PRICE PREDICTIONS PROJECT

ALY6050: Introduction to Enterprise Analytics

DHRUV VIJAY GUJRATHI

Northeastern University

DHRUV VIJAY GUJRATHI, College of Professional Studies, Northeastern University, Boston,

MA 02115.

This report is regarding price predictions project.

Dhruv Gujrathi is now a student at Department of Analytics, Northeastern University

Contact: gujrathi.d@northeastrn.edu

(NUID: 001029464)

# INTRODUCTION

Exponential smoothing is a period arrangement forecasting strategy for univariate information. Time arrangement strategies like the Box-Jenkins ARIMA group of techniques build up a model where the expectation is a weighted direct entirety of later past perceptions or slacks. Exponential smoothing forecasting strategies are comparable in that an expectation is a weighted entirety of past perceptions, yet the model expressly utilizes an exponentially diminishing load for past perceptions. Past perceptions are weighted with a geometrically diminishing proportion. Gauges created utilizing exponential smoothing strategies are weighted midpoints of past perceptions, with the loads rotting exponentially as the perceptions get more established. As it were, the later the perception the higher the related weight. Exponential smoothing is a dependable guideline procedure for smoothing time arrangement information utilizing the exponential window work. While in the straight forward moving normal the past perceptions are weighted similarly, exponential capacities are utilized to relegate exponentially diminishing loads after some time. It is an effectively learned and effectively applied methodology for making some assurance dependent on earlier suppositions by the client, for example, regularity. Exponential smoothing is frequently utilized for investigation of time arrangement information.

# Analysis

This report is regarding the prediction of stock price of the company – Honeywell International Incorporated. This project is focussed on finding the stock prices of the company and then predicting the stock price using regression analysis and smoothing forecasts. We basically used three types of techniques for predicting the stock price at a particular date which was 4/16/2018. The first method is known as exponential smoothing for time series forecasts – simple exponential smoothing. It is a method for short term predictions. Such forecasts are based on past data where the most recent observations are weighted more than the less recent observations. The second method that we used is known as Holt's method – Adjusted SES method. To work with a long term data, SES can create unwanted problems. Thus, we use this method where two smoothing parameters are used which corresponds to the level and trend components. The third method that we used is known as the simple regression analysis. The Simple Regression Analysis is concerned with specifying the relationship between a single numeric dependent variable (the value to be predicted) and one numeric independent variable (the predictor).


**Figure 1: Honeywell stock across required date range**

1.
In the part 1, we need to perform the exponential smoothing analysis for forecasting the stock price of Honeywell on the particular date that is given. For doing this, firstly we load the package "quantMod" which contains the stock prices of different companies. Using getSymbols, we call the stock price of the company which we need, i.e. - Honeywell. After getting the past results of the stock prices of the company, we define the classes and have a look at the top 6 results of the closing stock prices. The closing stock prices is the column in the table we are looking at is the variable we will be using for the forecasts. After setting the start date and the end date for deciding the range of forecasts, we load the required libraries – "forecast" and "metrics" so as to use the "ses" function to perform the simple exponential smoothing assumes that the time series data has only a level and some error (or remainder) but no trend or seasonality. For exponential smoothing, all past observations are part of the calculation for the forecasted value. The smoothing parameter $\alpha$ determines the distribution of weights of past observations and with that how heavily a given time period is factored into the forecasted value. If the smoothing parameter is close to 1, recent observations carry more weight and if the smoothing parameter is closer to 0, weights of older and recent observations are more balanced.

- $Weight = \alpha * (1-\alpha)$^$t$
- $Forecast = \Sigma t = 0$ to $n(\text{air}\_t * Weight\_t)$

After careful consideration and analysis, we came to a conclusion that the most accurate forecast is the one which has the lowest MSE value. MSE, basically stands for Meaned Square Error. It is a common measure of estimator quality, of the fitted values of a dependent variable. The lower the MSE the higher the accuracy of prediction as there would be excellent match between the actual and predicted data set. Thus, when alpha = 0.75, the corresponding MSE value is 2.437994 which is the lowest. Therefore, the 0.75 smoothing parameter is the most accurate forecast.

| **Sr. No.** | **Smoothing Parameters (alpha)** | **Corresponding MSE Values** |
|---|---|---|
| 1. | 0.15 | 6.91486 |
| 2. | 0.35 | 3.982149 |
| 3. | 0.55 | 2.906732 |
| 4. | 0.75 | 2.437994 |

