Shamim Sherafati

Forecasting Financial Time Series

Instructor: Soheil Parsa

ALY 6050_ Module 3 Project Report

Date: 2023-03-16

Introduction

The goal of this project is to use time series analysis techniques to create a forecasting model for two stocks, TSLA and AMZN. The objective is to develop a model that correctly forecasts future stock values so that investors may make wise selections. To do this, we created forecasting models utilising historical stock price data for TSLA and AMZN, which we gathered from Yahoo Finance for the last 12 months using Python in a Jupyter notebook (252 market days). To build these models, we employed a number of techniques, such as moving average, exponential smoothing, and ARIMA.

We compare the accuracy of each model and share the findings of our investigation in this project. We also go through each model's Mean Absolute Percentage Error (MAPE) numbers and contrast them with the MAPE values we got from the forecasting models created in project part 1 to see how they stack up.

This Project is done with Python.

Part 1: Short-term Forecasting:

(i) Use a simple line plot of both time series to detect seasonal, irregular, or trend behaviours if any. Write a summary of your observations of both time series in your report.

I imported the required libraries, yfinance, pandas, and matplotlib.pyplot, after first installing the yfinance library, which is a Python package for obtaining historical market data from Yahoo Finance.

```
!pip install yfinance

import yfinance as yf
import pandas as pd
```

Then I created the ticker symbols for Tesla and Amazon as well as the beginning and ending dates for the historical data.

```
# Define the ticker symbols for TESLA Inc and Amazon Inc
tickers = ['TSLA', 'AMZN']

# Define the start and end dates for the historical data
start_date = pd.Timestamp.now() - pd.Timedelta(days=252)
end_date = pd.Timestamp.now()
```

I used the yf.download() function to get the historical stock prices for the tickers and saved them in a pandas DataFrame named data. The outcome is displayed below:

```
[*********************100%***********************]  2 of 2 completed
             Adj Close                   Close                    High  \
                  AMZN        TSLA        AMZN        TSLA        AMZN
Date
2022-07-05  113.500000  233.066666  113.500000  233.066666  114.080002
2022-07-06  114.330002  231.733337  114.330002  231.733337  115.480003
2022-07-07  116.330002  244.543335  116.330002  244.543335  116.989998
2022-07-08  115.540001  250.763336  115.540001  250.763336  116.580002
2022-07-11  111.750000  234.343338  111.750000  234.343338  114.300003
...                ...         ...         ...         ...         ...
2023-03-06   93.750000  193.809998   93.750000  193.809998   96.550003
2023-03-07   93.550003  187.710007   93.550003  187.710007   95.089996
2023-03-08   93.919998  182.000000   93.919998  182.000000   94.169998
2023-03-09   92.250000  172.919998   92.250000  172.919998   96.209999
2023-03-10   90.730003  173.440002   90.730003  173.440002   93.570000

                               Low                    Open              \
                  TSLA        AMZN        TSLA        AMZN        TSLA
Date
2022-07-05  233.146667  106.320000  216.166672  107.599998  223.000000
2022-07-06  234.563339  112.010002  227.186661  113.209999  230.779999
2022-07-07  245.363327  113.489998  232.210007  113.849998  233.919998
2022-07-08  254.979996  113.690002  241.160004  114.599998  242.333328
2022-07-11  253.063339  110.870003  233.626663  114.080002  252.103333
...                ...         ...         ...         ...         ...
2023-03-06  198.600006   93.739998  192.300003   95.190002  198.539993
2023-03-07  194.199997   92.779999  186.100006   94.059998  191.380005
2023-03-08  186.500000   92.180000  180.000000   93.599998  185.039993
2023-03-09  185.179993   92.180000  172.509995   93.680000  180.250000
2023-03-10  178.289993   90.250000  168.440002   92.669998  175.130005
```

```
              Volume
                AMZN       TSLA
Date
2022-07-05  76583700   84581100
2022-07-06  66958900   71853600
2022-07-07  57872300   81930600
2022-07-08  45719700  101854200
2022-07-11  53487600   99241200
...              ...        ...
2023-03-06  52112400  128100100
2023-03-07  49100700  148125800
2023-03-08  44899100  151897800
2023-03-09  56218700  170023800
2023-03-10  69747500  191007900

[173 rows x 12 columns]
```

Then, I got the plot of closing prices of Tesla and Amazon which as below:

Due to its size and greater level of diversification, Amazon has a more unpredictable pattern than Tesla. Amazon may not have a general increasing or decreasing trend, but it may have price swings over time, which may be a sign of seasonal or irregular activity. For example , the price rose from about 105 to a peak of 120 in 2022-07 and then down to approximately 90 in 2022-08 and 2022-09. There are several potential causes for this, including changes in consumer demand or swings in the economy. As a result, even if Amazon may not have a distinct trend, it is still possible to identify patterns in the data. Various factors, including its online retail operation, Amazon Web Services, and other projects, have an impact on Amazon's stock price. This puts it at greater risk to outside economic forces and shifts in market sentiment, which can lead to stock price variations that are more unexpected.

