Introduction :

Given 4 time series datasets to analysis and provide predictions. The first section of this report will be data exploration after which a brief analysis will be done to select the best suited model and forecast values.

Data Exploration and Analysis :

1. Securities.xls

This dataset consists of the monthly rate % of securities. There are 647 rows in this dataset. The starting entry is from April 1953 to February 2007. Firstly, we will be checking for the time series plot. This will provide us with some insight about what kind of trend and seasonality this dataset has.

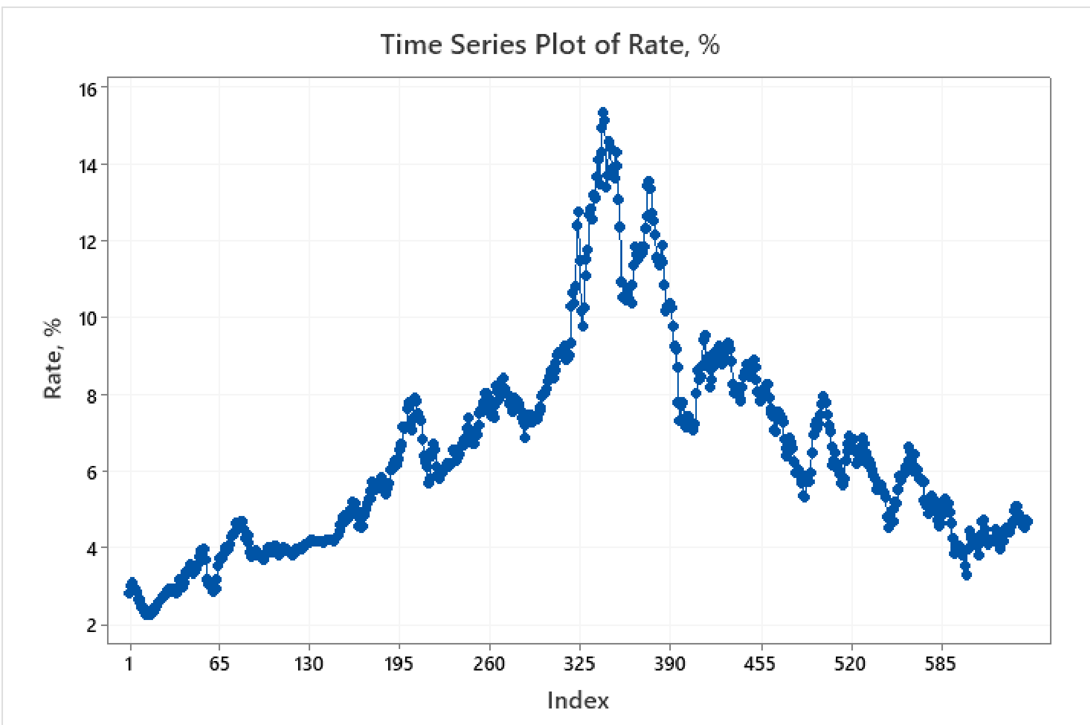


Figure 1 : Time series plot of Securities dataset

The above plot suggests that the monthly change in the Rate% may have a random trend. To verify this let us take a look at the Autocorrelation Function(ACF).

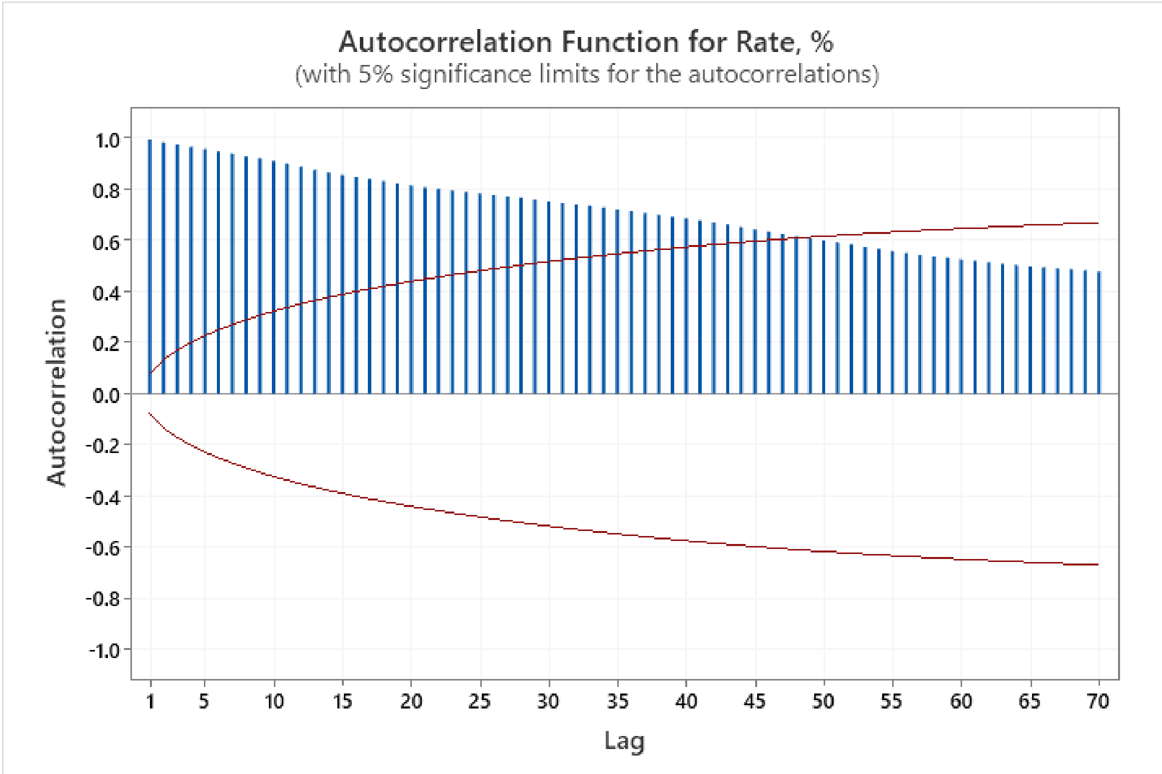


Figure 2 : ACF with default lag

The ACF with default lag is plotted and is shown in the figure above. We can observe that the plot does not get close to zero and we can conclude that there is no trend for this dataset. Let us take the ACF with lag 12 since we are given monthly records of Rate%.