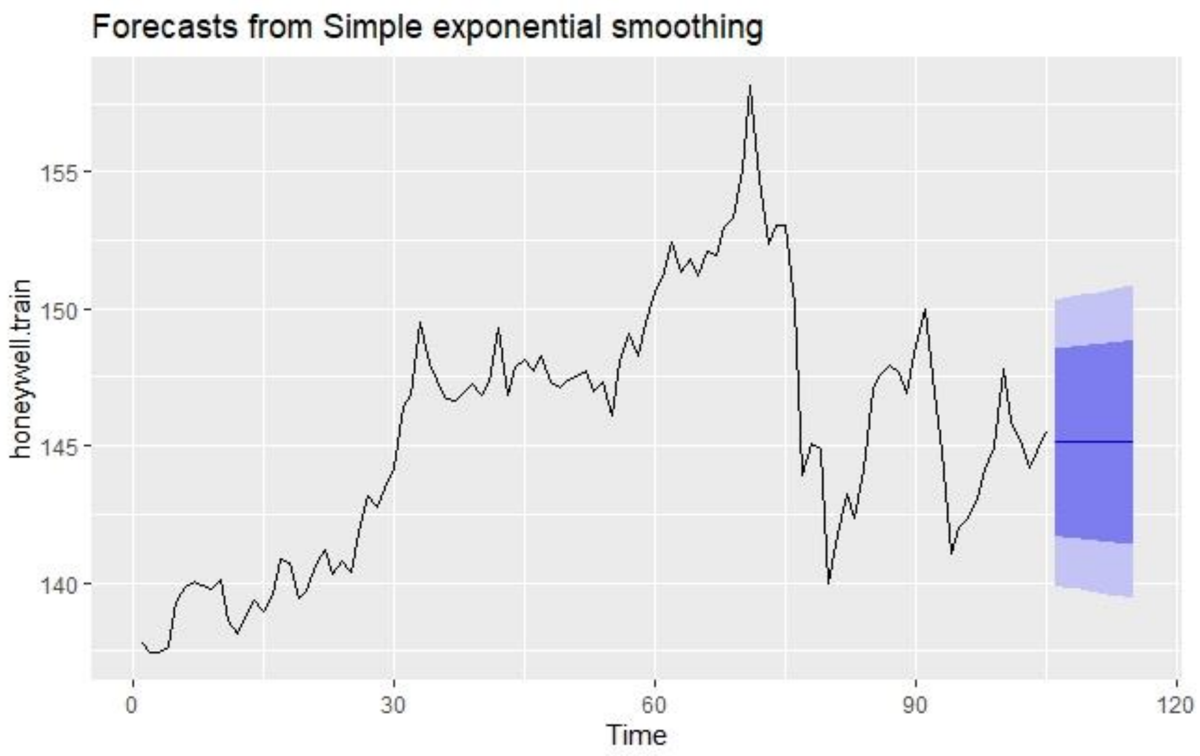



Figure 2: Alpha = 0.15

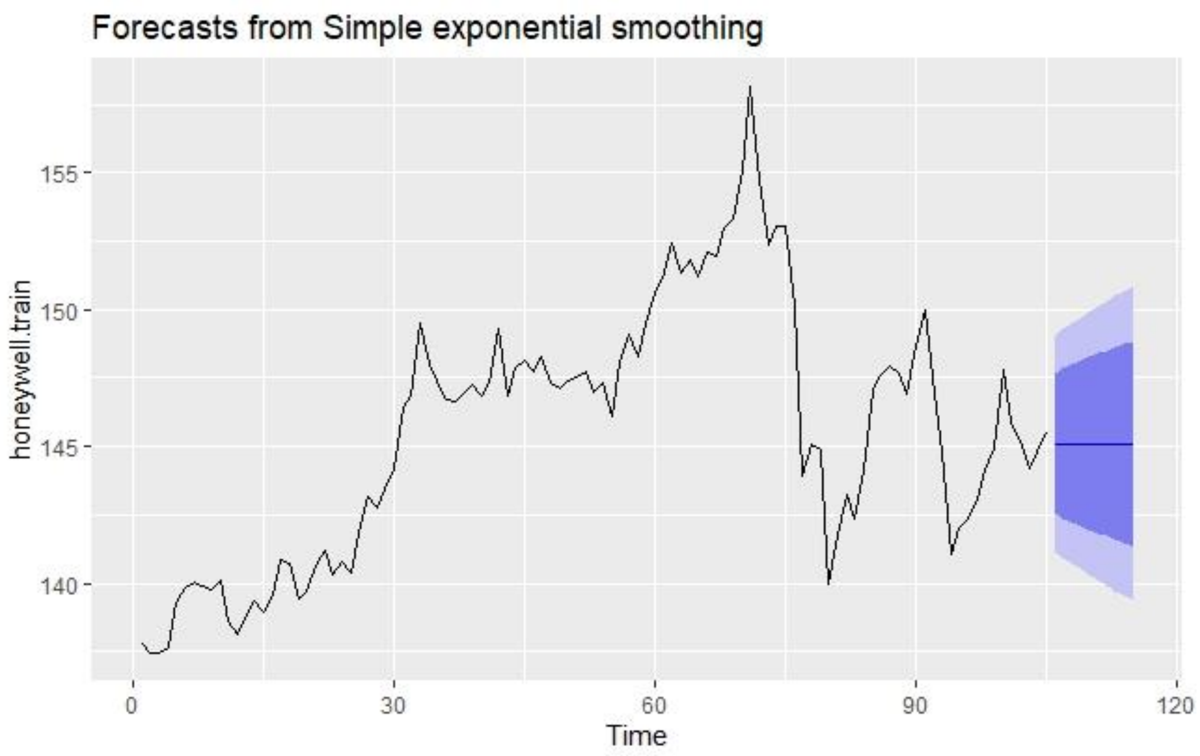



Figure 3: Alpha = 0.35

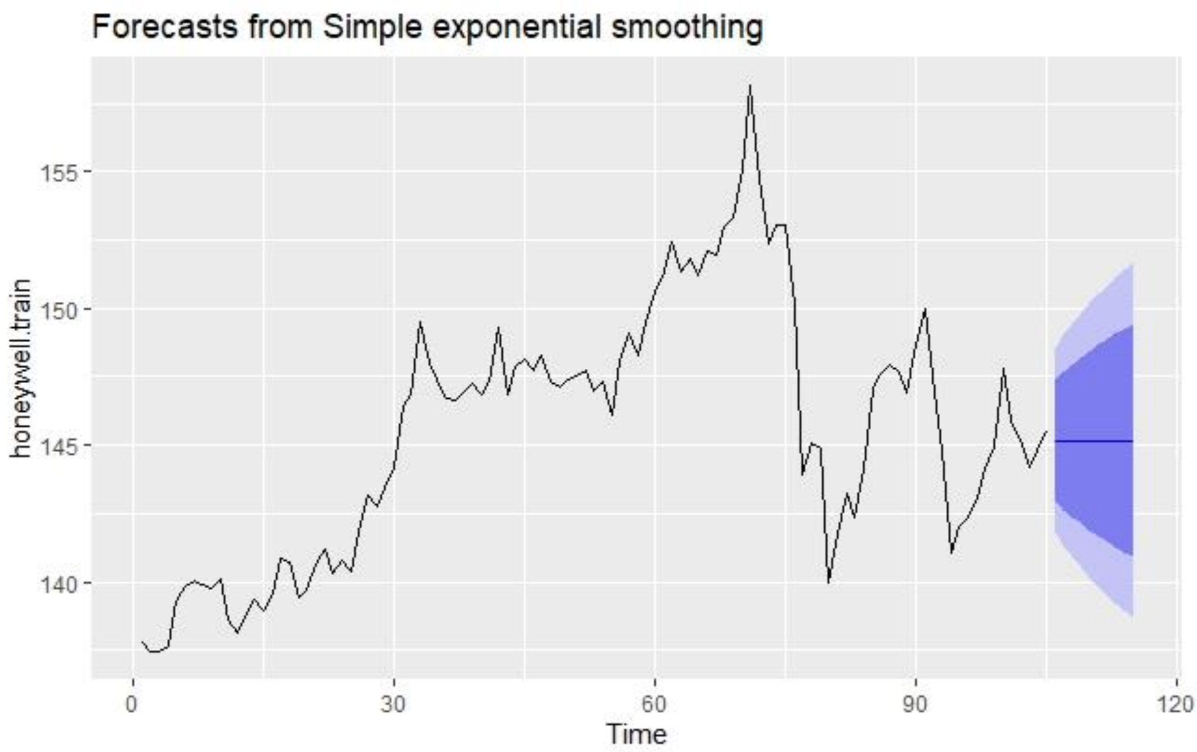



**Figure 4: Alpha = 0.55**

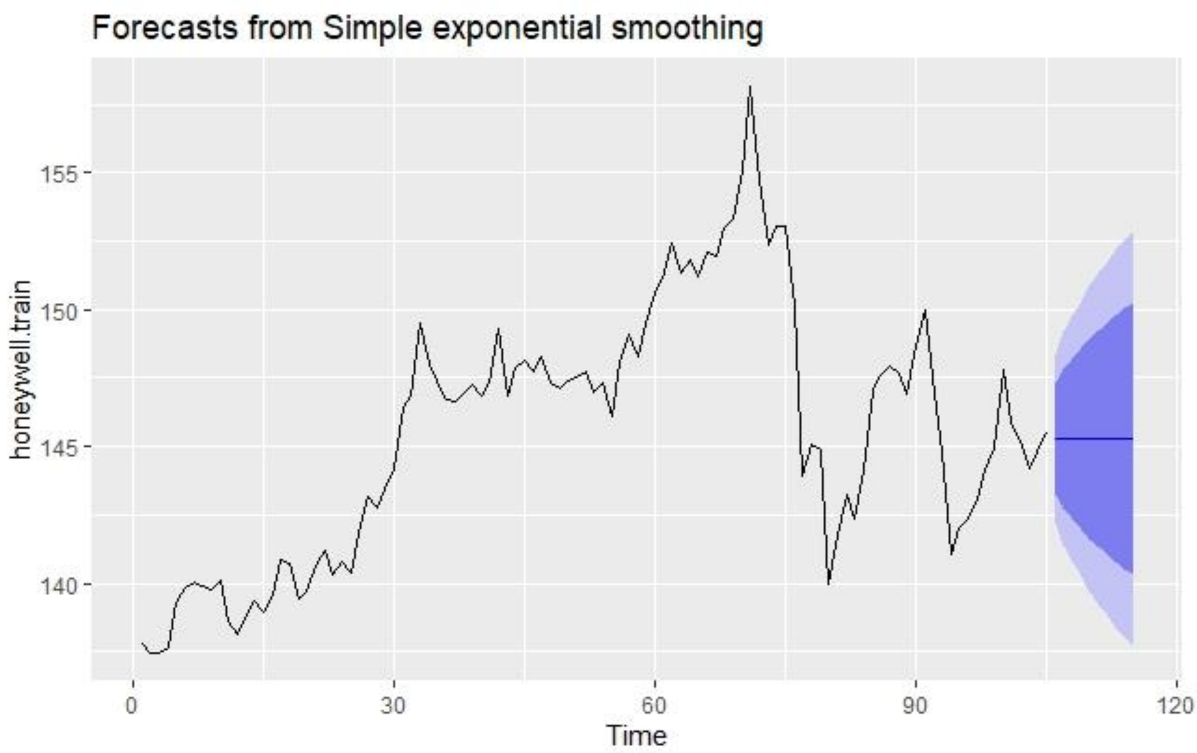



**Figure 5: Alpha = 0.75**

## **R-Code**:

```
#SES Method
#Installing Package "Forecast"
install.packages("forecast")
#Installing Package "Metrics"
install.packages("Metrics")
library(forecast)
library(Metrics)
#Performing the Simple Exponential Smoothing for alpha = 0.15
ses.honeywell1 <- ses(honeywell.train, alpha = 0.15, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.15
autoplot(ses.honeywell1)
#Determining the MSE for alpha = 0.15
ses.honeywell1$model$mse
#Performing the Simple Exponential Smoothing for alpha = 0.35
ses.honeywell2 <- ses(honeywell.train, alpha = 0.35, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.35
autoplot(ses.honeywell2)
#Determining the MSE for alpha = 0.35
ses.honeywell2$model$mse
#Performing the Simple Exponential Smoothing for alpha = 0.55
ses.honeywell3 <- ses(honeywell.train, alpha = 0.55, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.55
autoplot(ses.honeywell3)
#Determining the MSE for alpha = 0.55
ses.honeywell3$model$mse
#Plotting Simple Exponential Smoothing for alpha = 0.75
ses.honeywell4 <- ses(honeywell.train, alpha = 0.75, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.75
autoplot(ses.honeywell4)
#Determining the MSE for alpha = 0.75
ses.honeywell4$model$mse
```

2.

In the part 2, we again need to perform the exponential smoothing analysis for forecasting the stock price of Honeywell on the particular date that is given. For doing this, we will be using a different method called as the Holt’s method. SES does not perform well with data that has a long-term trend. In the last section we illustrated how we can remove the trend with differencing and then perform SES. An alternative method to apply exponential smoothing while capturing trend in the data is to use Holt’s Method. Holt’s Method makes predictions for data with a trend using two smoothing parameters, α and β, which correspond to the level and trend components, respectively. For Holt’s method, the prediction will be a line of some non-zero slope that extends from the time step after the last collected data point onwards. The methodology for predictions using data with a trend (Holt’s Method) uses the following equation with observations. The k-step-ahead forecast is given by combining the level estimate at time t (L t) and the trend estimate (which in this example is assumed additive) at time t.

