Business Project Microsoft Stock Price Prediction with ARIMA and ARIMAX	
Name-surname	Kutay Onlen
Module	BEMM466
Module Name	Business Project
Student Number	700059706
Word Count	9754

Contents

1 Introduction

Since the Babylonians and ancient Egyptians, based on the historical findings at least 5000 years ago, humankind has kept track of time (Andrewes, 2006). Later, based on a timeline, people started to keep records of various types of data, such as the degree of temperature, the number of sales and stock prices. This presented sample data over the same periodic timeline is called time series (Hayes, 2022). Ground-breaking statistical studies from the 20^{th} century provide us to apply prediction models with acceptable error ratios. These models use historical data to predict future values of the time series. Some pioneer prediction models are autoregressive (AR) models, moving average (MA) models, a combination of these two models ARMA, and integrated versions of it, ARIMA models (Jenkins & Box, 1976). There are other prediction models that I will mention in the literature review section.

With the dramatic increase in computational power of computers and convenient software languages like Python and R, even medium-level organisations produce more projections day by day. Some people even say the era of quantum computers has already started, and these computers might join our daily life sooner than expected. Martinis & Boixo (2019) claimed that Google had reached quantum supremacy, a machine that can use the power of quantum computing. They emphasised that this quantum computer finished the given task in 200 seconds, while the fastest supercomputer back in 2019 could complete the same task in 10.000 years, according to their calculations. Even today, we can see the traces of machine learning models every day once we open any application on our smartphones or computers from just the advertisements presented to us according to our identities, thoughts, likes, retweets, etc. At this point, I would like to discuss the chaos theory with an example from famous historian Harari's book Sapiens: A Brief History of Humankind. Harari (2022) argued that there are two levels of chaos in this World. First-level chaotic systems are the systems that do not react to the predictions like the weather. The second level of chaotic systems is the systems that would react to predictions like the share price of a company in a stock market. For example, if we assume today's share price is 90 pounds and tomorrow's price will be 100 pounds. Then everyone starts to buy that share for 100 pounds. The prediction with 100% accuracy gets stuck when it comes to the prediction of the second day. Harari (2022) gives this example of the second-level chaos. In my opinion, forecasting developments will change the fundamental of the future's second-level chaotic human-made systems. Moreover, we can even see the impact of non-certain predictions at present, which supports Harari's argument. However, until the discovery of prediction methods with 100% accuracy, most of the organisations must apply prediction models that we already have to survive against their rivals.

This extended report will examine to what extent basic ARIMA and ARIMA models with exogenous variables (ARIMAX) can predict the close price of Microsoft stocks and how reliable these models are based on my findings. In the application process of these models, I used Python software language on the Jupyter Notebook environment.

In the first section of this report, I will briefly talk about the stock market and the importance of making predictions in this field. I will present my literature view for ARIMA models in the second section. I will finalise this section with my reflection along with the risks and limitations of my study. In the third section, I will explain my dataset and provide an ethical aspect of it. Moreover, I will explain the libraries I used in Python and my pre-processing data stages. In the fourth section, I will present my findings with supported graphs and tables. The fifth section will be the conclusion part of the report. In the last and the sixth section of the report, I will give the recommendations that I extracted from my findings and application process.

1.1 Context and Background

Stock exchange platforms are significant for the business world and academics. The most basic explanation of how stock exchange platforms work would be allowing individuals to buy unit shares of the companies. As the company improve its profit, image or any asset that can affect the share prices, the shareholder's investment is affected in the same ratio. In addition to the companies, there are also different types of stock exchange platforms that focus on the prices of materials and currencies like steel, gold, Turkish lira, cryptocurrencies, euro, etc. These kinds of markets provide the nominal price of these assets. To illustrate, people can follow and analyse the price of steel from The London Metal Exchange, and the companies that use steel as a raw material can create their purchasing strategies based on these prices. An example for companies is that we can observe and buy shares of Microsoft from the National Association of Securities Dealers Automated Quotations (NASDAQ), a stock exchange platform. Chen (2022) states that the stock market is a term that refers to the overall of these stock exchange platforms.

Some advantages of the stock market:

- Provides valuable parameters to evaluate the economic situations of countries, regions, sectors, and companies. To demonstrate, if we want to check the private sector in the United States of America (USA), we can look at the S&P 500 index, which is a price-weighted average index of the largest 500 American companies in the stock market. Some studies emphasise that the S&P 500 index declined 25% in March 2020, which matches the pandemic's beginning (Curto & Serrasqueiro, 2022).
- Provides additional income channels to the companies by giving the chance to collect the investment from individuals.
- Provides profitable investment opportunities for people.

Some disadvantages of the stock market:

- Even components of the stock markets are inspected strictly; manipulations can happen in different regions of the world. Due to these manipulations, individual buyers can lose their money or might lose money from their profits. To demonstrate, Hayes (2022) claimed that the People's Banks of China, on the 5th of August in 2019, decreased the Chinese yuan's value against the American dollar to decrease the cost of exportation for Chinese companies against Trump's new tariffs for goods from China. Hayes also stated that manipulations, especially currency exchange markets, are doable by the states and do not have to be illegal.
- Trading in stock markets is riskier than traditional investment options such as buying property or putting savings in interest-bearing bank accounts.

While the stock market parameters give many insights and opportunities for profitable investment for individuals, the number of predictions, projections, and comments about them have a significant impact present. They show a ship's direction rather than where it is currently located. Even the direction is not certain and can change, and it significantly impacts the interpretation of whether the ship's current location is good or bad. These arguments are the main motivational points for choosing the time series forecasting model ARIMA and its application process as my final report's topic.

2 Literature Review

In the early years of the 20th century, George Udny Yule approached his time series study with a stochastic process and used the autoregressive (AR) model (Yule, 1927). By the time, being able to consider the time series as a stochastic process was an evolutionary perspective. Except for the few, most of the academics were struggling with time series problems as a deterministic process. When we remind ourselves that time series are the values over a timeline, m the academics were hustling with the problems by creating deterministic functions, which generalise structure of it $y = f(\text{ time })$. However, when we approach tto ime series problems as stochastic processes, generalise structure of functions is $\epsilon, y = f(time, \epsilon)$ (Parmezan et al., 2019). The difference between the two approaches is that when the deterministic approach has no room for errors, the stochastic process allows the prediction of future values of time series with errors. Since even today, almost 100 years later, with all the technological and scientific developments, we cannot make predictions with 100% accuracy. However, we can forecast the approximate future values by very small errors. De Gooijer & Hyndman (2006) stated that statisticians such as Slutsky, Yaglom, and Walker have essential studies in addition to Yule's studies in this matter. Mills (2019) said that, in 1938, Wold published the first paper in the literature about moving average (MA) models. In the same year, Wold also published a theorem, which will be called Wold's decomposition after him. Wold argued that weak stationary time series could be formulated as a sum of 2 time series, such as deterministic and stochastic processes (Nielsen, 2019). From this theorem, we can validate the statistical and mathematical background of autoregressive – moving average (ARMA). These models are fundamental components of many types of forecasting methods today. However, these models' development and implications were not progressive and easy to use until the seventies. Statisticians George Box and Gwilym Jenkins wrote together a book named "Time Series Analysis Forecasting and control" in 1970 (Jenkins & Box, 1976). This book introduced the integrated AR and MA model called ARIMA. They also suggested a generalised method that consists of three main steps for ARMA and ARIMA studies. This well-known method is called the Box-Jenkins method. I will explain this model and its components at the overview and methods section, since ARIMA is the applied method of this study. De Gooijer & Hyndman (2006) argued that statisticians had developed many models such as the vector ARIMA (VARIMA), vector autoregressions (VARs), autoregressive conditional heteroscedastic (ARCH), and generalised ARCH (GARCH).

2.1 Time Series Components and Stationarity

Time series studies increase its speed dramatically since early 20th century. Today we can decompose time series to their components. Major components of time series are trend, seasonality residue (residuals). Moreover, there is also cyclic patterns in time series. However, usage of this component in literature very rare compared to other components. Existence of these components does not mean that they are part of every time series. Except the residue, other components' existence varies to each time series. This decomposition ability provides couple of advantages to analyse and forecast time series. For example, we can visually explore and analyse time series in a more detailed way. We can decide which method would be more efficient to use according to these components. To illustrate, ARMA and many other models are not applicable for time series that have trend and seasonality. This condition leads us to stationary time series concept. In this section I will give brief definition of these terms.

There are two ways to mathematically define a time series, additive decomposition, or multiplicative decomposition.

$Z_t = T_t + S_t + R_t$ **(1)** $\qquad$ $Z_t = T_t \times S_t \times R_t$ **(2)**

Eq. (1) is formula of additive, and Eq. (2) is formula of multiplicative decomposition. Z_t denotes the original time series, T_t denotes trend component of time series and R_t denotes residue component of time series.

In additive decomposition all variables have the same focus, while in the multiplicative decomposition focused value of variable is trend. Other components of multiplicative decomposition show impact of residue and seasonality on trend.

2.1.1 Trend

Trend is a useful pattern that utilized in many fields. Parmezan et al. (2019) defined trend as upward or downward movement in the data within long periods. In time series, trend shows general direction of the situation. Trend can be in various behaviours, such as linear, damped and exponential. For example, time series with sales numbers shows the performance of employees, and if there is an upward linear trend means that employee performing improvement in a constant rate. Furthermore, people can visually observe relationships between variables. Trend is very important in linear regression. If the relationship between dependent and independent variables shows linear behaviour, we can estimate dependent variables for any independent variable that does not include in the dataset. It is also common application of transforming trend shapes to the linear trend shape to increase efficiency of the linear regression models. In this report, I will use similar method in addition to ARIMA model by adding exogenous variable. Many time series forecasting methods require time series without a trend, which is a condition of stationarity. Since most of the trends are easy to capture by human eye, plotting data is very useful as a first step of any analysis and forecasting.

2.1.2 Seasonality

Simple definition of seasonality is repeated similar patterns in the data with fixed time periods. We can see these patterns in different frequencies. An example for daily seasonal pattern can be density of the traffic in a city, which can be increase in every commuting hour. Example for weekly seasonal data can be number of customers of restaurant with the higher demands on weekend days. Monthly seasonal pattern can be observed at grocery spendings of white-collar employees in Turkey since most of them get their monthly salary on 15th day of the month. Lastly, Annual seasonal patterns can be seen in temperature time series. Frequency can be at any period with seasonal component, only condition, it must be fixed periods. There are many forecasting models that use this component of time series, such as seasonal ARIMA (SARIMA), recurrent neural networks (RNN) and Holt-Winter exponential smoothing (HWES) Winters (1960). In addition to seasonal pattens, time series can also have cyclic patterns, which is very similar to seasonal behaviour. Only difference is that cyclic patterns happen in different periods, which it makes unpredictable. We can see cyclic patterns on stock market time series.

2.1.3 Residue

Residue is the unpredictable fluctuations in the time series data, caused by randomness. Parmezan et al. (2019) stated that residue component is non-systematic dramatical changes within short durations in time series, caused by extreme incidents such as natural disasters, and terrorist attacks. In the finding section, I will give an example of stock price prediction of Microsoft's share from just before pandemic recognised worldwide and will compare its root mean square error (RMSE) with normal market prediction. When traditional statistical time series forecasting models considers numerical data, well trained ANN models can process texts, so they could be more responsive to these extreme incidents. However, it is very complex and expensive process to increase the speed of this responses to the predictions for today. In contrast, with the dramatical improvements in artificial intelligence (AI), academics have high expectations for ANN-RNN models that can capture these incidents and produce high accuracy predictions in shorter time in the future. Residue component is also very important to examine the data if it is predictable or not.

2.1.4 Stationarity

Many prediction models such as ARMA require stationary time series. Simple explanation of stationary time series has two major conditions, constant level of mean and variance of natural data over different periods of time series (Mills, 2019). If we consider time series with constant statistical metrics, it is easier to formulate prediction models on it. Nielsen (2019) argued that if we apply statistical prediction methods to non-stationary models, predictions will be dependent strongly with time since mean and variance will be different at each time. Thus, model produces biased and low accuracy results. Time series can be divided into two class as stationary and non-stationary time series. We can give white noise process as an example for stationary time series. If the time series has trend or seasonality, that time series is not stationary. Thus, we need to transform nan-stationary time series to the stationary process. General methods of these transformations are calculating the natural log, square root of the time series or applying differencing method. Sometimes, we can simply reach to a conclusion for non-stationary process by eyeballing visualization of the time series if we see clear trend or seasonality patterns. However, we need more evidence to evaluate stationarity. Another condition for stationary time series is that they should not have any unit root. We can use augmented Dicky-Fuller (ADF) test, which is using a null hypothesis method to detect unit root. If the null hypothesis can be rejected, we can assume time series as a stationary process.

2.2 Alternative Prediction Methods

In this section, I will give basic explanations for some of the alternative prediction models that been used instead of ARIMA in literature.

2.2.1 ARCH, GARCH, ARMA-GARCH Methods

ARCH modelling provides us to utilize of volatility in our predictions. Mills (2019) generally defined volatility as the levels of variability between the periods of time series. ARCH models consider the past values and standard deviations with the white noise. Tsay (2010) argued that ARCH models have some weaknesses, which are:

- ARCH models cannot distinguish positive and negative shocks that affect volatility of time series. This problem caused of the taking square of previous shocks.
- The model does not provide any new knowledge except statistical description of how conditional variance behaves without explaining the cause.
- There is big possibility for over-forecasting the volatility due to slow reaction large underlying shocks in return time series.

One step improved version of ARCH modelling is the GARCH model, which feeds the standard deviation with the past standard deviations. Nielsen (2019) argued that in high volatility time series such as financial time series GARCH models are very effective. In literature, we can also see that some researchers used GARCH model along with the ARMA. Arashi & Rounaghi (2022) argued that combination of ARMA and GARCH can be used for well resulted predictions. They stated their ARMA-GARCH prediction model error level is just 1%.

2.2.2 Exponential Smoothing Methods

An alternative statistical method to the ARIMA is exponential smoothing method. Formula of this method was created during the World War II by Robert G. Brown to forecast speed of the rocket and angle of the artillery according to depth in submarines (Gardner, 2006) for hitting the target with high accuracy. Moreover, Holt (2004), in 1957, found the similar formula and adjusted method for seasonal data. In 1960, Winters (1960) tested and adjusted Holt's study and found promising results. However, robustness and strong statistical background of ARIMA model is more reliable and convenient for most of the academics compared to exponential smoothing prediction methods.

2.2.2.1 Simple Exponential Smoothing Method (SES)

SES model has got relatively simple mathematical explanation, which it makes easy to understand and apply. This method is suitable for time series that has no trend and no seasonality. SES model just considers the level forecasting errors in the time series. We can think level as a interval of the value of the data.

$\mathrm{L_t} = \alpha \mathrm{Y_t} + \alpha(1-\alpha)\mathrm{Y_{t-1}} + (1-\alpha)^2 \mathrm{L_{t-2}}$ **(3)**

$F_{t+1} = L_t = L_{t-1} + \alpha(Y_t - L_{t-1})$ **(4)**

$F_{t+1} = F_t + \alpha E_t$ **(5)**

From the equations t denotes present, t-1 means first lag and so on. $\mathrm{L_t}$ denotes the value of level, Y_t denotes value of last data, α denotes smoothing constant, which is between 0 and 1, F_{t+1} denotes forecasted value of next period, and E_t denotes value of error. SES method consider the first value in the time series as first level value, then levels got updated according to eq.3. Because of α between 0 and 1, past values are less effective on the prediction. With SES method we can only make one-step-ahead predictions. Parmezan et al. (2019) argued that when the simplicity of method considered, accuracy performance is reasonable with the only problem, which is difficult to find efficient smoothing constant. We need to choose smoothing constant according to root mean square error (RMSE). Lastly, eq.5 is very similar with moving average method. However, in MA method, effect of the previous values does not decrease on prediction results.

2.2.3 Neural Network Methods

In addition to these methods, with the developments in machine learning, people have started to use artificial neural network (ANN) models to forecast. Gooijer & Hyndman (2006) also stated that ANN models could give more accuracy in some cases compared to ARIMA. However, one downside of these ANN models is the high complexity caused by deep learning. It causes a lack of explanation compared to ARIMA models. In contrast, Ho et al. (2002) argued that partially recurrent neural network (RNN) and ARIMA models produce more accurate predictions compared to other neural network and statistical models.

2.2.3.1 RNN

RNN model is a class of artificial neural network, that can forecast sequential data. This type of algorithms allows to the researchers create model with various types of inputs, such as text, image, video, and audio. Working structure is very different than statistical prediction methods such as ARIMA. RNN models also can store previous predictions to improve their knowledge for the future forecasts. Rumelhart et al. (1986) who introduced RNN models defined the algorithm as a learning procedure that can process the data repeatedly to minimize the error, by the help of hidden layers.

2.3 ARIMA and its Components

ARIMA models have been widely used by academics and professionals for time series forecasting. Compared to other methods, ARIMA has proven its reliability. Even some academics claim that ARIMA is the best method for prediction today (Sharma et al., 2022). In contrast, there is a good possibility that with the developments in AI, ANN-RNN models might be more robust since they can also learn from ARIMA model's prediction results in the future. ARIMA model consist of three parts, autoregressive (AR), integration (I) and moving average (MA). Only difference with ARMA model is integration part which shows the number of differencing methods, if time series is not stationary.

2.3.1 Autoregressive (AR)

AR models investigates the relationships between present time value and the lagged values directly. Stationarity is important concept for AR models. AR model assumes that mean and variance over time constant, and these conditions provide ability of prediction since expected value of time series will be same. In addition, if we check the formulas of AR (1) model, we can see the logic why coefficient of AR model ϕ needs to be smaller than 1, and why AR models are applicable just in stationary time series. Nielsen (2019) stated that in eq.8 derived from eq.7 based on stationary assumptions. As a condition of stationarity, variances over time need to be equal, and it needs to be positive. Thus, $-1 < \phi_1 < 1$.

$$y_t = \phi_0 + \phi_1 \times y_{t-1} + \phi_2 \times y_{t-2} + \cdots + \phi_p \times y_{t-p} + e_t \quad \textbf{(6)}$$

$$y_t = \phi_0 + \phi_1 \times y_{t-1} + e_t \quad \textbf{(7)}$$

$$\text{var}\,(y_t) = \frac{\text{var}(e_t)}{1-\phi_1^2} \quad \textbf{(8)}$$

Eq.6 is the general formula of autoregressive models. P denotes the order of the model, which it shows how many lagged values will be considered by the AR model. To decide the order, we can look at the partial autocorrelation function graph (PACF), which it shows significant values for prediction of y_t.

