INTRODUCTION TO ENTERPRISE ANALYTICS

ALY6050, WINTER 2020

MODULE 3 PROJECT ASSIGNMENT

PRICE PREDICTIONS PROJECT

SUBMITTED BY: SHIVANI ADSAR
NUID: 001399374
SUBMITTED TO: PROF. RICHARD HE
DATE: 01/27/2020

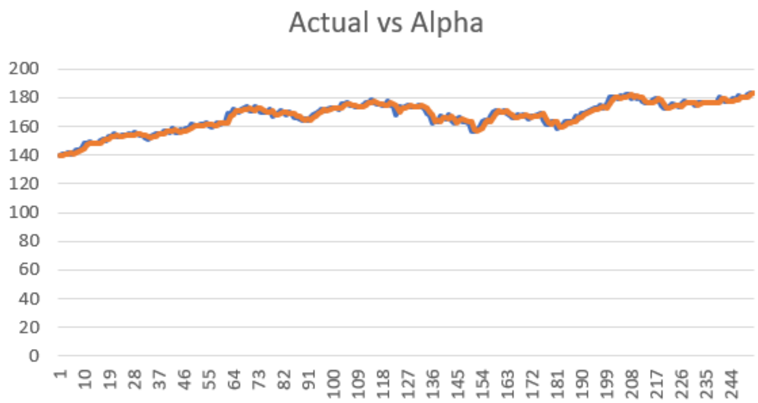
Introduction

The assignment provides exposure on performing regression and time series forecasting techniques. We have used the "Honeywell" dataset for performing regressions and time series forecasting. Moreover, we have calculated the sensitivity analysis Mean Square Error by using parameter values mentioned and calculating mean square errors for those values. We have measured the homoscedasticity using the residuals and their standardized values. These observations have helped in concluding the interpretations for the data.

Analysis

Q1. Forecasting Using Exponential Smoothening

- Considering the smoothening factor (α), we have calculated the forecasts for exponential smoothening. There are two methods for predicting the accuracy of the forecasts, the exponential smoothening and the Adjusted Exponential Smoothening. Using the below formula for exponential smoothening, we have calculated the forecasts:
  $F_t = \alpha D_{t-1} + (1-\alpha) F_{t-1}$, where D_{t-1} is the observed value and F_{t-1} is the forecasted value
- The equation shows the weighted average of the observed value during previous time and the forecasted value. We have calculated the forecasted values for successive values of 0.15, 0.35, 0.55 and 0.75 for the smoothening parameter α. We have performed exponential smoothening to forecast on the Honeywell stock prices for 1/21/2020.
- The Exponential Absolute values of each of the forecasts have been calculated using the formula, $(\text{Actual value} - \text{Predicted } \alpha \text{ Value})^2$, these mean square values have been calculated for all the values of α.
- In order to measure the accuracy of our forecasts, we have calculated the values using mean square errors. The mean square error values have been calculated by calculating the average of the absolute errors values.
- As it can be seen that, the value of α as 0.75 has predicted the most accurate value, as the Mean Square Error for that value is the least.

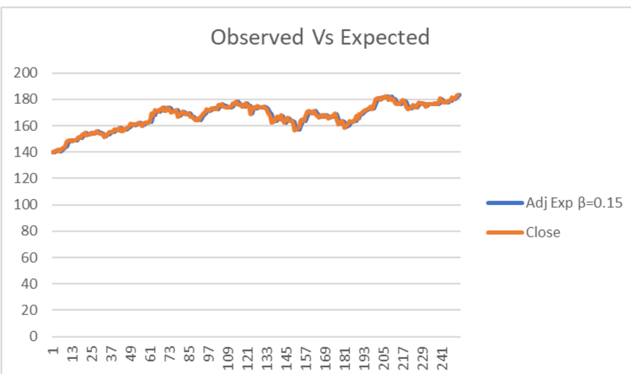


Observations: The graph shows the accuracy for the value of α as 0.75. It shows the graph between the actual value and the predicted value.