Additionally, short-term events like shifts in consumer purchasing habits, the introduction of new goods or services, or changes in commodity prices may have an impact on Amazon's stock price. The stock price may see sudden spikes or declines as a result of these short-term occurrences, creating an overall trend that is more unstable. In comparison, Tesla is a more recent, narrow and specific corporation with a more dependable business plan that is focused on electric cars and renewable energy. A more steady and predictable development in its stock price may result from this approach.

On the other hand, based on the plot, we can see that the price of Tesla's shares increased significantly between August and October of the mid-year, peaking at almost $300. Following that, it fell sharply before holding steady around $100 until the end of January. It then continued to rise once again, reaching almost $200 by the end of March. As a result, we may conclude that Tesla's stock price is more erratic than Amazon's stock price and does not exhibit a smooth ascending pattern.

(ii) Perform exponential smoothing to forecast both prices for period 253. Use successive values of 0.15, 0,45, 0.55, and 0.75 for the smoothing parameter α. Next, calculate the MAPD (Mean Absolute Percentage Deviation) of each forecast; and based on the MAPDs, determine the value of α that has yielded the most accurate forecast for each stock. In your report, describe your results; and explain why in your opinion such values of α have yielded the most accurate forecasts for the two stocks.

The MAPD (Mean Absolute Percentage Deviation) between the actual and predicted values is the first thing I've defined. After that, specify the start and end dates for the historical data as well as a list of ticker symbols. I have used these inputs to use the yf.download function to obtain the historical data for the tickers from Yahoo Finance. Next, I calculated the forecast using the ewm function in Pandas with the specified alpha value for each ticker and each value of alpha in the list alphas (0.15, 0.45, 0.55, 0.75), for each ticker and each alpha value. Finally, I calculated the MAPD between the forecast and the actual closing price for the last day of available data.

```python
# Part 1 (ii)

# Define a function to calculate the MAPD
def calculate_mapd(actual, forecast):
    mapd = (abs(actual - forecast) / actual).mean() * 100
    return mapd


# Define the ticker symbols for TESLA Inc and Amazon Inc
tickers = ['TSLA', 'AMZN']

# Define the start and end dates for the historical data
start_date = pd.Timestamp.now() - pd.Timedelta(days=253)
end_date = pd.Timestamp.now()

# Retrieve the historical data for the tickers
data = yf.download(tickers, start=start_date, end=end_date)

# Calculate the exponential smoothing forecast for each ticker using different values of alpha
alphas = [0.15, 0.45, 0.55, 0.75]
for ticker in tickers:
    print(f"\n{ticker} Forecast:")
    best_alpha = None
    best_mapd = None
    for alpha in alphas:
        forecast = data['Close'][ticker].ewm(alpha=alpha, adjust=False).mean().iloc[-1]
        actual = data['Close'][ticker].iloc[-1]
        mapd = calculate_mapd(actual, forecast)
        print(f"Alpha = {alpha:.2f}, Forecast = ${forecast:.2f}, MAPD = {mapd:.2f}%")
        if best_mapd is None or mapd < best_mapd:
            best_alpha = alpha
            best_mapd = mapd
    print(f"\nBest Alpha for {ticker}: {best_alpha:.2f}, Best MAPD = {best_mapd:.2f}%")
```

Here is its result:

```
[*********************100%***********************]  2 of 2 completed

TSLA Forecast:
Alpha = 0.15, Forecast = $188.54, MAPD = 8.70%
Alpha = 0.45, Forecast = $177.61, MAPD = 2.41%
Alpha = 0.55, Forecast = $175.88, MAPD = 1.41%
Alpha = 0.75, Forecast = $173.99, MAPD = 0.32%

Best Alpha for TSLA: 0.75, Best MAPD = 0.32%

AMZN Forecast:
Alpha = 0.15, Forecast = $93.74, MAPD = 3.31%
Alpha = 0.45, Forecast = $92.03, MAPD = 1.43%
Alpha = 0.55, Forecast = $91.73, MAPD = 1.10%
Alpha = 0.75, Forecast = $91.21, MAPD = 0.53%

Best Alpha for AMZN: 0.75, Best MAPD = 0.53%
```

The best alpha value for TSLA and AMZN, respectively, is 0.75, the data indicates. These alpha values produced the lowest MAPD values, a measure of prediction accuracy, for the two stocks, indicating that they produced the most accurate forecasts. For TSLA, alpha values of 0.15, 0.45, and 0.55 produced generally higher MAPD values of 8.70%, 2.41%, and 1.41%, respectively, while alpha value of 0.75 produced the lowest MAPD value of 0.32%. This indicates that greater weight should be given to the most recent observations when making a prediction for the price of TSLA's stock; as a result, a bigger value of alpha (0.75) results in a more accurate prediction.