2.3.2 Moving – Average (MA)

MA model can be express with similar linear formula with AR model. However, MA model works with past errors instead of values. Logic of MA model similar with SES model, however MA model can forecast multiple future values.

$$y_t = \mu + e_t + \theta_1 \times e_{t-1} + \theta_2 \times e_{t-2} \ldots + \theta_q \times e_{t-q} \quad \textbf{(9)}$$

From eq.9 as we can see that errors of previous lags added to the mean of the data. e_{t-q} element in eq.9 involves the term of q which means order of the MA model. This q value can be found from autocorrelation function plot (ACF). Another difference of MA model than AR model is that MA model creates the relationship with the data collectively. From its nature, MA model does not memorize the previous values. Nielsen (2019) argued that MA models are already stationary by its definition, because mean and variance of the errors are not infinitive and vary. The reason is that errors are assumed as independent and identically distributed with mean 0.

2.3.3 ARMA and Integration part

ARMA model basically AR model + MA model, which in a sense, intuitively, AR part will forecast the data and MA part will adjust the forecast according to previous errors.

$$y_t = c + \sum_{n=1}^{p} \alpha_n y_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \epsilon_t \quad \textbf{(10)}$$

One of the drawbacks of ARMA in practice, it requires stationary time series, except this feature, it is same model with ARIMA. One of the most popular to convert non-stationary time series to stationary time series is differencing, which is subtracting values in the time series with their first lagged values. Generally applying differencing once is enough to convert non-stationary processes to stationary time series. Number of differencing operations also called order of the integration I part of the ARIMA. We can see the effect of differencing method from ARIMA (1,1,0) in eq.11.

$y_t - y_{t-1} = c + \alpha_1(y_{t-1} - y_{t-2}) + \epsilon_t$ **(11)**

2.3.4 ARIMA

General mathematical equation of ARIMA model is shown in eq.12, which we can see that ARIMA model is ARMA model on differentiated time series. ARIMA models are described based on their orders, like ARIMA (p, d, q). First element p denotes the order of AR component, d denotes number of differencing, and q denotes order of MA component.

$d_t = c + \sum_{n=1}^{p} \alpha_n d_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \epsilon_t$ **(12)**

2.3.5 ARIMAX

ARIMA models can utilized from external factors. In literature these factors called exogenous variables. These variables do not affect ARIMA outcomes directly, like in a linear regression. Exogenous variables have impact on autoregression and moving-average components of the model. In real life, ARIMAX models can have hundreds of exogenous variables. Formula of ARIMAX shown down below (eq.13). Khairunnisa et al. (2020) argued that ARIMAX models have higher accuracy than ARIMA.

$d_t = c + \sum_{n=1}^{p} \alpha_n d_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \sum_{n=1}^{r} \beta_n x_{n_t} + \epsilon_t$ **(13)**

2.3.6 Seasonal ARIMA and Seasonal ARIMAX

In addition to non-seasonal time series, ARIMA also useful prediction method for time series that have seasonal behaviour. This type of ARIMA model is called seasonal ARIMA (SARIMA). Mills (2019) stated that ARIMA models are very good with expressing stochastic seasonality. SARIMA models can be expressed as SARIMA (p, d, q) X (P, D, Q) m, where p, d, and q same for ARIMA, m stands for order of seasonal patterns that model can use. Mills (2019) said that orders of P and Q elements in SARIMA can be extracted from sample autocorrelation function (SACF) and PACF graphs. Like ARIMA model, we can also add exogenous variables to the SARIMA model, this type of model is called SARIMAX, which carries similar features with ARIMAX. General formula of SARIMA and SARIMAX is shown down below as eq.14 and eq.15 respectively.

$y_t = c + \sum_{n=1}^{p} \alpha_n y_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \sum_{n=1}^{P} \phi_n y_{t-sn} + \sum_{n=1}^{Q} \eta_n \epsilon_{t-sn} + \epsilon_t$ **(14)**

$d_t = c + \sum_{n=1}^{p} \alpha_n d_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \sum_{n=1}^{r} \beta_n x_{n_t} + \sum_{n=1}^{P} \phi_n d_{t-sn} + \sum_{n=1}^{Q} \eta_n \epsilon_{t-sn} + \epsilon_t$ **(15)**

2.3.7 Finding Order of the Model

Deciding the p, d, and q parameters of ARIMA (p, d, q) requires different approaches. First approach is to check ACF graph for the order of MA component, and PACF graph for the order of AR component. Required order of differencing can be observed from ADF test. However, to prevent over differencing, trying taking log of the values or square root of them could be more beneficial in some cases. In contrast, Nielsen (2019) argued that when taking log or square root of the data large values would successfully compress, small values would not, and making these transformations caused positive data throughout the datasets. Another approach is checking Akaike information criterion (AIC), measurement type of goodness-of-fit of the model. AIC penalizes complex models to prevent over fit in the model, and lower AIC score means better fit for the model. Another approach to find the order of model is checking Bayesian information criterion (BIC), which penalizes complex models more than AIC. Like AIC, Lower BIC score means better model. Mills (2019) argued that when the BIC method sufficiently consistent to find right order of the model, AIC method most of the time shows higher order than the necessary one, by saying that this case may vary in finite sample. Thus, we should check both scores when we are deciding the order of model. Combination of these methods should be considered together.

2.3.8 Diagnostics of ARIMA

Examination of the model and prediction results are integrated concepts when it comes to the overall diagnosing of the study. Prediction results can be very promising when the goodness-of-fit of the model is not sufficient. We can check the model from diagnostic plots of residuals. In practice, good model should have distribution of residuals is expected to behave closely to the white noise. Mills (2019) argued that stationary time series can be considered as series consist of some signals and errors, which is demonstrated in eq.16.

$x_t = z_t + u_t$ **(16)**

In eq.16, x_t denotes stationary time series, z_t denotes the signal that is predictable, and u_t denotes the stationary noise. Signal component is important for predictions model because this is the component that we can predict. Thus, Mills (2019) stated that signal-to-noise ratio of the data is important factor in forecasting methods. After model is created, checking the similarity of distribution of residuals with white noise general visual practice, because it can be interpreted by eyeballing analysis, since distribution of white noise is certain and easy to define.

White noise process is consisting of normal distributed series of independent and identically distributed with zero mean and finite variance (Tsay, 2010). If we consider white noise in a time series, value of white noise at present is independent from its previous errors and values (Mills, 2019). Thus, white noise cannot be predicted. Moreover, white noise process is a stationary process, so variance not varying over time. There are 3 popular tests to decide if the process is white noise. First test is visual test to analyse the distribution of the process. Second test is to observing auto correlation function of the process, which relations should be 0 in theory. However, in practice, if the values of SACF are approximately zero, process assumed as white noise process (Tsay, 2010). Third test is to calculate mean and consistency of the variance of the process. This test has two approaches. First one is global testing, which it means finding the mean and variance of all data for all lags. Second approach is called local testing, which it is finding mean and variance period by period and comparing the similarity of these results.

Analysing the similarity between residuals and white noise process have similar procedure. First plot to check distribution of it standardized residual graph. From this graph we can visually observe the variance of the residuals. Second graph is histogram graph of residuals, which it is useful to exhibit how much close the distribution of residuals is to normal distribution. Third graph is called Q-Q graph, which is also show the difference between distribution of residuals and the normal distribution. Fourth graph is called correlogram, which is the easiest to read compared to other graphs. Desired result is approximately zero correlation between lagged values of residuals since they should be independent and identically distributed.

In addition to checking model with these methods, there are many forecast performance measures with residuals. De Gooijer & Hyndman (2006) argued that some of the most common measures are mean squared error (MSE), root mean squared error (RMSE), absolute error (MAE Mean), etc. Mathematically we can express residual of a forecast of data point as $e_t = Y_t - F_t,\ Y_t$ denotes actual value in time t and F_t denotes forecasted value in time t. Mentioned accuracy measures applied to total e_t of a time series. Some researchers also use $p_t = 100e_t/Y_t$ (percentage error) in their forecasting studies.

3 Overviews and Methods

In this section, I will give general methods and approaches about my application process of ARIMA. My intention is to examine the efficiency of ARIMA and ARIMAX models on predicting Microsoft's stock prices. In this project, I applied ARIMA and ARIMAX methods with Python software language on Jupyter Notebook. I used pandas, 'numpy', 'matplotlib.pyplot', 'statsmodels', 'datetime', 'pmdarima', and 'pandas_datareader' packages. All these packages are available for everyone and opensource libraries.

One of the important steps of my ARIMAX application process for forecasting Microsoft's stock prices, started with choosing relative and high correlated exogenous variable for Microsoft's stock price (MSFT) for ARIMAX method. My supervisor, Stephen M. Disney, advised me to check the relationship between MSFT and with Dow Jones industrial average index (DJIA). This index provides general performance observation of first 30 largest companies in USA. Hall (2022) argued that DJIA is significant factor that helps to evaluate economic progress of USA. Since these 30 companies' operations are worldwide, DJIA is also important indicator for economic situation of the world. One of the important reasons, these companies have tremendous impact on another countries' economy. To illustrate, Ausick (2017) claimed that Apple approximately caused directly or indirectly 4.8 million jobs just in China. Thus, any back-step of Apple's business would not just affect USA economy, it would also affect many businesses around the world. Besides Apple, other companies who is registered to the DJIA same kind of impacts on the global economy.

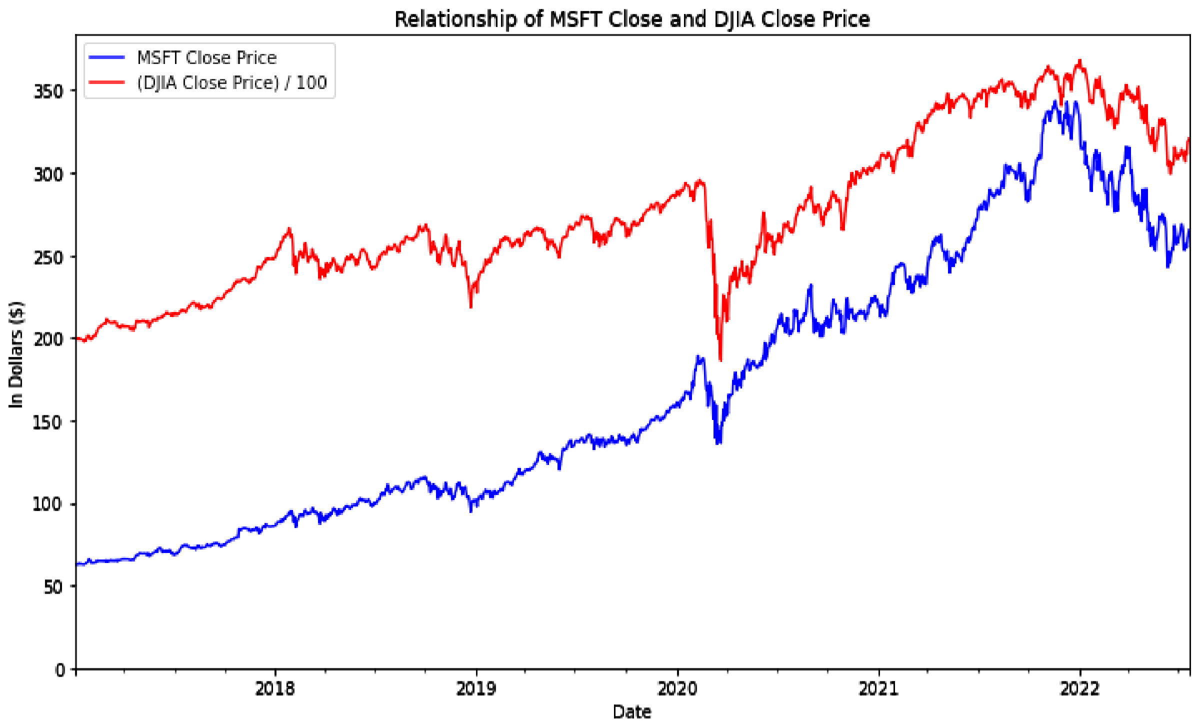


Figure 2 (author's own work.)

When I was analysing the relationship with MSFT close and DJIA close values, I found that correlation of two series is 0.94. To provide more clear visualization of this relationship, as we can see from **Figure 1,** I divided values of DJIA by 100. This result shows that for the prediction of MSFT, we can use DJIA values as exogenous variable in our ARIMAX models due to high correlation. I would like to remind that this correlation does not affect the model directly like in linear regression. It is going to affect AR and MA components instead.

My supervisor shared a document that prepared by him and Fotios Petropolus. In this document, Disney and Petropolus (2015) argued that average results of prediction models that start each turning point in the dataset might improve accuracy of the prediction. To illustrate, if we have 10 years of Apple's stock prices, we could apply many ARIMA models that starts every turning point of the time series. This method allows different trained models as number of dramatical changes in the time series. Disney and Petropolus (2015) also argued that some fluctuations can be explained and some of them cannot explained. Examples for explainable and dramatical changes could be new version of smart phone, good or bad investments, marketing operations, improvement or failure of the competitor companies' products, changes in customer behaviour, etc. Unexplainable changes could be incidents such as natural disasters and terrorist attacks as I mentioned earlier in this report. When we select different start date in the time series, prediction model trained differently. Disney and Petropolus (2015) argued that if the structures of out-of-sample prediction results from different trained models does not change significantly, predictions should be reliable and should have high accuracy. I applied this method in my study for business daily (BD) frequency data, which is presented with the results in findings section.

3.1 Retrieving Data and Ethics

I retrieved two datasets, stock prices of Microsoft and DJIA (Yahoo! Finance, 2022). The time interval of both datasets is from 3rd of January 2017 to 2nd of September 2022. I used 'pandas-datareader' library with a function that called 'get_data_yahoo'. This function provides to retrieve financial data from Yahoo! Finance.

This study did not require any ethical implication for retrieving datasets. The reason is that both datasets from Yahoo! Finance and it is an opensource platform. Every platform, software language and its libraries are also opensource tools, which I already mentioned them. In addition, I asked to my supervisor for any possible ethical implication, and we have signed a document that called ethics form, which is declaring that there is not any ethical violation for the datasets. Moreover, only ethical implication of this study is using APA7 reference style in this report. Lastly, I will explain the manipulations on the datasets in data pre-processing section. Being transparent in this kind of studies important behaviour because every manipulation might affect results of predictions.

3.2 Data Pre-processing

As I mentioned in previous section, I retrieved data with 'pandas-datareader' library, then I extract it to a local file as csv document. Afterwards, I upload it another notebook file as time series by converting date column to index of the datasets. In later stage, I extracted Close prices columns from both datasets and created new dataset. In this process only difference is that I divided DJIA Close column by 100 for visualizations purposes. The reason of this transformation, while the DJIA varies from 1900 to 32000, MSFT close prices varies from 62 $ to 343 $.

Date	Msft Close	Dow Jones Close	dj shifted	msft diffed
2017-01-03	62.580002	198.817598	NaN	NaN
2017-01-04	62.299999	199.421602	198.817598	-0.280003
2017-01-05	62.299999	198.992891	199.421602	0.000000
2017-01-06	62.840000	199.638008	198.992891	0.540001

Figure 3 (author's own work)

I created 2 additional columns in close price dataset. Name of first additional column is DJIA-shifted. In this column, I moved the values of DJIA to the next row. This manipulation was necessary for one-step-ahead forecasting with

ARIMAX. End of this process first value of the column became not-a-number (NaN) value. Name of the second additional column is MSFT-diffed, which it consists of differenced values of MSFT close prices column, by applying 'diff()' function in python. Adding differenced version of MSFT close price is convenient for applying ADF test. As you can see in **Figure 2,** 3rd of September 2017 has 2 NaN values. I removed this row by applying 'dropna()' function and assign this dataset to new variable which is called 'df'.

