**Introduction to Enterprise Analytics**

# ALY6050 Module 3 Assignment

# Faculty: Prof. Richard He

# 2020 Winter CPS Quarter

# Image result for neu cps

Shraddha Gopalakrishnan [NUID: 001376444]

**Introduction**

Time is known to be the most challenging concepts in the universe. In machine learning, time series refers to chain of observations taken in a timely order. Time series forecasting involves considering models and trying to fit them on previous data (data being analyzed) and using it to predict future values. It is just amazing for instance, how based on the current sales of a company, value for next 15 years can be predicted and this is the power of machine learning in data analytics. Remember in time series, time is the independent variable and the goal is to make estimates for the future.[1]

**Analysis**

We have been provided an excel sheet which has data for Honeywell International a multi-national company’s stock prices from 22nd January 20y = 19 to 17th January 2020. Based on the date, close and volume columns, we need to perform forecasting and prediction by making use of other aspects of time.

**Q.1**

In the question it says, we need to predict the stock estimates for 21st January 2020. So before going onto the lengthy calculations we add a row at the end and put date as 1/21/2020.

Exponential smoothing is a way to model a time series data. It is much like the moving averages in which every next value is the mean of all the previous values.[2] But here, different decreasing weight is assigned to each observation. Less importance is given to observations as we move further from the present. To find the values using exponential smoothing, we use the mathematical formula: ***yt = 𝜶xt + (1-𝜶)yt-1***  where α is the smoothing parameter and varies between 0 and 1. Using alpha values as 0.15, 0.35, 0.55, 0.75 and the yt-1 as the previous number, we obtain the figures till 21st January 2020. We take the square of the difference between the observed and expected *Close* values and drag till the end date. For getting the MSE (mean squared error), we take the average of all the exponential errors of each alpha value.



**MSE**



**y** As observed, the MSE is lowest i.e., 3.34 for alpha value 0.75. Now we match 0.75 with the stock price value of our data for 21st January 2020 which is 182.88 as this is the highest figure of all. The qualitative description for the same will be, the stock price prediction is inversely proportional to the mean squared error and directly proportional to the alpha. Lesser the error, high would be the value of alpha 0.75 and best will be the forecast. If alpha is 1, the prediction will be 100% correct, but errors are bound to be. Also 0.75 very much approaches 1 hence 182.88 will be the best prediction estimate for date 1/21/2020.

**Q.2**

Here, we consider the same 4 columns of original Honeywell.csv data and append exponential alpha and exponential error columns for 0.75 alpha only (as stated in the question). Now using α values, we find the trend parameter β. The adjusted exponential smoothing is basically the method to calculate the weighted average of the current scenario’s actual value and forecast but with a correction as per the trend. The trend-adjusted exponential smoothing forecast costs of two parts: The exponentially smoothed forecast **y** and the exponentially smoothed trend ***Tt = 𝛃(yt – yt-1) + (1 – 𝛃)Tt-1***. Lastly, the trend adjusted exponential smoothing is computed as ***yt + Tt***. β is the trend smoothing constant. Using successive values for parameter β, 0.15, 0.25, 0.45, 0.85, MSE’s are evaluated. The MSE value is found to be least 3.33 for β = 0.15 from the Excel.





Corresponding stock price for 21st January 2020 is 183.42.





The value 183.42 is the most accurate measure because, in this method the new forecast is equal to the old plus some proportion of the past forecasting error.

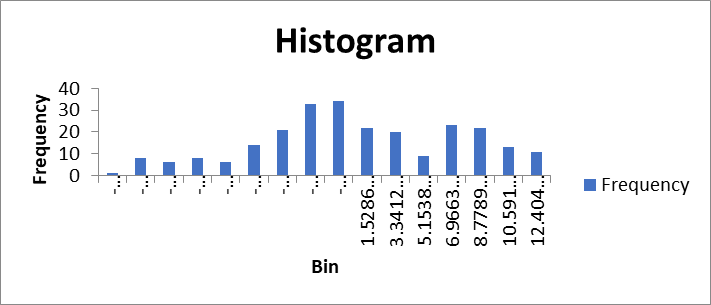
**Q.3**

Regression analysis is performed by taking period and close columns of the original data. We find the slope, intercept, correlation and determination values. By applying the general slope formula ***y=mx + c*** we get the Predicted Y column. Residuals are just the difference between Close and Y value. Exponential is the square of residual.

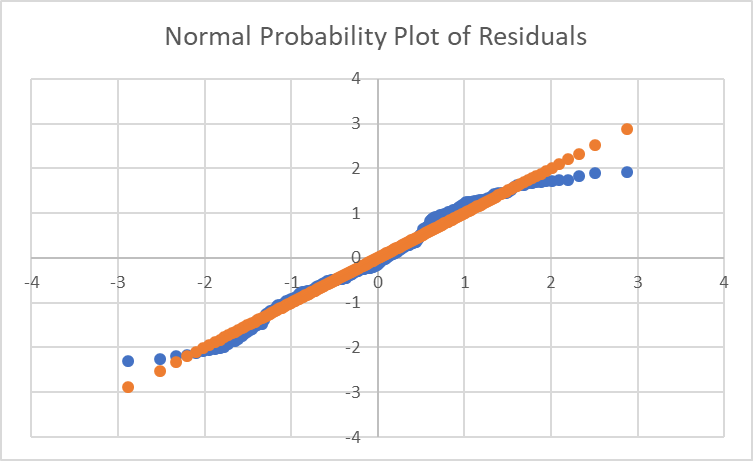
1. The coefficients of correlation and determination are found to be

Determination is the square of correlation value and hence always smaller since the correlation value is lesser than 1.