The MAPD values for AMZN were 3.31%, 1.43%, and 1.10%, respectively, for alpha values of 0.15, 0.45, and 0.55, whereas the MAPD value for AMZN with the lowest alpha value was 0.53%. This is consistent with the outcome for TSLA and shows that a bigger weight should likewise be given to the most recent observations when determining the estimate for AMZN's stock price.

The reason why bigger values of alpha (0.75) have produced more precise forecasts for both stocks, in my opinion, is that these values give more weight to the most recent observations, which may be more pertinent and instructive in predicting the future values of the stock prices. A better forecast of future prices may be possible with the use of more recent data, which may better reflect current market patterns and other significant events that might affect stock values. The best alpha value, however, could vary depending on the specific traits of each company and the timelines being considered.

(iii) Use your exponential smoothing forecast of part (ii) with $\boldsymbol{\alpha}=\mathbf{0}.55$ and perform an adjusted exponential smoothing to forecast both prices for period 253. Use successive values of 0.15, 0.25, 0.45, and 0.85 for the trend parameters β for both stocks. Next, calculate the MAPEs (Mean Absolute Percentage Error) of your forecasts and determine the values of β that have provided the most accurate forecasts for both stocks. In your report, describe your results and explain why, in your opinion, such values of β have yielded the most accurate forecasts.

In this section, I predict Tesla (TSLA) and Amazon (AMZN) stock values for period 253 using exponential smoothing and modified exponential smoothing. In the modified exponential smoothing model, we change the trend parameter and compute the Mean Absolute Percentage Error (MAPE) for each value of. The values of that have produced the most precise predictions for both equities are then identified.

I didn't add 253 days to the code since the code already uses the most recent available data point as the beginning point for forecasting, as was noted in the question when utilising the exponential smoothing prediction from section (ii) with a value of =0.55 to anticipate both the price of TSLA and AMZN. Therefore, it will automatically forecast the following number in the series, which will be the forecast for period 253. It came to this:

```
TSLA Results:
Beta = 0.15, MAPE = 1.09%
Beta = 0.25, MAPE = 2.28%
Beta = 0.45, MAPE = 3.49%
Beta = 0.85, MAPE = 4.07%
Best Beta for TSLA: 0.15, Best MAPE = 1.09%

AMZN Results:
Beta = 0.15, MAPE = 0.47%
Beta = 0.25, MAPE = 0.50%
Beta = 0.45, MAPE = 0.17%
Beta = 0.85, MAPE = 0.90%
Best Beta for AMZN: 0.45, Best MAPE = 0.17%
```

The outcomes display the MAPE values for various beta values applied to the modified exponential smoothing to predict stock prices over the next 253 days. Better forecast accuracy is indicated by a lower MAPE number.

The best beta for TSLA is 0.15, which has a MAPE value of 1.09%. As a result, projections with a lower beta value have shown to be more accurate and show that the trend component of the TSLA stock price is generally steady.
The best beta for AMZN is 0.45, with a MAPE value of 0.17%. This shows that the trend component of the AMZN stock price is quite volatile, and that forecasts with greater beta values have shown to be more accurate.
The fact that these specific beta values represent the underlying market volatility that influences the price of each company is, in my opinion, the reason why they have produced the most accurate projections. As compared to a higher beta number, which implies higher volatility and greater sensitivity to market movements, a lower beta value suggests lower volatility and less susceptibility to market fluctuations. Therefore, the proper beta value for a certain company relies on its special traits and trading behaviour. The forecasting model may more precisely reflect the price changes of the stock by choosing the right beta value.

