****

Week 3

**Price Prediction Project**

Course: ALY 6050 Intro to Enterprise Analytics

By,

Pavan Chand Garikapati

Spurthi Patnam

Shruti Avinash Pawar

Instructor: Joseph Manseau

26th April 2020

**Introduction:**

Any metric that is measured over regular intervals of time forms a time series. A time series can be broken down into components called trend, seasonality, cycle and noise. Analysis of time series is commercially significant because of industrial need and relevance especially for the purpose of forecasting.

Importance of Time series forecasting

* Time series helps in analyzing the past and forecasting the future
* It is used to understand the determining factors and structure behind the observed data, choose a model to describe the data thereby improves decision making.
* It utilizes historical data to analyze patterns and trends, cyclic fluctuations and seasonality to forecast the future.
* It brings out the pattern of situation reflecting the relation between the data point and the variable ("Time Series Analysis for Better Decision Making in Business", n.d).

Exponential smoothing Forecasts

Exponential smoothing is a time series forecasting method for univariate data. They are similar to that where the prediction is the weighted sum of past observations, but it explicitly uses an exponentially decreasing weights for past observations (Brownlee, 2020). To perform time series analysis, we have used R studio and library forecast.

**Analysis:**

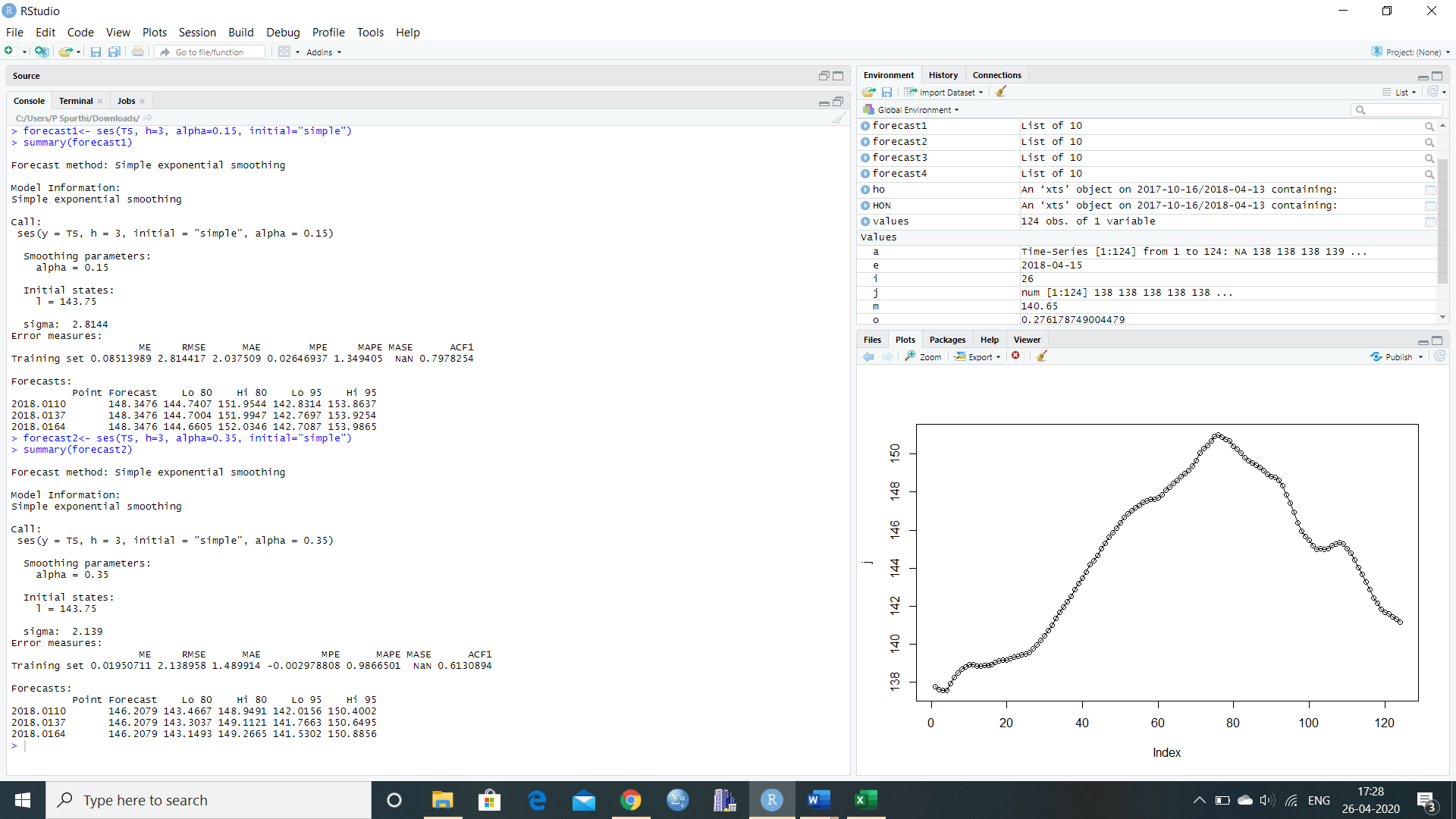
We are performing time series analysis on Honeywell dataset which is stored as an excel workbook. It consists of the closing stock prices in accordance with the date. We have the values from October 2017 to April 2018.