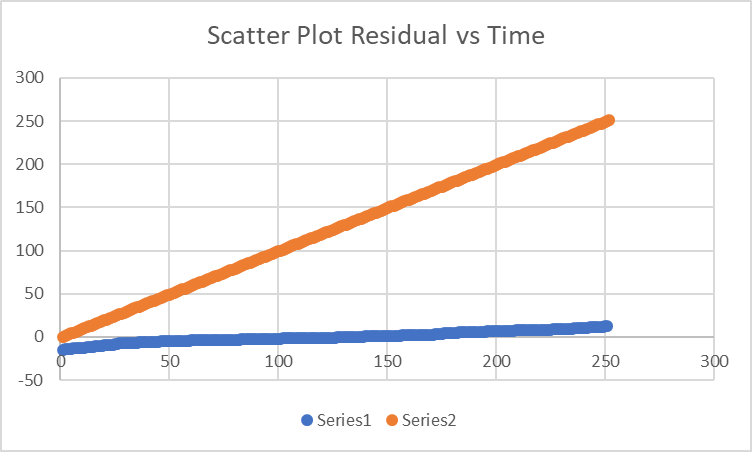
1. A histogram of regression residuals:

The shape of the histogram is of a normal curve depicting similar values between and closer to the mean.

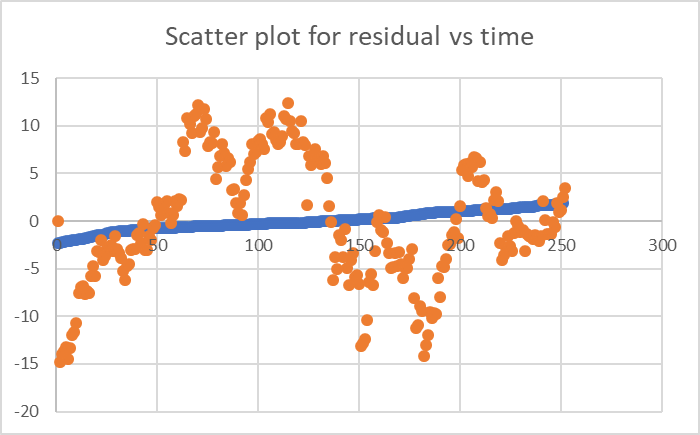
1. For carrying out regression, we follow the procedures for chi-square goodness of fit test for more accuracy in our results. P-value is very less than the alpha hence we reject the null hypothesis and our say that our data is not normally distributed.
2. A normal probability plot for our residuals:

We plot this by using data of residual mean, standard deviation to calculate column of standardized residuals and the standard z-value (***NORM.S.INV)***. The graph is progressive linear model which indicates that normal distribution is a good model for the dataset. The y-axis shows us the ordered response.

1. A scatter plot of residuals vs time:

From the graph there is clarity that both are independent of each other. Shape of the scatter plot is an acute angle. Residuals are constant throughout with increasing time.

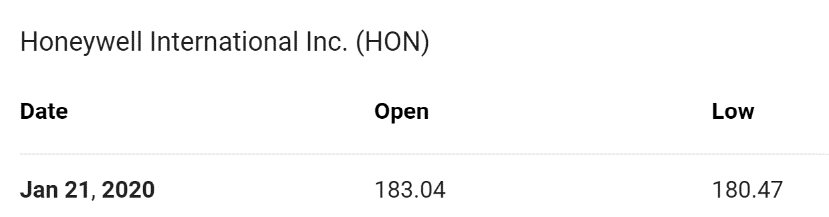
1. Scatter plot for residual vs time:



This graph indicates the homoscedasticity of residual variables and predicted stock prices. The shape cannot be defined as it fluctuates a lot. It shows estimates of prices have inflated variances.[2]

**Q.4**

The actual value of the stock price for Honeywell on 21st January 2020 was between 183.04 and 180.47 (estimated by google).

****

Our value in the project is 182.88 with Exponential smoothing and 183.42 with Trend adjusted smoothing. Among the two, it is clear that method 2 stands out well the trend adjusted smoothing in which we made use of parameter β for increased accuracy.

**Conclusion**

Through this project, we understood the importance of forecasting in data analysis with time as the main component. We actually did estimate a value which indeed turn out to be very similar to the real-world value which is not less than predicting the future. We learned how to robustly analyze and model time series and apply our knowledge in the desired project.[1] Regression was also used to compare with the time series prediction methods and did not turn out well in yielding correct results. Overall, this project was very informative and interesting.

**References**

[1] Pourya. (2019, April 6). Time Series Machine Learning Regression Framework. Retrieved from-<https://towardsdatascience.com/time-series-machine-learning-regression-framework-9ea33929009a>

[2] Forecast estimation, evaluation and transformation. (2010, November 10). Retrieved from <https://robjhyndman.com/hyndsight/forecastmse/>