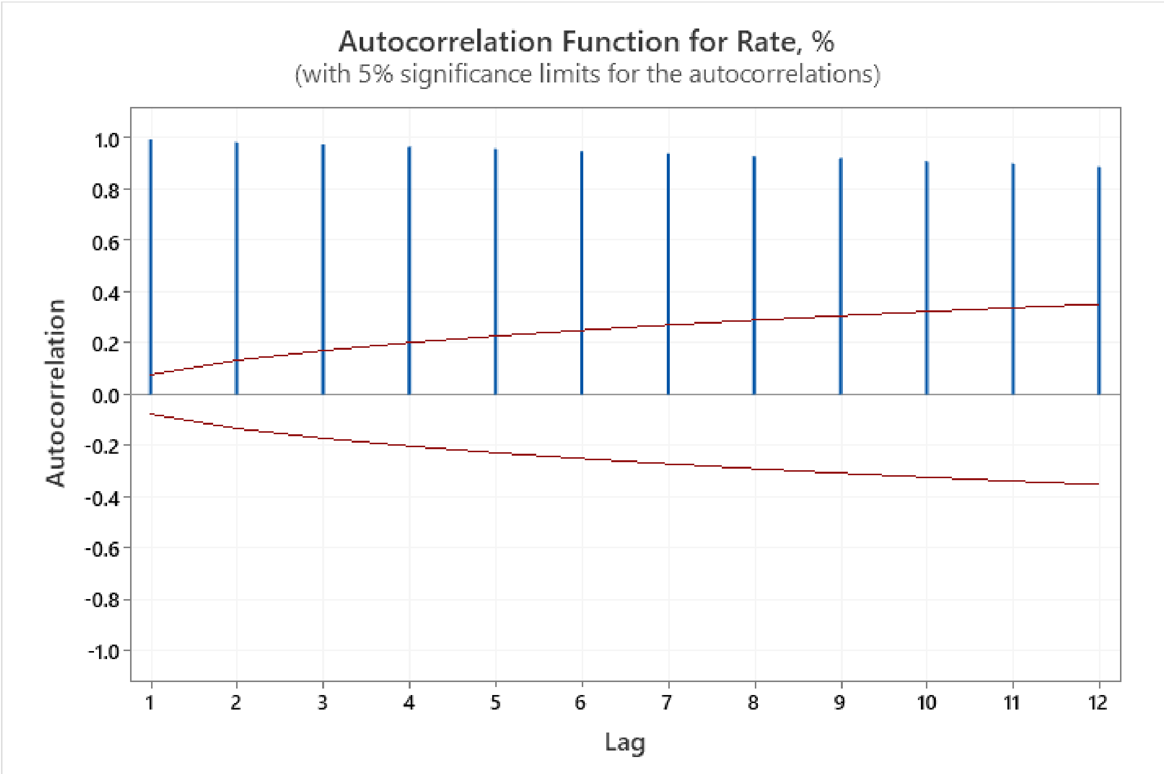
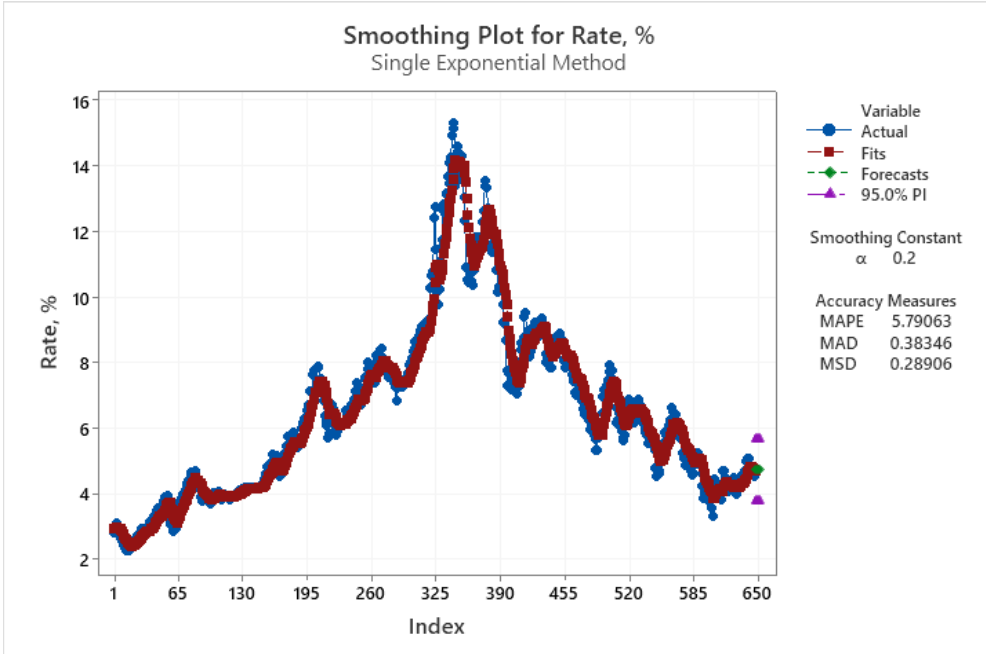
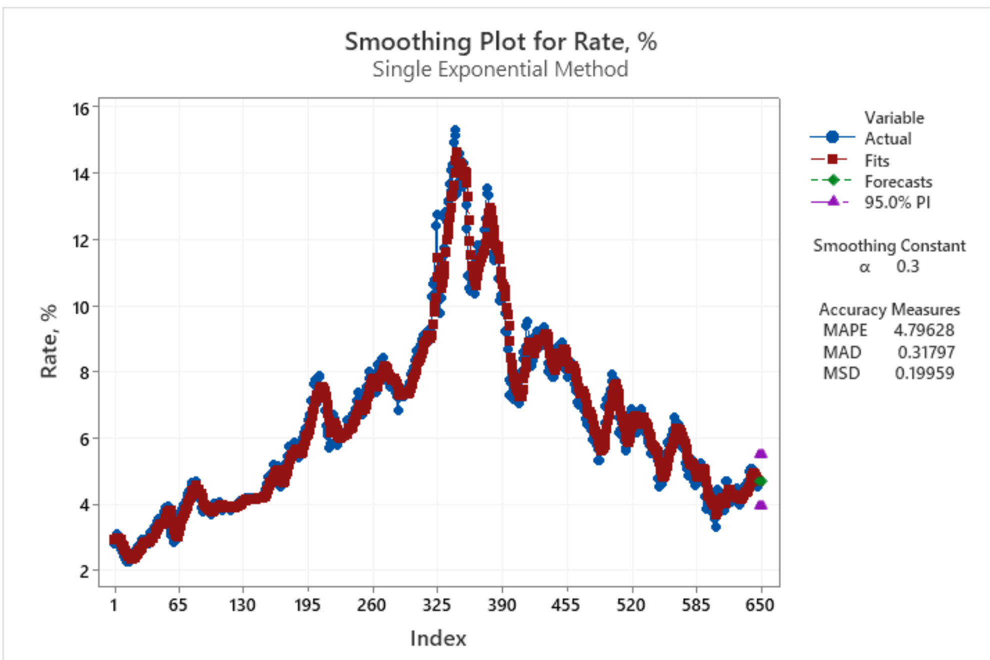


Figure 3 : ACF with lag 12

Since there is no trend we will be utilizing single exponential smoothing with varying alpha values to come up with a model.

We have found out that single exponential smoothing will be utilised since there is no trend that can be seen for this dataset. Given below are single smoothing plots with 3 different alpha values, 0.2, 0.3, and 0.4.


