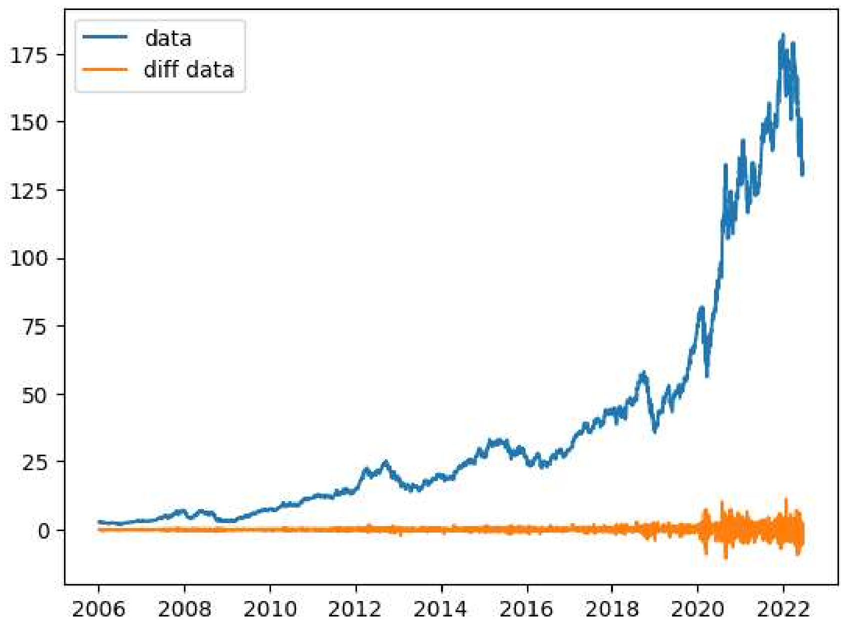
We then wanted to check if there’s a need for differencing and after plotting the ACF and PACF we got