After careful consideration and analysis, we came to a conclusion that the most accurate forecast is the one which has the lowest MSE value. MSE, basically stands for Meaned Square Error. It is a common measure of estimator quality, of the fitted values of a dependent variable. The lower the MSE the higher the accuracy of prediction as there would be excellent match between the actual and predicted data set. Thus, when alpha = 0.15, the corresponding MSE value is 2.662428 which is the lowest. Therefore, the 0.15 smoothing parameter is the most accurate forecast. Here, we calculated the MSE keeping the alpha value as 0.75 constant.

| **Sr. No.** | **Smoothing Parameters (alpha)** | **Corresponding MSE Values** |
|---|---|---|
| 1. | 0.15 | 2.662428 |
| 2. | 0.25 | 2.717874 |
| 3. | 0.45 | 2.88746 |
| 4. | 0.85 | 3.635233 |

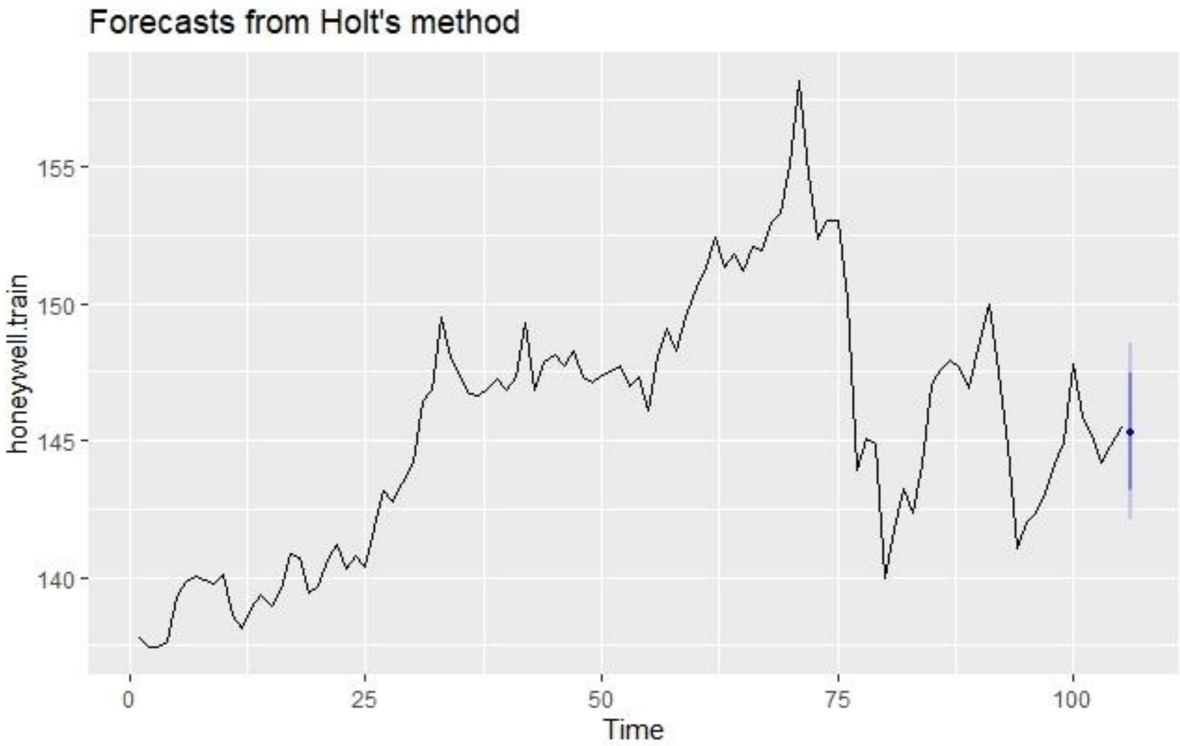



**Figure 6: Alpha = 0.75, Beta = 0.15**

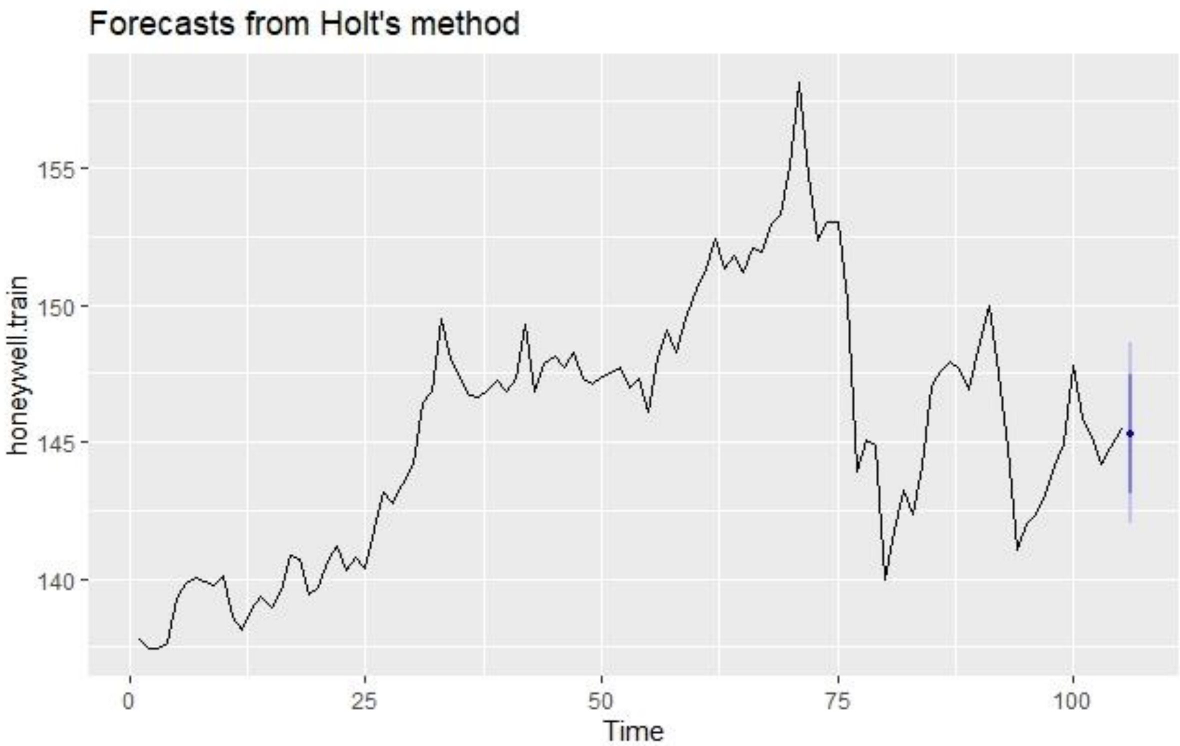