Smoothing Plot for Rate, %
Single Exponential Method
Rate, %
Index
Variable
Actual
Fits
Forecasts
95.0% PI
Smoothing Constant
α 0.2
Accuracy Measures
MAPE 5.79063
MAD 0.38346
MSD 0.28906


Smoothing Plot for Rate, %
Single Exponential Method
Rate, %
Index
Variable
Actual
Fits
Forecasts
95.0% PI
Smoothing Constant
α 0.3
Accuracy Measures
MAPE 4.79628
MAD 0.31797
MSD 0.19959

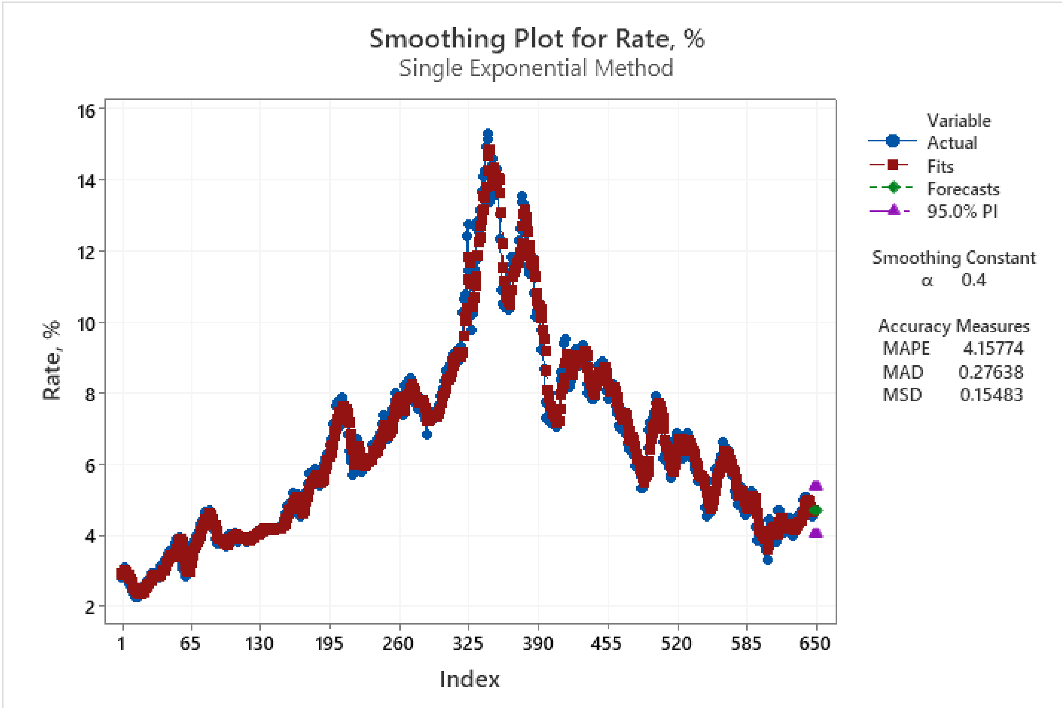


Figure 4,5,6 : Single exponential method with alpha values as 0.2,0.3,0.4 respectively

As the smoothing constant increases we can observe that the accuracy measures get better. And we can look at the graph and see that the range for the PI gets smaller as well. The difference between these models is small but we will be considering the model from Figure 6.

Given below are the next 5 forecasted values(648 to 652) for this model.

Forecasts

Period	Forecast	Lower	Upper
648	4.70817	4.03106	5.38528
649	4.70817	4.03106	5.38528
650	4.70817	4.03106	5.38528
651	4.70817	4.03106	5.38528
652	4.70817	4.03106	5.38528

2. Energybills.xls

This dataset consists of energy bills over a period of time. There are 40 entries in this dataset. The first thing we will be doing is plotting the time series plot of energy bills and see what we can find out.

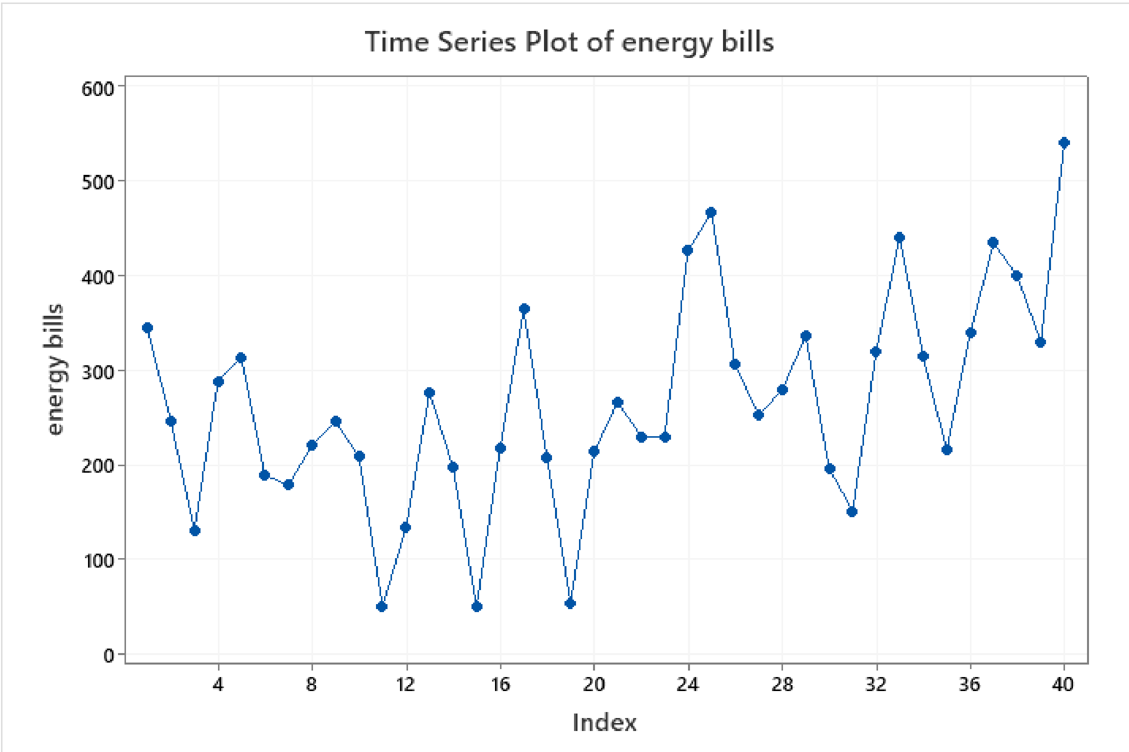


Figure 7 : Time series for energy bills

From figure 7, we can see that there may be a linear trend. Let us apply time series analysis to check which trend model works best for this dataset.