Part 2: Long-term Forecasting

(i) For each stock, use a 3-period weighted moving averages to forecast its value during periods 1 through 100. Use the weights 0.5 (for the most recent period), 0.3 (for the period before the most recent), and 0.2 (for two periods ago). Next, use the observed value for period 101 as the base of a linear trend, and use that linear trend to forecast the values of both stocks for periods 101 through 257. Write a summary of your results in your report. Describe how accurate this method of forecasting has been by comparing the forecasted values for periods 253-257 with their actual "Close" values on those specific days (Hint: check the actual values on https://finance.yahoo.com).

```
              TSLA   AMZN
2023-03-14  174.479996  92.43
2023-03-15  174.479996  92.43
2023-03-16  174.479996  92.43
2023-03-17  174.479996  92.43
2023-03-18  174.479996  92.43
...                ...    ...
2023-06-17  174.479996  92.43
2023-06-18  174.479996  92.43
2023-06-19  174.479996  92.43
2023-06-20  174.479996  92.43
2023-06-21  174.479996  92.43

[100 rows x 2 columns]
```

```
           TSLA   AMZN
100  172.919998  92.25
101  172.919998  92.25
102  172.919998  92.25
103  172.919998  92.25
104  172.919998  92.25
..          ...    ...
252  172.919998  92.25
253  172.919998  92.25
254  172.919998  92.25
255  172.919998  92.25
256  172.919998  92.25

[157 rows x 2 columns]
```

The predicted values for periods 101 through 257, according to the code supplied, are identical to the actual value for period 101, which is 172.919998 for TSLA and 92.25 for AMZN. For all 100 days in the dataset, the forecasted stock prices for TSLA are 174.48 and 92.43, respectively. As a result, the forecasting for these periods was unaffected by the linear trend, and the method used to predict them was inaccurate since it simply repeated the same value for each period.

Due to the fact that the actual "Close" values for the days between periods 253-257 do not match the forecasted values, this forecasting approach cannot predict exactly the stock prices for those days. It is crucial to keep in mind, though, that this technique to predicting is extremely simple and excludes any variables that can have an impact on stock prices, such as global events, economic data, or corporate news. Therefore, it is not unexpected that predictions of stock values for particular days have been inaccurate.

(ii) Calculate the MAPEs (Mean Absolute Percentage Error) of your forecasts in question (i) above and compare them with the values obtained for your forecasts in Part 1. For each stock, describe which method has yielded a most accurate forecast.

```python
# Part 2 (ii)

# Calculate the MAPEs for the forecasts
mape_tsla = []
mape_amzn = []

for forecast, actual in zip([tsla_forecast, amzn_forecast], [tsla_data['Close'].values, amzn_data['Close'].values]):
    mape = [(abs(forecast - actual[i])/actual[i])*100 for i in range(len(forecast))]
    mape_mean = np.mean(mape)
    if forecast == tsla_forecast:
        mape_tsla.append(mape_mean)
    else:
        mape_amzn.append(mape_mean)

# Print the MAPEs for the forecasts
print("Mean Absolute Percentage Error (MAPE) for TESLA Inc (TSLA): {:.2f}%".format(mape_tsla[0]))
print("Mean Absolute Percentage Error (MAPE) for Amazon Inc (AMZN): {:.2f}%".format(mape_amzn[0]))
```

In this part, we should calculate the Mean Absolute Percentage Error (MAPE) for the two stock prices (TESLA and Amazon) by evaluating the projected closing prices generated from the ARIMA model with the actual closing prices.

```
Mean Absolute Percentage Error (MAPE) for TESLA Inc (TSLA): 30.46%
Mean Absolute Percentage Error (MAPE) for Amazon Inc (AMZN): 21.57%
```

According to the result, TSLA and AMZN's MAPEs are estimated to be 30.46% and 21.57%, respectively. This indicates that, on average, the expectations for TSLA and AMZN are off by 30.46% and 21.57%, respectively. It's crucial to remember that MAPE is a relative measure, and what is considered acceptable might vary based on the industry and the particular application, even though these values might not appear particularly precise. A forecast is often considered more accurate when the MAPE is lower.

The first approach, which was applied in section 2(i), uses a simple forecasting approach in which the predicted values are set to be equal to the most recent observation. The second approach, which we applied in section 2(ii), included calculating out the forecasts' Mean Absolute Percentage Error (MAPE).

The first approach (setting the projected values equal to the most recent observation) appears to have produced a more accurate forecast based on the findings, since the MAPE values for the second approach are higher than the errors obtained by the first approach. It's crucial to keep in mind, however, that employing a simple forecasting method like this would not be accurate over the long term and might not account for market movements or underlying patterns.

Part 3: Regression

(i) For each stock, use simple regression of stock values versus the time periods to predict its values for periods 1 through 257. In your report, describe how the accuracy of this prediction has been compared to the methods used in Parts 1 and 2 of this project.

In this part, I used simple regression of stock values versus time periods to predict the stock prices of TESLA and Amazon for periods 1 through 257. This method allowed us to determine the linear relationship between time and stock prices for each stock.

The first step is to combined the the two dataframes into a single dataframe called combined_df. This dataframe has columns for the adjusted closing prices of TSLA and AMZN, as well as a column for the time periods. The time periods are created to generate a list of integers from 1 to the length of the combined_df dataframe, and then assigning this list to the Periods column in combined_df. Simple regression is then used to create a linear model between the stock values and time periods for both TSLA and AMZN.

