Forecasting Stock Prices
Apple Inc & Honeywell International

Heejae Roh
Northeastern University
ALY 6050: Intro to Enterprise Analytics
Azadeh Mobasher
March 12, 2023

1. INTRODUCTION

In this module, I will forecast stock prices with several methods. Based on the price of past periods I will use simple & adjusted exponential smoothing to predict stock prices of unknown periods which are 253 to 257. In addition to this, I execute the weighted moving average (WMA) and trend prediction based on linear trend. Lastly, I use the linear regression model to predict stock prices. Using the MAPD (Mean Absolute Percentage Deviation) and MAPE (Mean Absolute Percentage Error), I will compare the accuracy of each prediction method. After that, with the linear regression analysis, I will check the residuals' assumptions for regression.

2. ANALYSIS

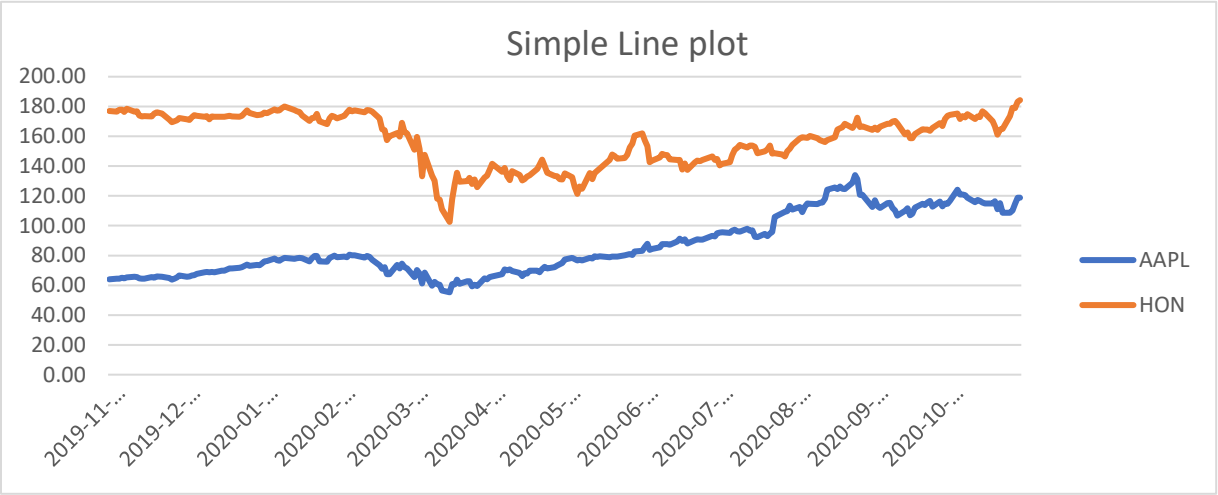
Understanding Key metrics

1. This is the forecasting of two stocks in time series.
2. The two stocks are Apple Inc (AAPL, NASDAQ ticker symbol for Apple Inc) and Honeywell International Inc (HON, NASDAQ ticker symbol for Honeywell International Inc)
3. Data for about 1 year contains 'close price' and 'volume' of each stock. This is a record for 252 market days which are from 2019-11-08 to 2020-11-06.
4. I predict the stock close price. In PART 1, I forecast the 253rd period through short-term forecasting. I use simple moving averages in this process.
5. In PART 2, forecast the 253rd period through long-term forecasting. I use adjusted exponential smoothing in the process. I also forecast until the 257th period using weighted moving averages and trend-based forecasting.
6. In PART 3, I predict data up to the 257th through simple regression. And let's see if the assumption is satisfied through the analysis of the residuals.

PART 1. Short-term Forecasting
(i) Data Analysis & Finding Trends

Understanding of simple moving averages

1. The simple moving average method is a smoothing method based on the idea of averaging random fluctuations in the time series to identify the underlying direction in which the time series is changing.
2. Because the moving average method assumes that future observations will be similar to the recent past, it is most useful as a short-range forecasting method.
3. Specifically, the simple moving average forecast for the next period is computed as the average of the most recent k observations. The value of k is somewhat arbitrary, although its choice affects the accuracy of the forecast.
4. The larger the value of k, the more the current forecast is dependent on older data and the forecast will not react as quickly to fluctuations in the time series (James R, 2013).



Understanding the dataset

1. The graph above is a simple plot of stocks.
2. The most distinctive part is that the shapes drawn by the two graphs are very similar. This is because many stocks track the Nasdaq or S&P 500 indices. Perhaps, in my opinion, it is expected that the stock prices were greatly influenced by the rise and fall of the Nasdaq at the time.
3. Taken together with the descriptive analysis below, Apple stock's price peaked at 133.95, which was recorded on September 1, 2020. The lowest was 55.29, which was recorded on March 23, 2020.
4. Honeywell stock's price peaked at 184.27, which was recorded on November 1, 2020. The lowest was 102.52, which was recorded on March 23, 2020.
5. Seasonal trends do not appear to be present in simple line plots. However, stocks usually tend to rise as the end of the year approaches.
6. The reason for the lowest share price in March 2020 is the impact of Covid-19. This can be regarded as an irregular component. The irregular component (sometimes also known as the residual) is what remains after the seasonal and trend components of a time series have been estimated and removed. It results from short term fluctuations in the series which are neither systematic nor predictable (ABC, 2005).
7. Two stocks can be seen that it rises following quantitative easing of U.S. government after March 2020.

Descriptive Analysis

	n	Mean	Sd	Median	Trimmed	Mad	Min	Max	Range	Skew	Kurtosis	se
Date*	257	129	74.33	129	129	94.89	1	257	256	0.00	-1.21	4.64
Period	257	129	74.33	129	129	94.89	1	257	256	0.00	-1.21	4.64
APPL Price	252	86.63	20.36	79.12	85.34	19.88	55.29	133.95	78.7	0.54	-1.10	1.28
Hon Price	252	157.68	17.21	161.95	159.06	18.53	102.52	184.27	81.8	-0.60	-0.65	1.08

(ii) Simple Exponential Smoothing for period 253

alpha	AAPL price	MAPD	HON price	MAPD
0.15	114.88	3.91%	175.33	2.84%
0.35	115.97	2.48%	179.97	2.19%
0.55	117.50	2.08%	182.58	1.90%
0.75	118.39	1.96%	183.72	1.77%

Interpretation

1. The exponential smoothing method was used to obtain the value of period 253 for 4 alphas.
2. The exponential smoothing formular is S*R + F*(1-S)
   R = most recent period's value
   S = the smoothing factor
   F = the most recent period's forecast
3. To analyze the accuracy of analysis, MAPD is used. MAPD is the sum of absolute difference of forecasted - actual values divided by sum of actual values.
4. For both the APPL and HON stocks, the predictions were most accurate when the alpha was 0.75, when the most recent value was heavily weighted. MAPD was 1.96% and 1.77% respectively, lower in HON.
5. This is because the closing price of a stock has the greatest impact on the opening price the next day. Perhaps, if alpha is given larger, the MAPD is expected to be lowered.

(iii) Adjusted Exponential Smoothing for period 253 with alpha 0.55

beta	AAPL price	MAPE	HON price	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
Actual value	116.32	Actual value	196.99	

Interpretation

1. If the time-series exhibits trend, forecasts based on simple exponential smoothing will lag the trend. In such cases, a variation of simple exponential smoothing called the trend-adjusted Exponential smoothing can be used as a forecasting technique (psu, n.d.).
2. Here, 4 Betas were used for analysis. MAPE was used as a method to check accuracy. MAPE is calculated by adding all the absolute percent errors together and divide the sum by the number of errors (Indeed, 2023).
3. APPL has the lowest MAPE when beta is 0.45 with a price of 119.24 for 253 periods

However, since I used the same alpha, almost all MAPEs came out similar.