**Problem 1**

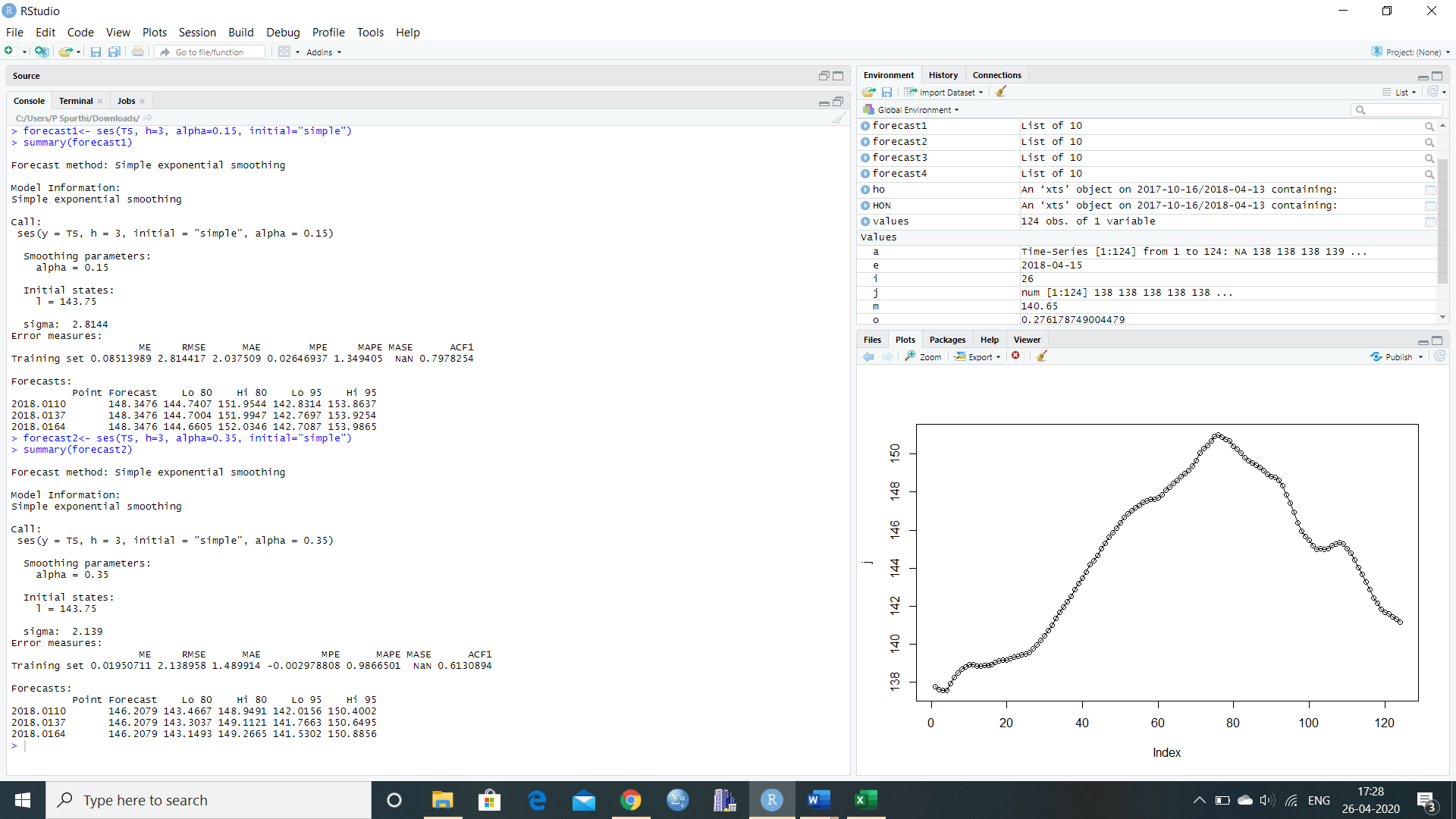
To forecast the Honeywell stock price for 4/16/2018, we have imported the dataset and converted into timeseries by using the ts() function. We are using the simple exponential smoothing which requires a smoothing parameter called alpha which controls the rate at which the influence of observations at prior time steps decay exponentially. We are observing the RMSE values by taking different alpha values