For further manipulations on the 'df' dataset, I change the frequency of the time series to the business days (BD) by using 'asfreq('B')' function. However, when the original dataset had 1396 rows, BD frequency dataset has 1447 rows. These additional days come from holidays, such as New Year's Day and Martin Luther King, Jr Day because NASDAQ closed in major holidays. I used interpolation method, which is one of the up-sampling methods, to fill those empty days. Longest training dataset for predictions with BD frequency starts from 4th of January 2017 and end up on 21st July 2022. Test data of this out-of-sample prediction models starts from 22nd July 2022 to 23 to 18th of August 2022, which is 20 days in total. Lastly, I needed to add 1 days to the dates for getting the one delayed DJIA value, as exogenous variable, in out-of-sample coding line. This method is not suitable for datasets that end with Friday since NASDAQ is closed on Saturdays. However last day of the dataset was Wednesday, which it makes this approach valid. An example image of 'df' dataset can be seen at **Figure 3.**

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 1447 entries, 2017-01-04 to 2022-07-21
Freq: B
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Msft Close        1447 non-null   float64
 1   Dow Jones Close   1447 non-null   float64
 2   dj shifted        1447 non-null   float64
 3   msft diffed       1447 non-null   float64
 4   msft sdiffed      1446 non-null   float64
dtypes: float64(5)
memory usage: 100.1 KB
            Msft Close  Dow Jones Close  dj shifted  msft diffed  msft sdiffed
Date
2017-01-04   62.299999       199.421602  198.817598    -0.280003           NaN
2017-01-05   62.299999       198.992891  199.421602     0.000000      0.280003
2017-01-06   62.840000       199.638008  198.992891     0.540001      0.540001
2017-01-09   62.639999       198.873809  199.638008    -0.200001     -0.740002
2017-01-10   62.619999       198.555293  198.873809    -0.020000      0.180000 None
```

Figure 4 (author's own work)

I did not mention for 'msft sdiffed', which it means two times differenced MSFT close price, because time series did not require second differenced process, as other predictions with different frequencies.

Second prediction is for an examination of ARIMA models to forecast MSFT prices of early pandemic days. Because of frequency of this training dataset is same with df training dataset, with BD frequency, creation of covid training dataset is sub-dataset of 'df' training dataset, which starts at 4th of January and terminated on 21st of February 2020. Key point here is end date. Impact of covid-19 started to be observed approximately around mid-February in 2020 on stock market prices. Nedelman (2020) stated that on 21st of February in 2020, 34 people tested positive for Covid-19. Purpose of this prediction is to examine effect of unknown changes on accuracy of predictions with ARIMA model. Size of the test dataset is 21 days. Test dataset starts from 24th of February and ends on 23rd of March 2020. Reason of this ending date, danger of Covid-19 was recognised worldwide. For example, Boris Johnson made his first restriction speech for Covid-19 pandemic on 23/03/2020 (Prime Minister's Office, 2020). We can see the fluctuations in MSFT stock prices and DJIA during early stages of the pandemic from **Figure 4**.

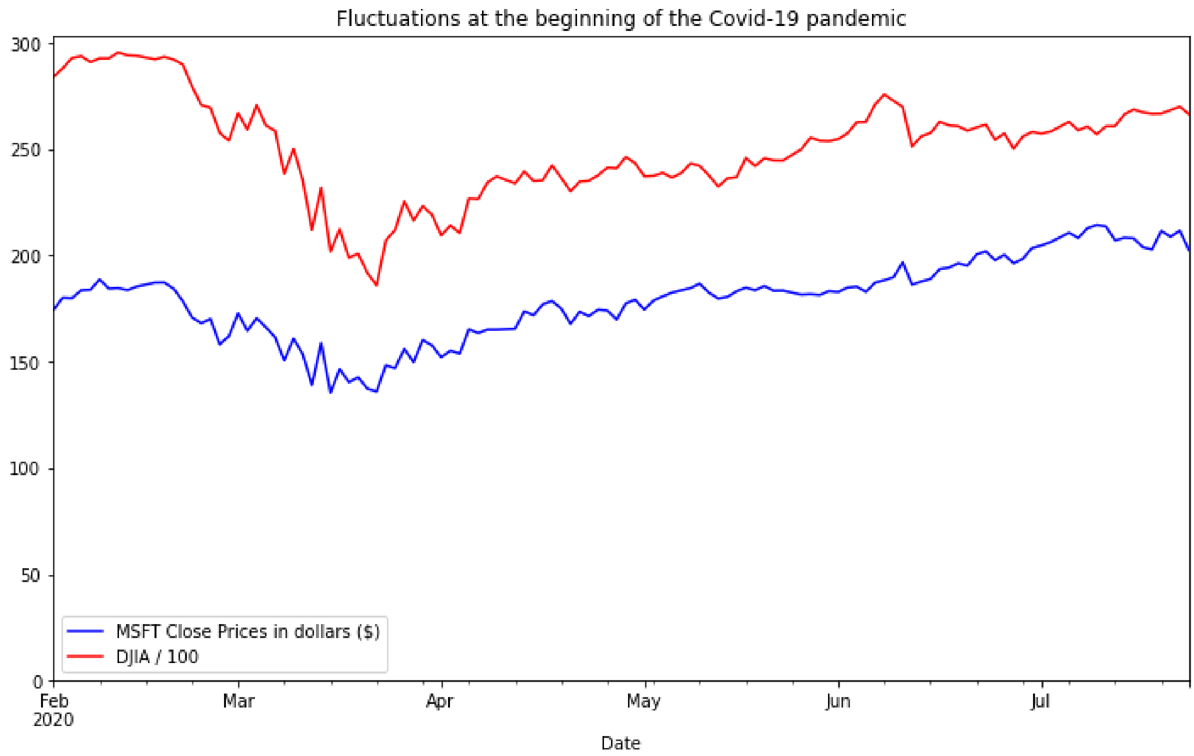


Figure 4 (author's own work)

Third training dataset has weekdays frequency. First, I created new dataset with BD frequency, by filling NaN values with interpolation method. This step is required to get raw values. The reason is that differencing and shifting methods should be applied after change of frequency. If these methods would be applied on df dataset, we could lose one more row, which it is not efficient way to lose data for forecasting algorithms.Afterwards, I changed the frequency from BD to the weekly frequency, which it called down-sampling. Since this study focuses on stock price prediction, I could not use built-in weekly function in python. Down-sampling requires aggerated calculations, which it depends on domain of the study. I used 'W-Fri' argument in asfreq() method, and instead of taking mean of it, I sum up the whole values in the week and divide them into the 5. In these cases, taking mean of the week divides values to the 7 automatically. Thus, this approach was significant step for maintaining integrity of the data. Moreover, 'W-Fri' argument assigns the mean values to Friday, which is the last active stock market day of the week. I changed the differenced values of MSFT price column with differenced values of weekly MSFT stock prices and applied shifting method for weekly DJIA values. These operations convert first raw values to the NaN. Thus, I deleted the first raw. Furthermore, I created weekly training dataset from 13th of January 2017 to 15th of July 2022, because I removed incomplete weeks from the dataset to prevent dramatical drops in mean prices of those weeks. Weekly test dataset starts from 22nd of July 2022 to 2nd September 2022, which it makes 5 weeks in total. Same manipulation processes applied to the test data. Lastly, date of last DJIA value of training dataset changed to next week's Friday date, for weekly exogenous variable. Demonstration of weekly training data set can be seen in **Figure 5**.

```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 287 entries, 2017-01-20 to 2022-07-15
Freq: W-FRI
Data columns (total 6 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Msft Close        287 non-null    float64
 1   Dow Jones Close   287 non-null    float64
 2   log_msft          287 non-null    float64
 3   log_dj            287 non-null    float64
 4   dj shifted        287 non-null    float64
 5   log_dif_msft      287 non-null    float64
dtypes: float64(6)
memory usage: 15.7 KB
            Msft Close  Dow Jones Close  log_msft    log_dj  dj shifted  \
Date
2017-01-20   62.537000       198.094781  4.135758  5.288746    5.293043
2017-01-27   64.041999       199.951520  4.159539  5.298075    5.288746
2017-02-03   64.042000       199.365062  4.159539  5.295138    5.298075
2017-02-10   63.693999       201.277637  4.154090  5.304685    5.295138
2017-02-17   64.592000       205.544500  4.168091  5.325663    5.304685

            log_dif_msft
Date
2017-01-20 -3.432065e-03
2017-01-27  2.378072e-02
2017-02-03  1.191311e-08
2017-02-10 -5.448757e-03
2017-02-17  1.400020e-02   None
```

Figure 5 (author's own work)

3.3 Box – Jenkins Method

Box-Jenkins method is one of the most popular iterative approaches for ARIMA models. De Gooijer & Hyndman (2006) argued that this iterative method is logical and can be adapted to various types of forecasting methods. This method consists of 3 main steps, which starts with identification, estimation and ends with model diagnosing. If one of these steps does not meet the requirements, process should restart from the beginning. Thus, Box-Jenkins methods is iterative process. **Figure 6** shows general process of Box-Jenkins method.

First step is identification of timeseries and ARIMA model. At the beginning of this step, stationarity of time series should be checked. We can look at the data to detect if there are any visual trends or seasonal patterns. Almost all the stock price datasets are non-stationary. In contrast, return values generally stationary. If any trend or seasonal pattern can be observed in time series, differencing could be applied directly. Stationarity condition of differenced time series hard to see with eyeballing analysis. 'Statsmodels' package has 'adfuller' function, which it used for applying ADF test. Desired array can be passed to this function for ADF test and should be assign to a new variable to be able to see the results of the test. Python presents the results in a tuple. Generally considering first two elements of this tuple gives sufficient information about stationarity. First element is value of test statistics. This value should be negative and as it becomes more negative, stationarity of time series would be more likely scenario. Second element of result tuple, which is p value, is very important for hypothesis test. Null hypothesis of ADF test claims that time series is not stationary. If p value smaller than 5%, null hypothesis can be rejected, and time series assumed as stationary process. Sometimes, if stationarity is not obtained with first differenced values, instead of applying second differencing, transformations, such as taking natural log or taking square root of the raw data would be more efficient. If converted data is not stationary, first order differencing can be applied. Both transformation methods can be applied by ‘numpy’ built-in functions. After each conversion ADF test must be applied, and both scenarios should be examined. Once stationarity of time series obtained, next course of action of identification step is finding order of the AR and MA models by checking PACF and ACF graphs, relatively. I used ‘plot_acf()’ and ‘plot_pacf()’ functions from ‘statsmodels’ library for plotting these graphs. Desired graphs for ARIMA model decaying patterns to the statistically not-significant area of the chart. However, real life data can behave differently. When the values of ACF and PACF tails off, older lags might be slightly over statistically not-significant area. In this case, those lags can be avoided. We can also observe the stationary condition of the data from ACF graph. If the values of ACF graph starts around 1 and get slightly lower, there is a high chance that time series is not stationary and needs to be differenced. Therefore, this incident called under-differencing. Second special condition with ACF values, if the first value of it very close to the -1, more than necessary differencing could be applied to the time series. This occurrence is called over-differencing, and it has negative impact on the model. The reason is that, too many differencing might erase the needed data for predictions.

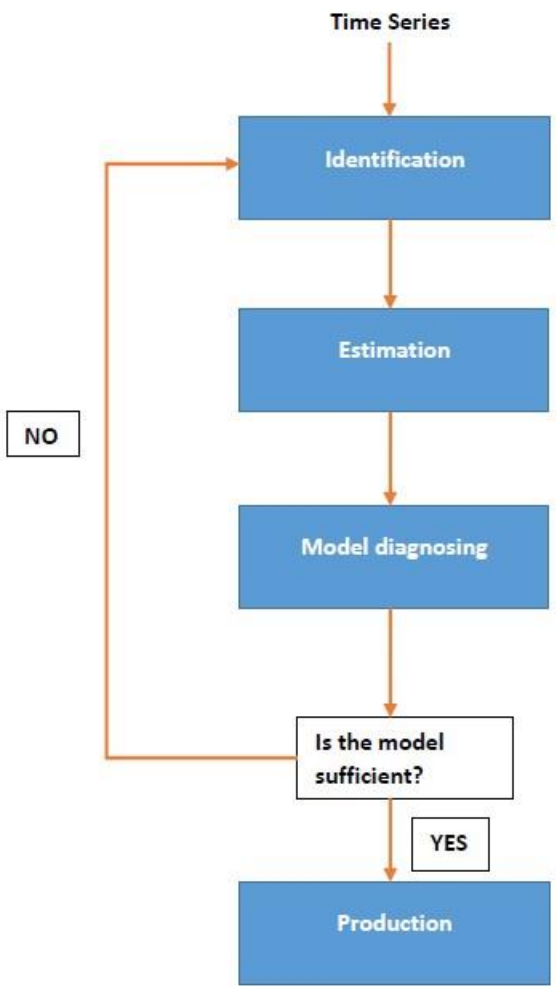


Figure 6 (author's own work)

In estimation step of Box-Jenkins method, determined ARIMA model is applied on the training dataset. I used SARIMAX function from ‘statsmodels’ library for this operation. In **Figure 7**, structure of the ARIMA code and from **Figure 8**, structure of ARIMAX code can be seen.

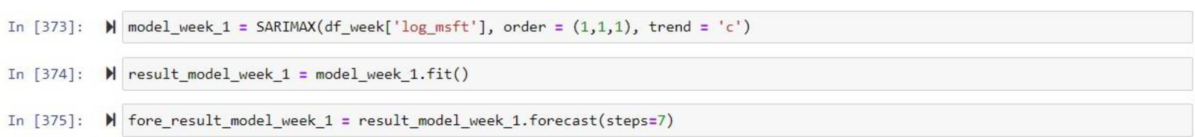


Figure 7 (author's own work)

As it can be seen from **Figure 7**, I was trying an ARIMA model on logarithmic version of MSFT close prices, which I will explain this kind of decisions on findings section. Second argument of function is orders of ARIMA components, AR, differencing, and MA, respectively. Trend argument has the value of ‘c’, which it means constant. If time series does not start from 0, this argument must be passed to the SARIMAX function. Trend argument is letting know the python that there is constant value in the

```
In [320]:  mod_arimax_1 = SARIMAX(df['Msft Close'], order = (1,1,0), exog = df['dj shifted'], trend = 'c')

In [321]:  result_mod_arimax_1 = mod_arimax_1.fit()

In [322]:  forecast_result_mod_arimax_1 = result_mod_arimax_1.forecast(steps=1, exog = dow_100.loc['2022-07-22':]['Close'])
```

Figure 8 (author's own work)

equation. Second line (374) fits the model on training dataset and assigned to a new variable. With this new variable, AIC and BIC values can be calculated by '.aic' and '.bic' methods. As these values discussed earlier in this report, lower AIC and BIC values means better model. Python allows us to get these values for many orders by loop method as you can see in **Figure 9**. Logic of this code is that allowing python to calculate all the models until reaching its max value, which in this case is 4 for each AR and MA models.

```
In [372]:  for p in range(4):
               for q in range(4):
                   model_1 = SARIMAX(df_week['log_msft'], order=(p,1,q))
                   results_aic_bic = model_1.fit()
                   print(p, q, results_aic_bic.aic, results_aic_bic.bic)
```

Figure 9 (author's own work)

Another method to see AIC values is to use 'auto_arima()' function from 'pmdarima' library. Coding structure of it can be seen in **Figure 10**.

```
In [511]:  results_daily_auto = pm.auto_arima(df['2020-03-16':]['Msft Close'],
                                              start_p=0,  # initial guess for AR(p)
                                              start_d=0,  # initial guess for I(d)
                                              start_q=0,  # initial guess for MA(q)
                                              max_p=4,    # max guess for AR(p)
                                              max_d=4,    # max guess for I(d)
                                              max_q=4,    # max guess for MA(q)
                                              trend='c',
                                              information_criterion='aic',
                                              trace=True,
                                              error_action='ignore'
                                              )

           Performing stepwise search to minimize aic
            ARIMA(0,1,0)(0,0,0)[0] intercept   : AIC=3560.755, Time=0.05 sec
            ARIMA(1,1,0)(0,0,0)[0] intercept   : AIC=3556.444, Time=0.15 sec
            ARIMA(0,1,1)(0,0,0)[0] intercept   : AIC=3556.490, Time=0.15 sec
            ARIMA(0,1,0)(0,0,0)[0]             : AIC=3560.755, Time=0.06 sec
            ARIMA(2,1,0)(0,0,0)[0] intercept   : AIC=3558.442, Time=0.25 sec
            ARIMA(1,1,1)(0,0,0)[0] intercept   : AIC=3558.438, Time=0.58 sec
            ARIMA(2,1,1)(0,0,0)[0] intercept   : AIC=3560.444, Time=0.38 sec
            ARIMA(1,1,0)(0,0,0)[0]             : AIC=3556.444, Time=0.14 sec