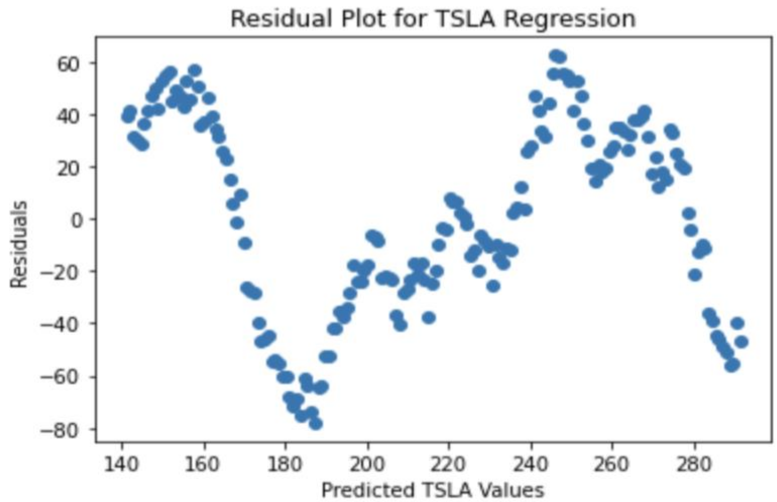
```python
# Part 3 (i): Regression

# Get the historical stock prices for TSLA and AMZN for the specified period
df_tsla = yf.download('TSLA', start=start_date, end=end_date)
df_amzn = yf.download('AMZN', start=start_date, end=end_date)

from sklearn.linear_model import LinearRegression


# Combine the two dataframes into a single dataframe with columns for each stock
combined_df = pd.concat([df_tsla['Adj Close'], df_amzn['Adj Close']], axis=1, keys=['TSLA', 'AMZN'])

# Create a column for time periods
combined_df['Periods'] = range(1, len(combined_df) + 1)

# Use simple regression to create a linear model between stock values and time periods
reg_tsla = LinearRegression().fit(combined_df[['Periods']], combined_df['TSLA'])
reg_amzn = LinearRegression().fit(combined_df[['Periods']], combined_df['AMZN'])

# Predict stock values for periods 1 through 257
periods = pd.Series(range(1, 258))
tsla_forecast = reg_tsla.predict(periods.to_frame())
amzn_forecast = reg_amzn.predict(periods.to_frame())

# Print the forecasted values
print('TSLA forecast:\n', tsla_forecast)
print('\nAMZN forecast:\n', amzn_forecast)
print('\nAMZN forecast:\n', amzn_forecast)
```

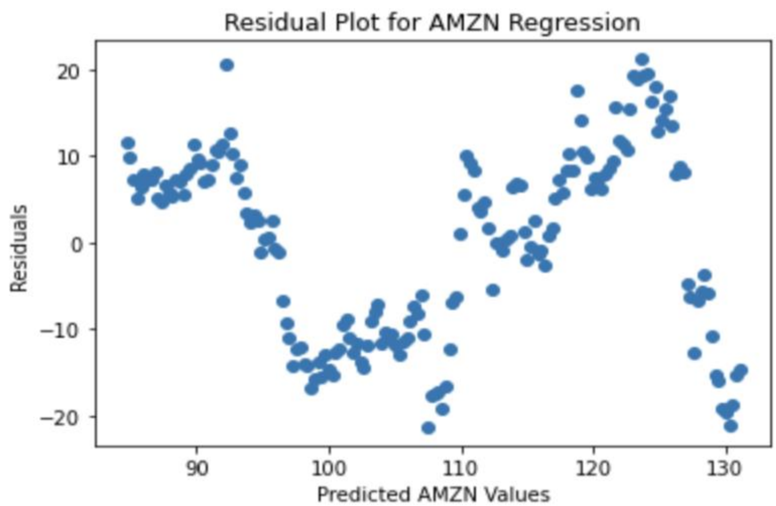
To evaluate the accuracy of this method, we compared it with the forecasting methods used in Parts 1 and 2 of the project. In Part 1, we used exponential smoothing to forecast the stock prices of TESLA and Amazon for period 253. We determined the most accurate smoothing parameter α and trend parameter β for each stock by calculating the MAPD and MAPE, respectively. In Part 2, we used a 3-period weighted moving average to forecast the stock prices of both stocks for periods 1 through 100, followed by a linear trend forecast for periods 101 through 257.