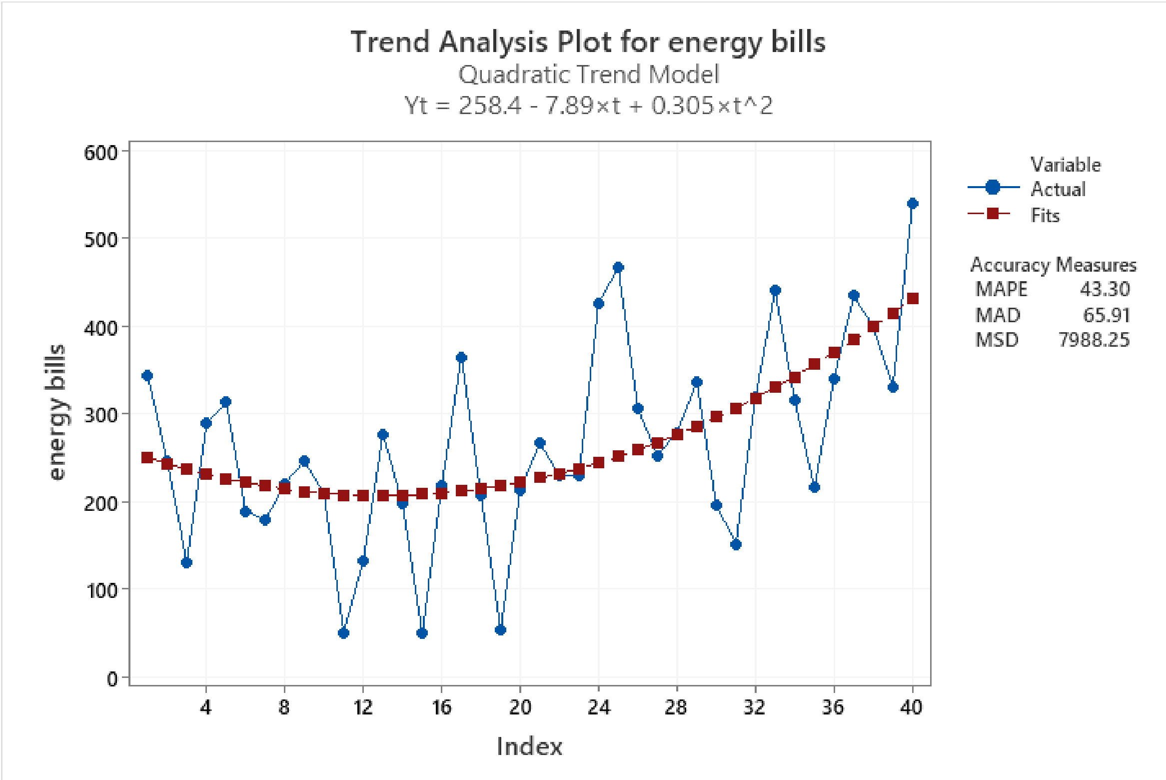


Figure 8 : Quadratic Trend model for energy bills

After checking with Linear, Quadratic, Exponential trend models, the quadratic model seems to be better than the other two.

Let us check for the ACF and partial ACF for energy bills with default lag to gain some insights.

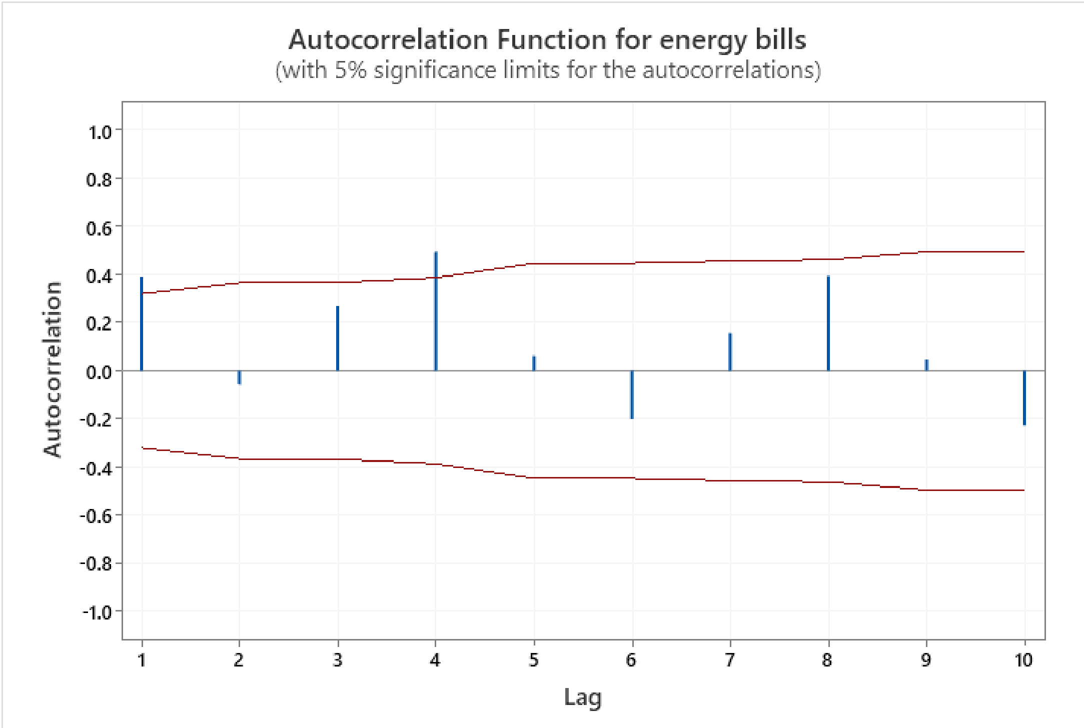


Figure 9 : ACF for energy bills with default lag

The above ACF shows us that the first and fourth entries are outside of the marked red region. This tells us that we probably need to take a difference of the energy bills.

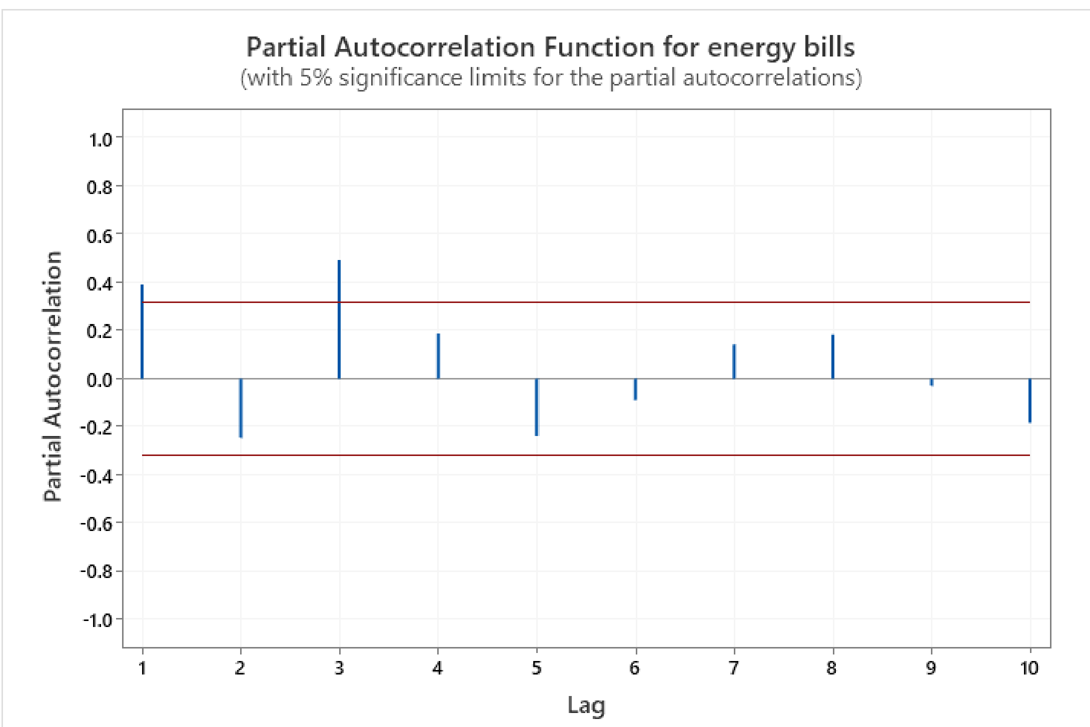


Figure 10 : Partial ACF for energy bills with default lag

The partial ACF shows us the first and third entries are out of bounds.
The second order difference is taken and the time series plot is checked to see if there is any change.