           Best model:  ARIMA(1,1,0)(0,0,0)[0]
           Total fit time: 1.763 seconds
```

Figure 10 (author's own work)

Important difference of 'auto_arima()' and finding AIC and BIC values with loop method is that 'auto_arima' function suggests an order whereas loop method does not suggest any. In addition, while 'auto_arima' function also suggests the order of difference, loop method assumes stationary time series.

Presented third coding line (375) in **Figure 7**, shows out-of-sample forecasting operation in python. This operation done by '.forecast(steps = …)' function on fitted data. Steps argument tells python to how many further periods will be forecasted.

Coding of ARIMAX model is just separated with ARIMA model in estimation step. As it shown in **Figure 8**, in addition to ARIMA arguments, there is an exog argument inside of SARIMAX function. 1 period shifted to forward DJIA values are registered to this variable. This shifting method is necessary, because DJIA value of 1 period later is not known. Therefore, I utilized correlation between MSFT close price of next period and present period's DJIA value. In SARIMAX function I registered DJIA values for training dataset. Furthermore, second difference occurs in 3rd line (322) inside of the '.forecast()' method. In addition to steps argument, there is also 'exog' argument for test data, which is the last value of unshifted DJIA column. Thus, out-of-sample '.forecast()' method adjusts its one-step-ahead prediction according to this relationship.

Model diagnosing step includes significant observations for verification of the model. Diagnostic plots, mentioned in literature review section, can be obtained from 'plots_diagnostics' method from 'statsmodels' library. This method applicable on fitted data. Code example of this process and plots can be seen in **Figure 11**.

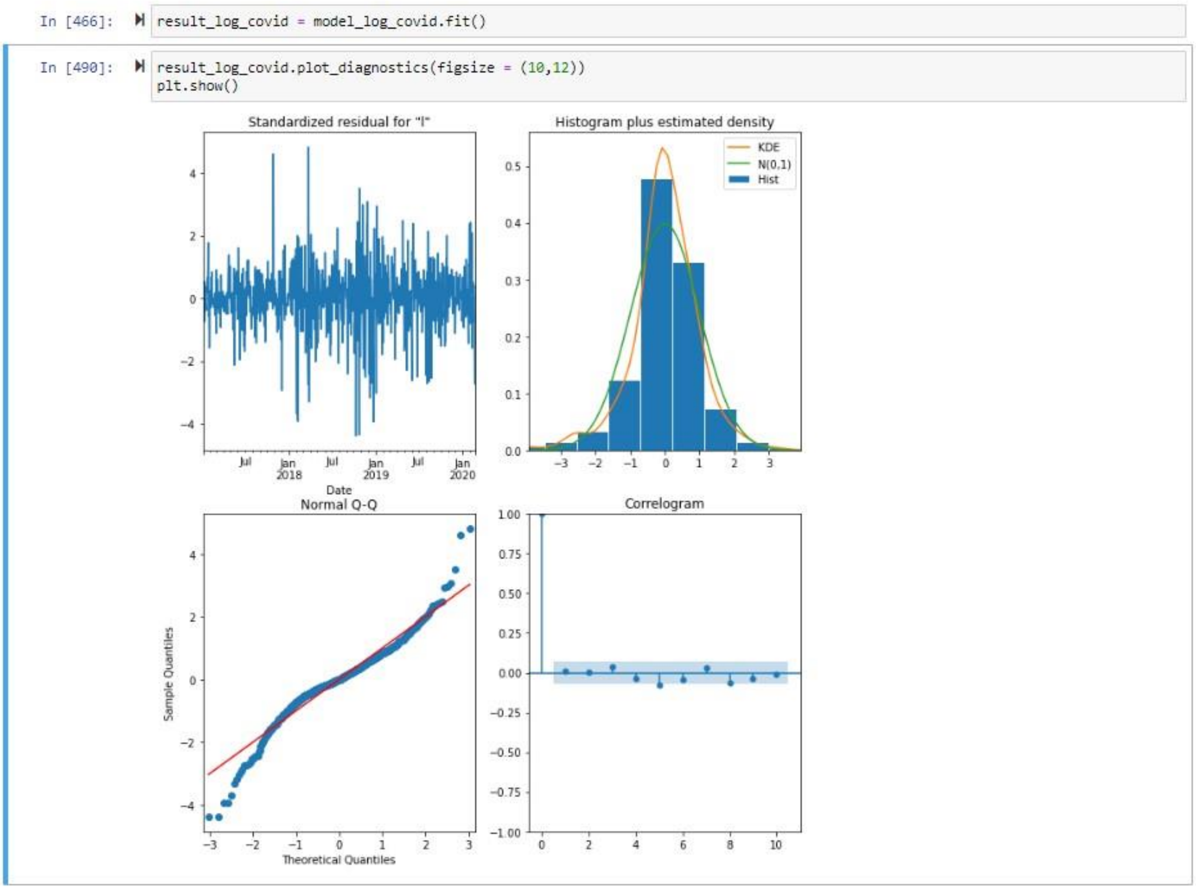


Figure 11 (author's own work)

Diagnostic plots show distribution of residuals and correlation between different lagged values. Good prediction model's residual distribution should have similar look with white noise behaviour. However, in practice, first few prediction trials without any transformations rarely give desired similarity. "Standardized residual for 'I'" plot provides easy interpretation for the variety of the residuals and compared with constant finite variance of white noise. If this plot considered in the example from **Figure 11**, most of the model's residuals are exhibits close variety, except several lags. This diagnose is important sign to improvement of the model is possible. Second graph is called 'Histogram plus estimated density', which it shows similarity of distribution of residuals between desired distribution. In this case, distribution of residuals slightly left skewed. 'Normal Q-Q' very useful graph to interpret for to what extend distribution of residuals exhibits normal distribution behaviour. If the points follow the line, it means that residuals are normally distributed. Because white noise is normally distributed, this is a desired scenario. However, residual points around both end parts of the red line are the residuals of outliers. In practice, those points expected to be distinctive from the line. If they are not separated, it is a sign of over-fitting. Shown 'Normal Q-Q' plot in **Figure 11** is showing that residuals of example model do not normally distributed. 'Correlogram' plot shows the relationship between residuals between different lags. In white noise process, values are independent and identically distributed. Thus, they should not be related and in 'Correlogram' plot they all should be statistically not-significant area, which is the blue are in the graph. In contrast, some of the significant values that very close to the statistically significant area, such as value at lag 5 in the example graph, can be avoided.

In conclusion, if one of these plots do not give the desired information, researchers should go back to step one, which is identification step, and applied some transformation techniques or different order of ARIMA or ARIMAX models. In general, there is not one concrete condition to reach best model. Researchers should approach these forecasting models by considering every detail collectively. This is the reason for why Box-Jenkins method is an iterative approach.

3.4 Risks and Limitations of the Study

This study has some risks about wrong information or lack of reasoning about its content. Only limitation was time, since this report should be submitted by 23rd of September 2022. Due to this limitation, some arguments and methods could not be reviewed, applied, and used. Moreover, findings of this study are obtained with some beginner and intermediate coding operations from python. Advanced transformation techniques like box-cox transformation were not applied for improving models due to lack of knowledge and experience. Thus, part of research question of this study instead of "examining ARIMA and ARIMAX", is written "examining **basic** ARIMA and ARIMAX ".

4 Findings

In this section, predictions models will be evaluated. I will compare datasets with different frequencies, different dates to find the impact of this dissimilarities on ARIMA and ARIMAX models. These comparisons will be based on structure of time series, diagnostic plots, ACF and PACF graphs, AIC, and BIC scores, and lastly RMSE values.

4.1 Predictions with Business Daily Frequency

For BD frequency prediction I determine the dates of 150 largest percentage changes compared to previous period to apply prediction mean method. To simplify the coding process, I wrote a function that called 'orderplot(dd)', which is shown in **Figure 12**. Instead of writing same code chunk for 150 times, I gather them into this function.

```
In [524]:  def orderplot(dd):
                fig, (ax1, ax2) = plt.subplots(2,1, figsize=(7, 7))
                plot_acf(df[dd:]['msft diffed'], lags=10, zero=False, ax=ax1)
                plot_pacf(df[dd:]['msft diffed'], lags=10, zero=False, ax=ax2, method='ywm')
                plt.show()
                adf_result = adfuller(df[dd:]['msft diffed'])
                for p in range(4):
                    for q in range(4):
                        model_1 = SARIMAX(df[dd:]['Msft Close'], order=(p,1,q), enforce_invertibility=True, enforce_stationary=True)
                        results_aic_bic = model_1.fit()
                        print('AR>',p, 'MA>',q, 'AIC>',results_aic_bic.aic, 'BIC>',results_aic_bic.bic)
                return plt.show(), print(p, q, results_aic_bic.aic, results_aic_bic.bic), print('ADF TEST >',adf_result)
```

Figure 12 (author's own work)

By just passing date argument in string format, which is start date of dataset, function plots ACF and PACF graphs of any sliced dataset of 'df'. Since 'df' dataset also has MSFT close prices diffed column, function can also apply ADF test simultaneously. Furthermore, I integrated identification step and estimation step by adding SARIMAX function and extracted the AIC and BIC values of variety orders

AR and MA components of ARIMA with first order of differencing. First, I found the orders from these values, and then, I checked them with 'auto_arima' function. I have extracted dates of 150 largest fluctuations, however, after 16th March of 2020, suggested order of models turned to the ARIMA (0,1,0), so I stopped adding new models and in total I decided to go with predicted mean of 66 models. I produced 66 forecasts with 'forval(dd)' function with one argument, which requires date of turning points as start date of training dataset. This forecasting function can be seen in **Figure 13**.