First, we take **alpha= 0.15**

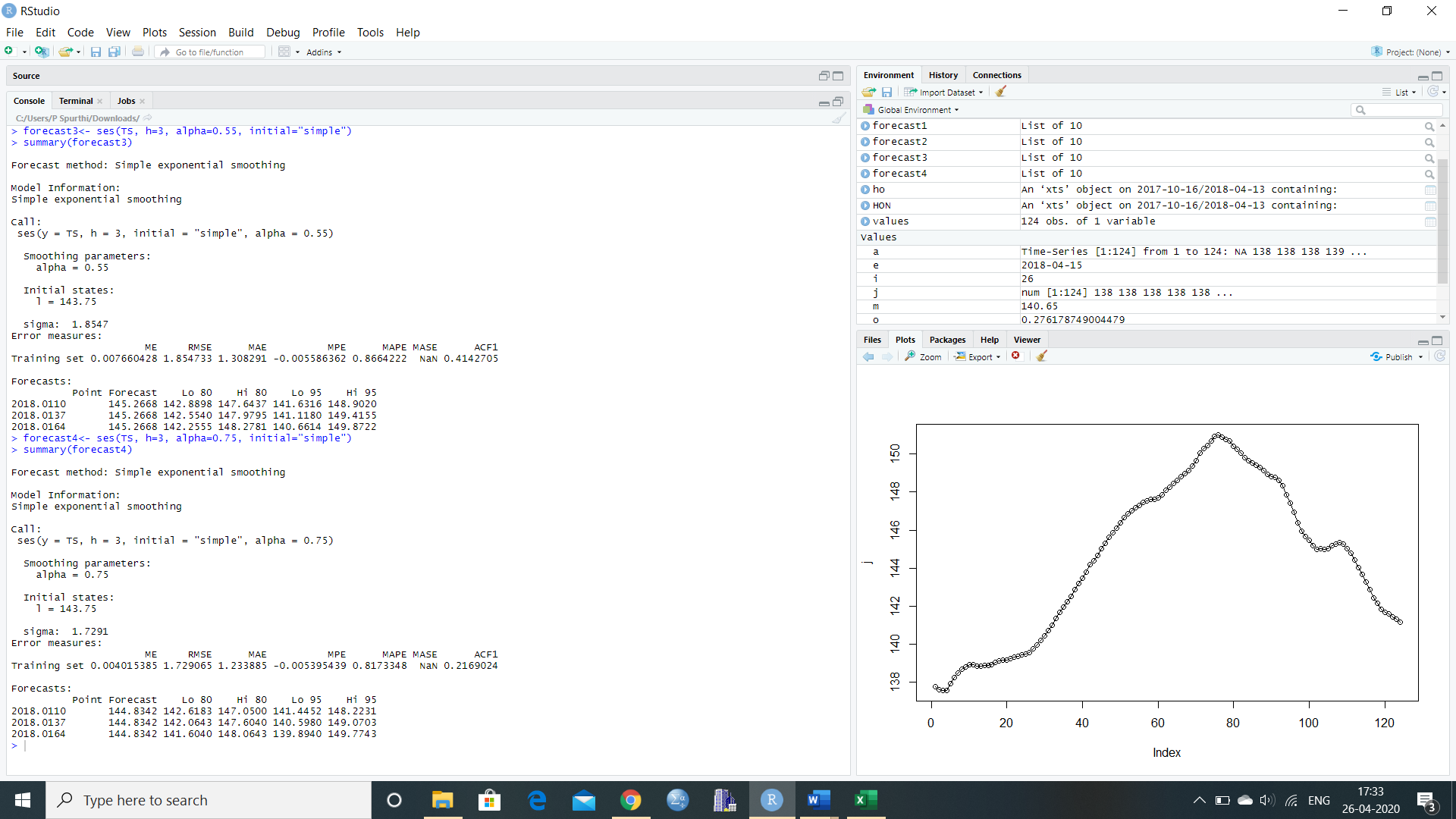
R output is on page 3, and in the output, we observe that the RMSE value is 2.814417. At Alpha 0.15 the forecasted closing price for 04/16/2018 is 148.3476.



