Module-2: Forecasting a Time Series

By: Nithin Reddy.P (NEU ID:002896440)

Subject:ALY6050: Introduction to Enterprise Analytics
Under the guidance of Shahram Sattar
Submission Date:10-Mar-2024

Introduction:

In this assignment, we are about to forecast the future share price of Apple and Honeywell using forecasting methods and further use regression model to forecast the data and lastly perform residual analysis of the simple regression model to verify appropriate use of regression.

Data Set:

We have the daily close price of AAPL and HON form 2019-11-08 to 2020-11-06, out aim to estimate the closing price at 2020-11-09 in three different phases. In phase-1, we use exponential smoothing to predict the future, in phase-2 we use weighted moving averages and linear trend and last phase, we use regression.

Part-1: Short-term Forecasting

(i): Line Plot to observe patterns:

From the given data, a line graph of corresponding share price and date has been plotted. Please refer to Fig-1.

For Apple, we can observe that there is an increasing trend, which can also spot the same using the trend line, it relates with the trend. However, we can observe a significant dip in the share price between 21 Feb to 23 Mar, thereafter the price increases and constitutes to sudden rise between July 28 to Sep 22, and a fall from 22nd Sep, since then the price has been fluctuating.

For Honeywell, overall the trend line shows a downward trend. The price has fluctuated from the beginning till 11 March, and then a sharp dip till end of March. However, the entity has revered form the loss and has shown consistent increase in the daily price with minor fluctuations in between.

(ii): Exponential smoothing to forecast both prices for period 253 and calculate Mean Absolute Percentage Deviation)of each forecast.

Using a successive values of 0.15, 0,35, 0.55, and 0.75 (smoothing parameters α), and historical share price, I predicted the share price for period 253 by using an inbuild excel function named, exponential smoothing in Data Analytics package, the formula is represented as follows:

$F_{t+1} = \alpha * Y_t + (1-\alpha) * F_t$

Where:

F_{t+1} : is the forecasted value for the next period (time t+1).

Y_t: is the actual value observed at time

F_t: is the forecasted value for the current period (time t) α: is s the smoothing parameter

Based on the formula we have arrived to the below values:

Exponential Smoothing at α= 0.15,

Period	AAPL	HON
253	118.6026449	184.02069

Exponential Smoothing at α= 0.35

Period	AAPL	HON
253	118.03	183.27

Exponential Smoothing at α= 0.55

Period	AAPL	HON
253	116.79	181.52

Exponential Smoothing at α= 0.75

Period	AAPL	HON
253	115.21	177.89

In the next Step to calculate MAPD, we first found the absolute percentage error by subtracting the forecasted share price form actual share price and dividing it by actual price; then by using Average function in excel we have identified the MAPD of each smoothing parameter, please refer to the below.

MAPD at α= 0.15,

MAPD At α=0.15	
AAPL_APE	HON_APE
1.96	**1.86**

MAPD at α= 0.35

MAPD at α = 0.35	
AAPL_APE	HON_APE
2.00	**1.95**

MAPD at α= 0.55

MAPD at α = 0.55	
AAPL_APE	HON_APE
2.23	**2.16**

MAPD at α= 0.75

MAPD at α = 0.75	
AAPL_APE	HON_APE
2.91	**2.64**

Observations:

From the derived MAPD values for two stocks (AAPL and HON) at different levels of α (0.15 to 0.75), a clear trend emerges: as α increases, the MAPD values tend to rise, suggesting decreased forecast accuracy. Notably, at 0.15 both stocks exhibit relatively low values, indicating higher forecast accuracy. This underscores the effectiveness of lower α values in providing accurate forecasts, likely due to their greater emphasis on historical observations, allowing the model to capture underlying trends more effectively. Thus, in this context, lower values of α appear to have yielded the most accurate forecasts for the two stocks.

(iii): at α=0.55 and perform an adjusted exponential smoothing to forecast both prices for period 253. Use trend parameters β of 0.15, 0.25, 0.45, and 0.85 for the for both stocks and, calculate the MAPEs of your forecasts

Using the below trend formula, the corresponding trend values for respective periods 1-252 have been calculated.

$T_t = \beta * (F_t - F_{t-1}) + (1-\beta) * T_{t-1}$

Tt represents the trend component of the time series at time

Ft is the actual value observed at time t

Ft−1 is the forecasted value for the previous period t−1

Tt−1 is the trend component for the previous period t−1

β is the trend value

To Adjusted Exponential Smoothing values for respective β is the sum of respective Exponential Smoothing at α = 0.55 and respective trend values. The respective values are populated for both the stocks. Since, the adjusted values have been derived, the absolute percentage error has been calculated in similar way as mentioned in part-1. Finally, MAPR is calculated by taking the mean of Absolute Percentage Error of corresponding beta values.

Based on the calculation we derived the below values.

MAPEs at β = 0.15

MAPEs at β = 0.15	
AAPL_APE	HON_APE
2.05	**2.09**

MAPEs at β = 0.25

MAPEs at β = 0.25	
AAPL_APE	HON_APE
2.01	**2.03**

MAPEs at β = 0.45

MAPEs at β = 0.45	
AAPL_APE	HON_APE
1.97	**1.94**

MAPEs at β = 0.55

MAPEs at β = 0.85	
AAPL_APE	HON_APE
1.95	**1.86**

Observations:

The MAPEs tend to decrease as β increases, indicating that higher values of β lead to more accurate forecasts. For instance, at 0.85 both AAPL and HON exhibit relatively low MAPEs of 1.95 and 1.86 respectively, which suggest higher forecast accuracy compared to lower values of β. This trend suggests that greater emphasis on recent trend information (achieved by higher values of β) enhances the model's ability to capture and forecast underlying trend patterns in the data.