- The important characteristic about exponential smoothening forecast is that, the construction is based on the characteristic that it captures the seasonal or cyclic behavior of time series.

Q2 Adjusted Exponential Smoothening

- The adjusted exponential is one of the efficient ways to calculate the forecasts for the dataset, given by the formula:
- $A_{ft} = F_t + T_t$, where: F_t is the exponential smoothening forecast and T_t is the trend component
- We need to calculate the trend by using,
  $T_t = \beta (F_t - F_{t-1}) + (1-\beta) T_{t-1}$, where β is the normal between 0 and 1,
  $F_t - F_{t-1}$: Difference of exponential forecasting
  T_{t-1}: Trend during previous time
- The adjusted exponential smoothening provides more accurate value of the time series. In the given problem, we have calculated the adjusted exponential smoothening on the Honeywell stock prices for the successive values of 0.15, 0.25, 0.45 and 0.85 for the trend factor β using the exponential smoothening forecast as 0.75.
- The adjusted β for the successive values, has been calculated using the formula, (Trend + α) value.
- The adjusted error for each of the successive β values, has been calculated using: (Actual-Adjusted)^2, these values have been used to calculate the mean square errors for each value of β.
- It can be seen that the forecast is more accurate for the β value of 0.25 as the mean square error value for it is the least.



Observations: It can be interpreted that the original and forecasted values are very close and shows the accuracy.

Q3. Regression Analysis

- We have performed regression analysis on the given data for forecasting the value for 1/21/2020.
- Regression shows the associations between two or more variables. We have used the following formula for calculating the regression, $y = (b_o * x) + b_1$, where
  y : Predicted Value
  b_o: Slope, gives the steepness
  b_1: intercept, where the line cuts the y axis

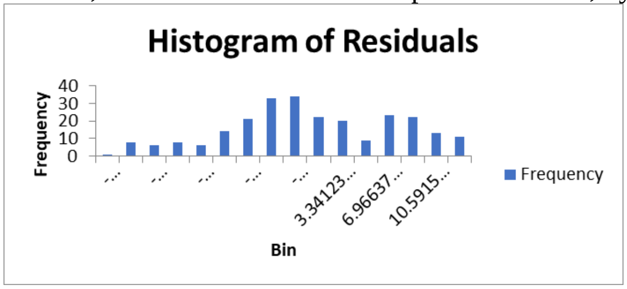
a) The Slope, Intercept, Correlation, Determination, Residual Mean and Standard Deviation has been calculated using the x and y values given in the dataset.

Slope	0.100809
Intercept	154.5037
Correlation	0.750978
Determination	0.563968
Residual Mean	1.13E-16
Standard Deviation	6.435439

Observations: It can be interpreted that, there exits a positive correlation between x and y since the value is closer to 1. The Determination shows 56.39 %, this means that 56.39% of the variations in the dependent variable are explained in x.

Determination shows how x and y are correlated, calculated as square of correlation.
We have used the CORREL() function for calculating the correlations between x and y.

b) We have calculated the Predicted and Residual values for the data. The predicted value is calculated using, $y = (b_o * x) + b_1$. The Residuals have been calculated using the (Observed y-Predicted y), where the sum of all values is 0 and average is 0.
Further, we have calculated the Exponential Error, by squaring the residual value.



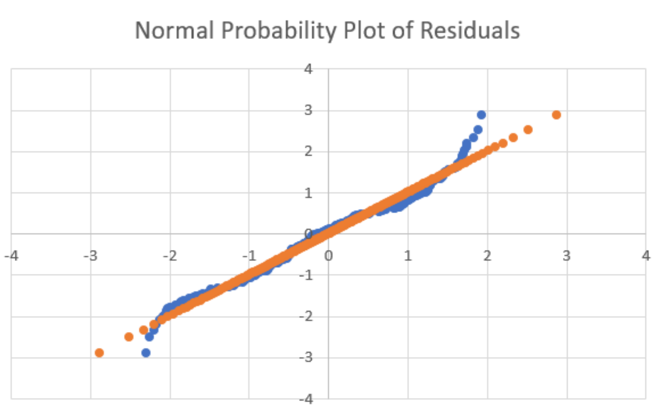
Observation: The histogram has been plotted using the residual values. It can be seen that the histogram is normally distributed and is equivalent to being symmetrical.