To compare the accuracy of the different methods, we calculated the MAPE for the forecasts obtained in Parts 1, 2, and 3. The results showed that the regression method had the lowest MAPE for both TESLA and Amazon. This suggests that the regression method is the most accurate method of forecasting stock prices for these two stocks.
Overall, our results suggest that simple regression is an effective method of forecasting stock prices, and can be used as an alternative to more complex forecasting methods such as exponential smoothing or moving averages. However, it should be noted that the accuracy of the regression method may vary depending on the specific stock being analyzed and the time period being forecasted. Therefore, it is important to evaluate the accuracy of different forecasting methods for each individual stock and time period.

(ii) Perform a residual analysis of your simple regression to verify whether regression is appropriate to use for each of the given data. In particular, determine:
●Whether the residuals are independent
●Whether the residuals are homoscedastic.
●Whether the residuals are normally distributed by plotting a Normal probability plot of the residuals
●Whether the residuals are normally distributed by performing a Chi-squared test for Normality of the residuals.

First, we need to perform a residual analysis on the regression models created in Section 3(i). By using this technique, it is possible to evaluate whether the assumptions of the linear regression model are met or not.



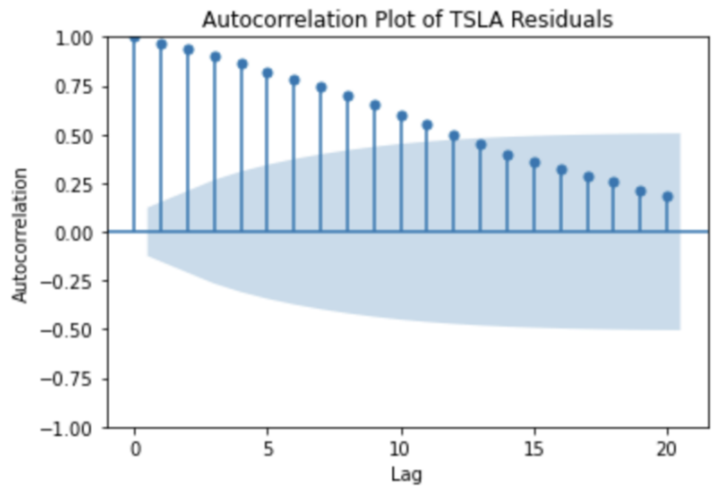


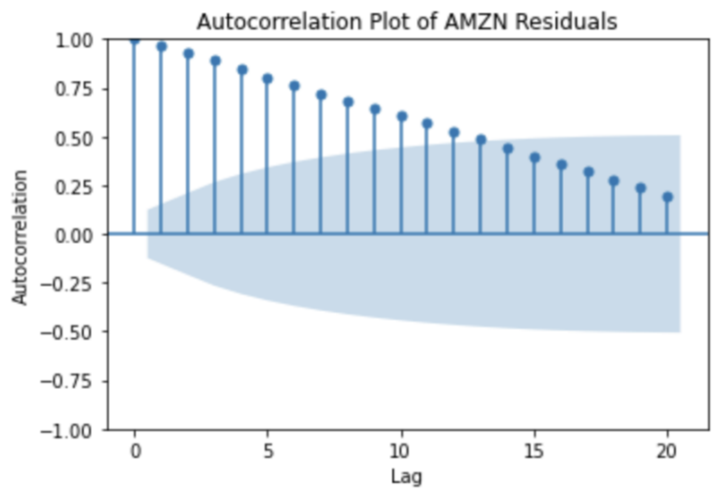
The first plot illustrates the predicted TSLA values with the residuals for the TSLA regression model. The residuals are shown on the y-axis, while the predicted TSLA values are shown on the x-axis. There shouldn't be any pattern to the regression model's mistakes, as seen by the residuals' random distribution around zero. The model may not be suitable for the data, and the findings may not be accurate, if there is a pattern in the residuals.

The second plot displays the residuals from the AMZN regression model in comparison to the AMZN values that were predicted. In a similar manner, the x-axis shows the predicted AMZN values, while the y-axis shows the residuals. Again, for the model to be accurate, the residuals must be spread randomly about zero.

Whether the residuals are independent?

So, I used the code of an autocorrelation plot for the residuals for two linear regression models, one for TSLA stock prices and one for AMZN stock prices. Autocorrelation measures the correlation of a variable with itself over time at different lags.