Part-2: Long-term Forecasting

3-period weighted moving averages to forecast

The 3-period weighted moving averages forecast is computed by assigning weights to the most recent period, the period before, and two periods ago. These weights are 0.5, 0.3, and 0.2 respectively. To perform the forecast, the first three values are set to zero. The current weighted average forecasted value is calculated by multiplying the most recent price by its weight, adding it to the product of the price of the period before and its weight, and further adding the product of the price of two periods ago and its weight.

In the next step the we calculated the absolute percentage error and mean of the error.

3-period weighted moving averages MAPE is as follows:

MAPE-APPL	MAPE-HON
2.11	2.07

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.

We constructed a table named "Table for linear trend to forecast" to predict stock prices based on observed values, with period 101 serving as the base for the linear trend. The prices were computed using the FORECAST.LINEAR formula in Excel, which requires inputs for period (x) to forecast and the corresponding share prices (y) for periods 1-100 (known ys, known xs). This approach allows us to project future stock prices by establishing a linear relationship between periods and share prices, with period 101 acting as the reference point for initiating the trend. Further we obtained the below MAPE by taking the mean of absolute errors.

MAPE-APPL	MAPE-HON
5.18	3.30

Observations:

We employed two methods for forecasting stock prices: linear trend and 3-period weighted moving averages. For the linear trend method, the predicted prices for Apple and Honeywell stocks over periods 253 to 257 were calculated. The Mean Absolute Percentage Error (MAPE) for Apple was found to be 5.18%, while for Honeywell, it was 3.30%. On the other hand, utilizing the 3-period weighted moving averages approach, the forecasted values were obtained. The MAPE for Apple was notably reduced to 2.11%, and for Honeywell, it was 2.07%.

Comparing the forecasted values with the actual "Close" prices on those specific days, we observed varying levels of accuracy. For instance, in the linear trend method, the predicted prices tend to deviate more significantly from the actual prices compared to the 3-period weighted moving averages. This suggests that the 3-period weighted moving averages method provided more accurate forecasts for both Apple and Honeywell stocks during the specified periods.

Part-3: Regression:

(i): use simple regression of stock values versus the time periods to predict its values for periods 1 through 257

In order to predict share values based on historical, we first pulled the Slope and Intercept of the historical by using corresponding SLOPE and INTERCEPT formula in excel. As we now have the values, we can find the value Y based on the below equation.

$Y = \beta 0 + \beta 1 * x1$

Y is the dependent variable (the variable we're trying to predict)

$\beta 0$ is the intercept term

$\beta 1$are the coefficients of the independent variables x1

x1 are the independent variables (predictors)

Please refer to Fig-2; you can observe the regression equation in the respective graphs of Apple and Honeywell, by using this equation, we have predict the share price for both companies from period 1 to 257.

Observations:

In Part 1, exponential smoothing and adjusted exponential smoothing were employed to forecast the stock prices of Apple and Honeywell. MAPD values were calculated for different smoothing parameters (α for exponential smoothing and β for adjusted exponential smoothing). Lower MAPD values indicate higher forecast accuracy. In Part 2, two forecasting methods were utilized: linear trend forecasting and 3-period weighted moving averages. The Mean Absolute Percentage Error (MAPE) was computed for each method, where lower MAPE values indicate better accuracy. Finally, in Part 3, regression analysis was conducted to predict the stock prices.

Comparing the accuracy of the prediction methods across the three parts, it appears that the regression analysis (Part 3) yielded the most accurate forecasts. This conclusion is drawn based on the comparison of predicted values with actual share prices for Apple and Honeywell. The predicted values in Part 3 exhibit smaller discrepancies from the actual prices compared to the forecasts generated in Parts 1 and 2.

(ii): **Perform a residual analysis of your simple regression to verify whether regression is appropriate to use for each of the given data.**

1. Whether the residuals are independent:

Residuals of the predicted values can be pulled by subtracting the corresponding predicted value from the actual share price. As we have the residuals, the Residuals plot for independence has been displayed for both enterprises using a scatter plot (residuals and periods). In Fig-3; if we have a close look at the patten in Apple, we could observe that there is some presence of a pattern or oscillation in the residuals, indicating a potential issue with autocorrelation, even though at the far end, we could see that the pattern disappears. For Honeywell, the plotted residuals appear randomly scattered around the horizontal axis with no discernible pattern or trend, it suggests that there is no autocorrelation present in the residuals. This is indicative of a well-fitted regression model, where the errors are independent and evenly distributed.

2. Whether the residuals are homoscedastic

The scatter plot in Fig-4 illustrates the relationship between residuals and predicted values. The "snake" pattern observed in the residuals plot indicates heteroscedasticity, where the variability of the dependent variable changes across different levels of the independent variables. The downward movement of the pattern signifies that the spread of residuals decreases as the predicted values increase, indicating a narrowing variance of the dependent variable as predicted values become larger. Conversely, the upward movement of the "snake" pattern after the initial decline suggests that the spread of residuals starts to increase as predicted values continue to increase. This reversal indicates a widening variance of the dependent variable, but in the opposite direction.