4. HON had the lowest MAPE at 1.85% when beta was 0.85. There was difference in MAPE according to beta rather than APPL. This is because the difference between the most recent 2 values has a greater effect as beta increases, so it can be said that the most recent ones are better reflected.
5. On the other hand, in the case of APPL, since the trend sometimes lags, the beta value of 0.45 predicts better than 0.85.

PART 2. Long-term Forecasting
(i) Weighted moving average & Trend forecasting

Headtail of Weighted moving average (WMA)

Date	Period	AAPL price	MAPE	HON price	MAPE
2019-11-13	4	64.33	1.06%	177.31	0.25%
2019-11-14	5	64.72	0.24%	177.55	0.67%
2019-11-15	6	64.67	1.02%	177.08	0.76%
…			…		…
2020-03-31	98	62.45	0.34%	131.02	0.79%
2020-04-01	99	62.38	5.07%	130.96	2.29%
2020-04-02	100	61.04	1.13%	129.65	1.11%

Interpretation

1. I used 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). The Weighted Moving Average (WMA) was calculated from period 4 to period 100.
2. Because of its unique calculation, WMA will follow prices more closely than a corresponding Simple Moving Average (Fidelity, n.d.).
3. Except for sudden stock price changes, MAPE remained within 2%. However, sometimes there was a period with a difference of more than 10% in the case of sudden change.

Headtail of Trend forecasting

Date	Period	AAPL Forecasted	Actual	HON Forecasted	Actual
2020-04-03	101	67.38	62.66	129.80	130.05
2020-04-06	102	67.78	59.37	130.10	132.06
2020-04-07	103	68.18	60.36	130.40	128.04
…		…			..
2020-11-09	253	128.02	116.32	176.05	196.99
2020-11-10	254	128.42	115.97	176.35	201.98
2020-11-11	255	128.81	119.49	176.66	199.29
2020-11-12	256	129.21	119.21	176.96	197.24
2020-11-13	257	129.61	119.26	177.27	201.54

Interpretation

1. I used TREND function in excel and calculated from period 101 to 257.
2. The TREND function returns values along a linear trend. It fits a straight line (using the method of least squares) to the array's known_y's and known_x's. TREND returns the y-values along that line for the array of new_x's that you specify (Microsoft, n.d.).
3. Since this calculates a linear trend, the MAPE is larger than the actual value than when using WMA. The TREND function does not fit well because stocks do not generally rise or fall linearly. And sometimes it stayed above 10% for several days.
4. On average, APPL was close to 5% during trend forecasting. HON recorded close to 3%. However, the last 5 days have been over 10%.

(ii) MAPE & finding most accurate forecast

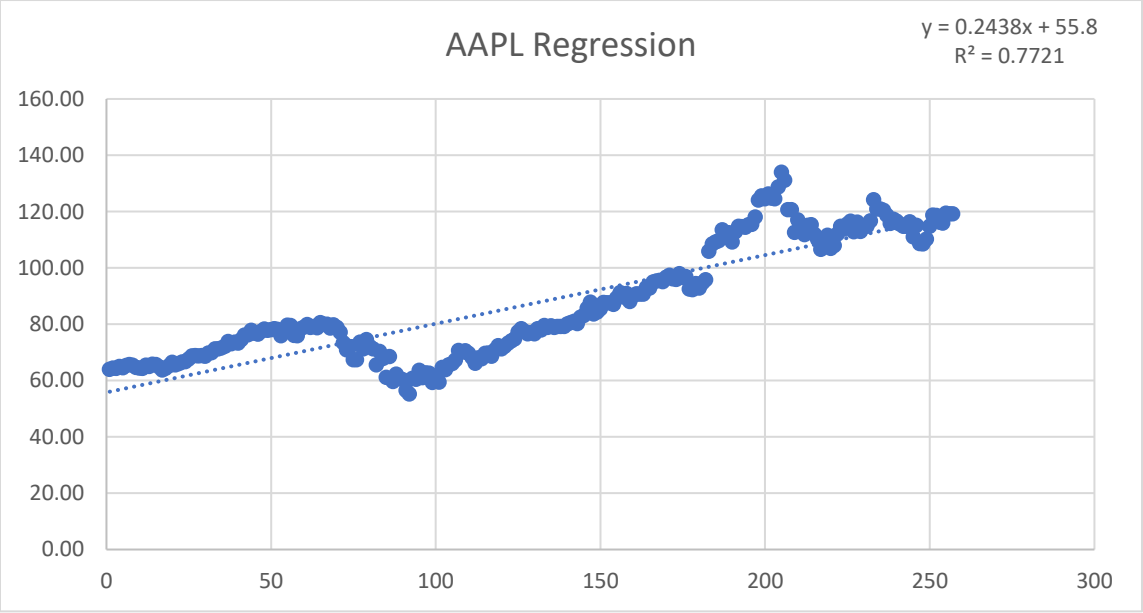
alpha	AAPL price	MAPE (MAPD)	HON price	MAPE (MAPD)
0.75	118.39	1.96% (1.96%)	183.72	1.89% (1.77%)
alpha= 0.55 beta=	AAPL	MAPE	HON	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
WMA + Trend	128.02	3.86%	176.05	2.85%
Actual value	116.32		196.99	

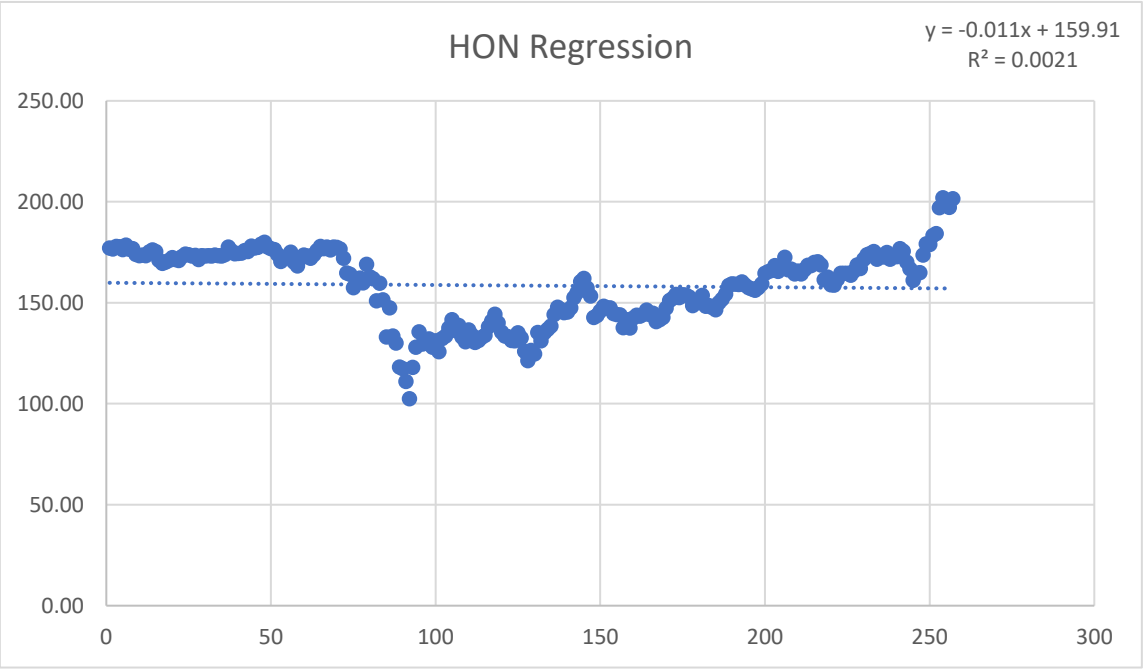
Interpretation

1. In the first exponential smoothing using only alpha, alpha = 0.75 recorded the lowest MAPD. If you recalculate the MAPD recorded the lowest MAPD, MAPE, is 1.96% and 1.89% for APPL and HON, which are equal to or higher than MAPD.
2. This is greater than 1.94% of the APPL calculated with alpha 0.55 and beta 0.45. It is also larger than 1.85% of HON calculated with alpha 0.55 and beta 0.85.
3. Therefore, in this case, the adjusted exponential smoothing method can be said to better to explain the stock price flow.