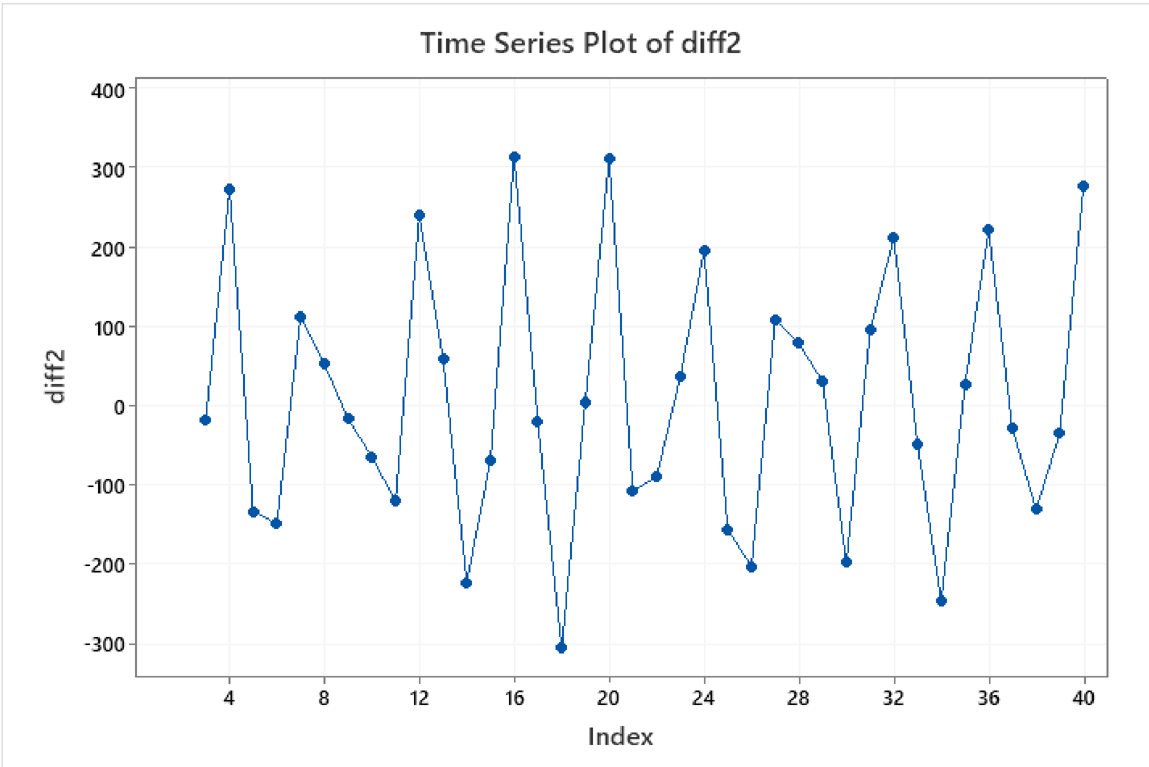


Figure 8 : Time series plot for energy bills with second order difference

As we can see there is a small linear trend and now we will check for the ACF and partial ACF for this second order difference.