3. Normal Probability plot of the residuals:

The histogram plotted from the Excel Data Analysis pack does not show a clear indication of a normal distribution for the residuals. To further investigate, the residuals are standardized using the STANDARDIZE function in Excel. After sorting the residuals in ascending order, standardized values are calculated using the mean and standard deviation of all residual values. Subsequently, ranks are assigned, and the cumulative distribution is computed. The Z-values are then obtained using the NORM.S.INV function. The scatter plot in Fig-6 illustrates the relationship between the residual values and standardized residuals. For Apple, a linear relationship between the two variables is observed, suggesting normality despite some variance. However, for Honeywell, the plot reveals a deviation from the red line representing the Z-score, indicating a lack of normality.

4. Chi-squared test for Normality of the residuals.

We are conducting a chi-square test to determine the normality of the data. Our hypotheses are as follows:

Null Hypothesis (H0): The data is normally distributed.
Alternative Hypothesis (H1): The data is not normally distributed.

To conduct this test, we need several variables:

- Sample Size: Number of periods or observations.
- Minimum (Min) and Maximum (Max) values of residuals.
- Range: Difference between the minimum and maximum values.
- Cell Length: Calculated using the formula Range / (1 + 3.32 * log(n)), where n is the sample size.

- Number of Cells: Range divided by the cell length.
- Residual Mean and Residual Standard Deviation (SD): Mean and standard deviation of the residuals.

Once we have these measures, we can compute the following variables to perform the chi-square test:

- Start and End values for each cell: Start of each cell is the minimum value of the residuals, and the end of each cell is the start value plus the cell length.
- Probability: Calculated using the NORM.DIST function, subtracting the end value from the start value with corresponding mean and standard deviation.
- Expected Frequency (E): Obtained by multiplying the probability by the sample size.
- Observed Frequency (O): Number of residuals falling within each cell interval, determined using the COUNTIF function.
- (E-O)^2/E: Squared difference between observed and expected frequencies, divided by expected frequency.
- Degree of Freedom: Number of cells minus 1 (df = 8).

The P-value of the test is calculated using the CHISQ.DIST.RT function, taking the sum of (E-O)^2/E values and degree of freedom as inputs. In our case, the resulting P-value is 5.567E-08, which is less than the significance level of 0.05. Therefore, we reject the null hypothesis, concluding that the data is not normally distributed based on the chi-square test.

Question: Suppose that you have decided to form a portfolio ***Π*** (Pi) consisting of the above two stock types (denote a share value of AAPL by X and that of HON by Y). You are however undecided as to what percentage of your investment should be allocated to the AAPL shares and what percentage should be allocated to HON 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?

Answer:

Let us consider that investor is a risk-taker with a long-term investment horizon and targets returns of 10-15%, they may consider a slightly aggressive approach while still diversifying their portfolio. Here's a suggested allocation strategy for the portfolio:

Allocation to AAPL (P): Given the risk tolerance and return expectations, a higher allocation to AAPL, which may offer higher potential returns but also higher volatility, could be suitable. Allocating around 60-70% of the portfolio to AAPL could be considered.

Allocation to HON (Q): Since HON may offer relatively lower volatility compared to AAPL, allocating around 30-40% of the portfolio to HON could provide diversification benefits while still capturing some upside potential.

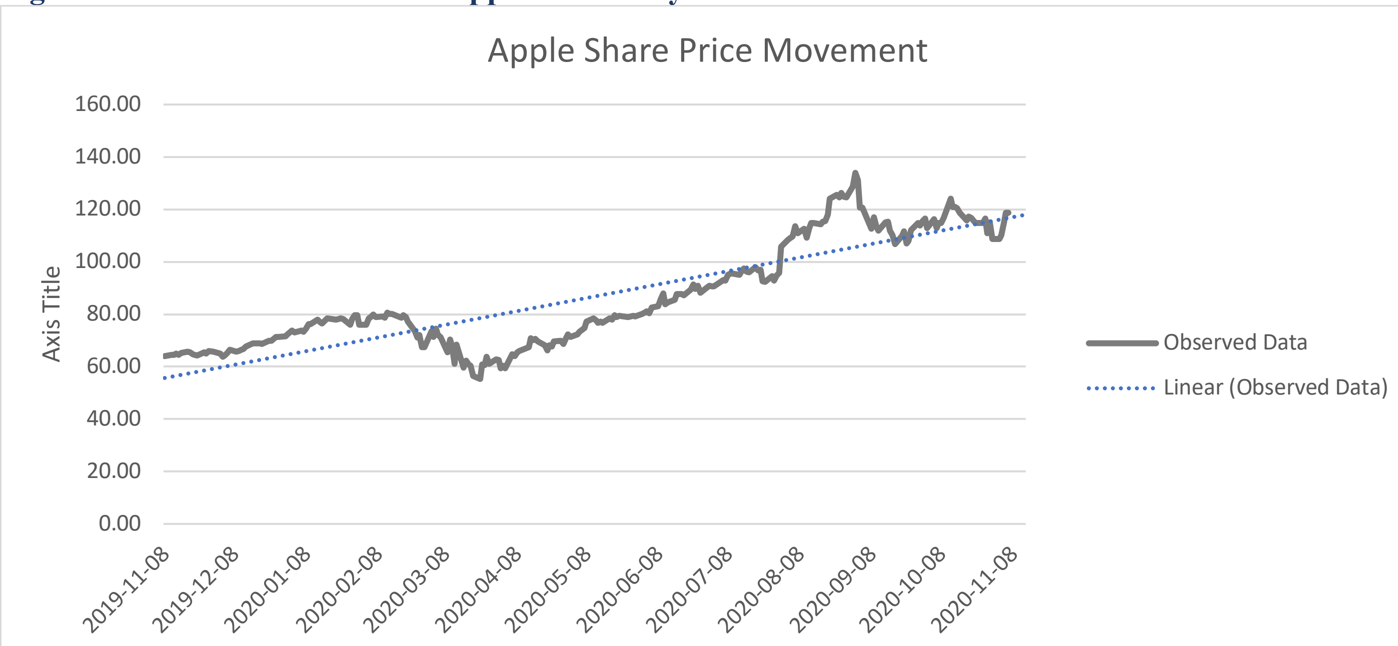
This allocation strategy aims to balance the desire for higher returns with risk management considerations. By allocating a larger portion to AAPL, the investor seeks to capitalize on its growth potential while acknowledging the associated risks. At the same time, allocating a portion to HON helps diversify the portfolio and mitigate overall risk.