PART 3. Regression
(i) Simple regression and predict value

Regression Statistics						
	SLOPE	INTERCEPT	Correlation R	Determination	Residual Mean	Residual SD
AAPL	0.2439	55.7850	0.8788	0.7723	0.00	9.8459
HON	-0.0110	159.9104	-0.0455	0.0021	0.00	17.9741





Interpretation

1. If you look at the value of R^2 for APPL, about 77% of the stock price can be explained through linear regression.
2. However, HON's R^2 value is 0.0021, which explains about 0.2%.
3. APPL's share price rises relatively linearly, while HON's regression is almost flat. SLOPE also shows -0.011, and there is not that much difference according to the input of X which is period.

Headtail of Regression forecasting					
Date	Period	AAPL Predicted	Actual	HON Predicted	Actual
2019-11-08	1	56.04	63.95	159.90	177.03
2019-11-11	2	56.29	64.46	159.89	176.66
2019-11-12	3	56.53	64.40	159.88	177.81
…		…			..
2020-11-09	253	117.47	116.32	157.13	196.99
2020-11-10	254	117.72	115.97	157.12	201.98
2020-11-11	255	117.96	119.49	157.10	199.29
2020-11-12	256	118.20	119.21	157.09	197.24
2020-11-13	257	118.45	119.26	157.08	201.54

Period 253 with alpha 0.55 exponential 4 betas + WMA & Trend + Regression

beta	AAPL price	MAPE	HON price	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
WMA + Trend	128.02	3.86%	176.05	2.85%
Regression	117.47	10.38%	157.08	10.03%
Actual value	116.32	Actual value	196.99	

Interpretation

1. Regression models show MAPEs greater than 10%. This indicates that although APPL can linearly explain 77% of stock prices, the error with actual values is large.
2. In HON, values that appeared to be unpredictable in regression also differed by more than 10% from actual values.
3. Personally, even though APPL's regression model showed high R^2, MAPE was high, so it seems difficult to explain APPL's stock price according to regression in the future.
4. Adjusted exponential smoothing method shows the lowest MAPE. Therefore, to

predict the stock price, a forecast using the values closest to the value we want to predict seems more appropriate than regression.

5. Among them, adjusted smoothing with a beta value that utilizes the adjustment of the trend rather than simple exponential smoothing seems most appropriate according to MAPE comparison.

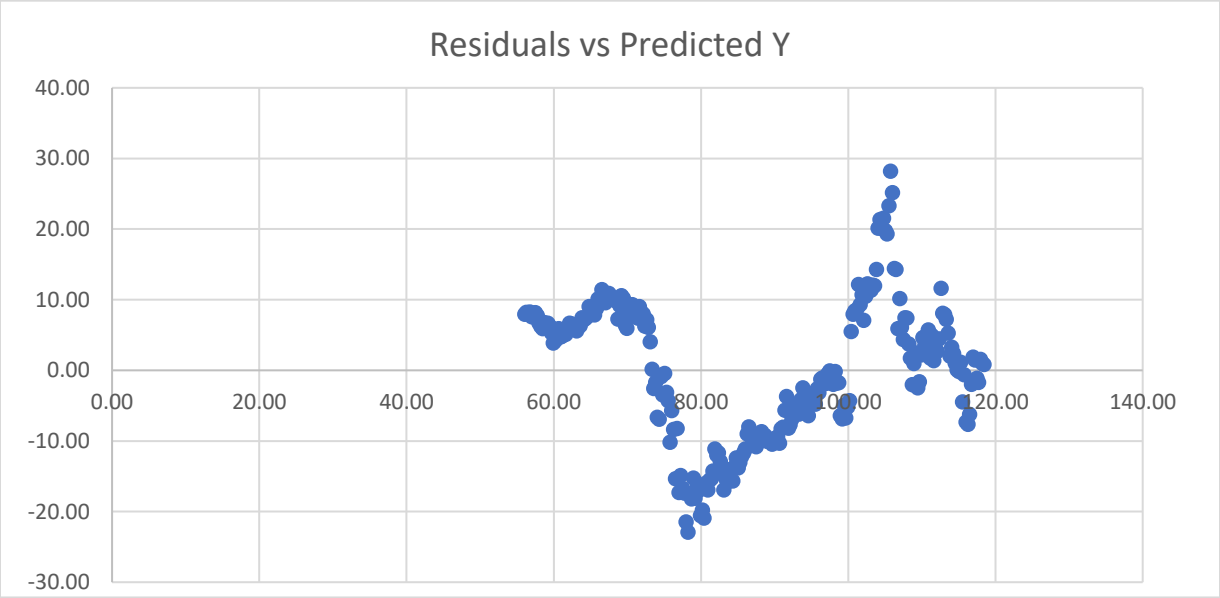
(ii) Residual analysis of Regression – with AAPL

Residual Analysis Understanding

1. We need to check 4 main things to check the regression model's assumptions.
2. They are Linearity, Normality of errors, Homoscedasticity, and Independence of errors.
3. For Linearity, If the model is appropriate, then the residuals should appear to be randomly scattered about zero, with no apparent pattern.

Homoscedasticity Understanding

1. There should not exist pattern in the plot, the residuals must be in the random formation.
2. Check with the plot of the residuals vs the predicted Y values
3. For Homoscedasticity, this can also be evaluated by examining the residual plot and looking for large differences in the variances at different values of the independent variable (James R, 2013).