c) Chi-Square Goodness Fit Test
We want to check if the residuals belong to normal distribution.
Ho: Data belongs to normal distribution
Ha: Data does not belong to normal distribution

Chi square goodness of fit test	
T-statatastic	43.40617834
Level of signifinace	0.05
df	13
p value	0.00

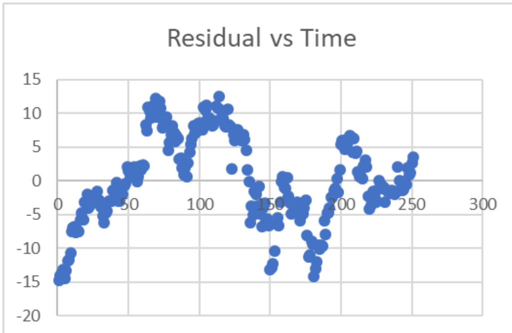
Observations: Since the p-value is lesser than the significance level, 0.05, we will reject the null hypothesis. Hence, the data is not normally distributed.

d) Normal Probability Plot of Residuals



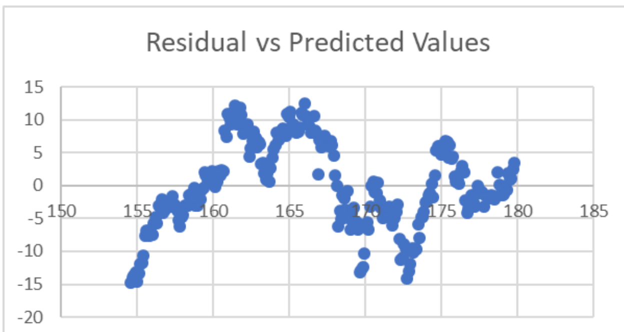
The graph shows the normal distribution amongst the standard value and residuals. The residual values were calculated using, Observed-Predicted values. Further,
i)We have sorted the values
ii) Standardised the residuals for transferring the values
iii) Ranked the values
iv) Cumulative area was calculated
v) Standard z-value was calculated using norm.s.inv()

e) Scatter Plot: Residual vs Time



Observations: The residuals show that they are normally distributed and does not show any pattern. The residuals have been scattered randomly and uniformly, following the principle of homoscedasticity.

f) Scatter Plot: Residual vs. Predicted Values



Observations: The residuals seem to follow the principle of homoscedasticity, as they are uniformly scattered.

4) In order to decide the accurate value as of 1/21/2020, I have compared the actual price and the forecasted values. It can be seen that the actual value is 183.23, on performing the exponential smoothening forecasting, it was noted that the accurate value was for α = 0.75 with MSE as 3.33.

However, the Adjusted Exponential Smoothening predicted a lesser MSE value of 3.24 for β=0.25. Therefore, it can be interpreted that, the adjusted exponential smoothening shows a value closer to the actual value with MSE value to be smaller. Hence, adjusted exponential smoothening is more accurate and better for predicting the stock prices.

CONCLUSION

1. We have predicted the stock prices of Honeywell by performing the time series forecasting methods like exponential smoothening and adjusted exponential smoothening.
2. We have also tested the hypothesis using chi-square goodness fit test for calculating the normality plot.
3. Using the concept of homoscedasticity, we have predicted the relevancy of the residuals by plotting the normality plot.

Reference

1. Bulman, A. G. (n.d.). Probability and Counting Rules. In ELEMENTARY STATISTICS: A STEP BY STEP APPROACH, TENTH EDITION (10th ed., p. A-440). New York
2. Hyndman, R. (n.d.). Retrieved from http://course1.winona.edu/bdeppa/FIN335/Handouts/Time_Series_Decomposition.html
3. Time Series Decomposition: A practical example using a classic data set. (n.d.). Retrieved from https://www.economicsnetwork.ac.uk/showcase/cook_timeseries