Next, we consider the **alpha value = 0.35**  and below is the R output. From the summary we can see that the RSME has reduced from 2.814417 to 2.138958 and predicted price is 146.2079.

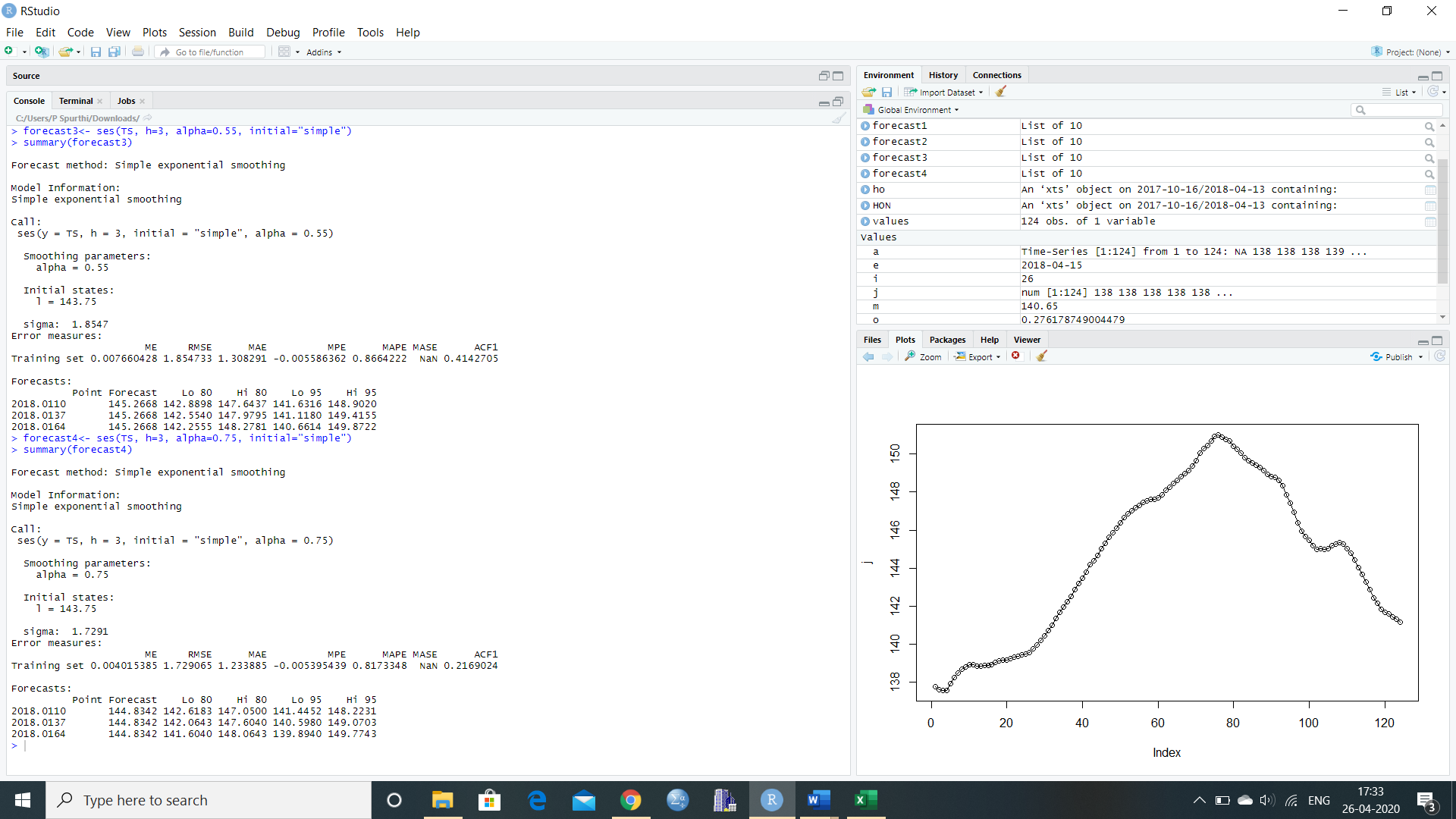


Further we continue our analysis using **alpha=0.55**



We observe from the above R output that the RMSE value further reduces to 1.854

Finally, we use **alpha=0.75**  and the RMSE value is 1.72



We have taken the alpha values and the corresponding RMSE, MSE ( square of RMSE) and prediction values into a tabular form

**Table 1:** Summary of Simple Exponential Smoothing forecasts at 4 levels of alpha

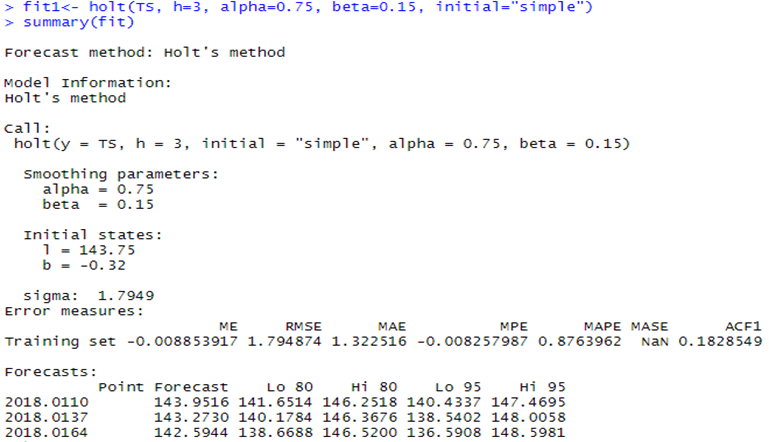
|  |  |  |  |
| --- | --- | --- | --- |
| **alpha** | **Predicted Price** | **RMSE** | **MSE** |
| 0.15 | 148.2076 | 2.814 | 7.9185 |
| 0.35 | 146.2079 | 2.138 | 4.5710 |
| 0.55 | 145.2668 | 1.854 | 3.4373 |
| 0.75 | 144.8342 | 1.729 | 2.9894 |

As per table 1, we observe that as the alpha value increases the RMSE as well as the predicted price has decreased. The lower the RMSE value, the higher will be the accuracy of the prediction. As the RMSE value is the least for **alpha value 0.75**, we take the alpha value to be 0.75. This value of alpha has yielded the most accurate results because it is closer to 1 i.e. the most recent values influence the forecasts with comparatively lesser weights to historical values.

**Problem 2**

In this task we are require forecasting the stock price using exponential smoothing method. We are using “holt” function in which we are doing Holt’s Two-Parameter Exponential Smoothing and the exponential smoothing contains a parameter called alpha (α) which is set as 0.75 and another parameter called beta (β) noted as smoothing parameter. The aim is to consider different values of this parameter as required (0.15,0.25,0.45 and 0.85) and identify the bets beta that forecasts the stock accurately by observing the RMSE value.