Conclusion:

In conclusion, after analysing various forecasting methods and conducting regression analysis for predicting stock prices, it is evident that the regression analysis yielded the most accurate forecasts compared to other methods. While exponential smoothing and adjusted exponential smoothing provided reasonable forecasts, regression analysis exhibited smaller discrepancies between predicted and actual prices for both Apple and Honeywell stocks. However, it's important to note that residual analysis revealed potential issues with the regression model, particularly concerning the independence and homoscedasticity of residuals. While the residuals for Honeywell showed no discernible pattern or trend, indicating independence and homoscedasticity, the residuals for Apple exhibited some oscillation, suggesting potential autocorrelation and heteroscedasticity issues. Despite these limitations, regression analysis remains a valuable tool for forecasting stock prices, provided that the assumptions of independence and homoscedasticity hold true.

Appendix:

Fig-1: Share Price Movement of Apple and Honeywell



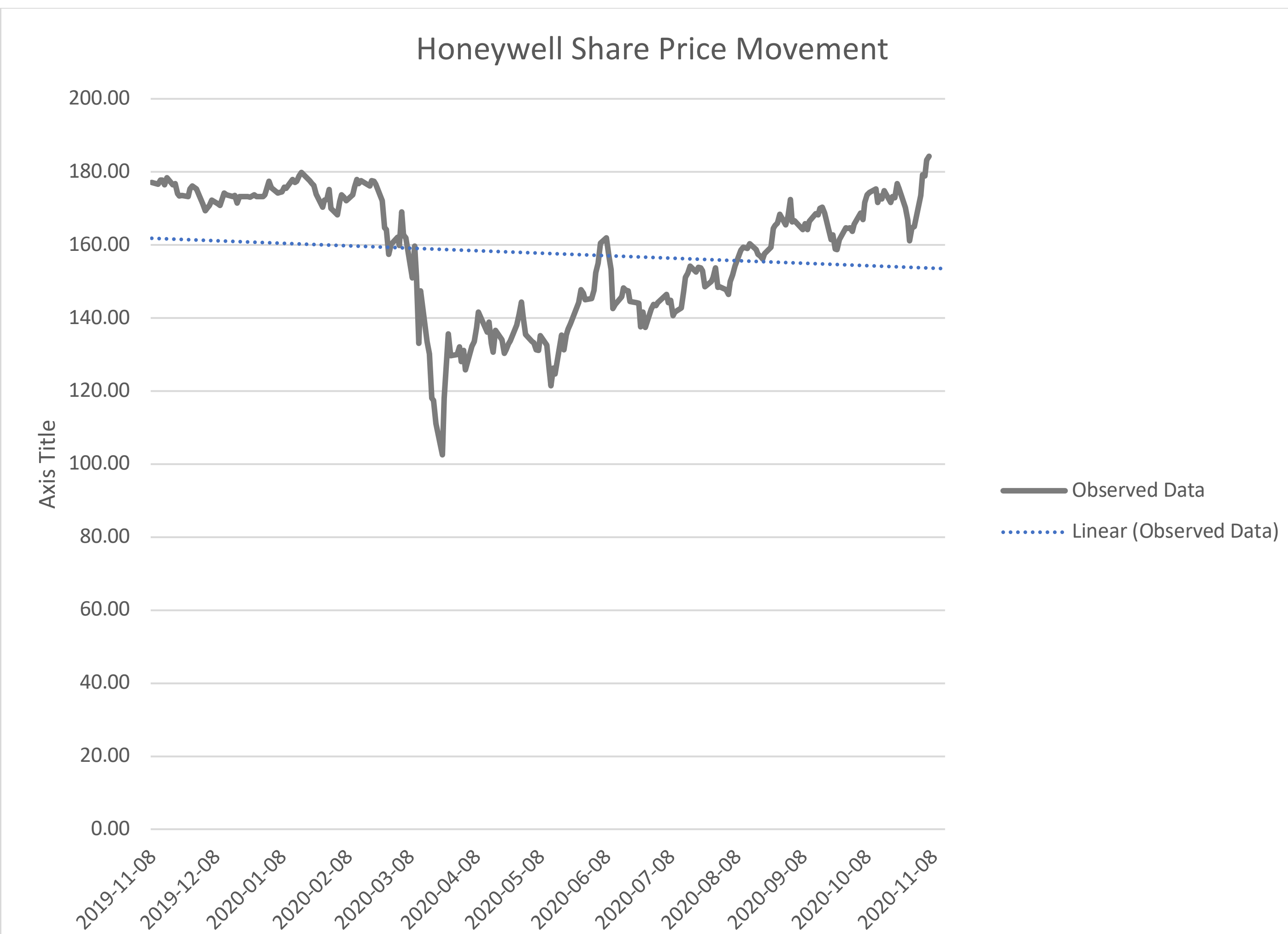
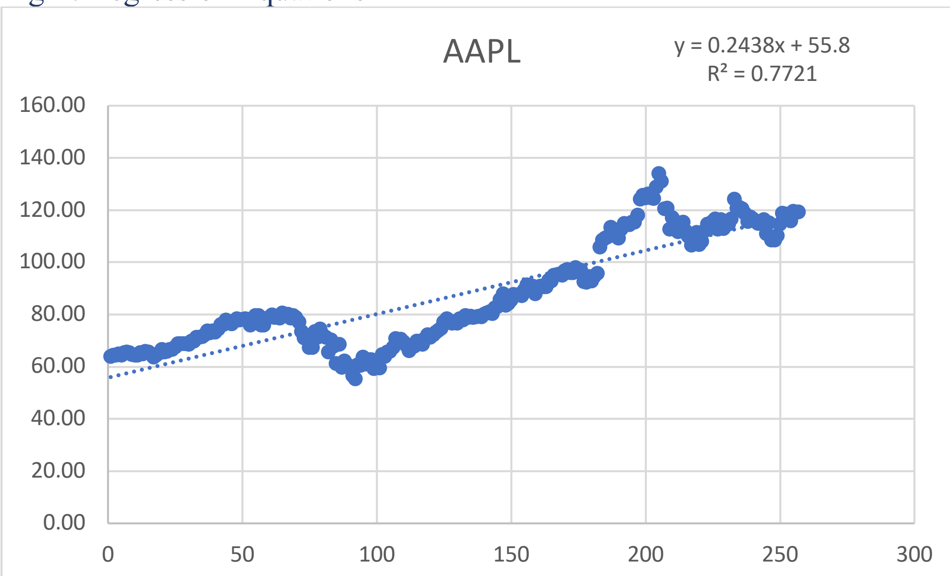