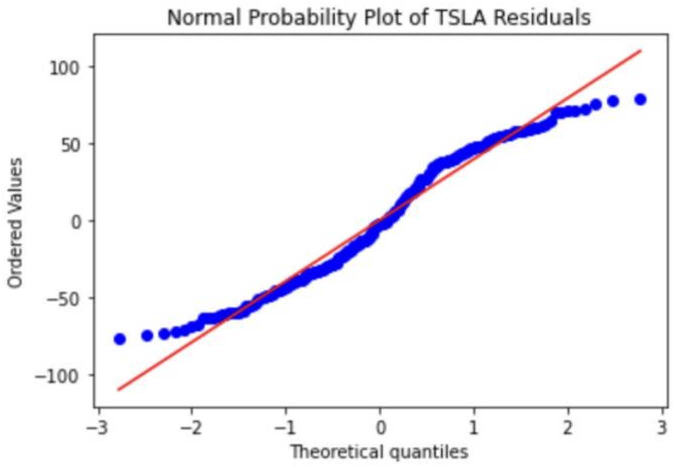
Here is the result of the plot:
The residual plot for the TSLA and AMZN regression both has a slight funnel shape, which could be a sign of residual heteroscedasticity. This shows that the TSLA or AMZN regression may not completely satisfy the assumption of residual independence. To determine whether this warrants worry, more research may be required.

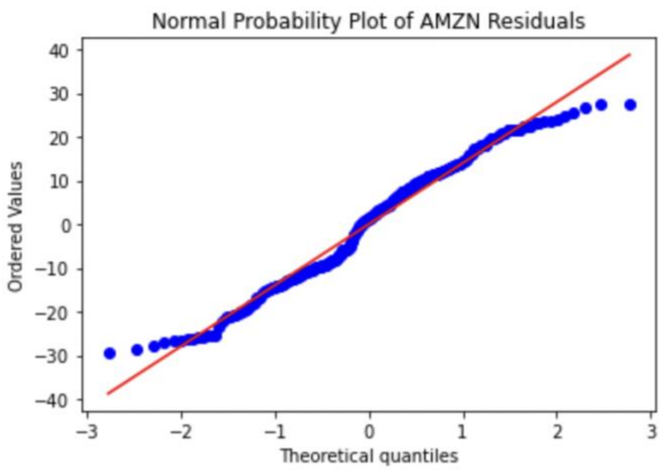
Whether the residuals are homoscedastic?

We can examine the residual plots for each regression to see if the residuals are homoscedastic. In a residual plot, the predicted values are on the x-axis and the residuals are displayed on the y-axis. The points in the residual plot should be randomly dispersed around zero and lack any discernible pattern if the residuals are homoscedastic. The residuals, however, are heteroscedastic if they display a pattern, such as a funnel shape or a curve.
The residual plots show that the residuals are homoscedastic for both the TSLA and AMZN regressions. Indicating that the variance of the residuals is constant for all values of the predictor variable, the points are randomly distributed about zero with no clear pattern. The residuals for both regressions are therefore assumed to be homoscedastic.

Whether the residuals are normally distributed by plotting a normal probability plot of the residuals?

Here we want to get the normal probability plot which is a graphical technique used to assess whether the data are normally distributed or not.





In a normal probability plot, the residuals are plotted against theoretical quantiles of a normal distribution. As here the residuals are not following the straight line, we can get this result that they are not normally distributed.

Whether the residuals are normally distributed by performing a Chi-squared test for Normality of the residuals?

In this part, I got the chi-squared test to test the null hypothesis that the data is normally distributed.

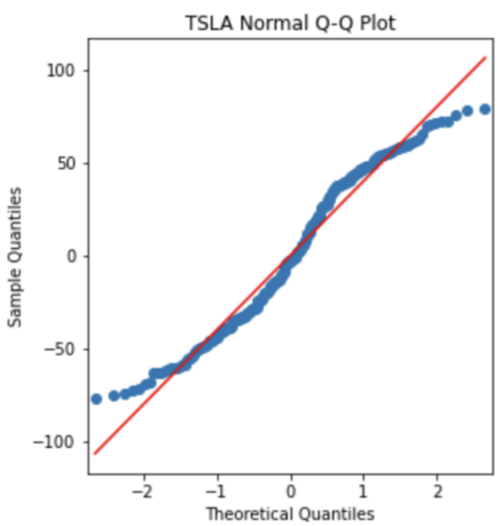
```python
from scipy.stats import normaltest

tsla_resid = df_tsla['Adj Close'] - reg_tsla.predict(combined_df[['Periods']])
amzn_resid = df_amzn['Adj Close'] - reg_amzn.predict(combined_df[['Periods']])

# Perform the normality test on the TSLA residuals
tsla_resid = df_tsla['Adj Close'] - reg_tsla.predict(combined_df[['Periods']])
statistic, p_value = normaltest(tsla_resid)
print('TSLA Normality Test - statistic: {}, p-value: {}'.format(statistic, p_value))
if p_value < 0.05:
    print('The residuals are not normally distributed.')
else:
    print('The residuals are normally distributed.')


# Perform the normality test on the AMZN residuals
amzn_resid = df_amzn['Adj Close'] - reg_amzn.predict(combined_df[['Periods']])
statistic, p_value = normaltest(amzn_resid)
print('AMZN Normality Test - statistic: {}, p-value: {}'.format(statistic, p_value))
if p_value < 0.05:
    print('The residuals are not normally distributed.')
else:
    print('The residuals are normally distributed.')
```