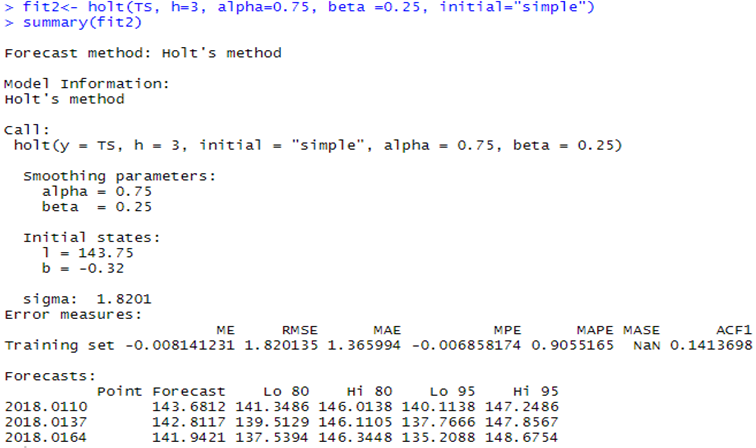
First, we take α = 0.75 and β = 0.15



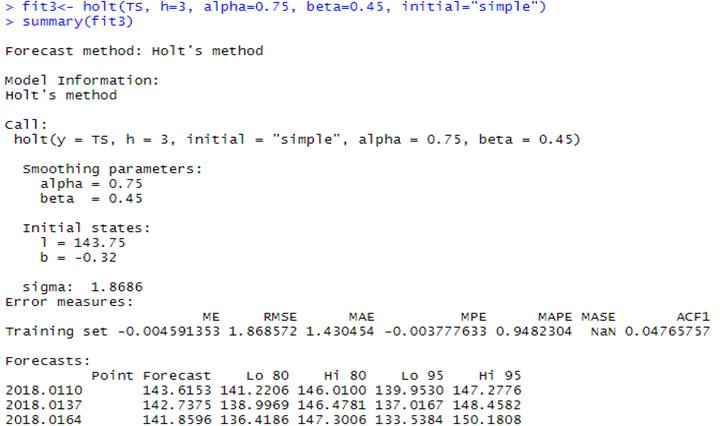
We observe that the RMSE value is 1.794874 and the forecasted price is 142.5944.

Next, we take α = 0.75 and β = 0.25

R output is on page 6 and we observe that the RMSE value is 1.820135 with the prediction value at 141.9421



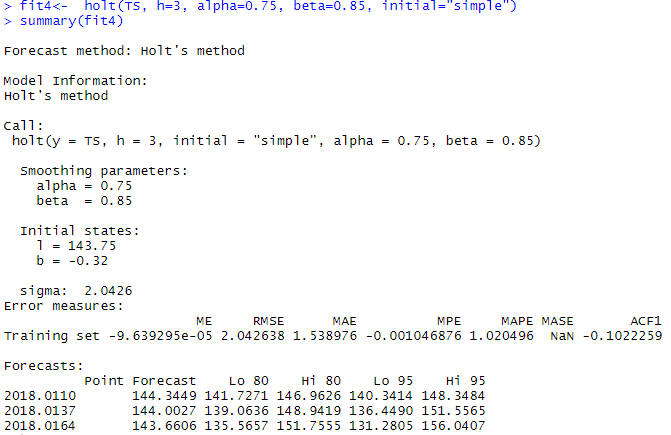
Further, we take α = 0.75 and β = 0.45



The RMSE value has increased to 1.86 with predicted price at 141.8596

Finally, we take α = 0.75 and β = 0.85

R output is on page 7 and we observe that the RMSE value is 2.042638 with the prediction value at 143.6606.



Now we take the alpha, beta values and the corresponding RMSE, MSE ( square of RMSE) and prediction values into a tabular form.

**Table 2:** Summary of Holt Winter forecasts at 4 levels of beta and 0.75 alpha

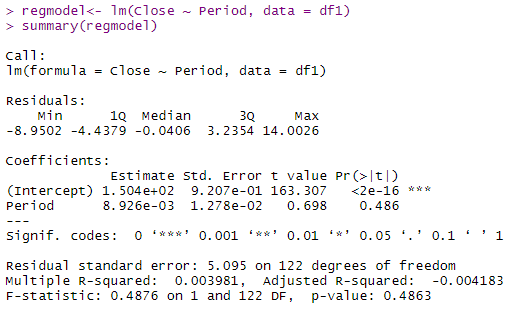
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Alpha(α)** | **Beta(β)** | **Predicted price** | **RMSE** | **MSE** |
| 0.75 | 0.15 | 142.5944 | 1.794874 | 3.22157 |
| 0.75 | 0.25 | 141.9421 | 1.820135 | 3.31289 |
| 0.75 | 0.45 | 141.8596 | 1.868572 | 3.49156 |
| 0.75 | 0.85 | 143.6606 | 2.042638 | 4.17236 |

As per table 2, we observe that as the beta value increases the RMSE value as well as the predicted price for 04/16/2018 is increasing. The lower the RMSE value, the higher will be the accuracy of the prediction. As the RMSE value is least for beta value 0.15, we take beta value to be 0.15. Alpha value is also known as the base value and beta value is known as the trend value. Alpha sets the magnitude of the series and beta sets the trend. Higher the values of these two higher will be the forecast if the time series has an increasing trend. This is because values close to 1 will empathize more weightage on recent data and values close to zero emphasize on historic data. (Snapp, 2020)

**Problem 3**

Now that we have forecasted price for 04/16/2018 using simple exponential smoothing and Holt winter’s adjusted smoothing, we will check the value predicted by a simple linear regression and if it has a MSE lesser than the ones predicted prior.