**Figure 7: Alpha = 0.75, Beta = 0.25**

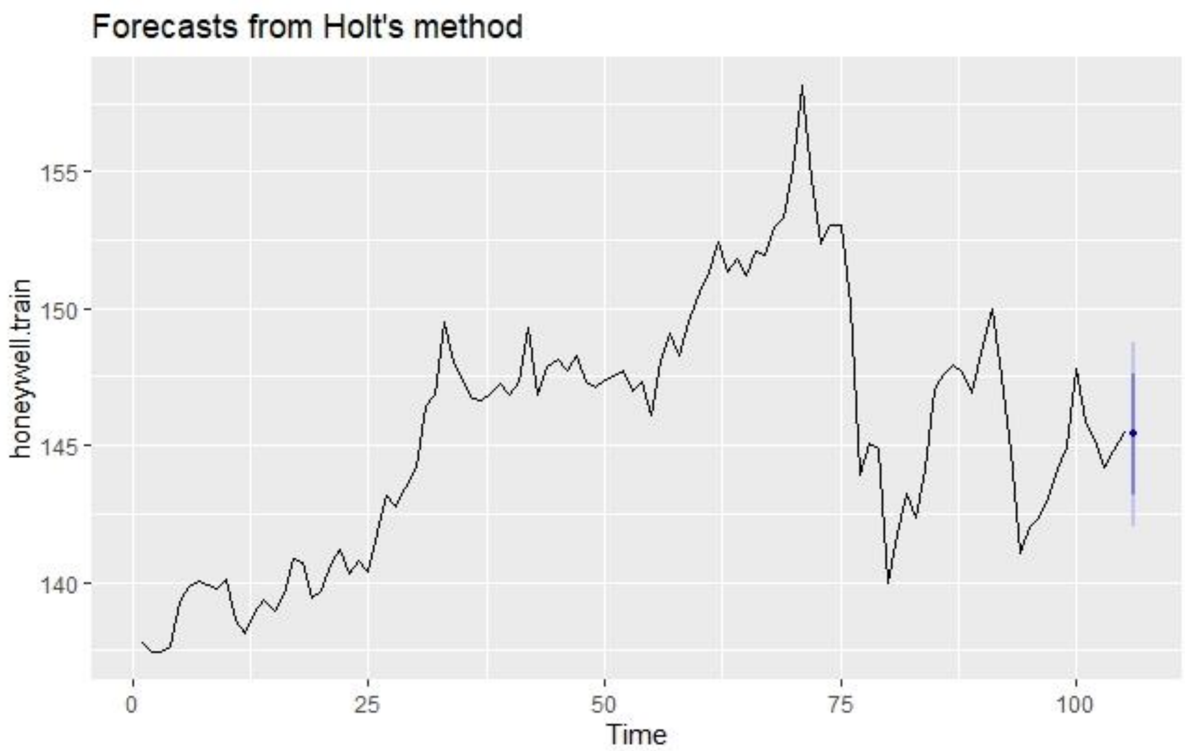


Figure 8: Alpha = 0.75, Beta = 0.45

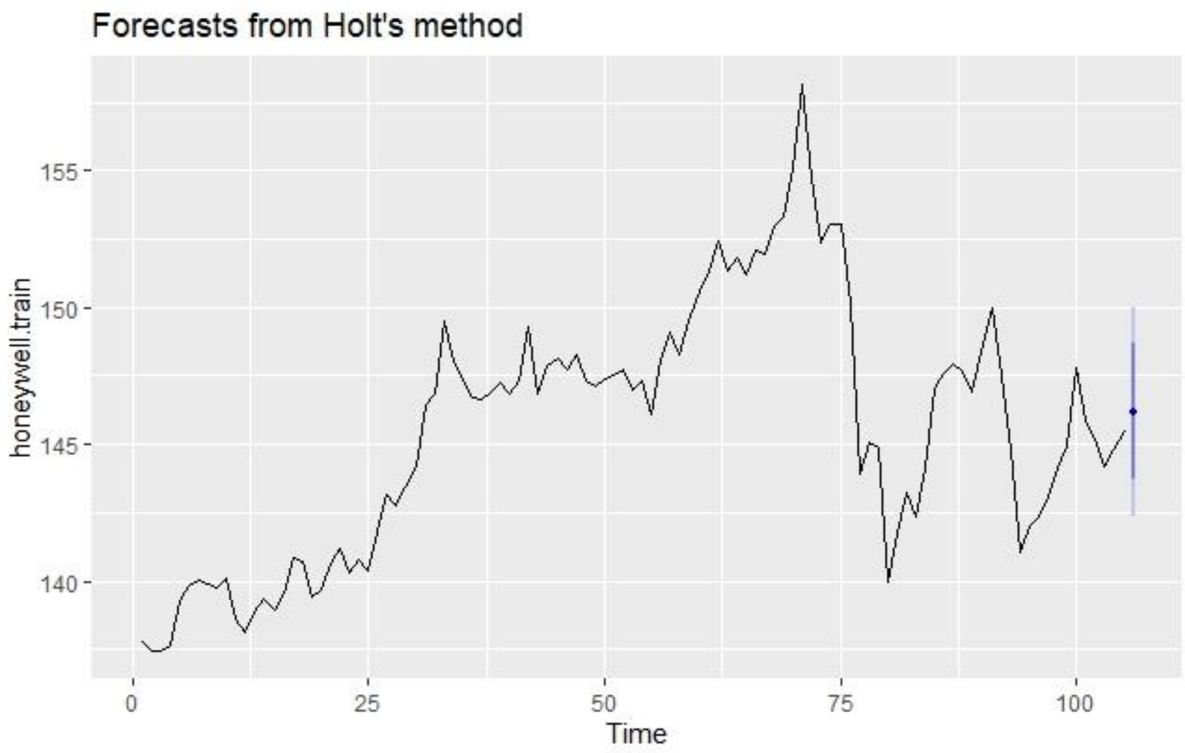


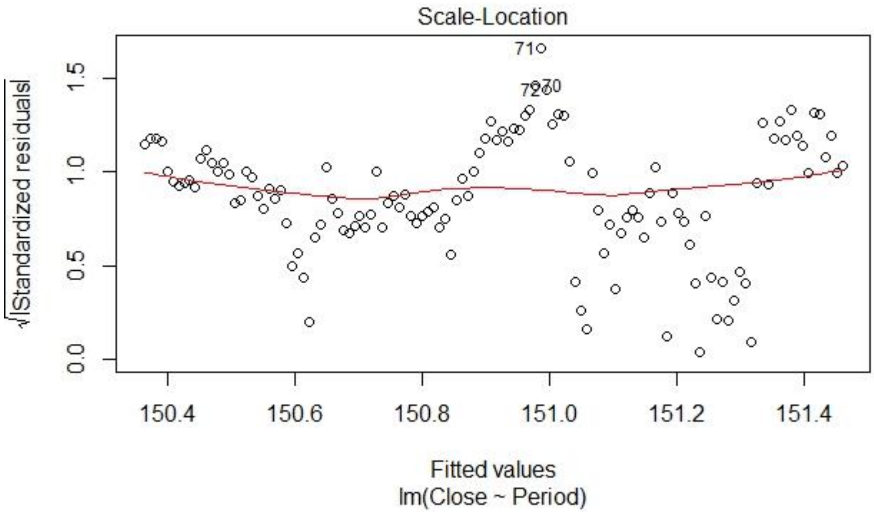
Figure 9: Alpha = 0.75, Beta = 0.85

## R-Code:

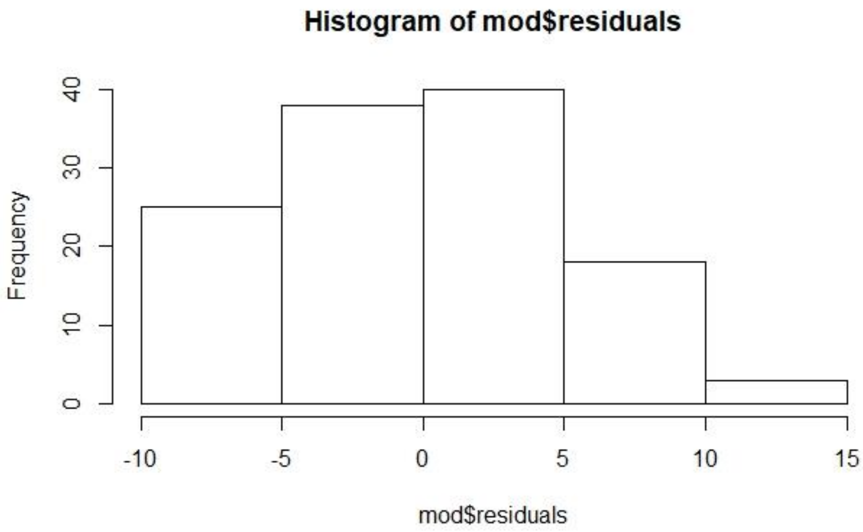
```
#HOLT's Method

#Assigning the values for beta and keeping the alpha value constant

b1 <- holt(honeywell.train, alpha = 0.75, beta = 0.15, h = 1)

b1

#Plotting the Holt's method plot for beta = 0.15

autoplot(b1)

#Determining the MSE for beta = 0.15

b1$model$mse

#Assigning the values for beta and keeping the alpha value constant

b2 <- holt(honeywell.train, alpha = 0.75, beta = 0.25, h = 1)

b2

#Plotting the Holt's method plot for beta = 0.25

autoplot(b2)

#Determining the MSE for beta = 0.25

b2$model$mse

#Assigning the values for beta and keeping the alpha value constant

b3 <- holt(honeywell.train, alpha = 0.75, beta = 0.45, h = 1)

b3

#Plotting the Holt's method plot for beta = 0.45

autoplot(b3)

#Determining the MSE for beta = 0.45

b3$model$mse

#Assigning the values for beta and keeping the alpha value constant

b4 <- holt(honeywell.train, beta = 0.85, h = 1)

b4

#Plotting the Holt's method plot for beta = 0.85

autoplot(b4)

#Determining the MSE for beta = 0.85

b4$model$mse
```

3.

a. In part 3, we use the simple linear regression method to forecast the stock for Honeywell. We know that the stock prices on the dates are divided into different periods (1, 2, 3...). We use this period to determine the fitting linear models. Since the simple line regression method is used to determine the forecasts, X-axis is the Honeywell stock prices (Closing prices) and on the Y-axis are the periods. Lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance. P-value of the both intercept and slope are lesser than 0.05 we conclude that intercept (8.92) and slope (1.504) are not equals to zero.