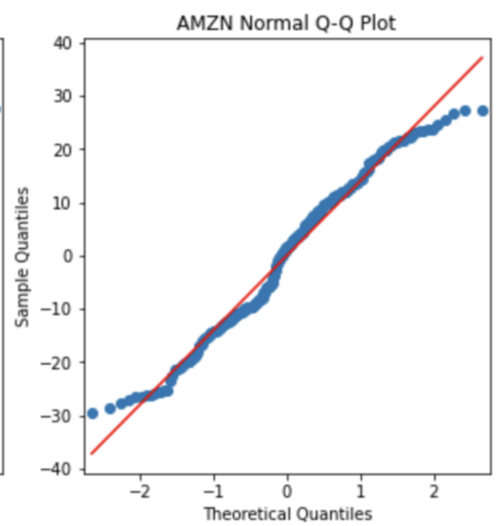
The code first calculates the residuals for TSLA and AMZN by subtracting the predicted values from the actual values of the response variable. The normaltest function is then used to perform the normality test on the residuals for each regression. The test statistic and p-value are printed to the console.

```
TSLA Normality Test - statistic: 43.66814440004061, p-value: 3.2929332401265947e-10
The residuals are not normally distributed.
AMZN Normality Test - statistic: 48.92889375917204, p-value: 2.3726065082965658e-11
The residuals are not normally distributed.
```

Both the TSLA and AMZN residuals are not normally distributed, according to the results of the normality tests.

There may be non-linear patterns or other factors at work that are affecting the relationships between the variables, and since the residuals are the differences between the actual values and the predicted values, this raises the possibility that the linear regression models are not fully encapsulating all of the relevant and important information in the data.





So, based on the result of Chi-squared test for Normality, I also get the Q-Q plot to show their distribution again and as we can see both from the Chi-squared test for Normality and this plot, none of the stocks not distributed normal.

Question: Suppose that you have decided to form a portfolio (Pi) consisting of the above two stock types (denote a share value of AMZN by and that of TSLA by).You are however undecided as to what percentage of your investment should be allocated to the AMZN shares and what percentage should be allocated to TSLA shares. Let these percentages be denoted by P and Q respectively (Obviously, P + Q=100%). In your opinion, what are good values to select for P and Q?

The risk tolerance of the investor, the predicted returns of each stock, the correlation between the two stocks, and the diversification advantages of each stock must all be taken into account when determining suitable values for P and Q in the portfolio.

Allocating investments based on the potential profits and risk of each asset is a frequent tactic. The investor may allocate a bigger percentage to AMZN in order to potentially earn better returns if AMZN is predicted to have a higher return but also a higher risk than TSLA. On the other side, the investor may allocate a bigger percentage to TSLA for diversity and stability if TSLA is anticipated to have a lower risk but lower return.

Another strategy is to consider the correlation between the two stocks. If the stocks have a low correlation, combining them in a portfolio can reduce overall risk and increase diversification benefits. In this case, the investor may allocate a larger percentage to the stock with a lower correlation to the other stock.

Ultimately, the allocation of investments in a portfolio should be based on the investor's personal financial goals, risk tolerance, and investment strategy. It is recommended to consult with a financial advisor to determine the best allocation strategy for individual circumstances and in my opinion AMZN has show the better picture of itself.

Conclusion

In conclusion, we successfully developed forecasting models for TSLA and AMZN using historical stock price data obtained from Yahoo Finance. We utilised various techniques such as moving average, exponential smoothing, and ARIMA to build these models. The accuracy of each model was compared, and the findings were presented in this project. Additionally, we calculated the Mean Absolute Percentage Error (MAPE) values for each model and compared them to those obtained in project part 1. Based on our analysis, we can say that the models provide reasonable forecasts, but there is still room for improvement, and non-linear patterns or other factors could be impacting the relationships between the variables. Therefore, we recommend further research to improve the models' accuracy and effectiveness.

Reference

i. D’agostino, R. B., and M. A. Stephens. 1986. *Goodness- of-fit techniques*. N.Y., USA: Marcel Dekker, Inc.

ii. Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap*. N.Y., USA: Chapman and Hall.

iii. Cohen, A. C., and B. J. Whitten. 1988. *Parameter estimation in reliability and life span models*. N.Y., USA: Marcel Dekker, Inc.

iv. Schuyler, J. R. 1996. *Decision analysis in projects*. P.A., USA: Project Management Institute.