```
In [292]:  def forval(dd):
               modelfct = SARIMAX(df[dd:]['Msft Close'], order = (1, 1, 0), trend='c')
               resultfct = modelfct.fit()
               forecastfct = resultfct.forecast(steps = 20)
               return forecastfct
```

Figure 13 (author's own work)

Forecasting results of next 20 business days from 66 models shown in **Figure 14** alongside with the original MSFT close prices and average of 66 models. All of the model's orders are ARIMA(1,1,0).

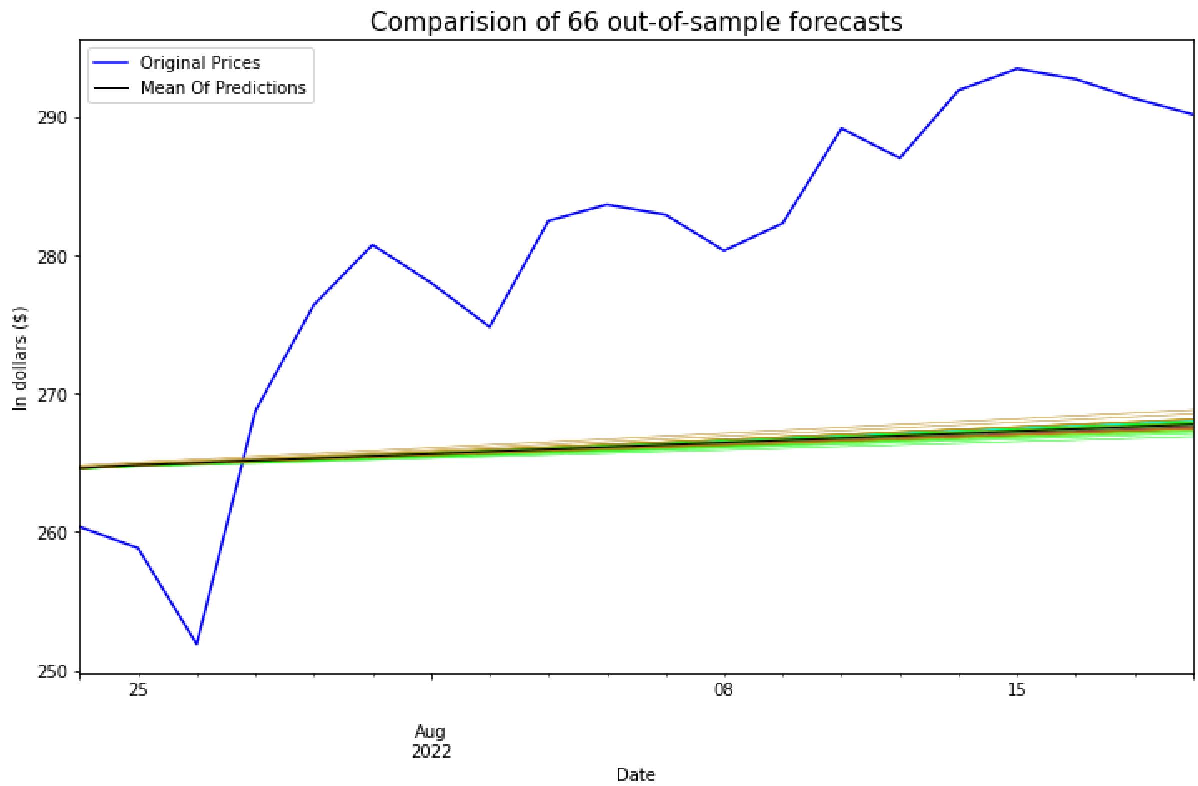


Figure 14 (author's own work)

When RMSE values of the 66 predictions are compared in descending order, ranking of mean prediction's RMSE is 35 with 17.362 RMSE. 5 models with highest accuracy shown in **Figure 15**.

	model	rmse
65	pred_66	16.746724
63	pred_64	16.929383
62	pred_63	17.140618
33	pred_34	17.143329
61	pred_62	17.149684

Figure 15 (author's own work)

In **Figure 14**, structure of out-of-sample prediction results of 66 models are very similar. According to Disney and Petropolus (2015), this similarity is important indicator of reliability for out-of-sample forecasts. Diagnostic plots of pred_66 model is shown in **Figure 16**. Standardized residuals graph shows that variety of residual values indicates changes, which it is not desired scene. However, it can be acceptable. Distribution of residuals shows high similarity with normal distribution. Moreover, 'Normal Q-Q' plot supports this argument. Residuals follow the red line very closely, and residuals of outliers are getting further away from red line, which is normal behaviour. Correlogram graph shows slightly statistically significant 2 residuals at lag 6 and lag 10.

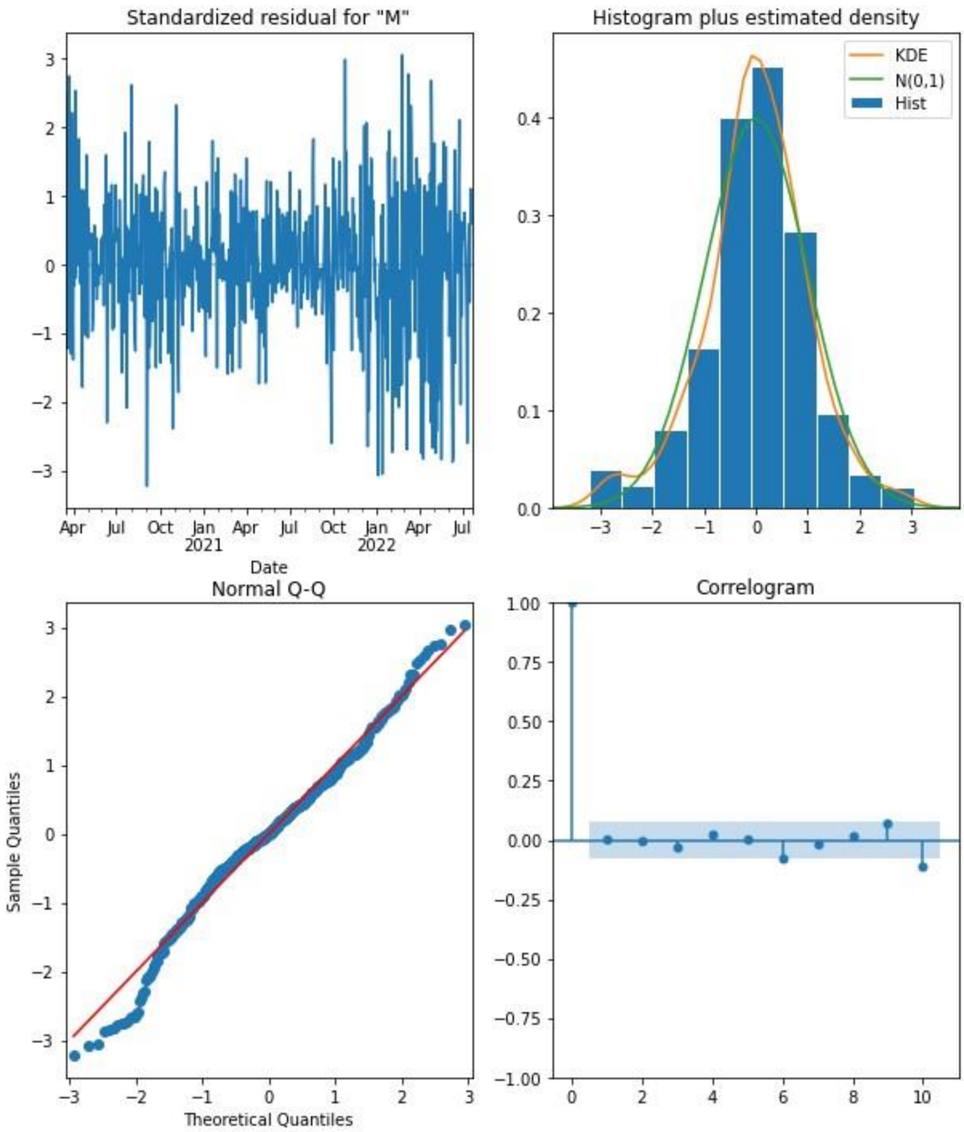


Figure 16 (author's own work)

Correlation between MSFT prices and 1 period shifted DJIA Close values is 0.937. One-step-ahead prediction with ARIMAX version of pred_66's RMSE value is 4.426, when ARIMA's RMSE is 4.449. This difference shows that by adding DJIA, model's accuracy increases 3.6%.

4.2 Predictions with Business Daily Frequency for Beginning of Covid-19

When I was interpreting daily forecast results, starting dates of successful models' training datasets were interesting. Because, most accurate model's training dataset, which is pred_66, starts from 16th of March 2020 and fifth accurate model's dataset, pred_62, starts from 9th of March 2020. Pred_63 and pred_64 starts between these dates. These dates were couple of minimum values during early pandemic.

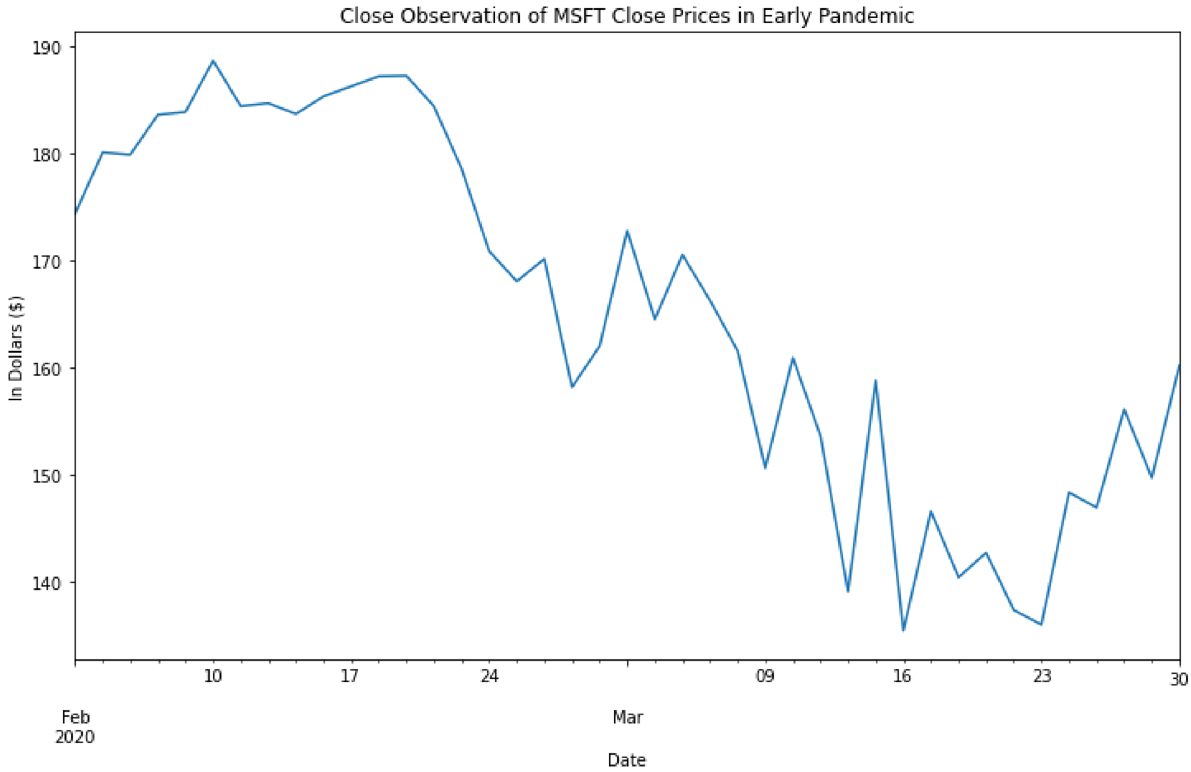


Figure 17 (author's own work)

Start date of pred_66 model was 16th of March 2020, which is the lowest value of early pandemics. There is high probability of eliminating the fluctuations until that date, made pred_66 and other models that started similar dates more successful than other models. Thus, I decided to predict early pandemic days to observe RMSE value of it compared to post-pandemic predictions' RMSE values. After couple of trials, I could obtain the best possible model by taking natural logarithm of MSFT values. ACF and PACF graphs and 'auto_arima' functions were not suggesting same orders, after trying each of model, I picked 'auto_arima' function's suggestion, which is ARIMA(3,1,0). Diagnostic graphs of the model are shown in **Figure 18**.

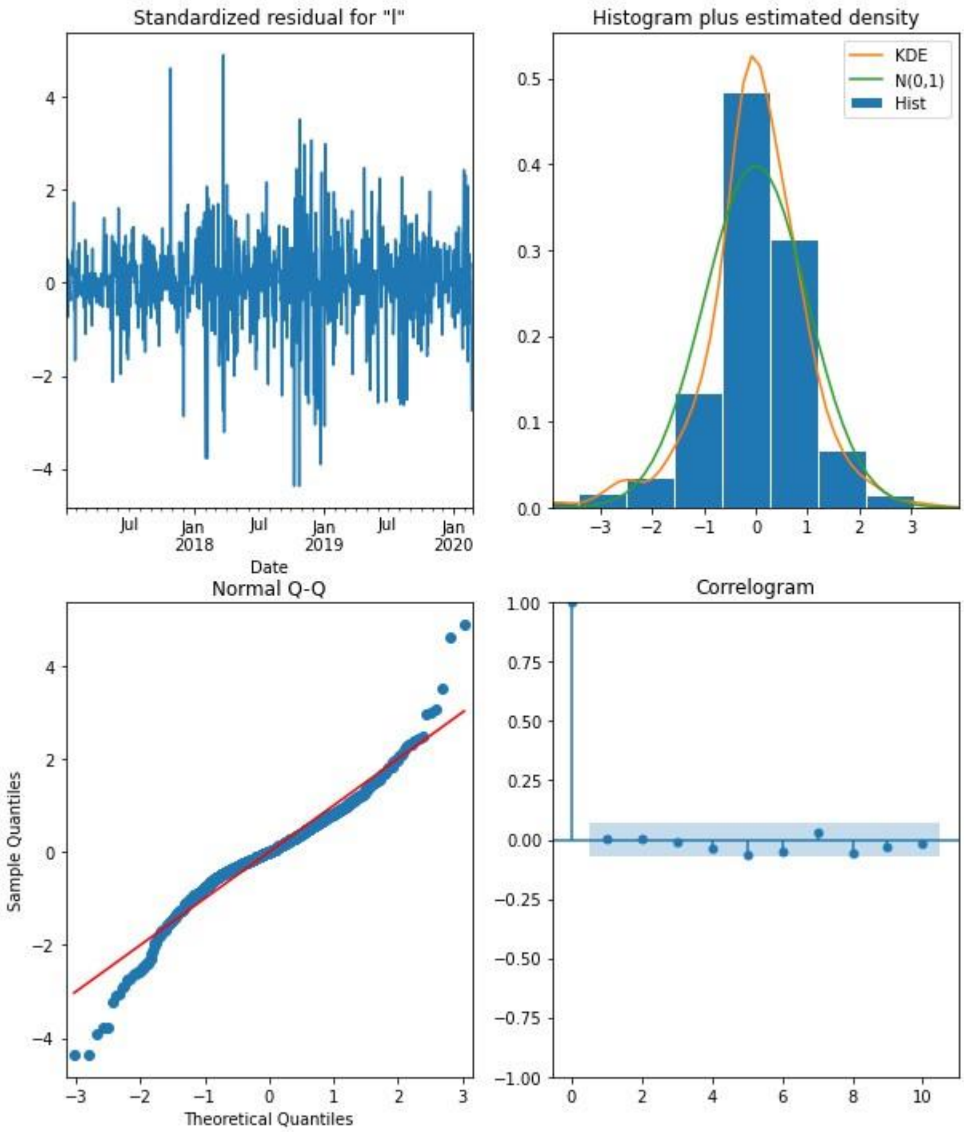


Figure 18 (author's own work)

After forecasting, I converted log values back to the MSFT close price version by applying 'np.exp()' function. Moreover, for a fair comparison, I increase the test size of pred_66 model to the 21 and applied forecast again. RMSE value of pred_66 model's forecast was surprisingly decrease to the 16.765, when prediction of covid model's RMSE is 30.225. Thus, dramatical structure change, which generally caused by rare incidents, in data can significantly decrease performance of ARIMA models. This lack of accuracy also can be observed from **Figure 19 and Figure 20**.

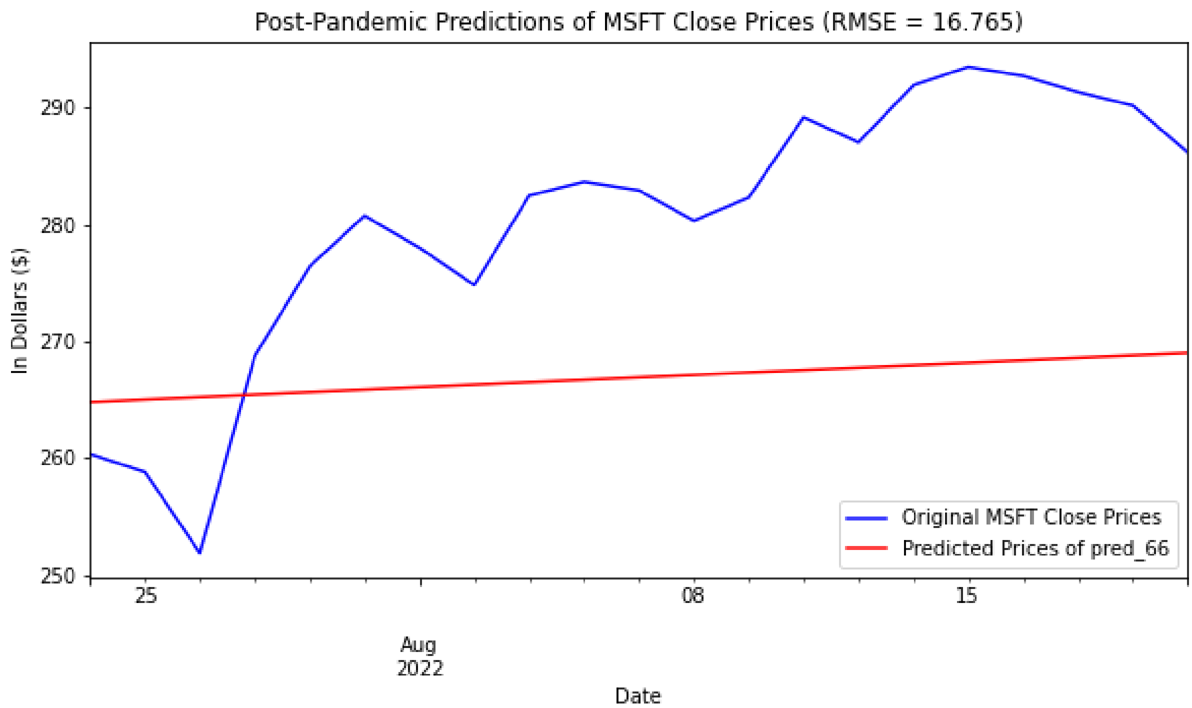


Figure 19 (author's own work)

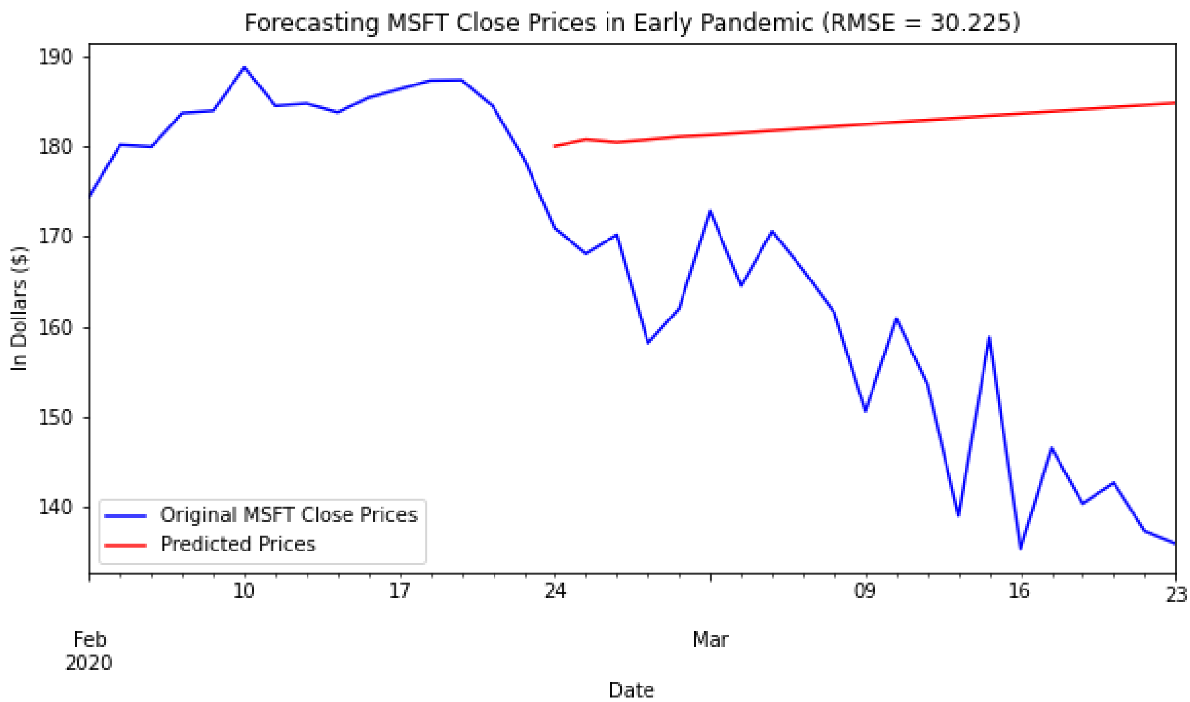


Figure 20 (author's own work)

4.3 Predictions with Business Weekly Frequency

I tried creating model based on original MSFT close prices. However, even data set passed ADF test, variance of first order diffed weekly data was a bit odd, which it can be observed in **Figure 21**.

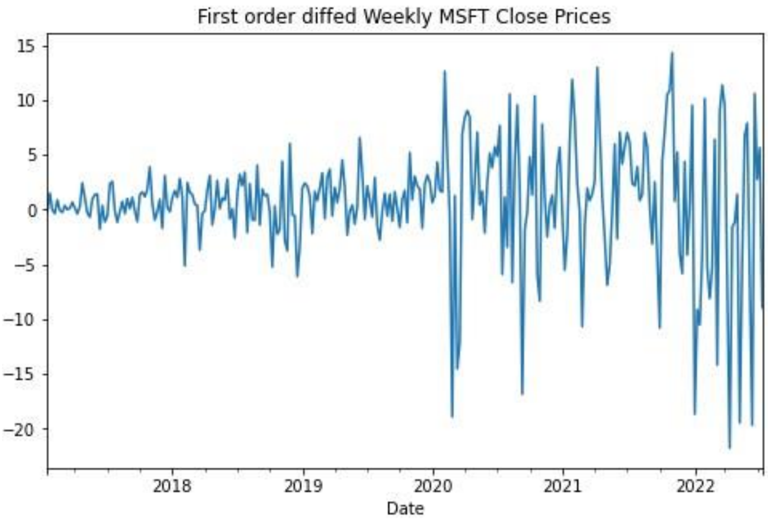


Figure 21 (author's own work)

From the figure, we can clearly see that variance of the data does not exhibit any consistency. As a result of down-sampling operation, data becomes smoother, which this impact might cause difficult to obtain healthy stationary time series. Thus, I transformed MSFT weekly close prices by taking natural logarithm and then I applied differencing operation. Graph of transformed time series can be seen in **Figure 22**. According to graph, it can be said that variance of time series exhibits less variety.

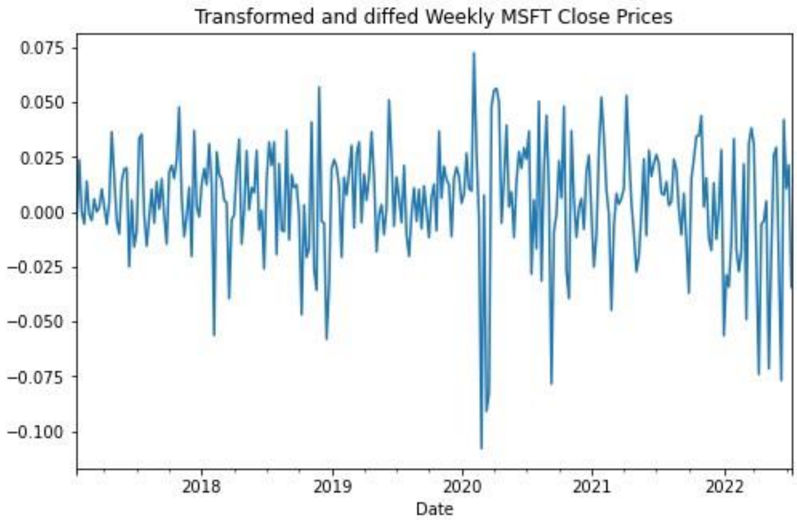


Figure 22 (author's own work)

ACF and PACF graphs were suggesting ARIMA(1,1,1) model. However, 'auto_arima' function suggested ARIMA(0,1,1) model, which was giving lowest AIC score. I compared both models' diagnostic plots, and they were very similar. Thus, I picked ARIMA(0,1,1) since order of model is less complex. Diagnostic graphs of ARIMA(0,1,1) can be seen in **Figure 23**.

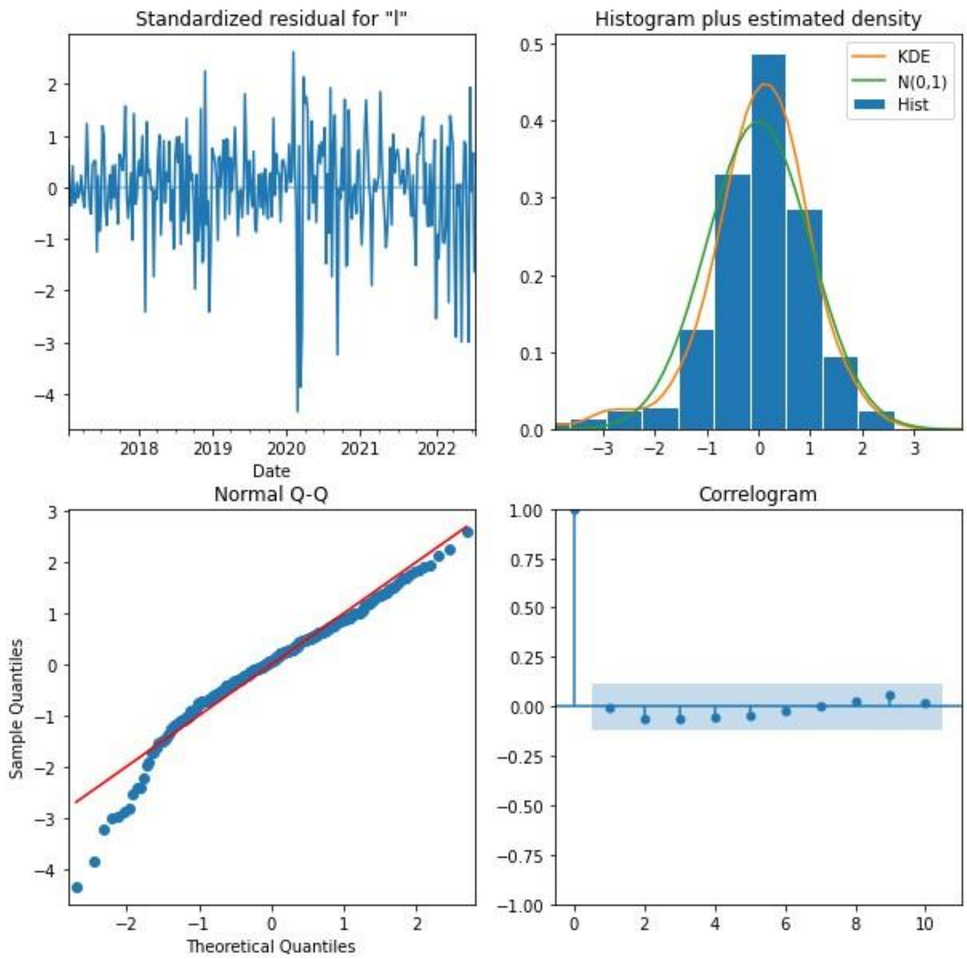


Figure 23 (author's own work)

RMSE value of weekly prediction for 7 next periods is 18.264. For fair comparison, I predict MSFT close prices of next 7 days by using all 'df' dataset, which it was not most efficient dataset for daily prediction. In addition, time intervals of training datasets are equal. However, RMSE value of this prediction is 10.391. Even, accuracy of pred_66 model for next 20 days was 16.746, which is lower than 7 weeks prediction. This difference might come from down-sampling operation, which has negative impact on integrity of data. To illustrate, both prediction graphs are shown in **Figure 24.**

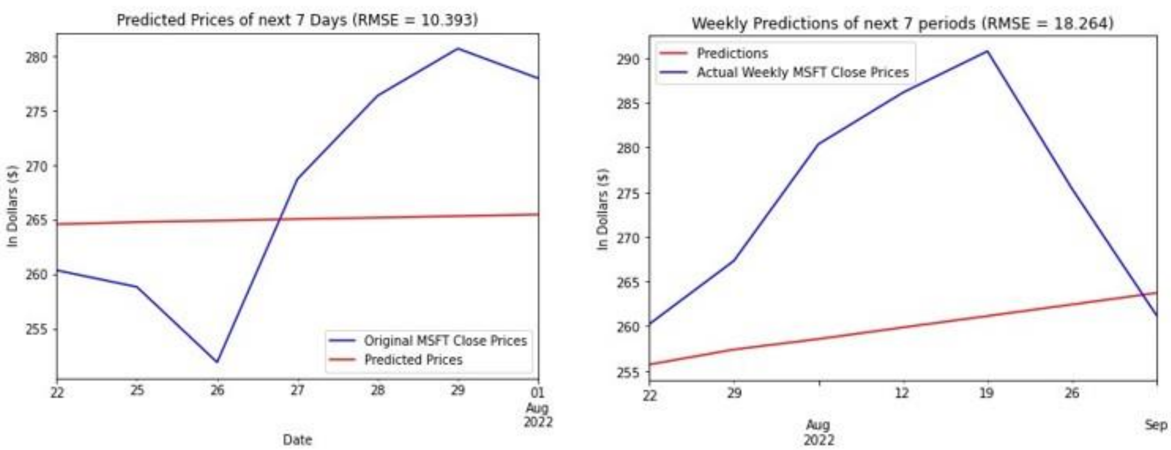


Figure 24 (author's own work)

Weekly one-step-ahead predictions performed better than daily prediction based on weekly test dataset. Correlation between weekly DJIA and MSFT close prices is 0.924. RMSE value of ARIMA(0,1,1) was 4.361, whereas ARIMAX(0,1,1) was 4.270. Both accuracy scores higher than daily one-step-ahead predictions. Weekly average DJIA and MSFT close prices relationship improved accuracy 2%. However, test dataset created by assigning average MSFT close prices of all business days to Friday. If real MSFT price considered for predicted day, which is 22nd of July 2022, RMSE values of weekly ARIMA and ARIMAX models are 4.471 and 4.555, respectively. RMSE value of ARIMAX bigger than ARIMA, but it does not important, because price of specific Friday of predicted week could be close to the average or could not be, which is depended on luck. As it presented, RMSE values according to test dataset, ARIMAX gives more accurate forecast than ARIMA.