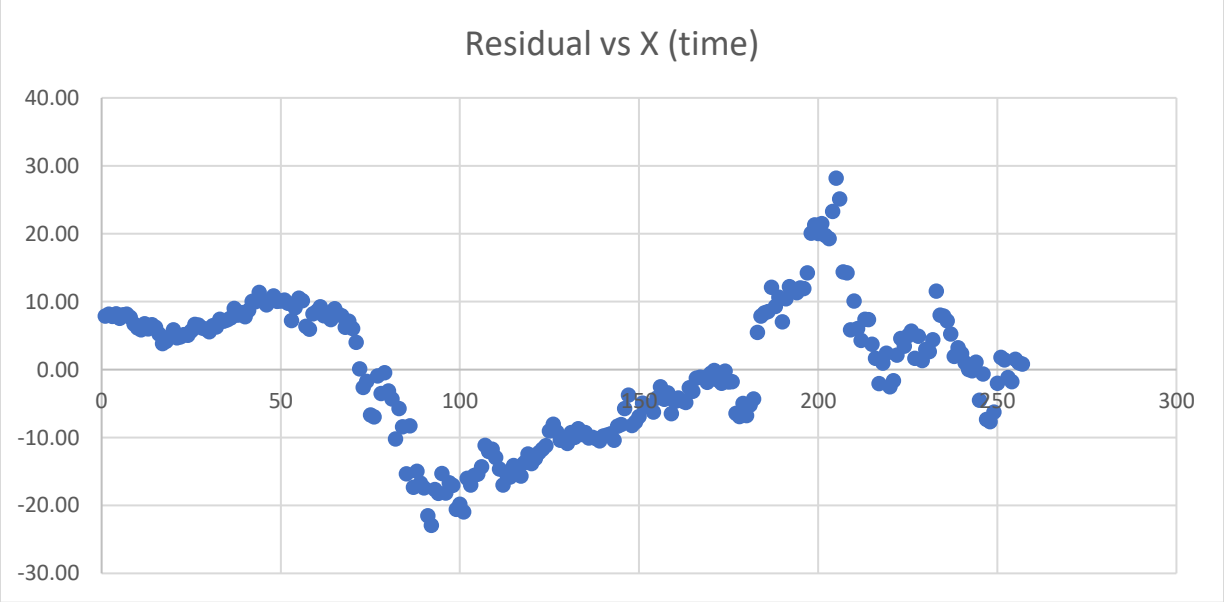


Homoscedasticity interpretation

1. This is a graph showing the relationship between residuals and predicted Y comparing the expected and actual values of APPL stock.
2. There must be no pattern between the two to satisfy the assumption. However, the graph above shows a linear distribution when divided into sections.
3. Therefore, it can be said that homoscedasticity is not satisfied.
4. Homoscedastic (also spelled "homoscedastic") refers to a condition in which the variance of the residual, or error term, in a regression model is constant. That is, the error term does not vary much as the value of the predictor variable changes (Will, 2022).
5. In other words, the variance of the residual, or error term, in a regression model is not constant. Also, the error term varies much as the value of the predictor variable changes.

Independency Understanding

1. There should not be any distinguishing pattern.
2. Check with the plot of the residuals vs the independent variable (in this case, time which is 'period').
4. For Independence of errors, residuals should be independent for each value of the independent variable (James R, 2013).



Independency interpretation

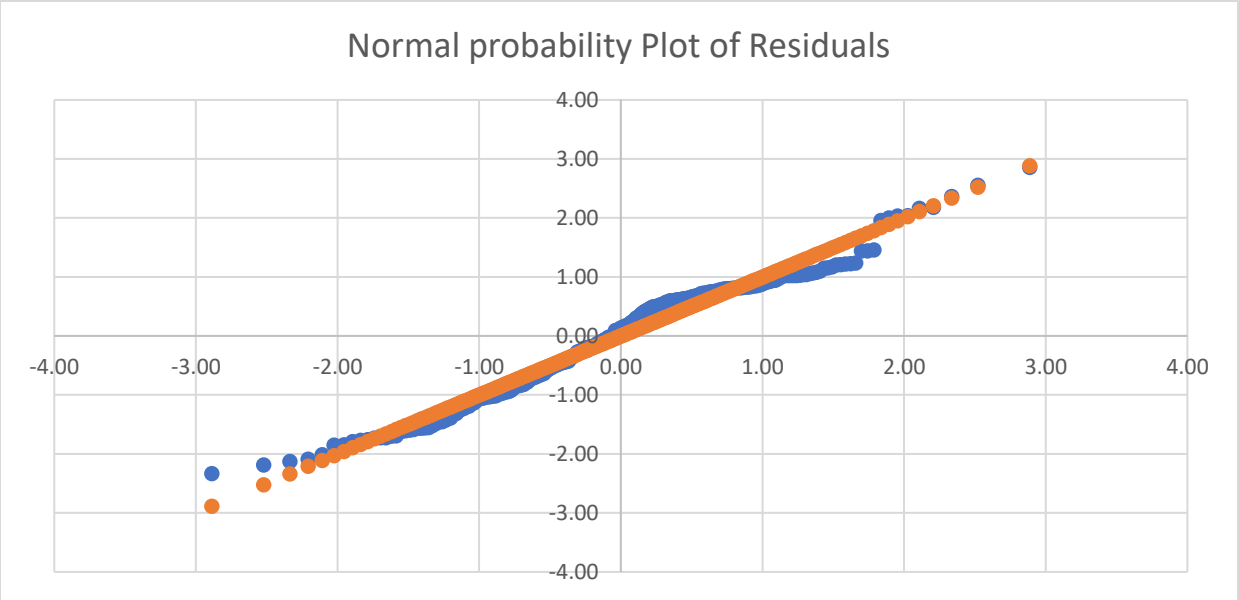
1. This is a graph showing the relationship between X (time) and residuals which is calculated by the expected and actual values.
2. There must be no pattern between the two to satisfy the assumption. However, the

graph above shows a linear distribution when divided into intervals.

3. Therefore, it can be said that independency is not satisfied.
4. The residual is not randomly distributed according to the variable called time.

Normality Understanding

1. Using Normal Probability plot of residuals
2. Chi-square Goodness of fit test for normality
3. For Normality of errors, this can be verified by examining a histogram of the standard residuals and inspecting for a bell-shaped distribution or using more formal goodness-of-fit tests. I use the chi-square goodness-of-fit test (James R, 2013).



Normal probability interpretation

1. I first check the normal probability. This is to standardize residuals to see if they follow normality, and then change them to Z values using rank to look at the tendency.
2. If the blue points are distributed similarly to the orange points, then it follows normality.
3. Looking at the graph above, normality of residuals is satisfied.
4. A straight, diagonal line in a normal probability plot indicating normally distributed data. A straight, diagonal line means that you have normally distributed data. If the line is skewed to the left or right, it means that you do not have normally distributed data (StatisticsHowTo, n.d.).

Chi-squared Goodness of fit test for normality

Step 0 Finding Key-Metrics
This test is to check which the normality of residuals with normal distribution

Step 1 Hypothesis.
Null: The residuals are normally distributed
Alternative: The residuals are not normally distributed

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

Chi-squared Test Statistic:	64.58
Chi-squared P-value:	8.70092E-08
DF	16

- The p-value with 1-CHISQ.DIST() and CHISQ.DIST.RT() in excel is 0.00

Step 4 Make the decision.

- There is enough evidence to reject null hypothesis since 0.00 (p-value) < 0.05

Step 5 Summarize the results.

- Don't accept that the residuals are not normally distributed

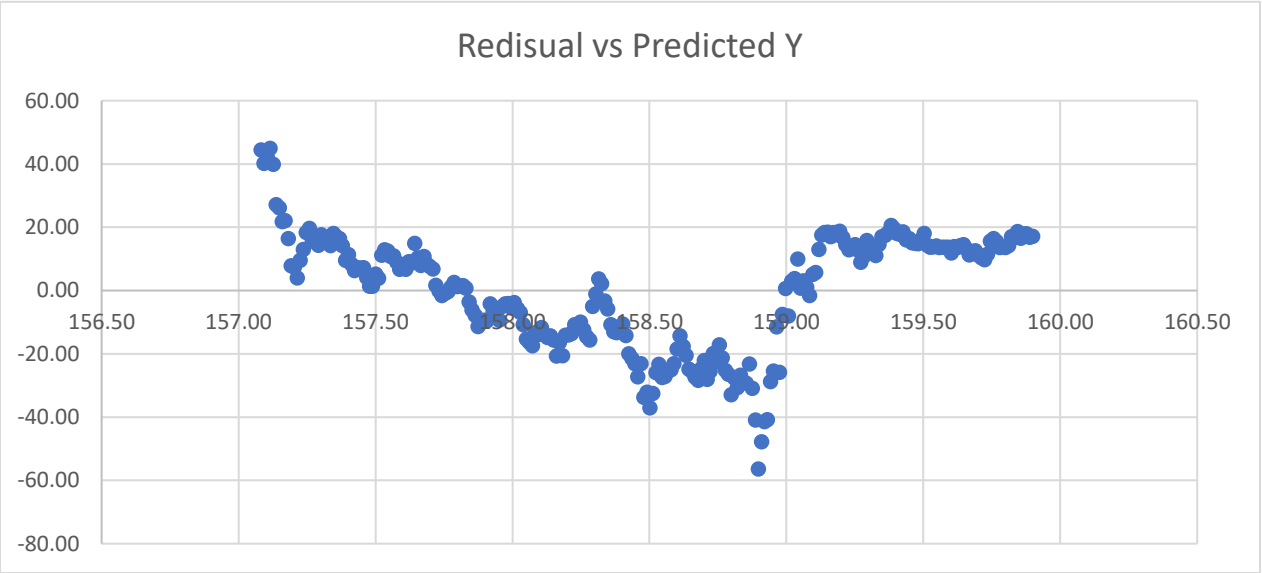
Interpretation

1. Chi-squared test statistic is SUM of 17 bins (Expected – oberseved)^2/ Expected
2. Df = Number of Bins -1
3. Number of observations are 17. Therefore, DF is 17 minus 1 is 16
4. P-value calculated by 1-CHISQ.DIST() is 0.00