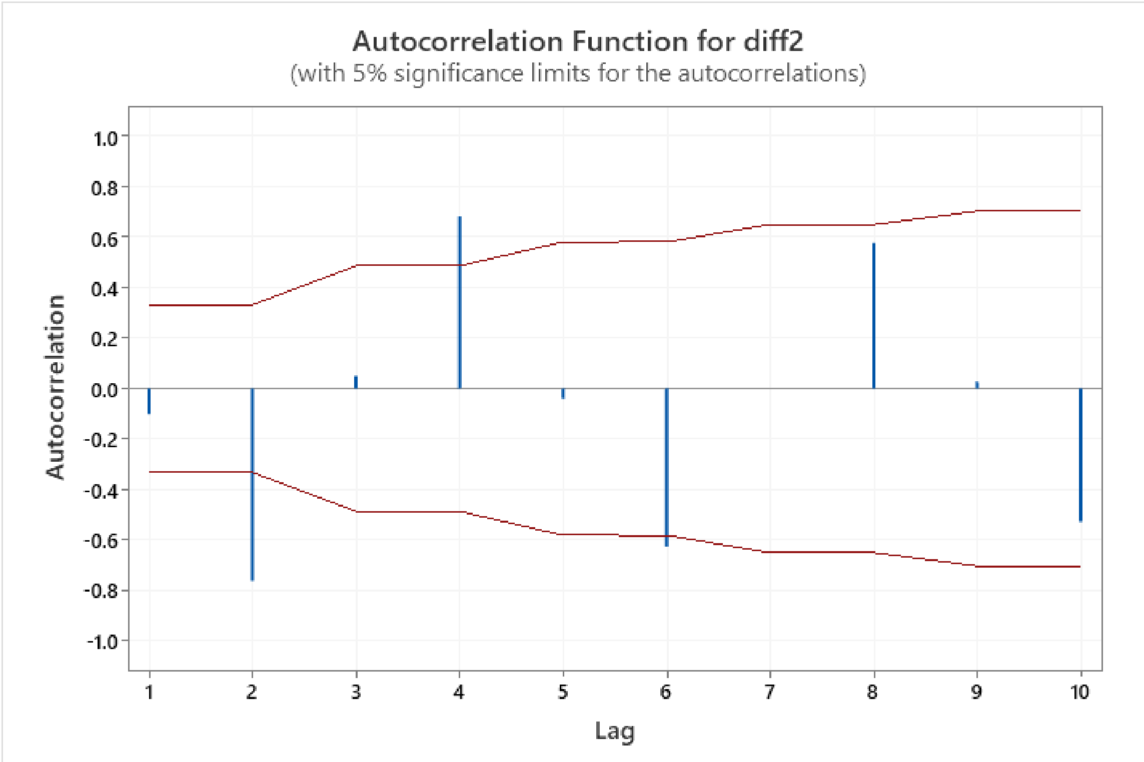


Figure 9 : ACF of second order difference

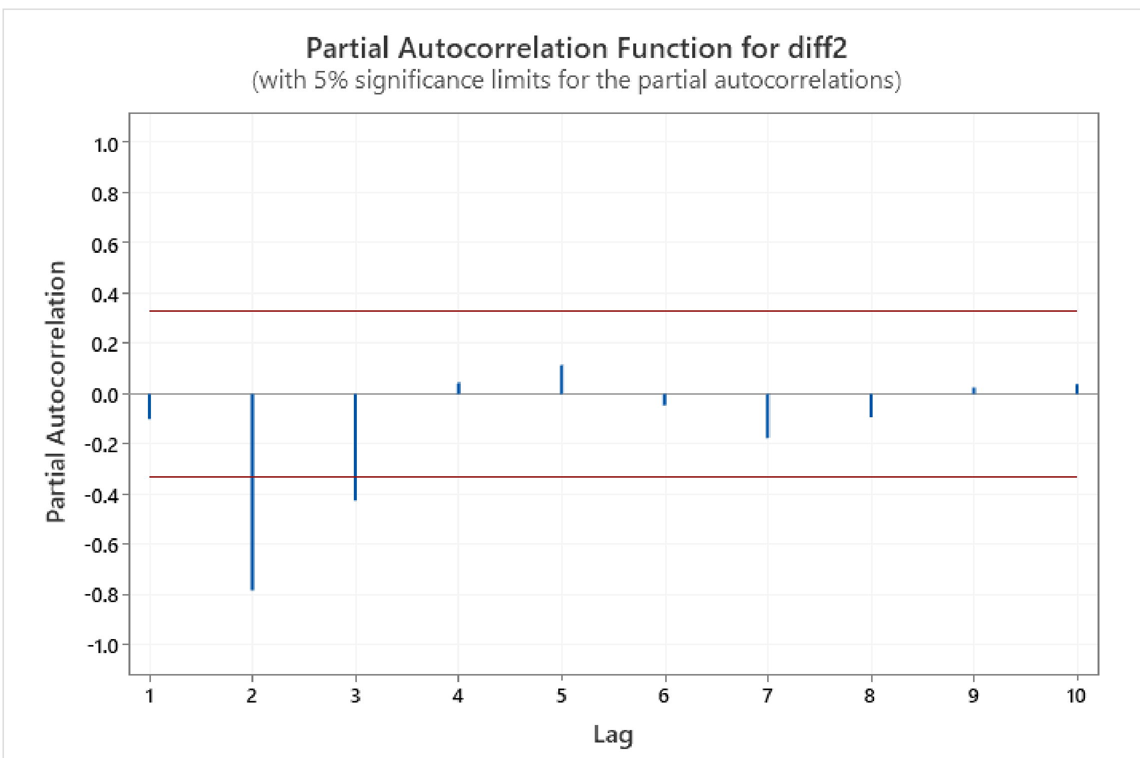


Figure 10 : Partial ACF of second order difference

The second entry in both of these ACF and partial ACF tells us that we could run ARIMA with an autoregressive value of 2 and seasonal difference of 2. Moving average is taken as zero.

ARIMA is utilized to come with a model with autoregressive value and the seasonal difference of 2. The given picture below is for forecasts from the last row(40). These values are the next 5 forecasted values for energy bills.

Forecasts from period 40

		95% Limits		
Period	Forecast	Lower	Upper	Actual
41	729.93	557.608	902.26	
42	672.48	316.667	1028.28	
43	684.09	219.657	1148.52	
44	913.55	334.579	1492.51	
45	1041.28	262.646	1819.90	

We will be storing the forecasted values and upper and lower bound values. A time series plot is utilized to check and verify forecasted values lie in the range of the lower and upper bounds.

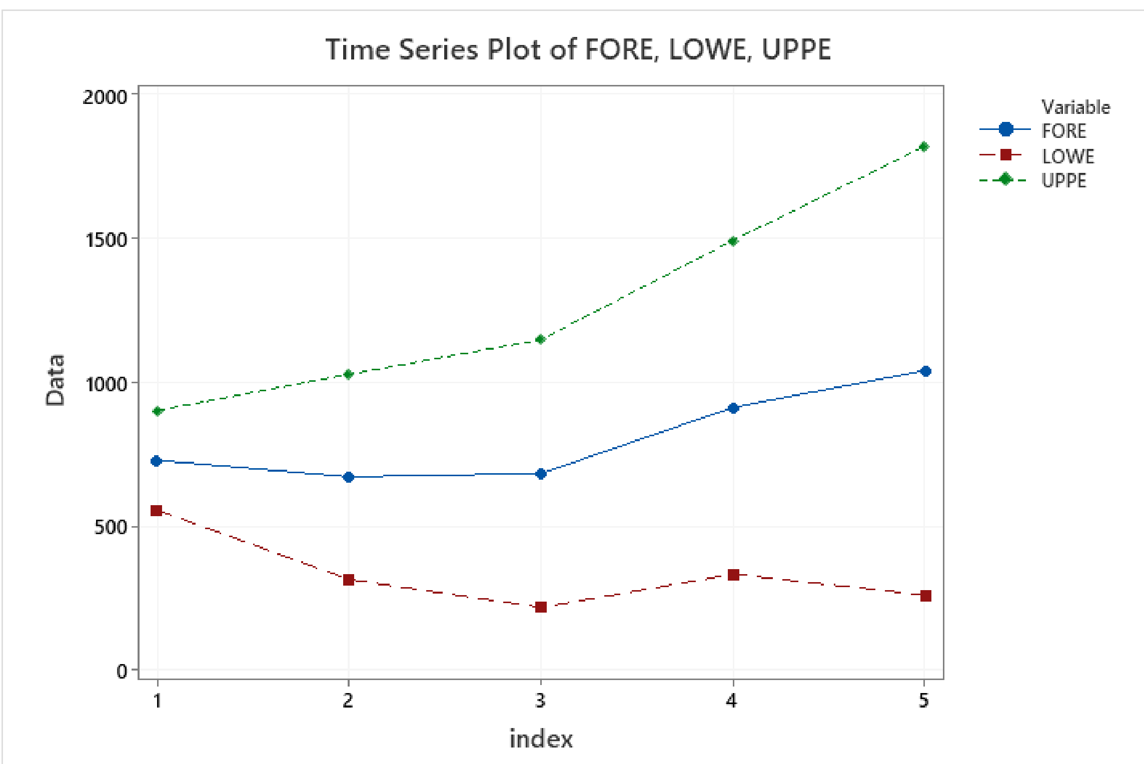


Figure 11 : Plot for forecast, lower and upper bound values

3. AirCanada.xls

This dataset consists of Air Canada miles flown with 96 entries. The question tells us that the year is from 1998 to 2005. We will be adding a year column with 1998 to 2005 for each month (12times).

Let's look at the time series plot.