Fig-2: Regression Equations



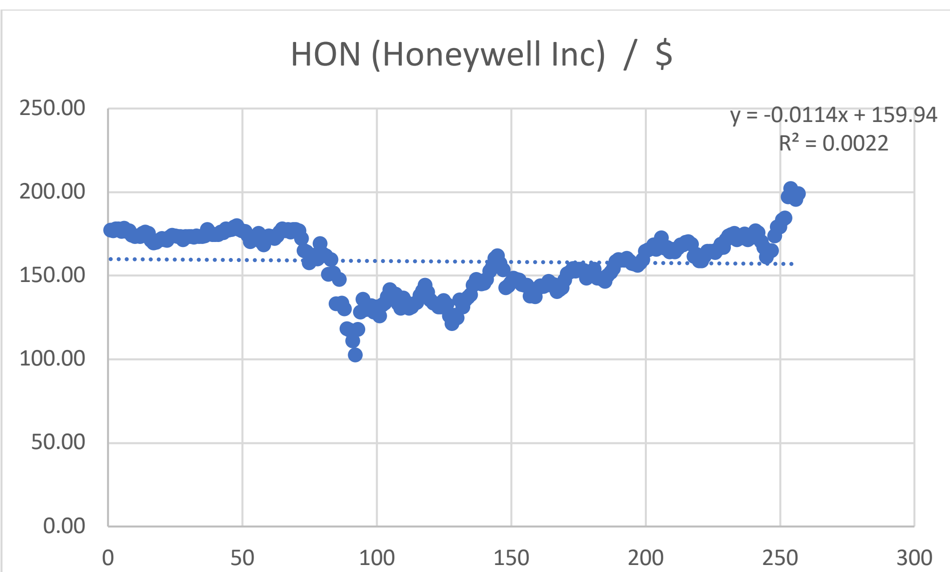
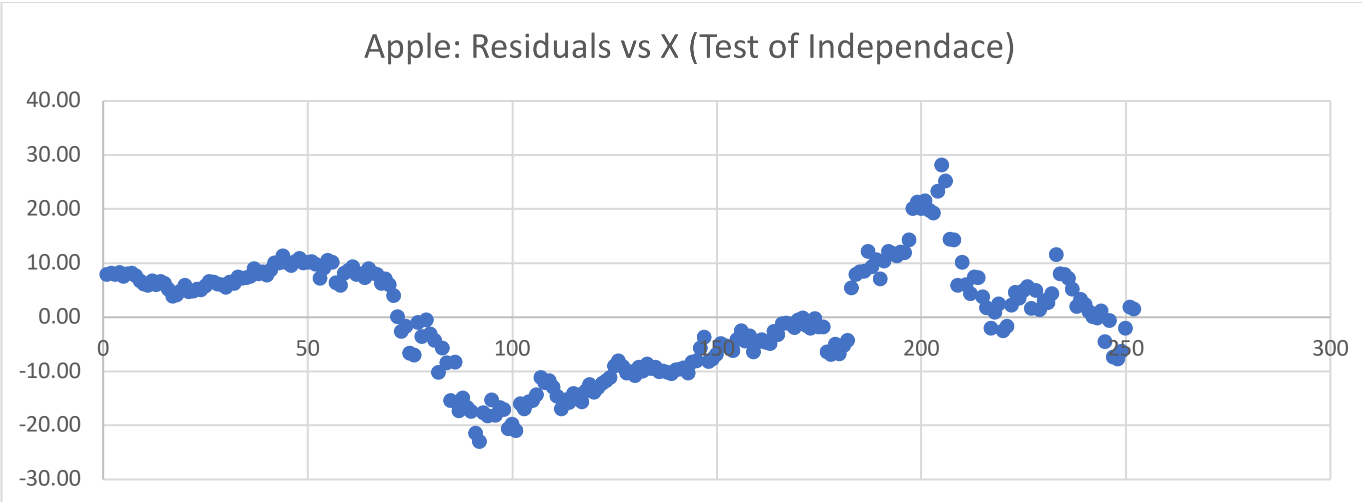