(ii) Residual analysis of Regression – with HON

Homoscedasticity Understanding

1. There should not exist pattern in the plot, the residuals must be in the random formation
2. Check with the plot of the residuals vs the predicted Y values

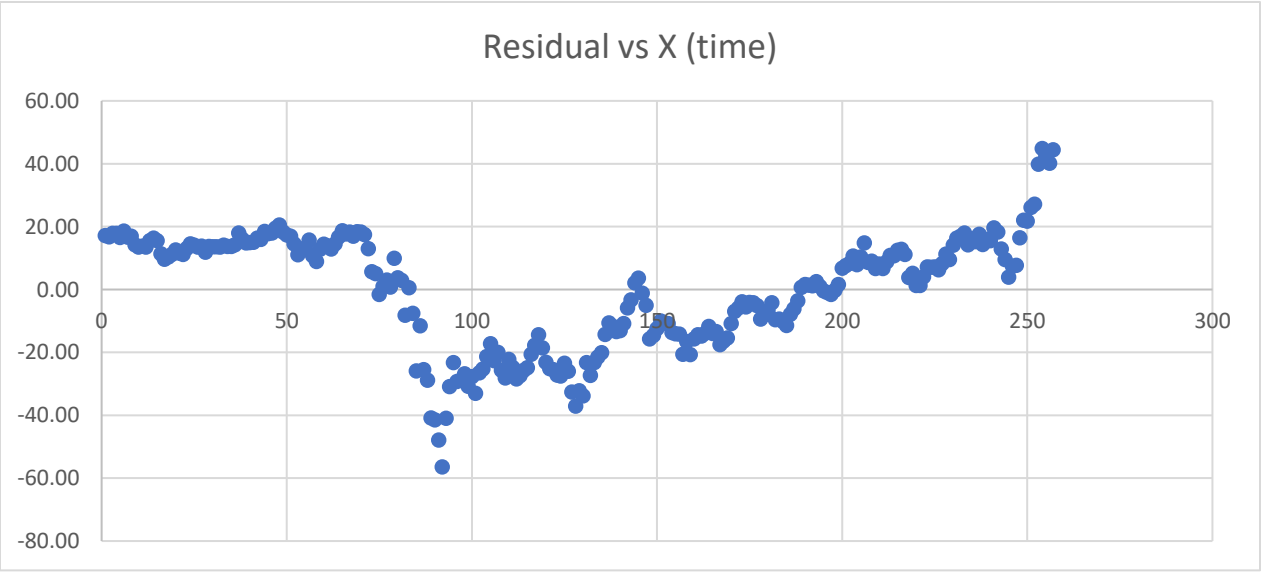


Homoscedasticity interpretation

1. This is a graph showing the relationship between residuals and predicted Y comparing the expected and actual values of APPL stock.
2. There must be no pattern between the two to satisfy the assumption. However, the graph above shows a linear distribution when divided into sections.
3. Therefore, it can be said that homoscedasticity is not satisfied.

Independency Understanding

1. There should not be any distinguishing pattern.
2. Check with the plot of the residuals vs the independent variable (in this case, time which is 'period').

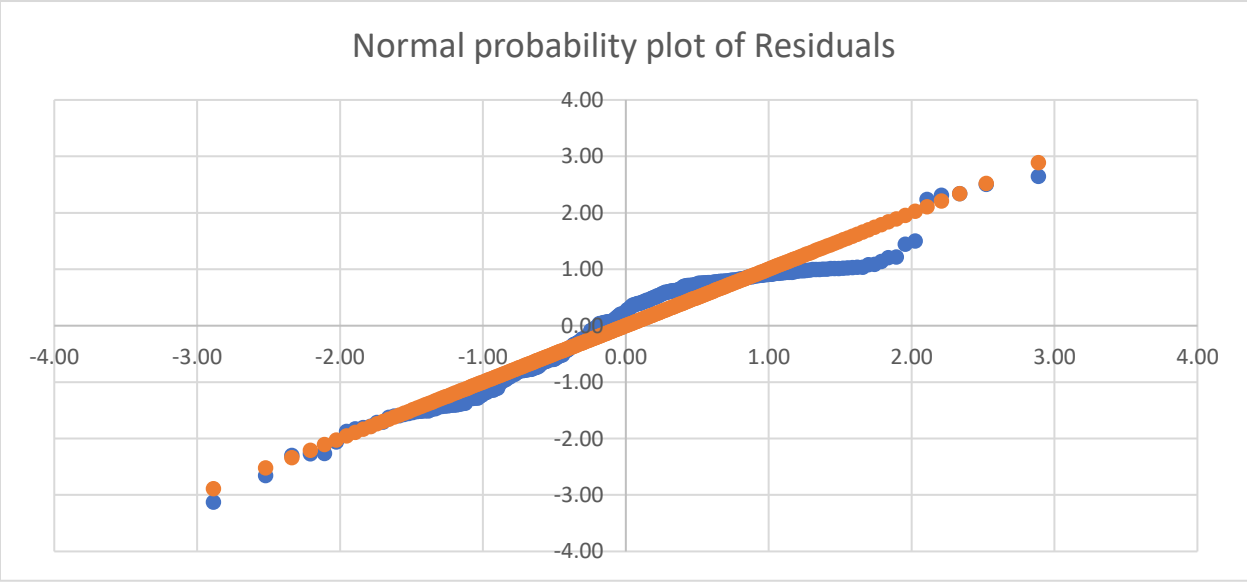


Independency interpretation

1. This is a graph showing the relationship between X (time) and residuals which is calculated by the expected and actual values.
2. There must be no pattern between the two to satisfy the assumption. However, the graph above shows a linear distribution when divided into intervals.
3. Therefore, it can be said that independency is not satisfied.
4. The residual is not randomly distributed according to the variable called time.

Normality Understanding

1. Using Normal Probability plot of residuals
2. Chi-square Goodness of fit test for normality



Independency interpretation

1. This is to standardize residuals to see if they follow normality, and then change them to Z values using rank to look at the tendency.
2. If the blue points are distributed similarly to the orange points, then it follows normality.
3. Looking at the graph above, normality of residuals is satisfied.

Chi-squared Goodness of fit test for normality

Step 0 Finding Key-Metrics

This test is to check which the normality of residuals with normal distribution

Step 1 Hypothesis.

Null: The residuals are normally distributed
Alternative: The residuals are not normally distributed

Step 2 Find the critical value.

- The p-value is $\alpha = 0.05$

Step 3 Compute the test value.

Chi-squared Test Statistic:	95.12
Chi-squared P-value:	2.82245E-13
DF	16

- The p-value with 1-CHISQ.DIST() and CHISQ.DIST.RT() in excel is 0.00

Step 4 Make the decision.

- There is enough evidence to reject null hypothesis
  since 0.00 (p-value) < 0.05

Step 5 Summarize the results.