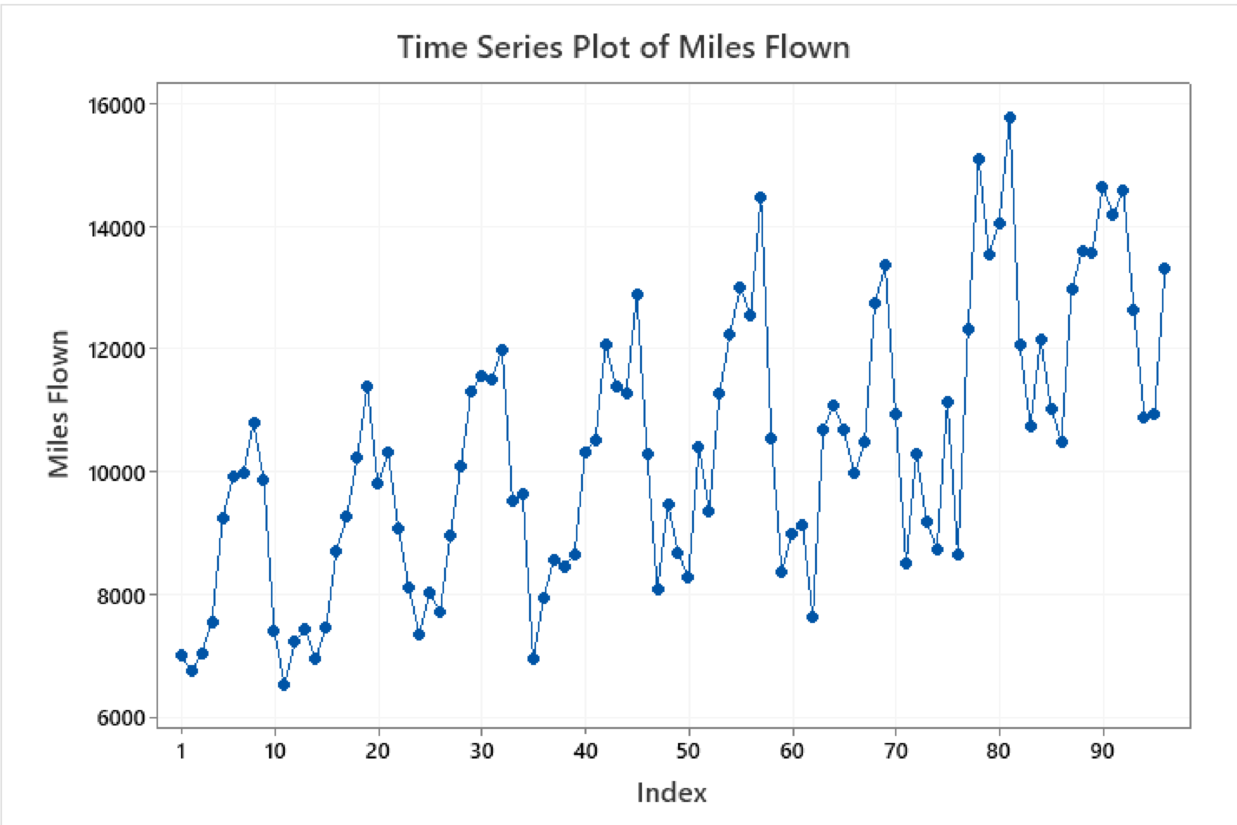
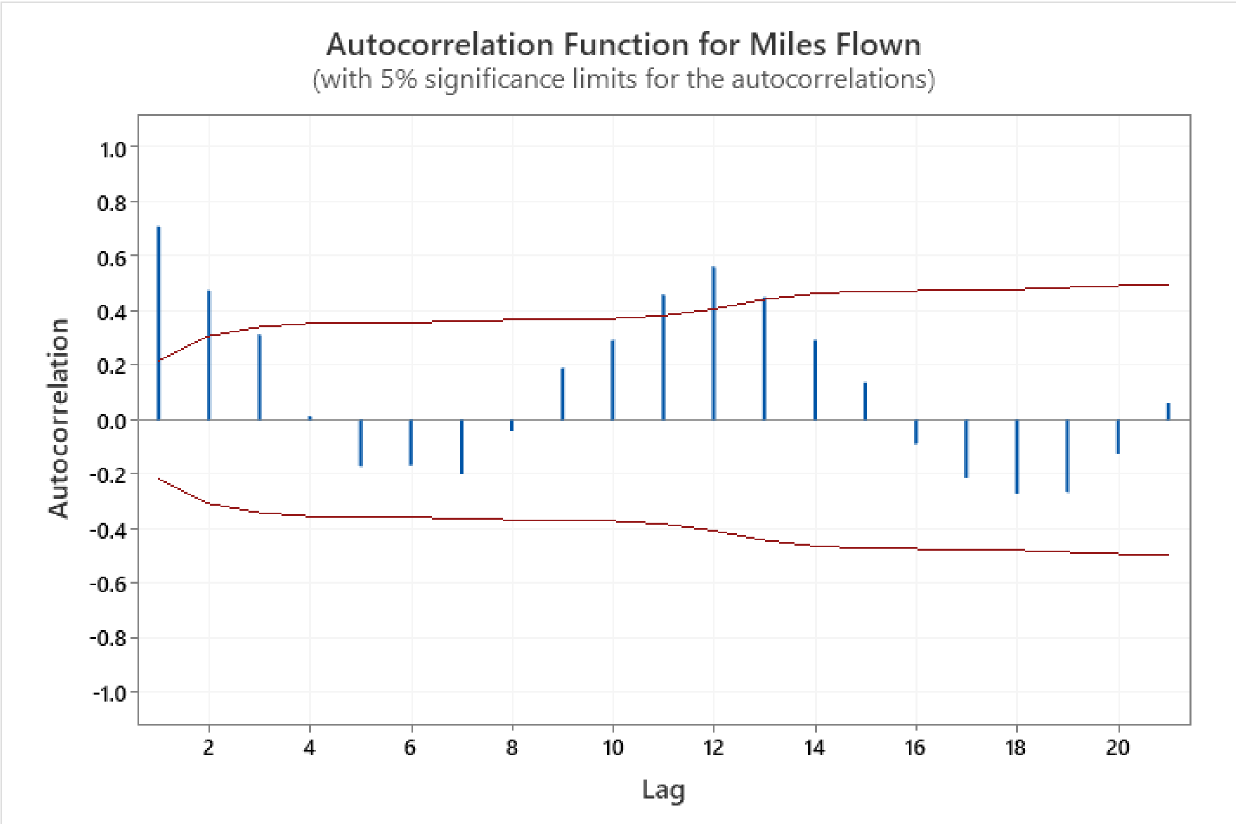


Figure 12 : Time series plot of miles flown

The below acf shows us that this dataset (aircanada) has a seasonal component.



There seems to be a trend and seasonality to this dataset. We know that Holt-Winters method is going to be utilized for this time series(given in the question) but let us try to use single and double exponential smoothing to get an idea of what the alpha, beta, gamma values we can use to come up with a good model.