5 Conclusion

Findings of this study indicates that accuracy of ARIMA and ARIMAX models increase as the forecasting closer future. In addition, adding exogenous variables to the ARIMA method, increase its accuracy. According to findings, error of one-step-ahead forecast results approximately gather around 4.4 dollars ($), which is relatively good amount to provide supportive argument for investments. These forecasts can be applied iteratively in every period and maintain their accuracy. In addition, new models can be tested and replaced to obtain better results in this process. However, basic ARIMA and ARIMAX models should not be the only consideration for investments. Moreover, these models do not produce good forecasting results during infrequent extreme times, such as pandemic. As a result of several predictions and their comparisons, basic ARIMA and ARIMAX models are reliable to some extent to forecast MSFT close prices.

6 Recommendations

My first recommendation to improve accuracy of ARIMA and ARIMAX models would be increasing number of exogenous variables. All the prediction trials show ARIMAX models have got higher accuracies compared to ARIMA models. Second suggestion is that predictions with created lower frequency datasets by down-sampling method are good indicators as expected value of that period. However, they should be interpreted with predictions from frequency of natural data, because down-sampling operations have got negative impact on quality of the data points. Moreover, since these models are not very effective to predict the dramatic volatile future values, prediction results should be considered alongside worldwide news to increase the chance of preventing any possible risks. Furthermore, these models' prediction ability much better in one-step-ahead predictions rather than long-term predictions. Thus, existing model should predict just next period and its training dataset should be updated if new model is better compared to existing model.

7 References

Andrewes, W. J. H. (2006, February 1). *A chronicle of timekeeping*. Scientific American. Retrieved September 11, 2022, from https://www.scientificamerican.com/article/a-chronicle-of-timekeeping-2006-02/#:~:text=ACCORDING%20TO%20archaeological%20evidence%2C%20the,cycles%20of%20planting%20and%20harvesting.

Arashi, M., & Rounaghi, M. M. (2022). Analysis of market efficiency and fractal feature of NASDAQ Stock Exchange: Time Series Modeling and forecasting of stock index using ARMA-GARCH model. *Future Business Journal*, *8*(1). https://doi.org/10.1186/s43093-022-00125-9

Ausick, P. (2017, March 17). *Apple supports 4.8 million jobs in China, more than double US total*. Yahoo! Finance. Retrieved September 18, 2022, from https://finance.yahoo.com/news/apple-supports-4-8-million-135220928.html

Chen, J. (2022, July 11). *What is the stock market and how does it work?* Investopedia. Retrieved September 12, 2022, from https://www.investopedia.com/terms/s/stockmarket.asp#:~:text=The%20stock%20market%20allows%20buyers,barometer%20for%20the%20overall%20economy.

Curto, J. D., & Serrasqueiro, P. (2022). The impact of COVID-19 on S&P500 Sector Indices and fatang stocks volatility: An expanded APARCH model. *Finance Research Letters*, *46*, 102247. https://doi.org/10.1016/j.frl.2021.102247

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 years of Time Series forecasting. *International Journal of Forecasting*, *22*(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Disney, S.M. and Petropoulos, F., (2015), Forecast combinations using multiple starting points. Logistics and Operations Management Section Annual Conference, 9th January, Cardiff, UK.

Gardner, E. S. (2006). Exponential smoothing: The State of the art—part II. *International Journal of Forecasting*, *22*(4), 637–666. https://doi.org/10.1016/j.ijforecast.2006.03.005

Hall, M. (2022, June 7). *Dow Jones: What it is and why it's important in stock investing*. Investopedia. Retrieved September 18, 2022, from https://www.investopedia.com/ask/answers/who-or-what-is-dow-jones/

Harari, Y. N. (2022). The Secret of Success. In *Sapiens: A brief history of humankind* (pp. 263–264). essay, Signal.

Hayes, A. (2022, August 26). *Time series definition*. Investopedia. Retrieved September 11, 2022, from https://www.investopedia.com/terms/t/timeseries.asp

Hayes, A. (2022, February 18). *Manipulation definition*. Investopedia. Retrieved September 13, 2022, from https://www.investopedia.com/terms/m/manipulation.asp

Ho, S. L., Xie, M., & Goh, T. N. (2002). A comparative study of neural network and Box-Jenkins Arima modeling in time series prediction. *Computers & Industrial Engineering*, *42*(2-4), 371–375. https://doi.org/10.1016/s0360-8352(02)00036-0

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Jenkins, G., & Box, G. (1976). *Time Series Analysis Forecasting and control: Rev. ed*. Holden-Day.

Khairunnisa, S., Sa'dah, N., & Artika, R. (2020). Forecasting and effectiveness analysis of domestic airplane passengers in Yogyakarta Adisutjipto Airport with autoregressive integrated moving average with exogeneous (ARIMAX) model. *Proceeding International Conference on Science and Engineering*, *3*, 365–369.https://doi.org/10.14421/icse.v3.529

Martinis, J., & Boixo, S. (2019, October 23). *Quantum supremacy using a programmable superconducting processor*. Google AI Blog. Retrieved September 13, 2022, from https://ai.googleblog.com/2019/10/quantum-supremacy-using-programmable.html

Mills, T. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Academic Press.

Nedelman, M. (2020, February 22). *February 21, 2020 coronavirus news*. CNN. Retrieved September 20, 2022, from https://edition.cnn.com/asia/live-news/coronavirus-outbreak-02-21-20-intl-hnk/index.html

Nielsen, A. (2019). *Practical time series analysis: Prediction with statistics and machine learning*. O'Reilly Media.

Parmezan, A. R., Souza, V. M. A., & Batista, G. E. A. P. A. (2019). Evaluation of statistical and Machine Learning Models for time series prediction: Identifying the state-of-the-art and the best conditions for the use of each model. *Information Sciences*, *484*, 302–337. https://doi.org/10.1016/j.ins.2019.01.076