Below is the R output of the linear regression model:



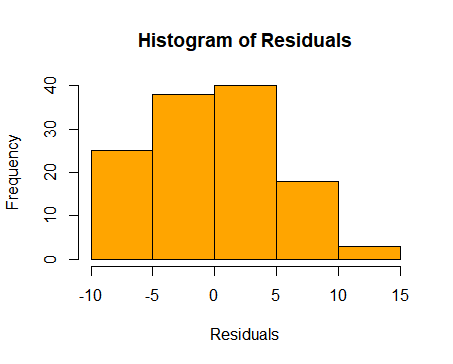
Using the values generated by the regression model summary, our prediction model is Price = 1.504e+02 + 8.926e-03 \* Period. Using a code to run this regression model in R, we predicted the closing price to be 151.5336.



The MSE of the prediction is 25.54194. This is way higher than the mean squared error of those calculated in the exponential smoothing forecasts. Mean squared error is the average of squared difference of actual and predicted values. Thus, a lower MSE would mean that the predicted values are similar to the actual values with a very small error.



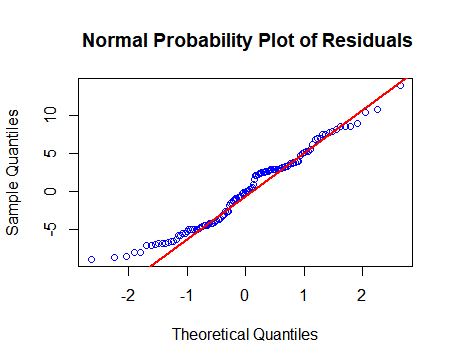
1. We calculated the coefficient of correlation in R for period and Closing price using the cor() function and the coefficient of correlation is 0.06309489. The R- squared also known as the coefficient of determination is 0.003981. These values are considered good if these are approaching 1. There is a very week correlation between period and closing price and the regression model is not a perfect fit for the data as the R squared is not approaching 1.
2. Further we plotted a histogram of residuals and below is the plot:



**Fig 1:** Frequency Histogram of Residuals of Linear Regression Model

Looking at figure 1, we assume that residuals are approximately following normal distribution. To validate that, we performed a qqplot analysis and chi squared goodness of fit test.

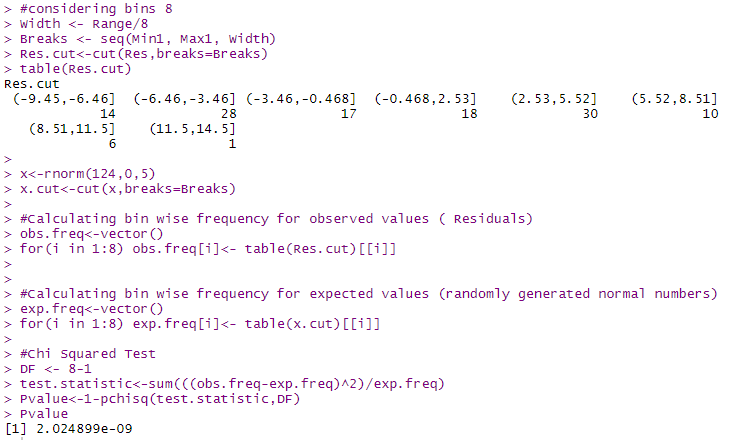
1. Below is the r output of qqplot



**Fig 2:** QQplot for Residual with sample and theoretical quantiles

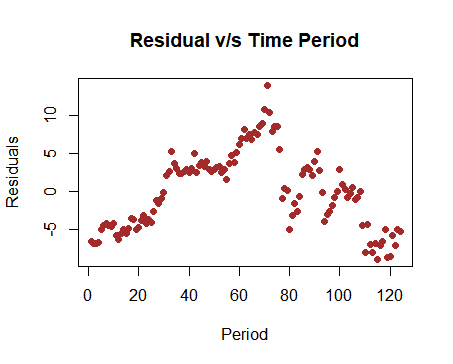
In figure 2, if the data points were close to the qq line seen in red, it would give a gist that the residuals follow normal distribution. However, the curve diverts from the qqline. This scenario might happen if there are outliers thus, we are not sure about the normality yet. A chi squared goodness of test will give more clarity whether the residuals follows normal distribution or not.