Fig-3: Test of independence:



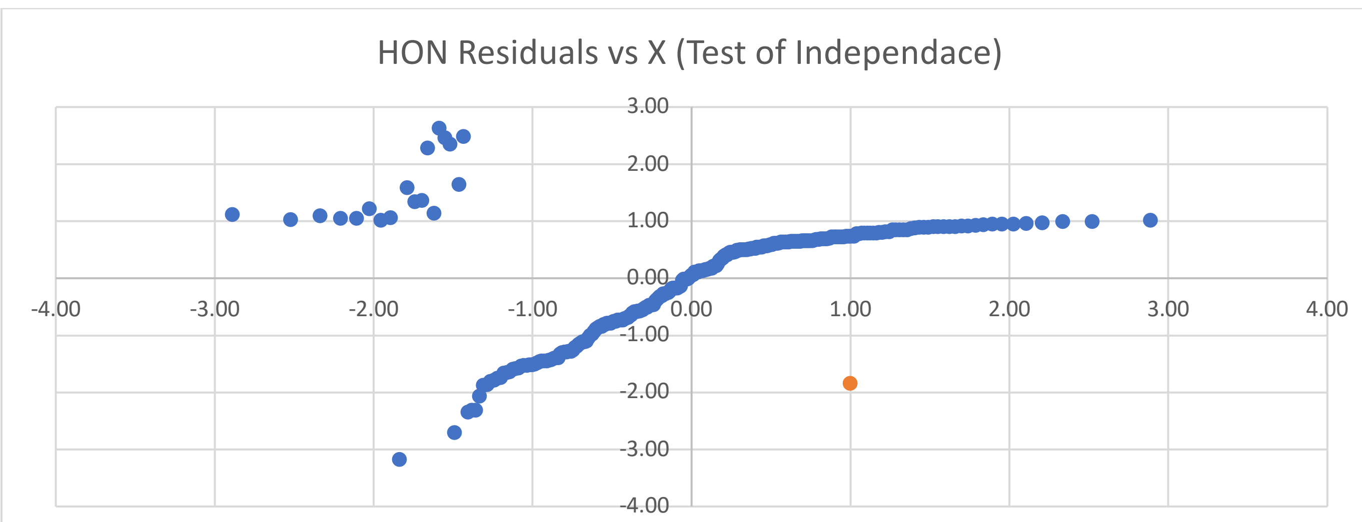
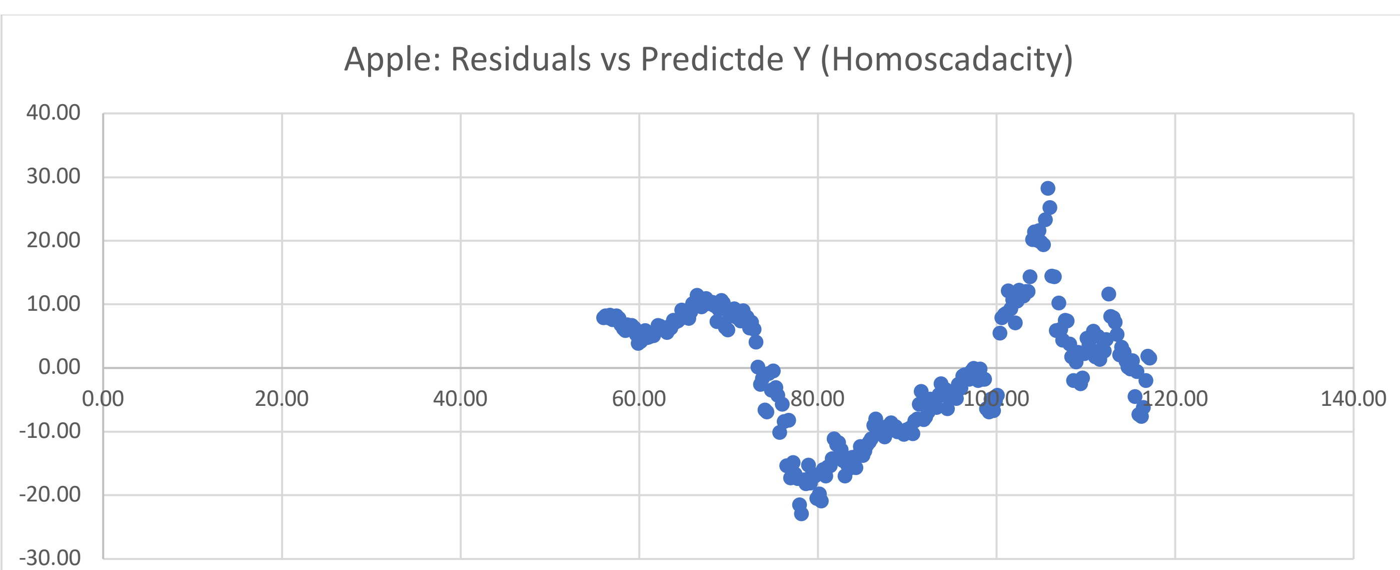


Fig-4: Test of homoscedastic :