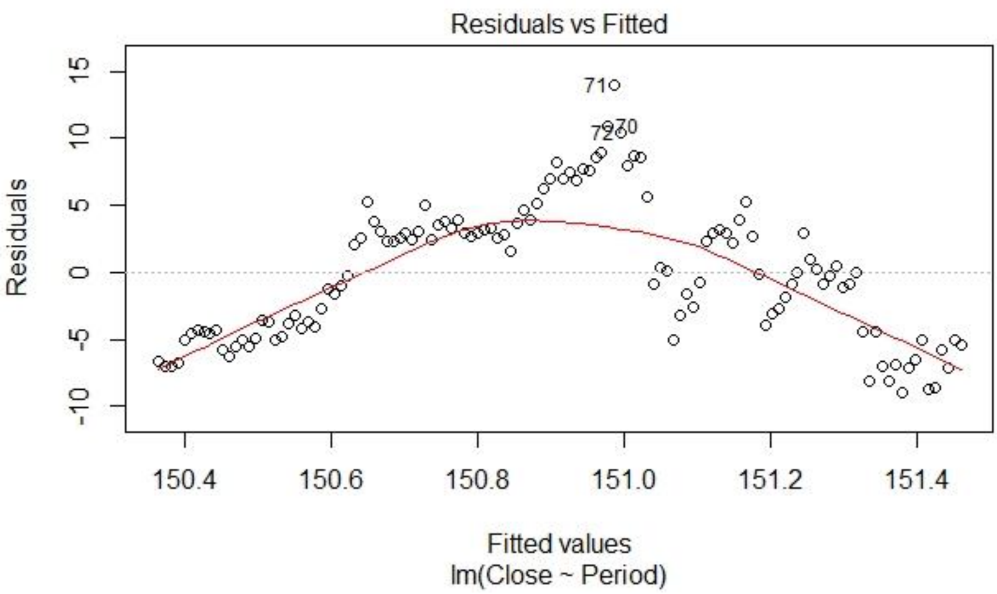



b. A residual is the vertical distance between a data point and the regression line. Each data point has one residual. They are positive if they are above the regression line and negative if they are below the regression line. An important way of checking whether a regression, simple or multiple, has achieved its goal to explain as much variation as possible in a dependent variable while respecting the underlying assumption, is to check the residuals of a regression. In the above values, we can see the mean absolute percentage error and the accuracy is also good. From the above observations, we can conclude that the model is good.

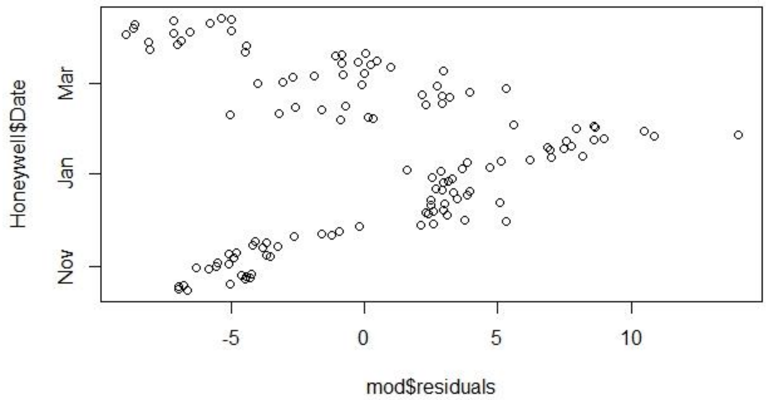



c. For a Chi-square test, a p-value that is less than or equal to your significance level indicates there is sufficient evidence to conclude that the observed distribution is not the same as the expected distribution. You can conclude that a relationship exists between the categorical variables.

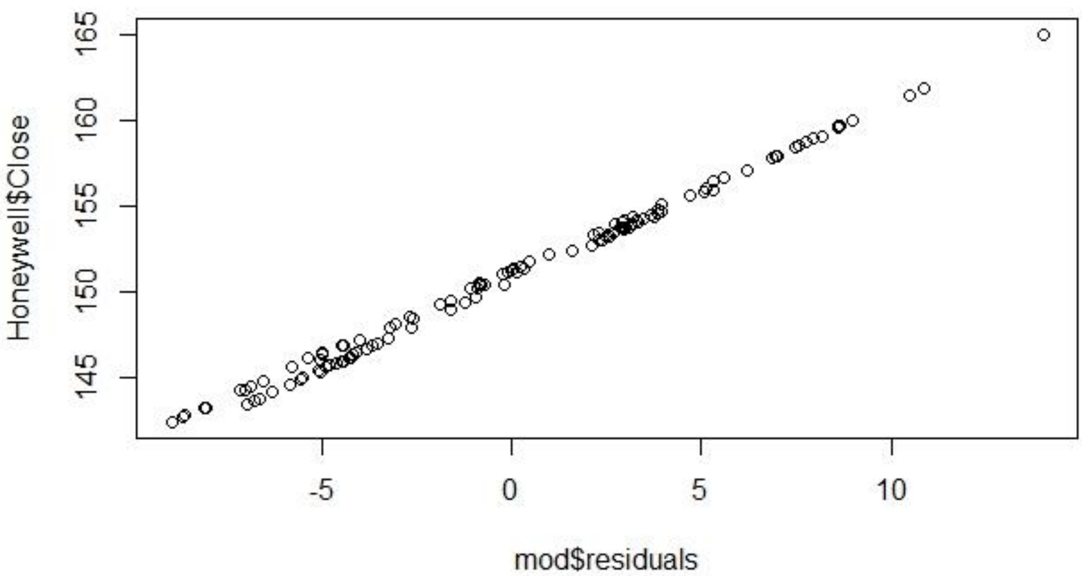
d. The normal probability plot is a graphical tool for comparing a data set with the normal distribution. We can use it with the standardized residual of the linear regression model and see if the error term ϵ is actually normally distributed. We can observe that the probability is more than 0.05, thus we accept the null hypothesis saying that the residual data is normally distributed. Therefore, normality is met on residuals.



e. Scatterplot – Residuals vs Time: Here, we can observe that the probability is greater than 0.05. So, we accept the null hypothesis saying that there is no correlation among the residuals. Thus, the independency is met.




f. Scatterplot – Residuals vs Predicted stock values: In the normal q-q plot drawn, the slope is equal to the residuals which are linearly distributed. In sales-location plot, all the residuals are scattered which means that none of the points are clustered at a single spot. Therefore, homoscadescity is met on residuals.