Prime Minister's Office, 10 D. S. (2020, March 23). *Prime minister's statement on Coronavirus (COVID-19): 23 March 2020*. GOV.UK. Retrieved September 20, 2022, from https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Sharma, S., Saxena, A. K., & Bansal, M. (2022). Forecasting of GHG (greenhouse gas) emission using (ARIMA) data driven intelligent time series predicting approach. *2022 7th International Conference on Communication and Electronics Systems (ICCES)*. https://doi.org/10.1109/icces54183.2022.9835888

Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*, *6*(3), 324–342. https://doi.org/10.1287/mnsc.6.3.324

Yahoo! (2022, September 18). *Dow Jones industrial average (^DJI;) historical data*. Yahoo! Finance. Retrieved September 18, 2022, from https://finance.yahoo.com/quote/%5EDJI%3B/history?p=%5EDJI%3B

Yahoo! (2022, September 18). *Microsoft Corporation (MSFT) stock historical prices & data*. Yahoo! Finance. Retrieved September 18, 2022, from https://finance.yahoo.com/quote/MSFT/history?p=MSFT

Yule, G. U. (1927). Vii. on a method of investigating periodicities disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, *226*(636-646), 267–298. https://doi.org/10.1098/rsta.1927.0007

8 Appendix 1) Proposal

Outline:

My project I stock market predictions. I have chosen to Microsoft's stock prices for forecasting. I am planning to approach in different ways to predict future stock prices. First way is predicting stock prices from only historical prices of the Microsoft. Second way is to add impact of some stock market indexes, such as Dow Jones Industrial Average index and NASDAQ 100 composite index. Stock markets are important tool for investment since 1611 (Hwang, 2022). In addition, stock markets are beneficial to the companies with the earnings from individual partners from public. It is not possible to exactly forecast the future values. However, we have enough theories and tools to forecast the price range in a short term with no guarantee, because any major incident can happen any time. Thus, businesses could be affected significantly, and prices could go down very fast. According to Mallikarjuna & Rao (2019), there are efficient methods to predict close future prices with least possible error. However, they also argue that, predicting long term future prices nearly impossible due to randomness. If we consider the negative effect of the Covid-19 pandemic on stock markets from many different countries (Ganie et al., 2022), no one could predict the pandemic before.

Data:

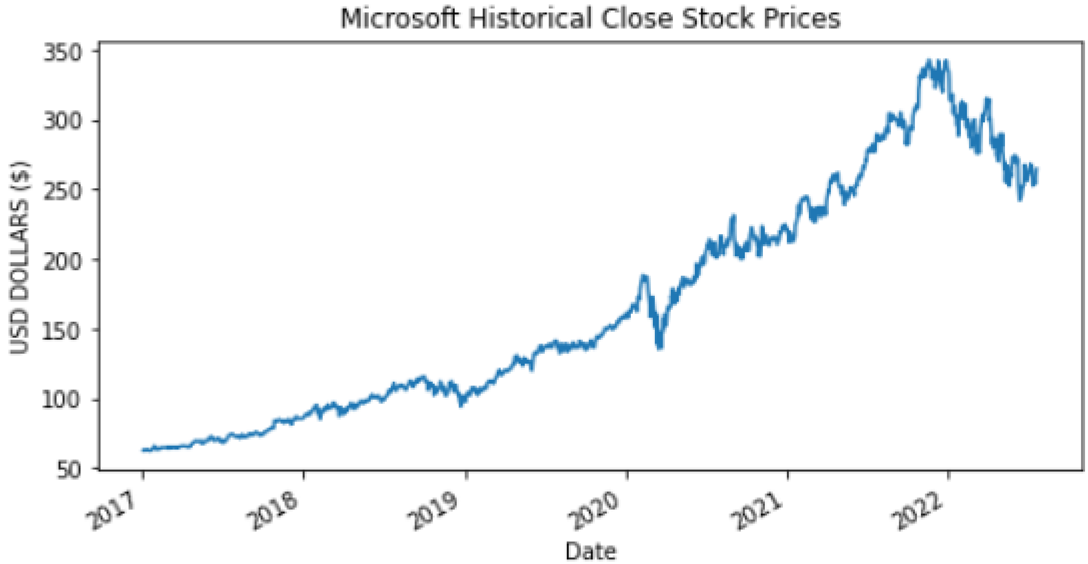
I retrieved the Microsoft's stock market prices, Dow Jones Industrial Average index and Nasdaq 100 composite index from Yahoo Finance, which is an open source, so everyone can use the data. Data starts from beginning of the 2017 to 22/07/2022.

Limitations:

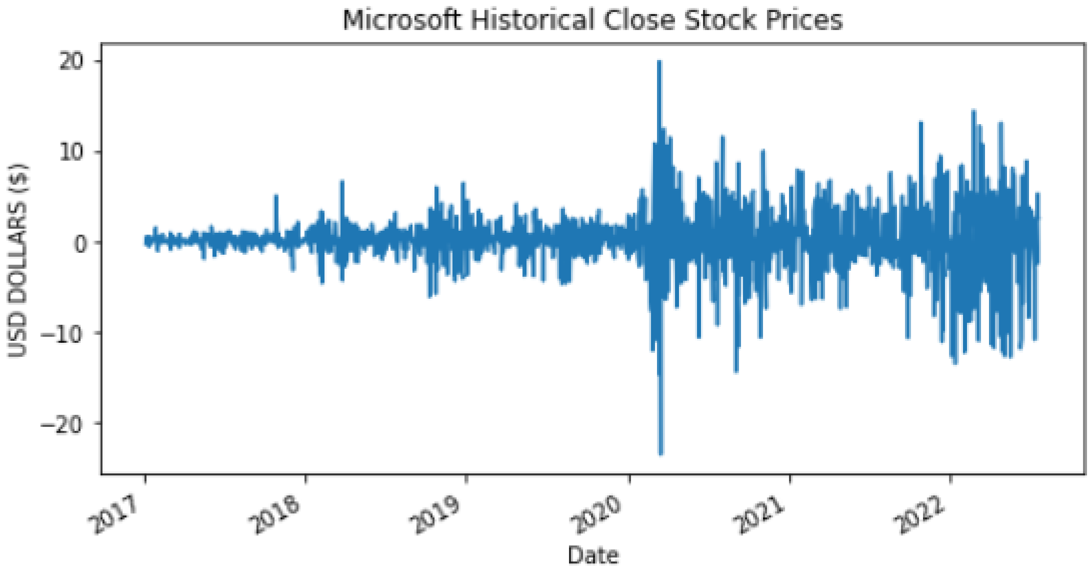
There are few possible limitations with the economical situation of the world, that I am not able to solve it. After the pandemic, many central banks have increased the interest rate, and inflation has been increasing in many countries too (Inman, 2022). In addition to the pandemic, Russia-Ukraine war also has significant effects specially in energy and food industries. Thus, stock prices may vary very quickly. However, with the methods we will use, we are not going to do long term predictions, to find closest predictions more possible in short term predictions.

Methodology:

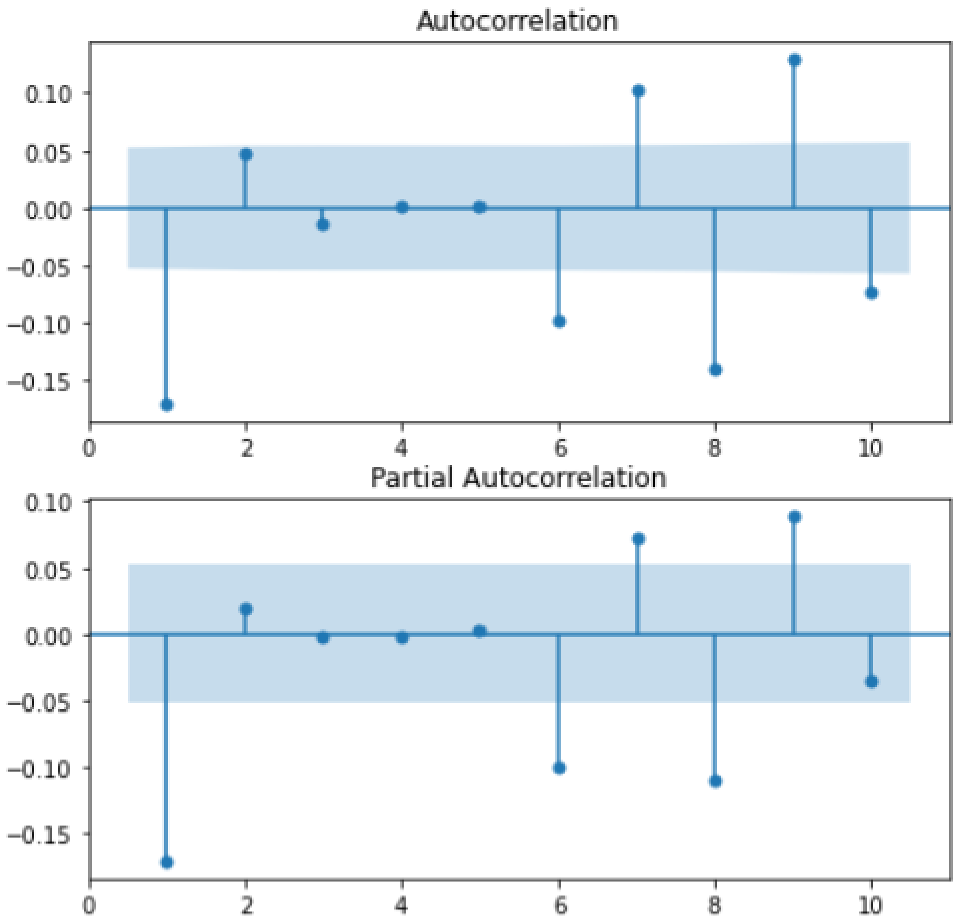
I am planning to use ARIMA and ARIMAX models for my project. I am going to use ARIMA for prediction from just Microsoft historical share prices. In addition, I am planning to use ARIMAX time series analysis, which provide us to add explanatory variables such as Dow Jones Industrial Average. To apply ARIMA models, we need to check our dataset to check if the timeseries of the data meet some conditions. First we need to check if the timeseries of the Microsoft's share price data stationary or not. We cannot apply ARIMA model if the timeseries non-stationary. To check if it is stationary first we can look at if there is any trend on the data.



As we can see from the graph, this timeseries has clear trend. Thus, it is non-stationary timeseries, we need to convert it to the stationary data. To accomplish this, there are many approaches, such as taking root square of the values, extracting differentiated values, and converting data to log values. If there is no trend, we can check the white noise, which considers if the data has many outliers or not. Once we checked these conditions, we can apply augmented Dicky-Fuller (ADF) test. If the test statistic negative number and the p is less than 0.5, we can reject the null hypothesis, which is timeseries is stationary. When we applied this test to the natural version of Microsoft stock price data, test statistic is -0.741 and p value is 0.835, which is way higher than 0.5. When I applied diff method in python our dataset converted to this:



When we look at this graph, no clear trend seen. However, around march of 2020 we can se few white noises for some periods (days). When I applied ADF test, I found that, test statistic is -12.7 and p value is a very close number to the 0. This version of the timeseries has passed from ADF test but need to do some research to understands If the white noise is acceptable or not. After reaching the stationary data, next step is to specify order for our ARIMA model. Finding right order is crucial for our model's ability to predict future prices. To determine order, we need to look at 4 different terms, which are Autocorrelation Function (ACF), partial autocorrelation function (PACF), Akaike information criterion (AIC), and Bayesian information criterion (BIC). Plots of ACF and PACF are down below until 10 lags (10 days before).

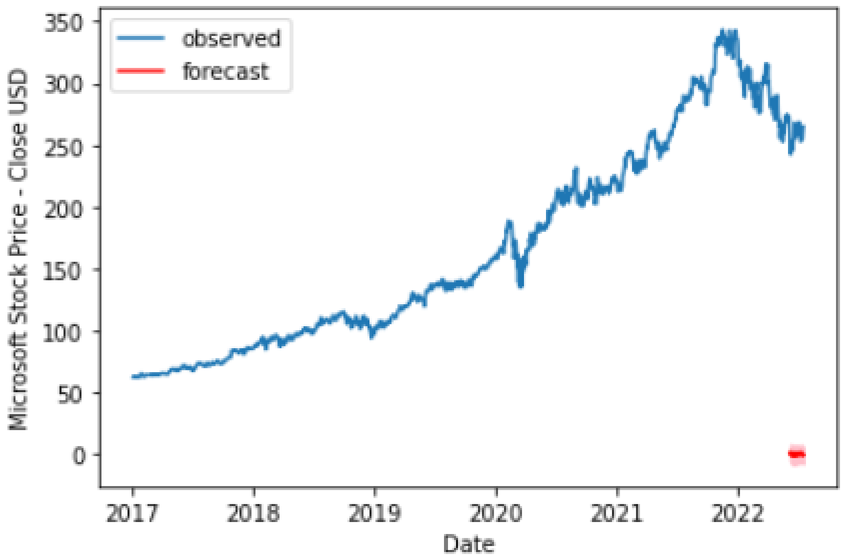


I am not capable of completely reading this graph, but it does not seem normal since critical test statistic values are not in order. Thus, we can look at AIC and BIC values.

		AIC VALUES	BIC VALUES
0	0	8557.820638005302	8563.097194408021
0	1	7360.437472748776	7370.990585554213
0	2	7320.978212365564	7336.80788157372
0	3	7321.514596586001	7342.620822196875
1	0	7958.8384970077295	7969.391609813167
1	1	7319.65134538834	7335.481014596496
1	2	7321.64153384879	7342.747759459664

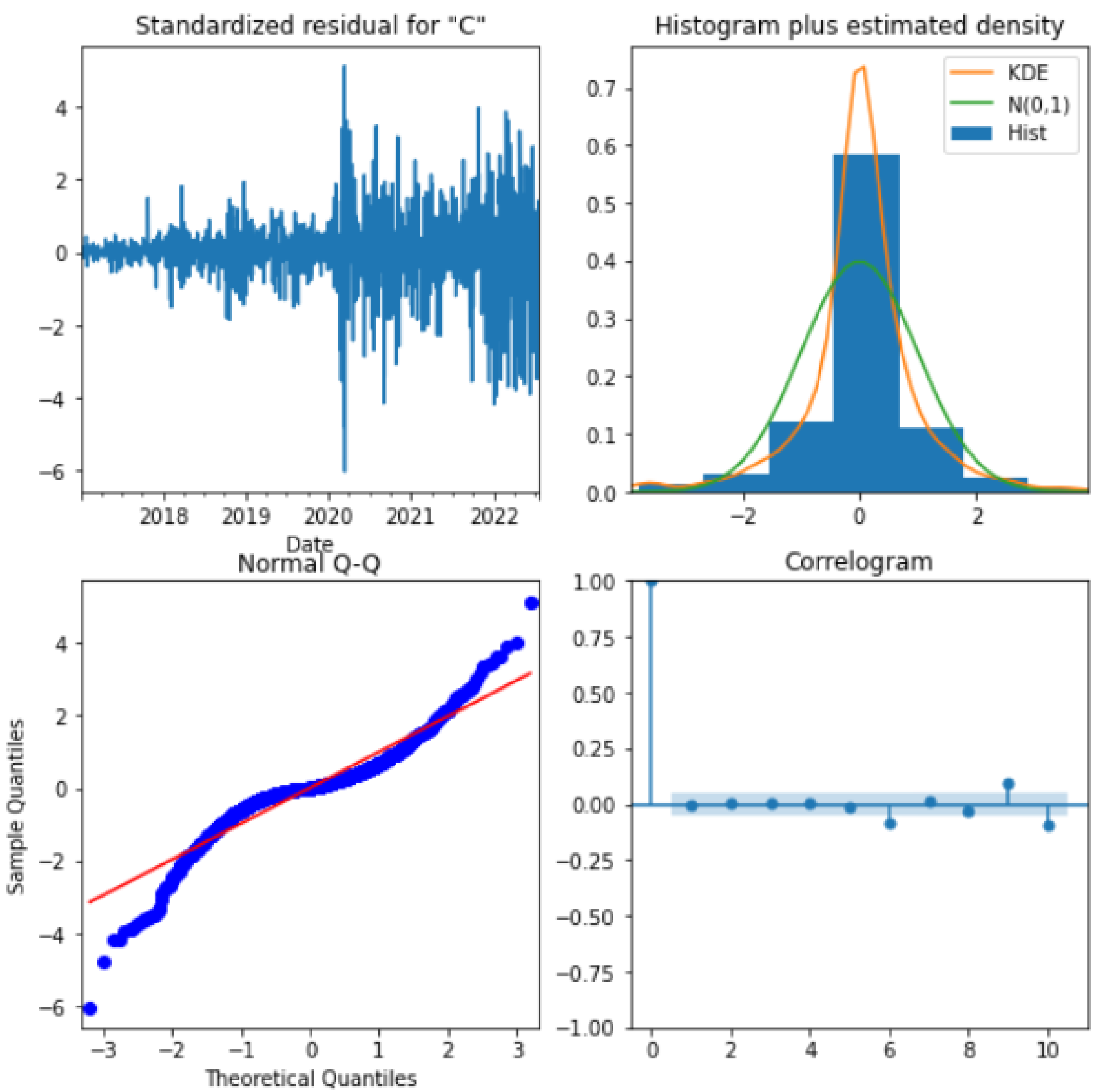
```
1 3 7319.621494707537 7346.004276721131
2 0 7750.304561328342 7766.134230536498
2 1 7321.641395309685 7342.74762092056
2 2 7317.14245369236 7343.52523570593
2 3 7315.21961089087 7346.878949307183
3 0 7641.140986312048 7662.247211922922
3 1 7323.529259228211 7349.912041241805