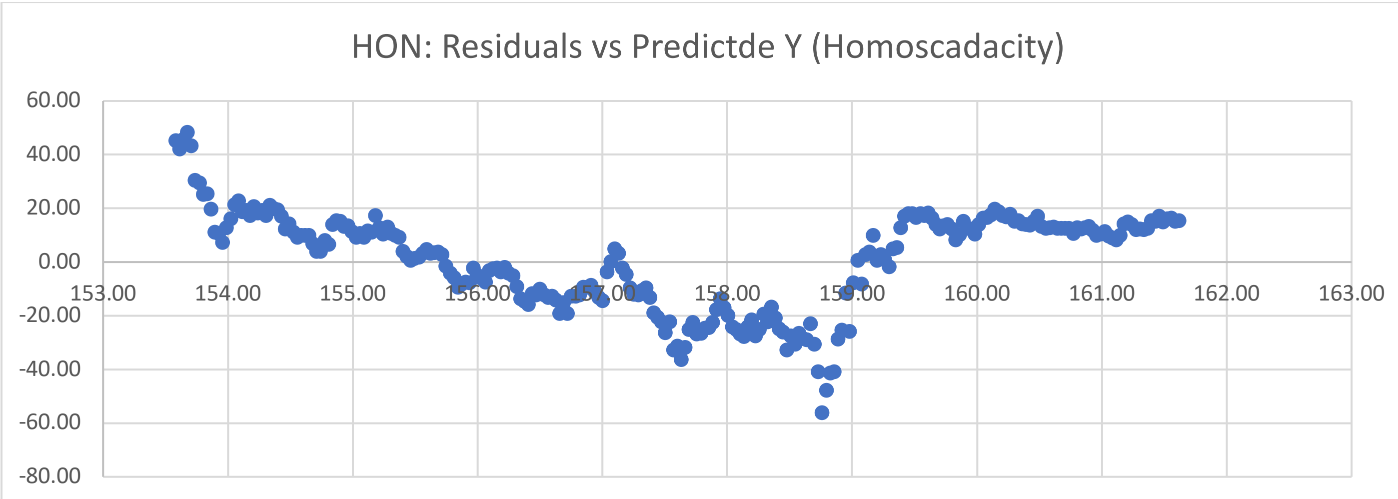
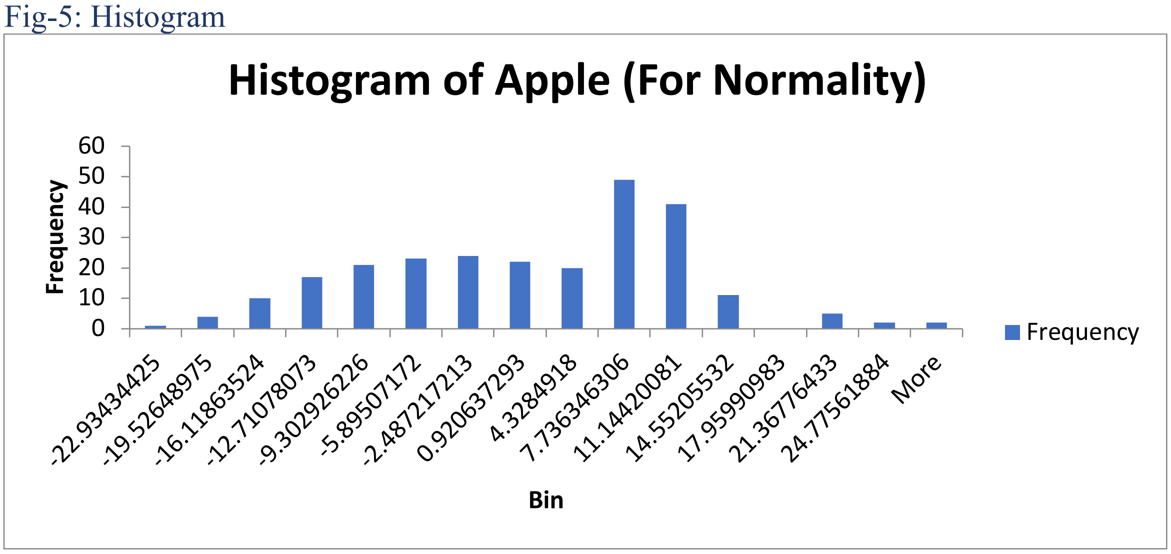