- Don't accept that the residuals are not normally distributed

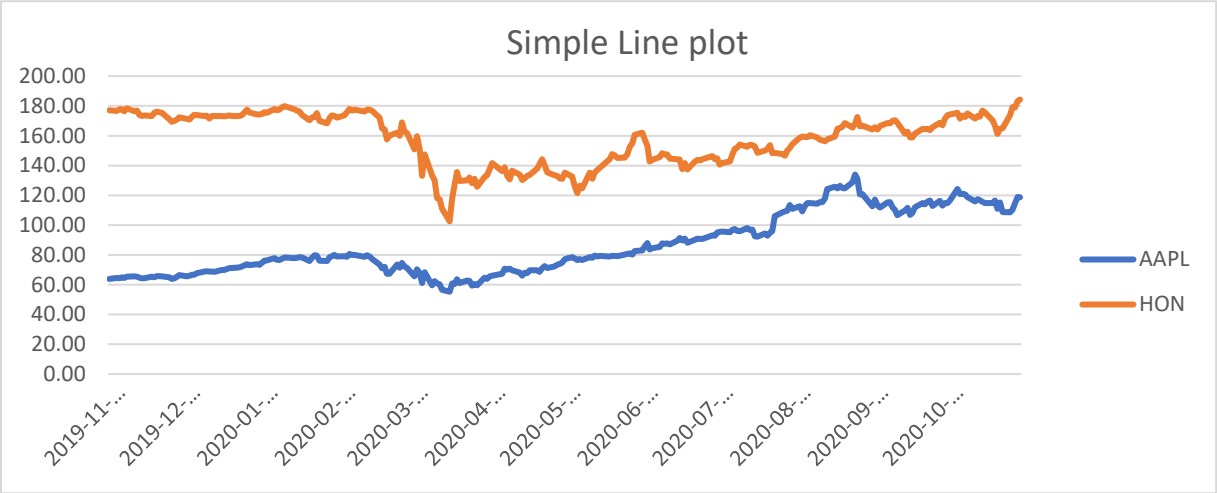
Interpretation

1. Chi-squared test statistic is SUM of 17 bins (Expected – oberseved)^2/ Expected
2. Df = Number of Bins -1
3. Number of observations are 17. Therefore, DF is 17 minus 1 is 16
4. P-value calculated by 1-CHISQ.DIST() is 0.00

3. Answers for Questions

PART 1: Short-term Forecasting

1. Use a simple line plot of both time series. Write a summary of your observations.



Apple stock's price peaked at 133.95, which was recorded on September 1, 2020. The lowest was 55.29, which was recorded on March 23, 2020. 4. Honeywell stock's price peaked at 184.27, which was recorded on November 1, 2020. The lowest was 102.52, which was recorded on March 23, 2020. Seasonal trends do not appear to be present in simple line plots. However, stocks usually tend to rise as the end of the year approaches. The reason for the lowest share price in March 2020 is the impact of Covid-19. This can be regarded as an irregular component. Two stocks can be seen that it rises following quantitative easing of U.S. government after March, 2020.

2. Perform exponential smoothing to forecast both prices for period 253. Next, calculate the MAPD (Mean Absolute Percentage Deviation) of each forecast; determine the value of α that has yielded the most accurate forecast for each stock

(ii) Simple Exponential Smoothing for period 253

alpha	AAPL price	MAPD	HON price	MAPD
0.15	114.88	3.91%	175.33	2.84%
0.35	115.97	2.48%	179.97	2.19%
0.55	117.50	2.08%	182.58	1.90%
0.75	118.39	1.96%	183.72	1.77%

The exponential smoothing formular is S*R + F*(1-S). R = most recent period's value, S = the smoothing factor, and F = the most recent period's forecast. To analyze the accuracy of analysis, MAPD is used. MAPD is the sum of absolute

difference of forecasted - actual values divided by sum of actual values. For both the APPL and HON stocks, the predictions were most accurate when the alpha was 0.75, when the most recent value was heavily weighted. MAPD was 1.96% and 1.77% respectively, lower in HON. This is because the closing price of a stock has the greatest impact on the opening price the next day. Perhaps, if alpha is given larger, the MAPD is expected to be lowered.

3. Use an adjusted exponential smoothing forecast for both prices for period 253. Next, calculate the MAPEs. Explain which values of β have yielded the most accurate forecasts.

(iii) Adjusted Exponential Smoothing for period 253 with alpha 0.55

beta	AAPL price	MAPE	HON price	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
Actual value	116.32	Actual value	196.99	

APPL has the lowest MAPE when beta is 0.45 with a price of 119.24 for 253 periods However, since I used the same alpha, almost all MAPEs came out similar. HON had the lowest MAPE at 1.85% when beta was 0.85. There was difference in MAPE according to beta rather than APPL. This is because the difference between the most recent 2 values have a greater effect as beta increases, so it can be said that the most recent ones are better reflected. On the other hand, in the case of APPL, since the trend sometimes lags, the beta value of 0.45 predicts better than 0.85.

PART 2: Long-term Forecasting

1. Describe how accurate this method of forecasting with Weighted moving average & linear trend function has been by comparing the forecasted values for periods 253-257 with their actual “Close” values on those specific days.

Headtail of Weighted moving average (WMA)					
Date	Period	AAPL price	MAPE	HON price	MAPE
2019-11-13	4	64.33	1.06%	177.31	0.25%
2019-11-14	5	64.72	0.24%	177.55	0.67%
2019-11-15	6	64.67	1.02%	177.08	0.76%
…			…		…
2020-03-31	98	62.45	0.34%	131.02	0.79%
2020-04-01	99	62.38	5.07%	130.96	2.29%
2020-04-02	100	61.04	1.13%	129.65	1.11%

Used 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). The Weighted Moving Average (WMA) was calculated from period 4 to period 100. Except for sudden stock price changes, MAPE remained within 2%. However, sometimes there was a period with a difference of more than 10% in the case of sudden change.

Headtail of Trend forecasting

Date	Period	AAPL Forecasted	Actual	HON Forecasted	Actual
2020-04-03	101	67.38	62.66	129.80	130.05
2020-04-06	102	67.78	59.37	130.10	132.06
2020-04-07	103	68.18	60.36	130.40	128.04
…		…			..
2020-11-09	253	128.02	116.32	176.05	196.99
2020-11-10	254	128.42	115.97	176.35	201.98
2020-11-11	255	128.81	119.49	176.66	199.29
2020-11-12	256	129.21	119.21	176.96	197.24
2020-11-13	257	129.61	119.26	177.27	201.54

Used TREND function in excel and calculated from period 101 to 257. Since this calculates a linear trend, the MAPE is larger than the actual value than when using WMA. The TREND function does not fit well because stocks do not generally rise or fall linearly. And sometimes it stayed above 10% for several days. On average, APPL was close to 5% during trend forecasting. HON recorded close to 3%. However, the last 5 days have been over 10%.

2. Calculate the MAPEs (Mean Absolute Percentage Error) of your forecasts and compare them with the values obtained for your forecasts in Part 1. Describe which method has yielded a most accurate forecast.

MAPE & Finding most accurate forecast for 253 values

alpha	AAPL price	MAPE (MAPD)	HON price	MAPE (MAPD)
0.75	118.39	1.96% (1.96%)	183.72	1.89% (1.77%)
alpha= 0.55 beta=	AAPL	MAPE	HON	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
WMA + Trend	128.02	3.86%	176.05	2.85%
Actual value	116.32		196.99	

In the first exponential smoothing using only alpha, alpha = 0.75 recorded the lowest MAPD. If you recalculate the MAPD recorded the lowest MAPD, MAPE, is 1.96% and 1.89% for APPL and HON, which are equal to or higher than MAPD. This is greater than 1.94% of the APPL calculated with alpha 0.55 and beta 0.45. It is also larger than 1.85% of HON calculated with alpha 0.55 and beta 0.85. Therefore, in this case, the adjusted exponential smoothing method can be said to better explain the stock price flow.