3 2 7318.966811350145 7350.626149766457
3 3 7311.160243230981 7348.096138050012
```

Normally, we should have chosen the order with least AIC and BIC values, which is (3,1) for AIC value and (1,1) for BIC value. Price unit need to convert to another form, which I am going to look at it, but for now to show the rest of the steps we can choose order of the model (3,1).

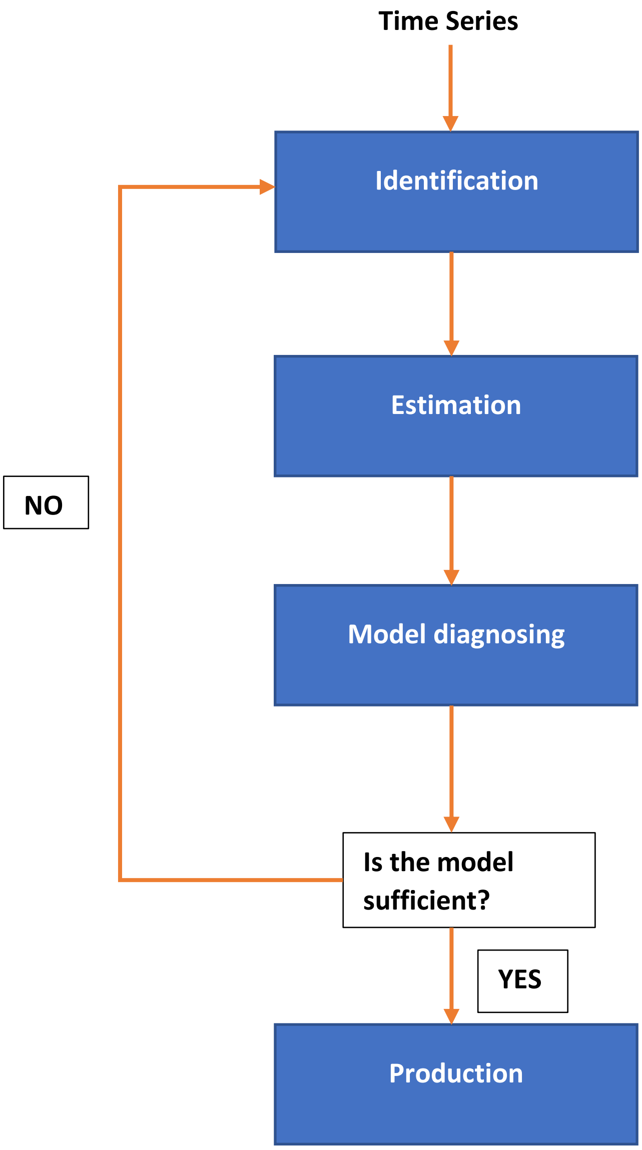


As we can see from the graph, I could not fit the model. The reason of it, first order diff() method not sufficient to make the timeseries predictable. If this forecast would be seemed sufficient, I would pass into the next stage which is inspecting the performance of the module.

As we can see from the graph down below, residuals' normal distribution does not match with its partner. In addition, from Q-Q test we can tell that blue line does not follow the red line, which it means our model should fail to predict future values which it does. Q-Q plots shows after histogram plus density graph, residuals are not normal distributed. Since we have very poor performance with this model, we must change it.



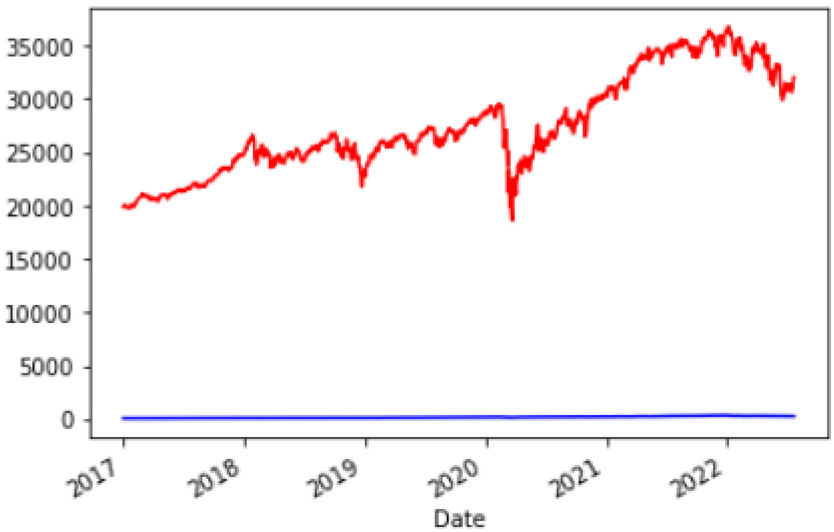
Second part of my project is to create ARIMAX model to add Dow Jones Industrial Average to enhance the model and the predictions. I am following steps of Box Jenkins methodology, which is special methodology for timeseries analysis (Brownlee, 2017).

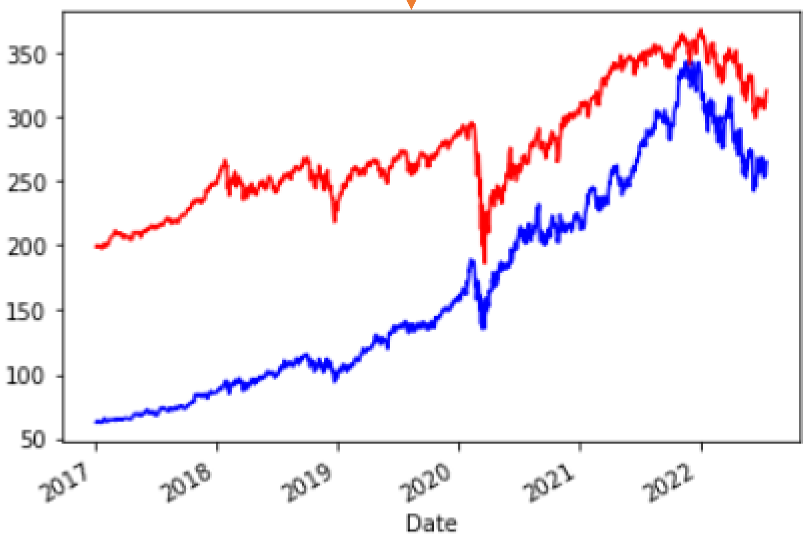


In this case, according to Box Jenkins, I need to return start.

As a second part of my project, once I found a good fit with the model, I will add Dow Jones Industrial Average index to the prediction by the ARIMAX type model, which is going to enhance the model to predict future values.

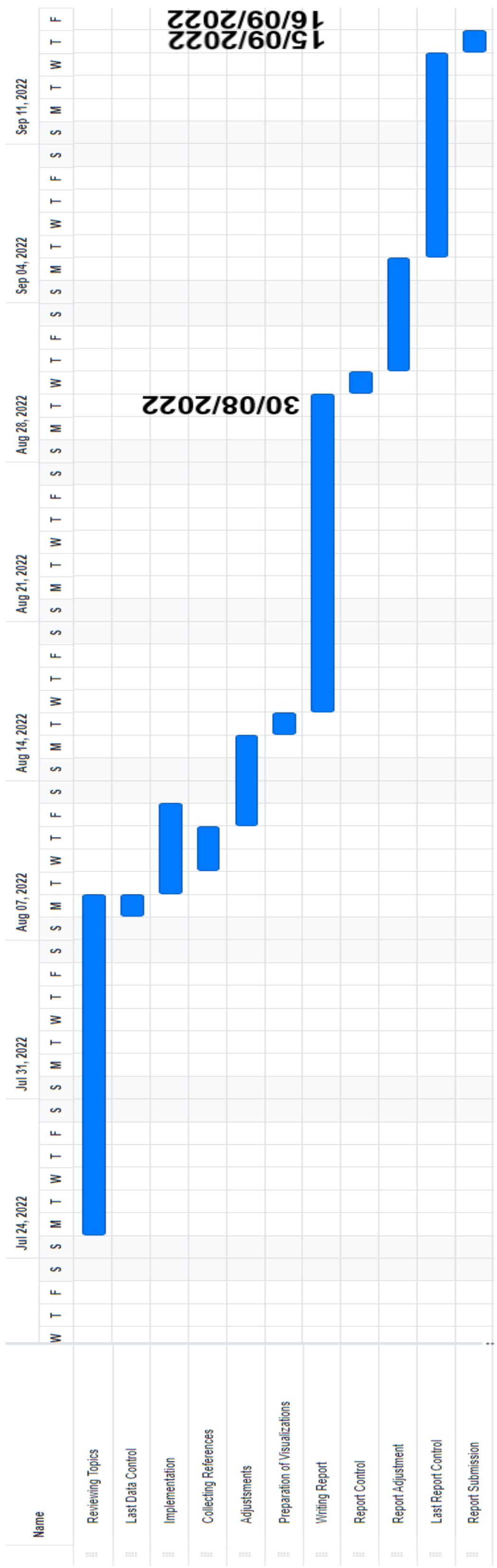
We can see the closing prices of Microsoft and the Dow Jones index down below.





If we look at the second graph shown above, we can see the similarity in the pattern of data. Red shows Dow Jones average index and blue shows Microsoft's stock prices. In the first graph since average index far way higher than Microsoft's stock share prices, I divided the Dow Jones average by 100 and reached the second graph. If after I just divide the average index by 100 and see the similarities between two patterns, I can improve the fit more by finding best conversion formula of the average index.

Gantt Chart of the Project:



References

Brownlee, J. (2017, January 13). *A gentle introduction to the box-jenkins method for time series forecasting*. Machine Learning Mastery. Retrieved July 27, 2022, from https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/

Ganie, I. R., Wani, T. A., & Yadav, M. P. (2022). Impact of covid-19 outbreak on the stock market: An evidence from select economies. *Business Perspectives and Research*, 227853372110736. https://doi.org/10.1177/22785337211073635

Hwang, I. (2022, July 11). *A brief history of the stock market*. A Brief History of the Stock Market. Retrieved July 27, 2022, from https://www.sofi.com/learn/content/history-of-the-stock-market/

Inman, P. (2022, June 18). *We face a global economic crisis. and no one knows what to do about it*. The Guardian. Retrieved July 27, 2022, from https://www.theguardian.com/business/2022/jun/18/we-face-a-global-economic-crisis-and-no-one-knows-what-to-do-about-it

9 Appendix 2) Codes

I have sent the codes to my supervisor with e-mail.

10 Executive Summary

Introduction

Autoregressive integrated moving average (ARIMA) is a well-known time series forecast method, which can be used in many fields, such as health sector, finance, military, and public management. This method consists of 3 parts which is mentioned. Because of statistical and mathematical reasons Autoregressive (AR) models requires **stationary** time series, which has constant mean and finite variance. Based on conditions of stationarity, AR component forecasts future values from the significant relationships with previous values. Second component is number of times needed differencing to convert non-stationary data to the stationary. Last component is moving average (MA). When AR model produce prediction from lagged values, moving average utilizes from past errors and mean of the data. ARIMA model with exogenous variables is called ARIMAX model, which is type of regression. However, instead of direct affect, these variables affect the coefficients of AR and MA models. **This report examined to what extent basic ARIMA and ARIMAX models are reliable to forecast Microsoft's close prices**.

Stock markets are significant indicators for how well-being economies all around the world. Thus, Dow Jones industrial average index (DJIA) used as exogenous variable for predicting MSFT close price. Individuals use stock market to increase their money or preventing it to value loss of their money against inflation. For this matter, having high accuracy of predictions are significant, which it emphasises importance of ARIMA models. I used python software language to apply predictions with ARIMA and ARIMAX models.

Methodology

Box-Jenkins method (Jenkins & Box, 1976) is very famous and one of the best iterative processes for ARIMA and ARIMAX models. This procedure consists of 3 parts. First part is identification process that stationary of data is checked here by the augment dicky-fuller (ADF) test. If data is not stationary, necessary transformations can be applied. Order of the model determined by the visual help of autocorrelation (ACF) and partial autocorrelation (PACF) graphs. Second step is estimation section. In this section, 'SARIMAX' function is used in this step to create model and then model is fitted. From fitted data variable, AIC and BIC scores can be reached. Order with lowest of these scores is chosen for prediction model alongside with the ACF and PACF graphs. Last and third sections is diagnosing models with graphs. In these graphs analysts trying to get white noise futures, as residuals cannot predicted part in time series. If any steps got failed, analysist should start from the beginning. Lastly to measure accuracy of the model, root mean square error (RMSE).

Findings

In this section first prediction will be presented with its model diagnostics graphs, graph of predictions and RMSE values for first prediction. Result of other predictions will be presented as summary.

Predictions with Business Daily Frequency

This finding obtained by creating largest first 66 dramatically changes and then trained 66 different data at those turning points. All models suggested ARIMA(1,1,0) model. Result graph can be seen in **Figure 25**.

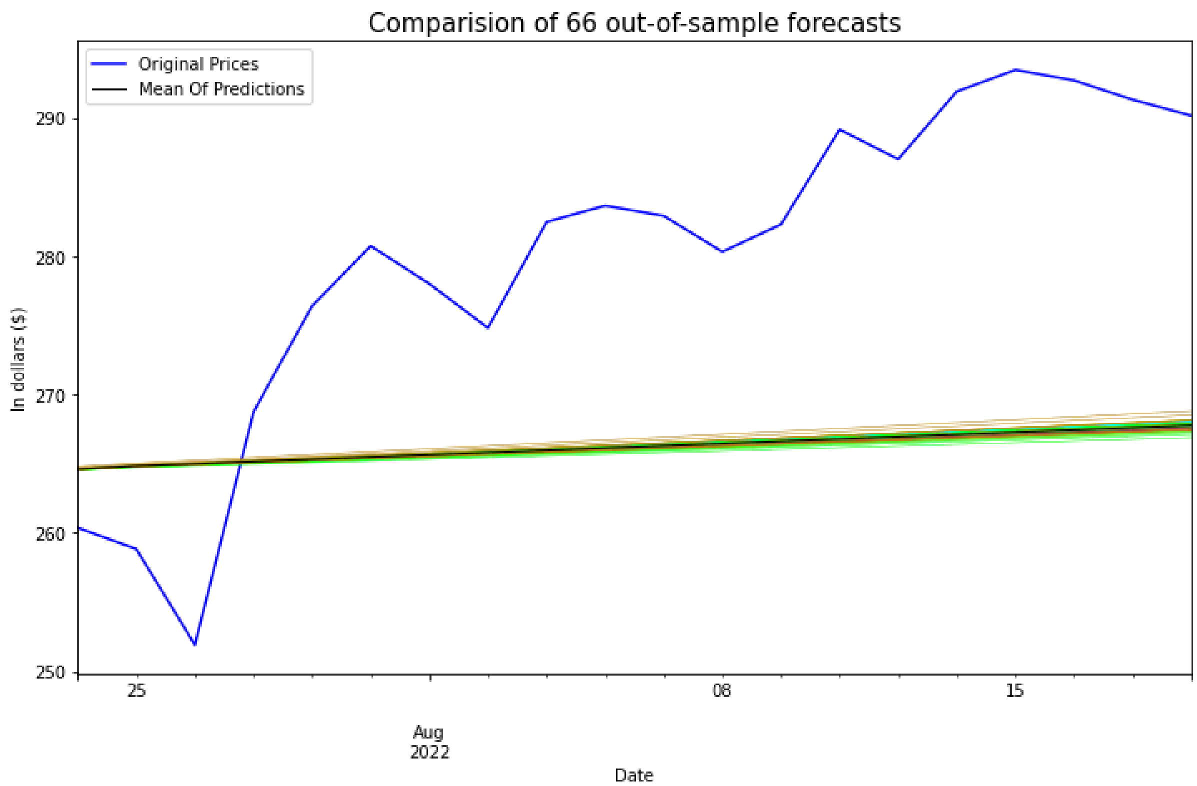


Figure 25 (author's own work)

Best model here pred_66 which its train dataset started 16.03.2020. Disney and Petropolus (2015) argued that if the structure of the out-of-sample forecasts does not vary, prediction accuracy should be high. Diagnosing graph of pred_66 shown in **Figure 26.** Only problem is in the correlogram at lags 6 and 10.

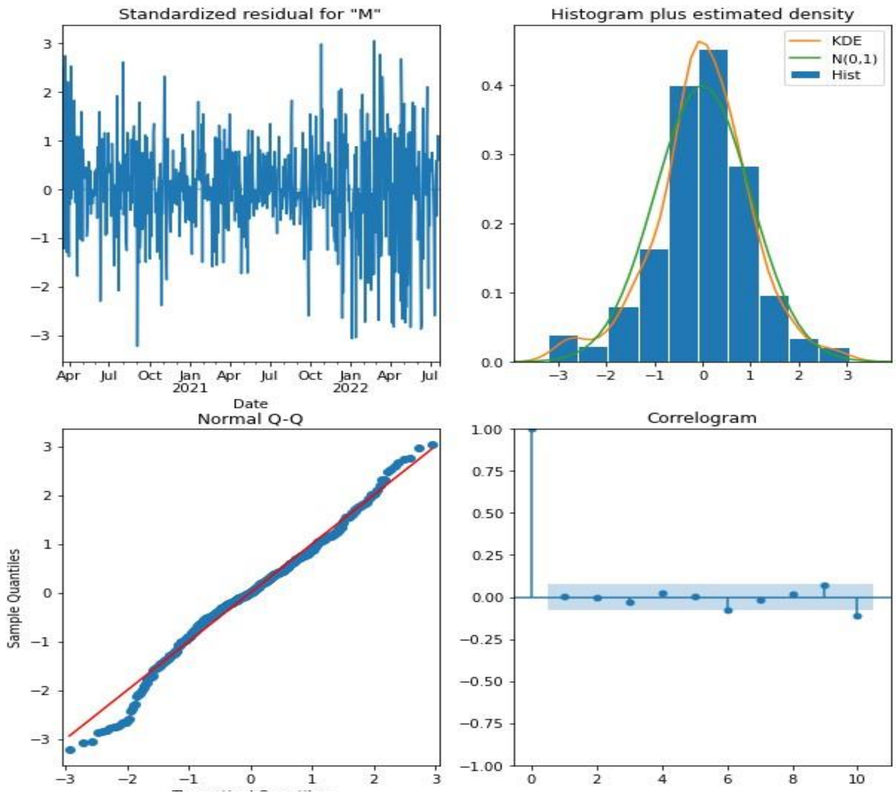


Figure 26 (author's own work)

However, these statistically significant values can be avoided since they are close to the 0. Accuracy of pred_66 model is 16.746 as it can be seen **Figure 27**.

	model	rmse
65	pred_66	16.746724
63	pred_64	16.929383
62	pred_63	17.140618
33	pred_34	17.143329
61	pred_62	17.149684

Correlation between MSFT prices and 1 period shifted DJIA Close values is 0.937. One-step-ahead prediction with ARIMAX version of pred_66's RMSE value is 4.426, when ARIMA's RMSE is 4.449. This difference shows that by adding DJIA, model's accuracy increases 3.6%.

Figure 27 (author's own work)

<u>Predictions with Business Daily Frequency for Beginning of Covid-19</u>

Starting dates of the first 5 models during March of 2020 which is early pandemic except pred_34. Thus, training data without Covid-19 fluctuations would be better trained model. To test this argument, I predicted MSFT prices of 21 days during early pandemic times. RMSE value of 21 days prediction during post-pandemic from pred_66 is 16.765. However, RMSE value of early-pandemic is 30.225.

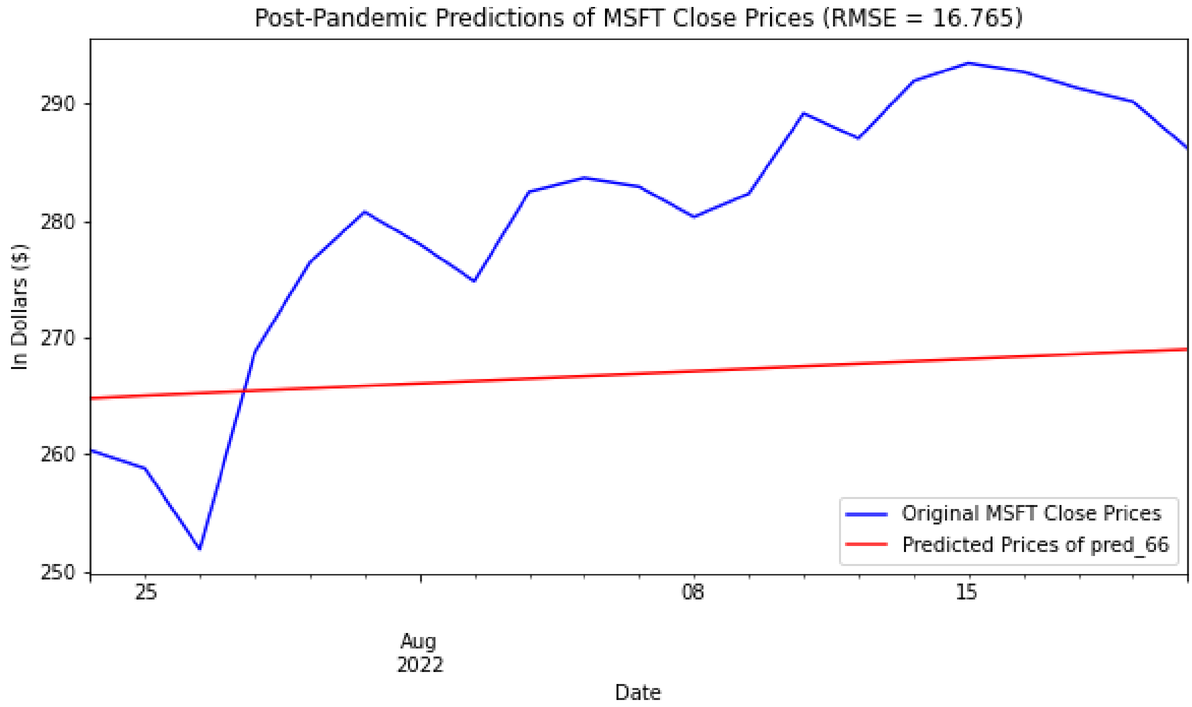


Figure 28 (author's own work)

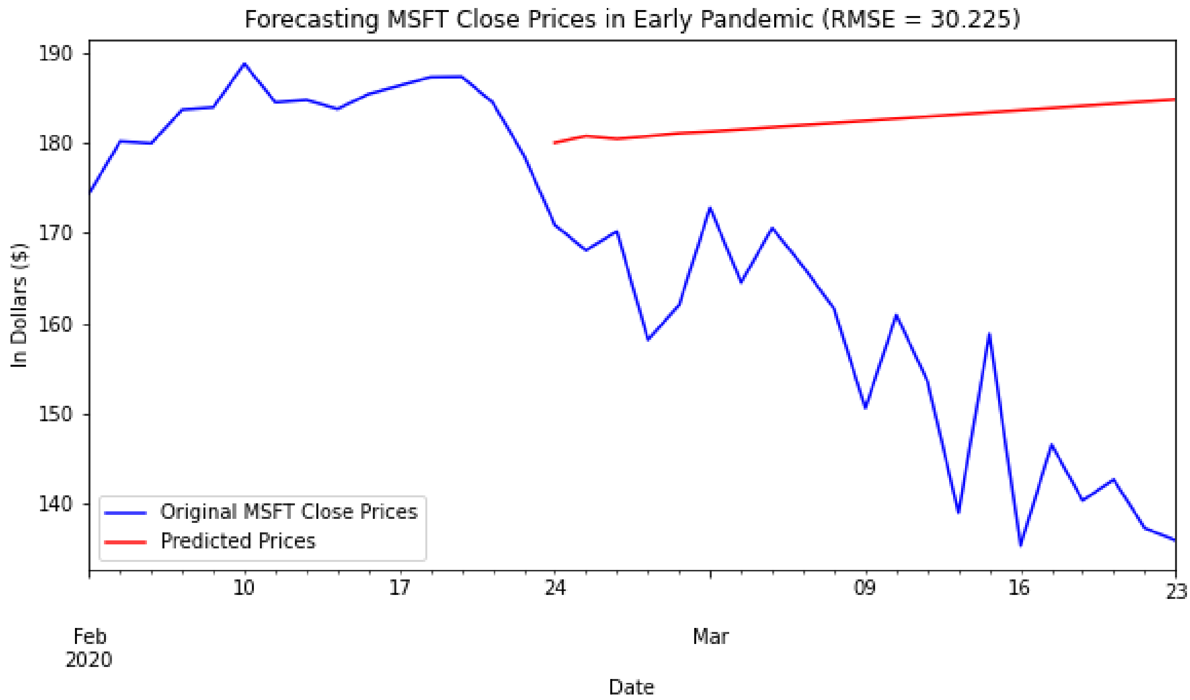


Figure 29 (author's own work)

<u>Predictions with Business Weekly Frequency</u>

Training weekly data affected negatively by down-sampling operation. Difference between weekly and daily data shown in **Figure 30**.

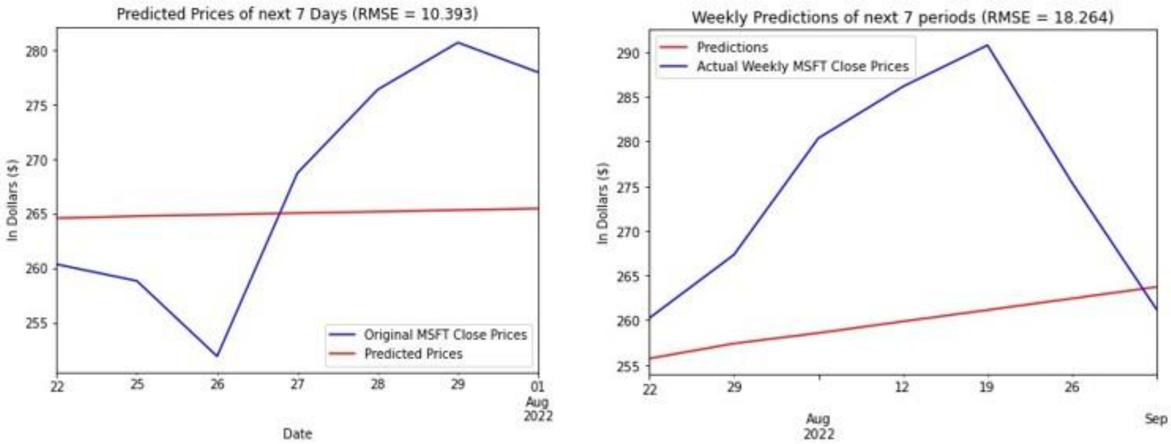


Figure 30 (author's own work)

Correlation between weekly DJIA and MSFT close prices is 0.924. One-step-ahead predictions of weekly trained dataset more successful than daily trained data. However, this case just for the compared to the test datasets. Weekly test dataset created by calculating average price of the week. When RMSE values with original prices compared, predictions of weekly trained dataset produced less accurate model.

Conclusion

Findings show that, ARIMA and ARIMAX models perform better results as horizon of forecast decrease. Moreover, in trials, ARIMAX always gave more accurate results with one-step-ahead prediction and RMSE values of these forecasts close to 4.4 Dollars ($). Any further developments could not be achieved. Individuals should not make any investments by just relying on predictions with 4.4 Dollar ($) error. However, these forecasts give strong intuition about the level of future prices, which it can be useful, when it combined with another analysis, such as following news. Furthermore, performance of these models was tested in early pandemic periods, which is March of 2020. When RMSE values of ARIMA is 16.765 for post pandemic days, same accuracy measurement is 30.225 in early pandemic due to lack of ability to forecast high volatility time series. To conclude, basic ARIMA and ARIMAX models are reliable to some extent.

Recommendation

There are many approaches to improve models to obtain more accurate results. First, observing and understanding raw data is very important to anticipate effects of any kind of manipulations on training data set or on the model. Goal should not just be got lower RMSE values. Because, with a little luck even worst models can predict very close values. Fellow analysists should use iterative one-step-ahead prediction, which is a constant prediction process that requires changes in the model and in the training dataset. When ARIMA models applied both lower and higher frequency of data, closer frequency to the raw data should be considered first. Other frequency can be used for increasing intuition of the data. However, Operations like down-sampling corrupts the integrity of data. Furthermore, all ARIMA forecast results should be considered with 'What is going on in the World' question to prevent any dramatic failure of ARIMA, when some extreme scenario becomes real. Lastly, some manipulation techniques such as box-cox transforming very important to have for an analyst who use ARIMA. If with these kinds of transformations convert stationary time series to be more normally distributed stationary processes, performance of the prediction would be increase.