## **R-Code**:

```
#Part 3

#Performing Simple Regression Analysis

mod <- lm(Close ~ Period, data = Honeywell)

mod

plot(mod)

#Determining the coefficients of corelations and determinations

summary(mod)

#Determning the residual values

mod$residuals

#Plotting a histogram of regression residuals

plot(mod$residuals)

abline(lm(Close ~ Period, data = Honeywell), col = "red")

hist(mod$residuals)

#Performing the Chi-squared test

x <- abs(mod$residuals)

x

chisq.test(x)

#Plotting a normal probability plot of the residuals

plot(mod)

#Plotting a ascatter plot of residuals vs time

plot(mod$residuals, Honeywell$Date)

#Plotting a scatter plot of residuals vs predicted stock values

plot(mod$residuals, Honeywell$Close)
```

4. From the above three methods, namely SES method, Holt's method and the simple regression method we can find which method is the most accurate in its results. The method that gives the forecasted value which is near to the actual value is considered as the most accurate in forecasting or predicting the Honeywell stock value.

| Sr. No. | Method | Forecasted Values (04/16/2018) | Actual Value (04/16/2018) |
|---|---|---|---|
| 1 | SES | 145.2844 | 140.65 |
| 2 | Holt’s | 145.3163 | 140.65 |
| 3 | Simple Regression | 150.9115 | 140.65 |

As you can observe from the above table, that the forecast from the SES method is most nearest to the actual value, thus we can say that the SES method is the most accurate to predict the stock value of Honeywell on the given date.

# Conclusion

The assignment gives the real – world situation of analysing the Honeywell stock based on the past values. We can conclude that the stock values that were forecasted or predicted were pretty near to the actual stock value of the company. This concludes that the forecasting methods that were used – SES, Holt's and simple regression are all tested procedures and can be used in the future as well.

## References:


1. http://uc-r.github.io/ts_exp_smoothing
2. http://www.learnbymarketing.com/tutorials/linear-regression-in-r/
3. https://www.datacamp.com/community/blog/r-xts-cheat-sheet
4. https://rpubs.com/Abhilash333/396604

- **R-Code (Full Code) : -**

```
#ALY6050_Project3_Gujrathi_D
#Installing package "quantmod" for getting the stock values
install.packages("quantmod")
library(quantmod)
#Assiging the start dates
start <- as.Date('2017-10-15')
#Assiging the end dates
end <- as.Date('2018-04-16')
#Calling the stock prices of Honeywell
getSymbols("HON", src = "yahoo", from = start, to = end)
#Determining the classes of the dataset
class(HON)
#Determinig the first values of the set
head(HON)
#Assigning a vector to the column
honeywell <- HON$HON.Close
honeywell
#Plotting the stock price chart
plot(honeywell)
#Assigning the training values
honeywell.train <- window(honeywell, end = as.Date('2018-03-16'))
honeywell.train
#Assigning the testing values
```

```
honeywell.test <- window(honeywell, start = as.Date('2018-03-15'))
honeywell.test

#Part 1
#SES Method
#Installing Package "Forecast"
install.packages("forecast")
#Installing Package "Metrics"
install.packages("Metrics")
library(forecast)
library(Metrics)
#Performing the Simple Exponential Smoothing for alpha = 0.15
ses.honeywell1 <- ses(honeywell.train, alpha = 0.15, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.15
autoplot(ses.honeywell1)
#Determining the MSE for alpha = 0.15
ses.honeywell1$model$mse
#Performing the Simple Exponential Smoothing for alpha = 0.35
ses.honeywell2 <- ses(honeywell.train, alpha = 0.35, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.35
autoplot(ses.honeywell2)
#Determining the MSE for alpha = 0.35
ses.honeywell2$model$mse
#Performing the Simple Exponential Smoothing for alpha = 0.55
ses.honeywell3 <- ses(honeywell.train, alpha = 0.55, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.55
autoplot(ses.honeywell3)
```

```
#Determining the MSE for alpha = 0.55
ses.honeywell3$model$mse
#Plotting Simple Exponential Smoothing for alpha = 0.75
ses.honeywell4 <- ses(honeywell.train, alpha = 0.75, h = 10)
#Plotting Simple Exponential Smoothing for alpha = 0.75
autoplot(ses.honeywell4)
#Determining the MSE for alpha = 0.75
ses.honeywell4$model$mse
#Part 2
#HOLT's Method
#Assigning the values for beta and keeping the alpha value constant
b1 <- holt(honeywell.train, alpha = 0.75, beta = 0.15, h = 1)
b1
#Plotting the Holt's method plot for beta = 0.15
autoplot(b1)
#Determining the MSE for beta = 0.15
b1$model$mse
#Assigning the values for beta and keeping the alpha value constant
b2 <- holt(honeywell.train, alpha = 0.75, beta = 0.25, h = 1)
b2
#Plotting the Holt's method plot for beta = 0.25
autoplot(b2)
#Determining the MSE for beta = 0.25
b2$model$mse
#Assigning the values for beta and keeping the alpha value constant
b3 <- holt(honeywell.train, alpha = 0.75, beta = 0.45, h = 1)
b3
```

```
#Plotting the Holt's method plot for beta = 0.45
autoplot(b3)
#Determining the MSE for beta = 0.45
b3$model$mse
#Assigning the values for beta and keeping the alpha value constant
b4 <- holt(honeywell.train, beta = 0.85, h = 1)
b4
#Plotting the Holt's method plot for beta = 0.85
autoplot(b4)
#Determining the MSE for beta = 0.85
b4$model$mse
#Part 3
#Performing Simple Regression Analysis
mod <- lm(Close ~ Period, data = Honeywell)
mod
plot(mod)
#Determining the coefficients of corelations and determinations
summary(mod)
#Determning the residual values
mod$residuals
#Plotting a histogram of regression residuals
plot(mod$residuals)
abline(lm(Close ~ Period, data = Honeywell), col = "red")
hist(mod$residuals)
#Performing the Chi-squared test
x <- abs(mod$residuals)
x
```

```
chisq.test(x)
#Plotting a normal probability plot of the residuals
plot(mod)
#Plotting a ascatter plot of residuals vs time
plot(mod$residuals, Honeywell$Date)
#Plotting a scatter plot of residuals vs predicted stock values
plot(mod$residuals, Honeywell$Close)
```