PART 3: Regression

1. Each stock use simple regression of stock values versus the time periods to predict its values for periods 1 through 257. Describe how the accuracy of this prediction has been compared to the methods used in Parts 1 and 2 of this project.

Regression Statistics

	SLOPE	INTERCEPT	Correlation R	Determination	Residual Mean	Residual SD
AAPL	0.2439	55.7850	0.8788	0.7723	0.00	9.8459
HON	-0.0110	159.9104	-0.0455	0.0021	0.00	17.9741

Headtail of Regression forecasting

Date	Period	AAPL Predicted	Actual	HON Predicted	Actual
2019-11-08	1	56.04	63.95	159.90	177.03
2019-11-11	2	56.29	64.46	159.89	176.66
2019-11-12	3	56.53	64.40	159.88	177.81
…		…			..
2020-11-09	253	117.47	116.32	157.13	196.99
2020-11-10	254	117.72	115.97	157.12	201.98
2020-11-11	255	117.96	119.49	157.10	199.29
2020-11-12	256	118.20	119.21	157.09	197.24
2020-11-13	257	118.45	119.26	157.08	201.54

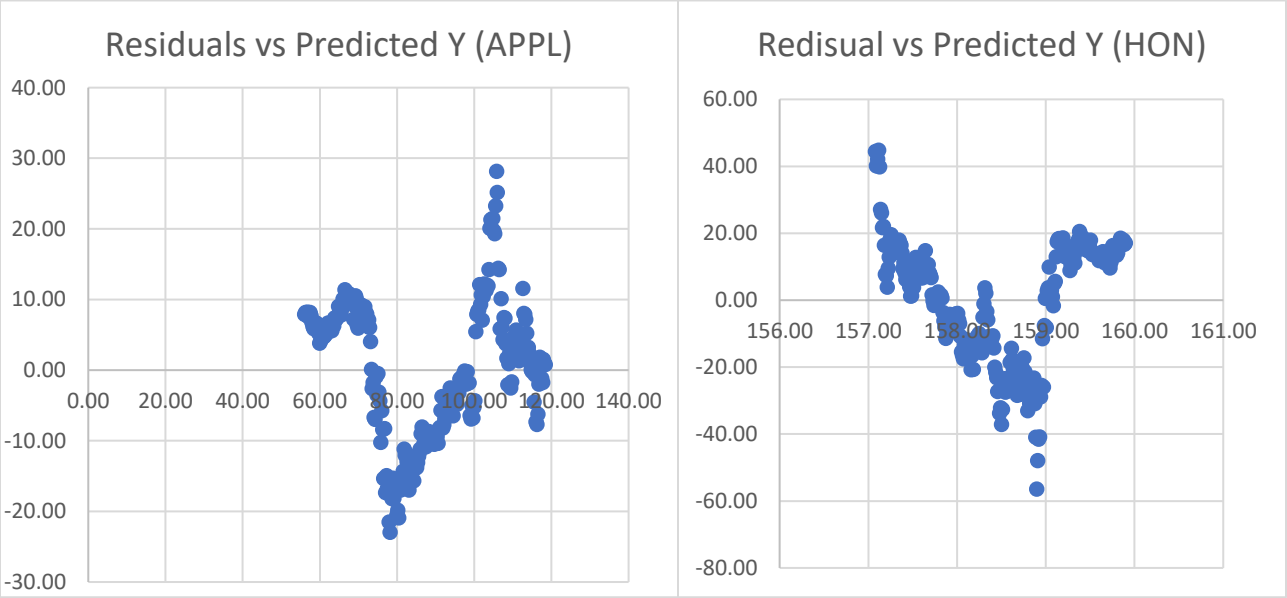
Period 253 with 4 betas (alpha 0.55) + WMA & Trend + Regression

beta	AAPL price	MAPE	HON price	MAPE
0.15	118.04	1.97%	184.07	1.98%
0.25	118.54	1.95%	184.72	1.94%
0.45	119.24	1.94%	185.19	1.89%
0.85	119.22	1.98%	184.83	1.85%
WMA + Trend	128.02	3.86%	176.05	2.85%
Regression	117.47	10.38%	157.08	10.03%
Actual value	116.32	Actual value	196.99	

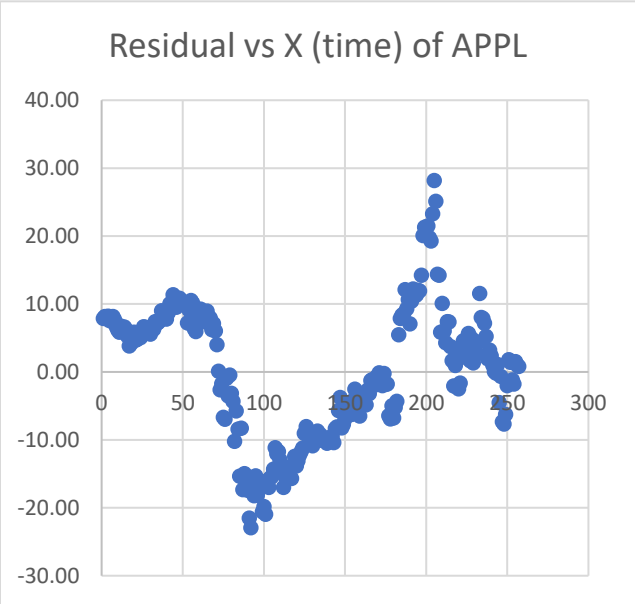
Regression models show MAPEs greater than 10%. This indicates that

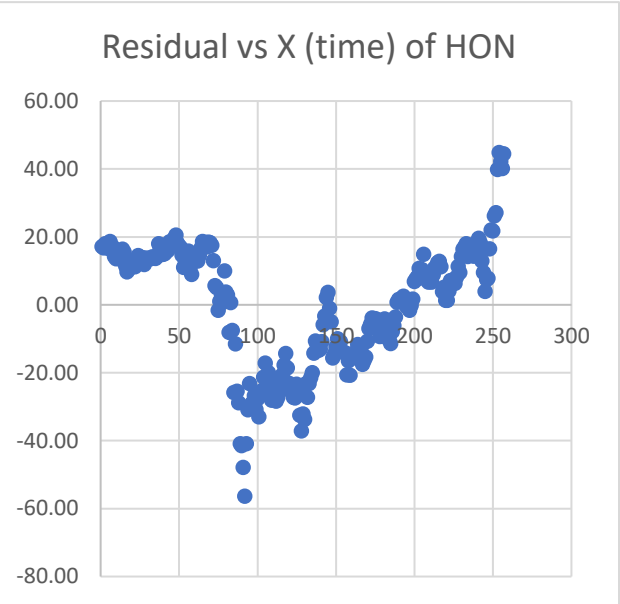
although APPL can linearly explain 77% of stock prices, the error with actual values is large. In HON, values that appeared to be unpredictable in regression also differed by more than 10% from actual values. Personally, even though APPL's regression model showed high R^2, MAPE was high, so it seems difficult to explain APPL's stock price according to regression in the future. Adjusted exponential smoothing method shows the lowest MAPE. Therefore, to predict the stock price, a forecast using the values closest to the value we want to predict seems more appropriate than regression. Among them, adjusted smoothing with a beta value that utilizes the adjustment of the trend rather than simple exponential smoothing seems most appropriate according to MAPE comparison.