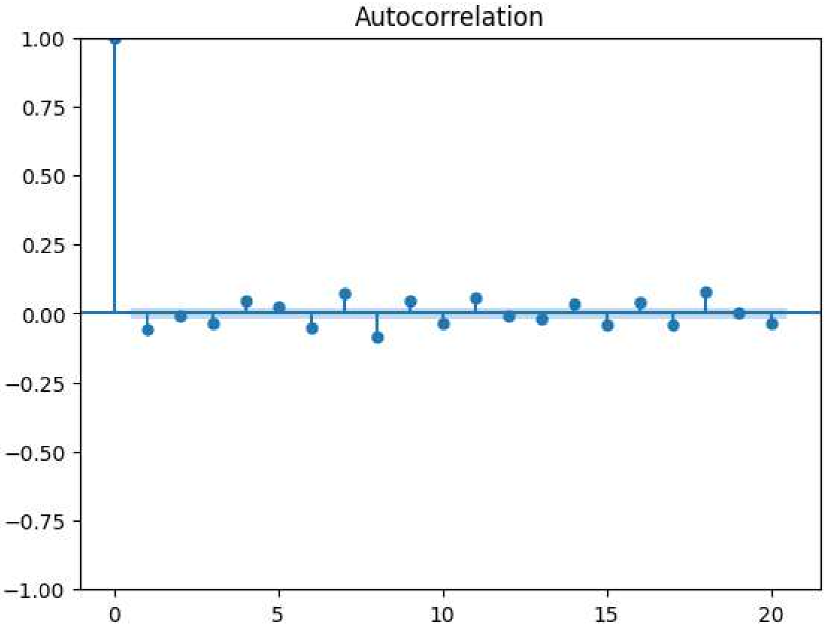


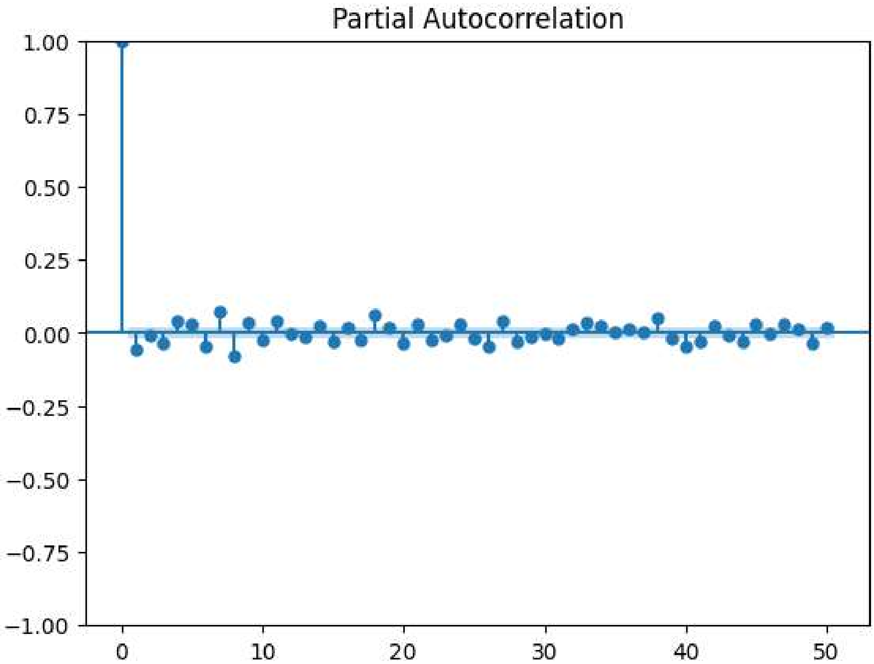


The ACF plot shows the correlation between the time series and its lagged values.

The ACF plot provided has a gradual decline in correlation values as the lag increases. This pattern typically suggests that the underlying time series is not stationary and may contain a trend or seasonal component. and so we will be needing for differencing.







Components:

Non-seasonal:

Since the PACF showed a sharp cutoff after lag 1, that suggest AR(1) model then we set p=1. The quick drop-off of ACF could suggest a MA(1) could be tested, particularly if after differencing the ACF shows a spike at lag 1, the q=1. Then we set differencing to 1 due to non-stationarity.

Seasonal:

In stocks market and for daily information, we want to set seasonality to be the number of trading days in a year which is 252 as mentioned in [1] and the rest of the data would be the same as the non-Seasonal.

And so we start with two model SARIMA(1, 1, 1)(1, 1, 1, 242), and ARIMA(1, 1, 1) that were trained on the whole dataset using the model [2], and tested on another dataset that contained future data points, but due to time complexity we switched to SARIMA(1, 1, 1)(1, 1, 1, 242 // 4) to split the data into quarterly data, we found that ARIMA(1, 1, 1) performed similar to SARIMA with less time complexity, and so we started optimizing the parameters of ARIMA, and the best parameters that we found considering also the time complexity, were ARIMA(2, 0, 2).

We tested another model in addition to the models mentioned above and it was the Prophet model, which yielded great results, but the ARIMA model performed better.

We know that stock prices could be significantly influenced by many indicators, including Apple products sales, new devices launch, GDP growth, interest rates and so on. After conducting research in exploring datasets that could be helpful for our data, we found that NASDAQ-100 index [3][6] might be an interesting insight into predicting the Apple stock Close price. (The dataset we found regarding this index is partial in relation of the stock data we're currently working on, so we adjust the dataframes we're working on to fit the index's dataset)

When adding the Exogenous variable the differencing had bigger effect and the model with differencing performed better than that with no differencing.

And so our last and best model that we found was the ARIMA(2, 1, 2) with the Exogenous variable which is the NASDAQ-100 index stock price, which improved the AIC of the model by a factor of 100.

Results and Discussion:

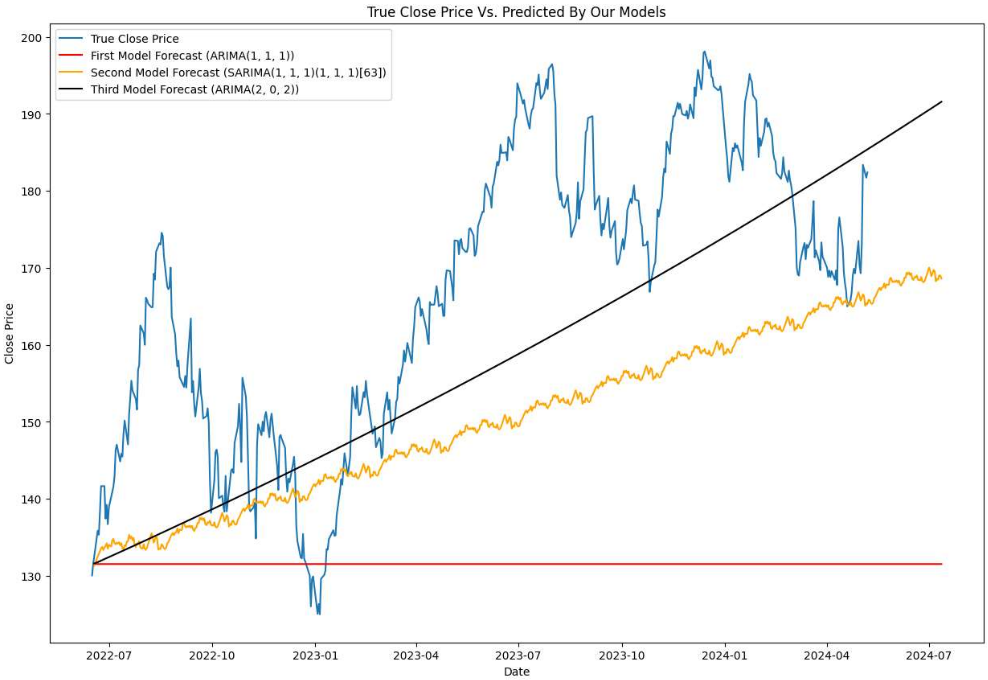
Our model evaluation based on information criteria such as AIC and BIC yielded the following scores:

Model	**AIC**	**BIC**
SARIMA(1, 1, 1)(1, 1, 1, 63)	12458	12489
ARIMA(1, 1, 1)	12565	12584
ARIMA(2, 0, 2)	12557	12588
ARIMA(2, 1, 2) w. Exogenous variable	97	133
SARIMA(1, 1, 1)(1, 1, 1, 63) w. Exogenous variable	96	120
Prophet	* 4150	-
Prophet w. Exogenous variable	* 3000	-

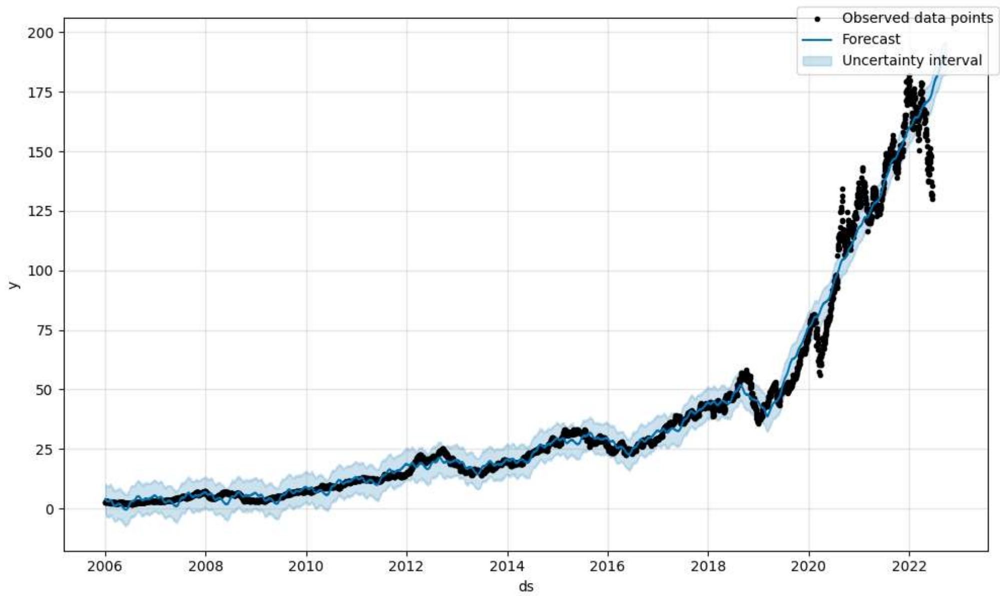
* Note that the value in the AIC column for the Prophet model Is the WAIC value, since it estimates the same value the comparison between them is valid. [4]

Our analysis suggests that the ARIMA(2, 1, 2) model with an exogenous variable, namely the NASDAQ-100 index, is the most suitable for predicting Apple stock prices. This model offers simplicity, effectiveness, and improved performance compared to SARIMA and other ARIMA configurations. Incorporating exogenous variables can enhance forecasting accuracy and capture external factors influencing stock prices.

Visual comparison of our models' predictions (Prophet wasn't included in the graph since it had very large predicted values in relation to our models):



Prophet forecast:



Conclusion

In conclusion, our analysis suggests that the SARIMA(1, 1, 1)(1, 1, 1, 63) model with an exogenous variable, namely the NASDAQ-100 index, is the most suitable for predicting Apple stock prices. This model offers simplicity, effectiveness, and improved performance compared to SARIMA and other ARIMA configurations. Incorporating exogenous variables can enhance forecasting accuracy and capture external factors influencing stock prices.

References

[1] https://en.wikipedia.org/wiki/Trading_day

[2] https://www.statsmodels.org/stable/tsa.html

[3] https://www.investopedia.com/terms/n/nasdaq100.asp

[4] https://en.wikipedia.org/wiki/Watanabe%E2%80%93Akaike_information_criterion

[5] https://www.kaggle.com/datasets/meetnagadia/apple-stock-price-from-19802021

[6] https://www.kaggle.com/datasets/lp187q/ndxt-index-until-jan-202018