1. A chi squared test was performed with residuals as observed values and 124 values simulated from normal distribution as expected values. We used 8 bins to calculate the frequency and p-value was calculated as 2.024899e-09 at 7 degrees of freedom. The p-value is remarkably close to zero and less than the significance level 0.05. We thus reject the null hypothesis that the data follows normal distribution.



The qq plot gave a rough idea that the residuals are not fitting the qq line thus they do not follow normal distribution. However, the chi squared test gave evidence the residuals do not follow normal distribution. ("Understanding Q-Q Plots | University of Virginia Library Research Data Services + Sciences", 2020)

1. To check the relationship between residuals and time period we plotted them in a scatterplot and below is the output



**Fig 3:** Scatter plot of Residual v/s time

Residuals are the difference between the actual and predicted values in the regression model. The residuals will not be related to time thus there is no correlation seen in the figure 3.

1. To check for Homoscedasticity of residuals we plotted close price and residuals in a scatterplot and below was the output.



**Fig 4:** Scatter plot to check homoscedasticity of Residual

In simple words homoscedasticity means having the same scatter. The points should be close to each other at a same distance in a linear line. From figure 4, the residuals seem to follow the homoscedasticity rule. ("Homoscedasticity / Homogeneity of Variance/ Assumption of Equal Variance - Statistics How To", 2020)

**Problem 4**

As per data provided on Yahoo Finance, the closing price of Honeywell was 140.65 on 04/16/2018. The value calculated using Holt Winter’s which was adjusted exponential smoothing with alpha value at 0.75 and beta value at 0.45 was 141.8596 with a RMSE of 1.868572. This predicted price is the closest to the actual closing price. The error here is actual 140.65 – predicted 141.8596 which is 1.2096 which is less than the RMSE. This model has given a closely accurate prediction as both Alpha and Beta are not close to 0 with less weightage on historical values.

**Conclusion:**

In the report, we saw that regression was the worst predictor model amongst all with an MSE of 25.54194 and an actual error of 10.88 (151.53-140.65). So why is linear regression model a bad predictor? There was absolutely no correlation between period and closing price. Besides, time series data is autoregressive and there is internal correlation in the datapoints between two different times. Also, the residuals do not follow all the assumptions of linear regression model and with a low R-squared value, linear regression is not the best fit model for time series data.

**Reference:**

Time Series Analysis With R - r-statistics.co. (n.d.). Retrieved April 26, 2020, from <http://r-statistics.co/Time-Series-Analysis-With-R.html>

Time Series Analysis and Forecast Service for Your Business. (n.d.). Retrieved April 26, 2020, from <https://www.researchoptimus.com/article/what-is-time-series-analysis.php>

Brownlee, J. (2020, April 12). A Gentle Introduction to Exponential Smoothing for Time Series Forecasting in Python. Retrieved April 26, 2020, from <https://machinelearningmastery.com/exponential-smoothing-for-time-series-forecasting-in-python/>

Snapp, S. (2020). How to Make Sense of The Natural Confusion with Alpha Beta Gamma • Brightwork | Demand Planning. Retrieved 27 April 2020, from <https://www.brightworkresearch.com/demandplanning/2011/03/alpha-beta-and-gamma-in-forecasting/>

Understanding Q-Q Plots | University of Virginia Library Research Data Services + Sciences. (2020). Retrieved 27 April 2020, from <https://data.library.virginia.edu/understanding-q-q-plots/>

Homoscedasticity / Homogeneity of Variance/ Assumption of Equal Variance - Statistics How To. (2020). Retrieved 27 April 2020, from <https://www.statisticshowto.com/homoscedasticity/>

Closing price reference:

<https://finance.yahoo.com/quote/HON/history?period1=1429920000&period2=1587772800&interval=1d&filter=history&frequency=1d>