Fig-5: Histogram



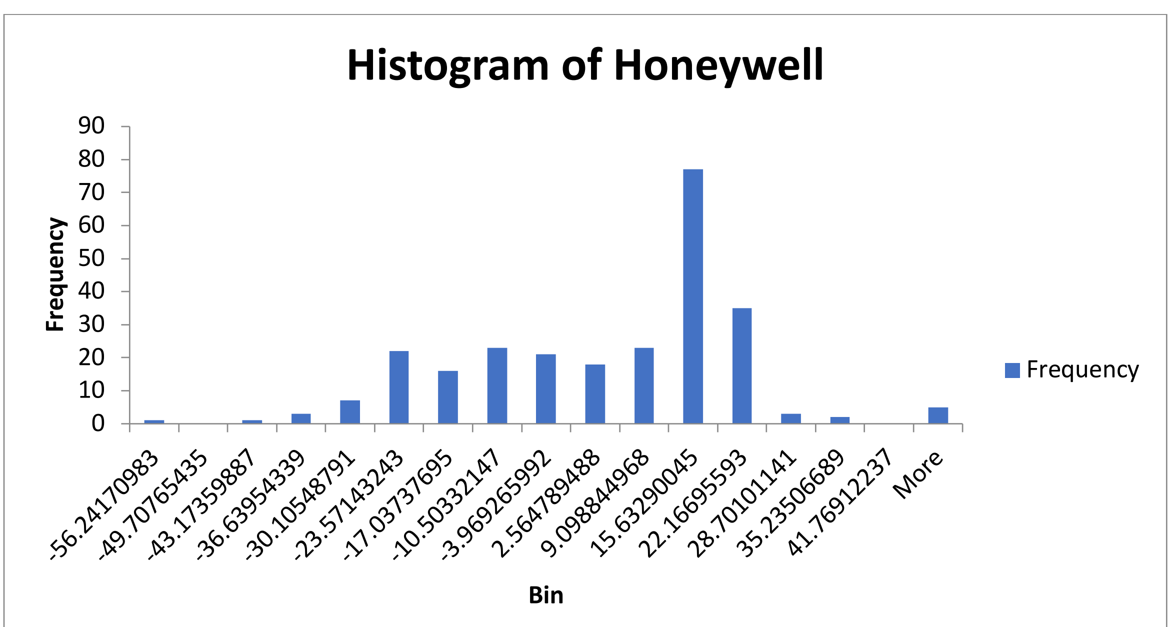
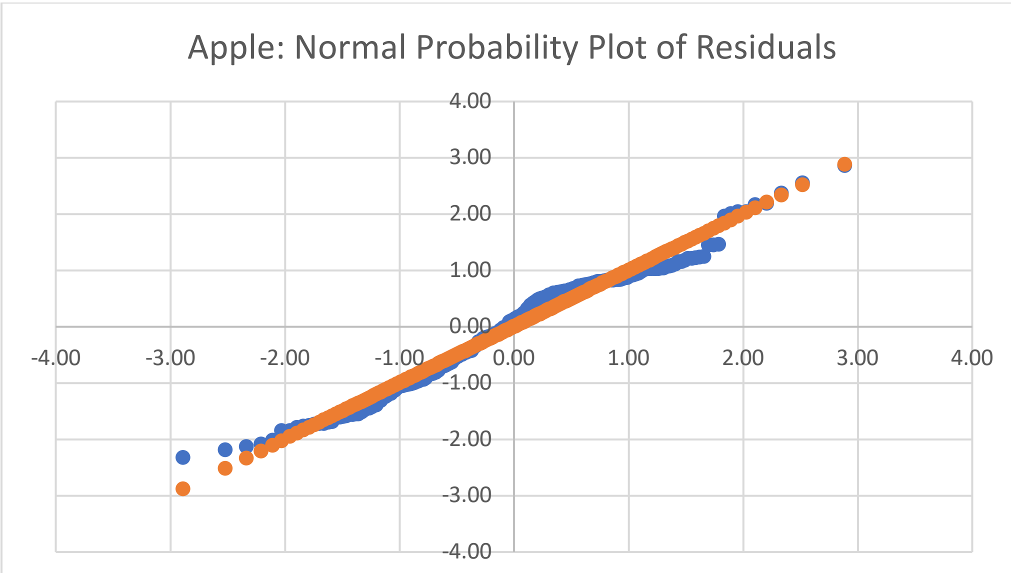
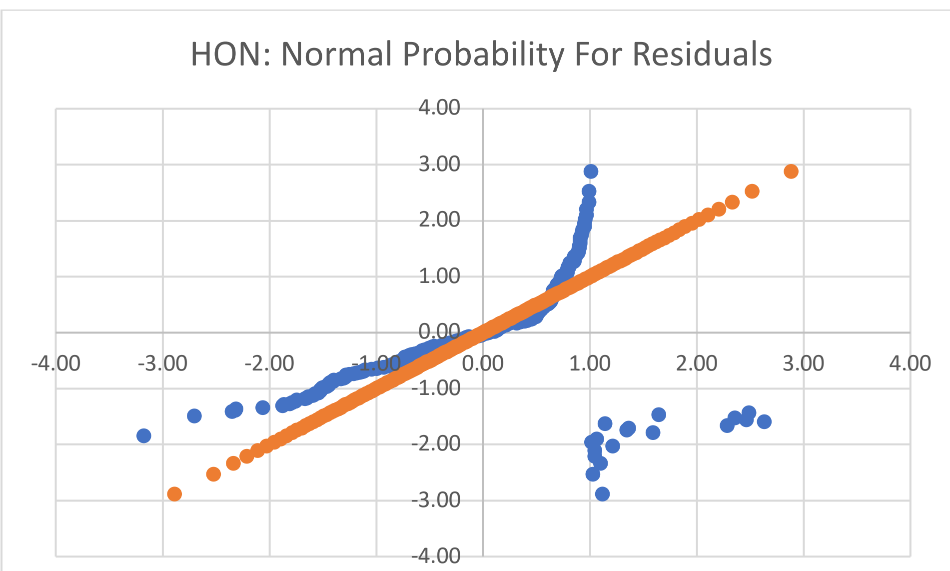


Fig-6: Normal Distribution plots





References:

ALY 6050: Module 3 — Forecasting of Time Series & Regression Analysis

The Open Educator: Normality Test Chi Square Goodness of fit MS Excel; [online]: https://www.youtube.com/watch?v=Zs8IA3K1-4I

Math with Neilu: Weighted Moving Average Forecast in Excel; [online]: https://www.youtube.com/watch?v=YGjYK8swPX8

Nanditha Saravanakumar: Homoscedasticity; [online]: https://www.wallstreetmojo.com/homoscedasticity/

Stats Solutions: Homoscedasticity; [online]: https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/homoscedasticity/