2. <u>Perform a residual analysis of simple regression to verify whether regression is appropriate to use for each data (Homoscedasticity, and Independency)</u>



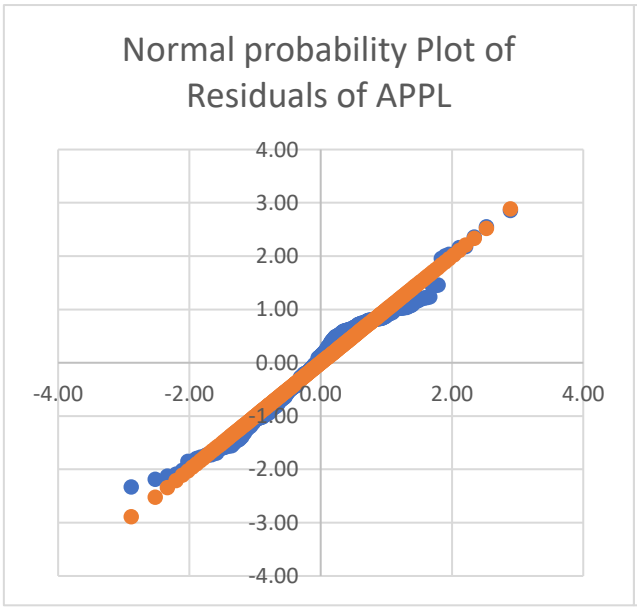
This is a graph showing the relationship between residuals and predicted Y comparing the expected and actual values of APPL & HON stocks. There must be no pattern between the residuals and predicted Y to satisfy the assumption. However, the graph above shows a linear distribution when divided into sections. Therefore, it can be said that homoscedasticity is not satisfied in both stocks.
In other words, the variance of the residual, or error term, in a regression model is not constant. Also, the error term varies much as the value of the predictor variable changes. This is a graph showing the relationship between X (time) and residuals which is calculated by the expected and actual values. There must be no pattern between the two to satisfy the assumption. However, the graphs above show a linear distribution when divided into intervals. Therefore, it can be said that independency is not satisfied. The residual is not randomly distributed according to the variable called time.

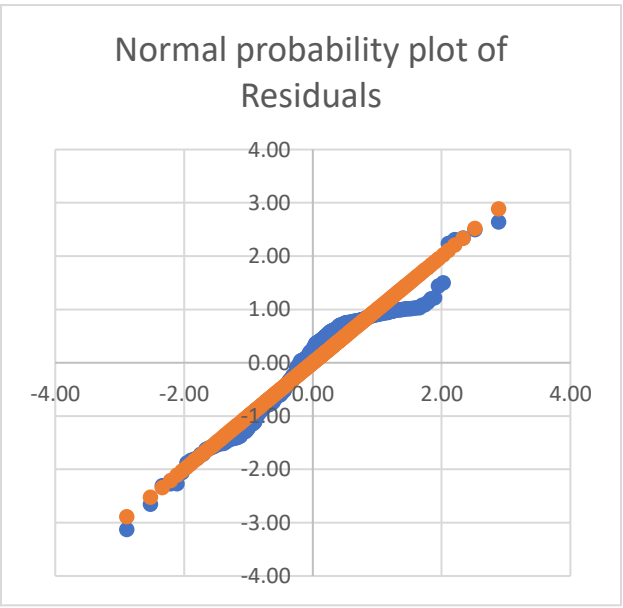




This is a graph showing the relationship between X (time) and residuals which is calculated by the expected and actual values. There must be no pattern between the two to satisfy the assumption. However, the graphs above show a linear distribution when divided into intervals. Therefore, it can be said that independency is not satisfied. The residual is not randomly distributed according to the variable called time.

3. <u>Perform a residual analysis of simple regression to verify whether regression is appropriate to use for each data (Normality with probability plot and chi-square test)</u>





This is to standardize residuals to see if they follow normality, and then change them to Z values using rank to look at the tendency. If the blue points are distributed

similarly to the orange points, then it follows normality. Looking at the graphs above, normality of residuals is satisfied.

	APPL	HON
Chi-squared Test Statistic:	64.58	95.12
Chi-squared P-value:	8.70092E-08	2.82245E-13
DF	16	16

P-value calculated by 1-CHISQ.DIST() is 0.00 in both stocks' prediction residuals. Therefore, there is not enough evidence to reject the null hypothesis that the residuals are not normally distributed.

Additional question: Portfolio

1. Suppose that you have decided to form a portfolio 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?

   I finally decided that when the alpha of adjusted exponential smoothing is 0.55, the beta value is 0.45 as the forecast value of APPL. Also, the beta value of 0.85 was set as the HON stock value. I will write a portfolio assuming that I only know the actual value up to the 252nd period. The price of Apple in the 252nd period is 118.01. The price of HON is 183.78. APPL has a difference of -0.58 from the actual stock price of 118.69, and HON has a difference of -0.49 from the actual stock price of 184.27. As for the ratio of the difference to the total share price, APPL is 0.49% more undervalued and HON is 0.27% more undervalued. Taking the two undervalued ratios is 64.5 (APPL): 35.5 (HON). Therefore, I will propose a portfolio investing 64.5% in Apple and 35.5% in HON on day 252.

Final Portfolio of Two Stocks		
	APPL	HON
Portfolio	64.5%	35.5%
Disparity rate against actual value	-0.49%	-0.27%

4. CONCLUSION

I learned about various forecasting methods in this module. Stock price was predicted through simple & adjusted exponential smoothing. And predicted values were obtained by using WMA and Trend function. Through MAPE, it was found that adjusted exponential smoothing is the most accurate method in this case. This is due to the nature of data ‘stock’. For stock, the close price of the previous day has the greatest impact on the next day's price. However, in the case of APPL, a lag trend was found in adjusted exponential smoothing. I also found that the regression analysis was not accurate in these two stocks’ analysis. This is presumably because stocks rarely have a steady rise or fall.

Finally, I thought about how to construct a portfolio of two stocks. I assumed that I had actual values only up to the 252nd and calculated the disparity rate based on the most accurate forecast method which was adjusted exponential smoothing. And through this, based on period 252, I finally concluded that we should invest 64.5% and 35.5% in APPL and HON, respectively.

REFERENCE

James R, Evans. (2013). Statistics, Data Analysis, and Decision Modeling, 5/e. Pearson.

ABC. (2005). Time series analysis: the basics. ABS (Australian Bureau of Statistics). Retrieved from https://www.abs.gov.au/websitedbs/d3310114.nsf/home/time+series+analysis:+the+basics

ZACH. (March 6, 2020). How to find weighted moving averages in excel. STATOLOGY. Retrieved from https://www.statology.org/weighted-moving-average-excel/

Dave, Piasecki. (n.d.). Exponential smoothing explained. INVENTORYOPS.COM. Retrieved from https://www.inventoryops.com/articles/exponential-smoothing.html

psu. (n.d.). Mangt 515: cost and value management. Retrieved from https://courses.worldcampus.psu.edu/welcome/mangt515/lesson02_13.html

Indeed. (February 3, 2023) What Is MAPE? a guide to mean absolute percentage error. Retrieved from https://www.indeed.com/career-advice/career-development/what-is-mape

Fidelity. (n.d.). Weighted moving average (WMA). Retrieved from https://www.fidelity.com/learning-center/trading-investing/technical-analysis/technical-indicator-guide/wma

Microsoft. (n.d.). TREND function. Retrieved from https://support.microsoft.com/en-us/office/trend-function-e2f135f0-8827-4096-9873-9a7cf7b51ef1

Will, Kenton. (December 31, 2022). Homoskedastic: what it means in regression modeling, with example. Investopedia. Retrieved from https://www.investopedia.com/terms/h/homoskedastic.asp

StatisticsHowTo. (n.d.). Normal probability plot: definition, examples. Retrieved from https://www.statisticshowto.com/normal-probability-plot/