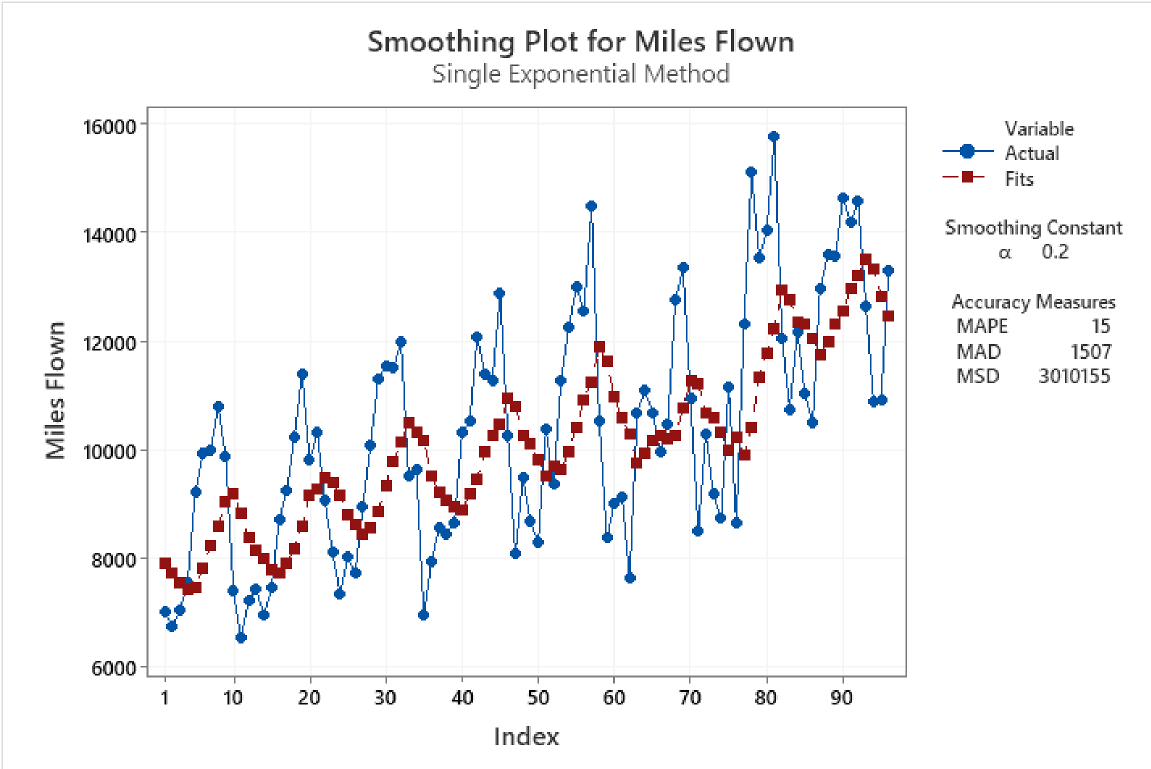


Figure 13: Alpha value 0.2 Single exponential

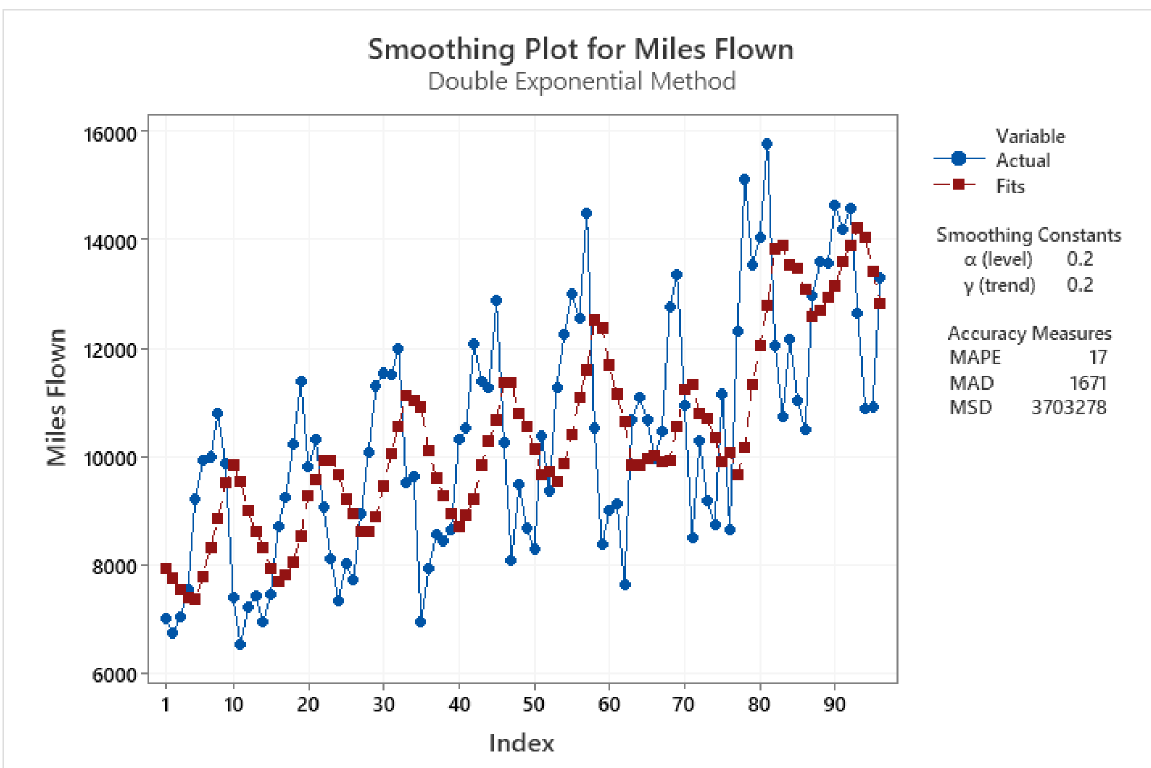


Figure 14: Double exponential with level and trend as 0.2

In the two above plots, the alpha value is taken as 0.2. For figure 14, the trend is taken as 0.2. We can observe in both of these graphs that the fit values are following a similar pattern to the actual values but are not accurate enough. This could be because these two techniques do not take into consideration the seasonality into consideration(gamma value).

Let us check for alpha, beta and gamma values as 0.2,0.2,0.2 using the Holt-Winters method.

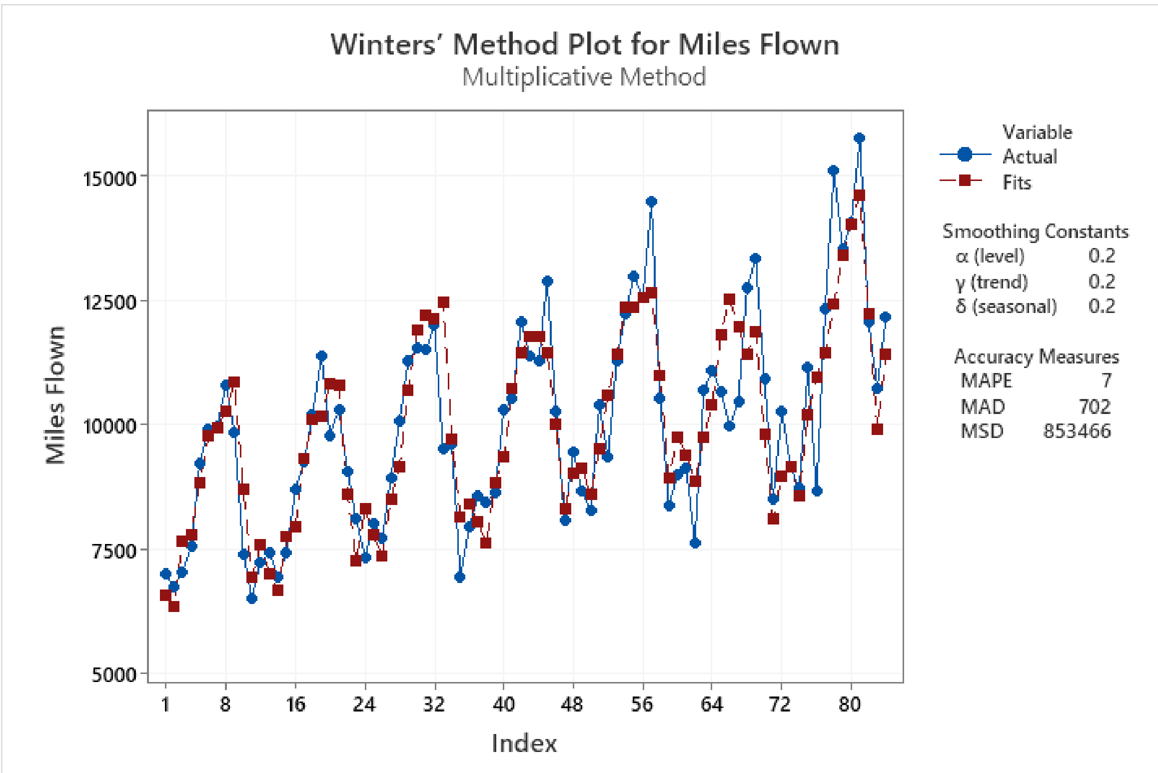


Figure 15 : Winters method using multiplicative method

In figure 15, we can see that when we used the default 0.2 value for all three smoothing constants the graph seems to follow the same cyclic pattern but does not get the values right. The accuracy measures when compared to other values of smoothing constants have similar accuracy measures except for the MSD which is the lowest in the above model.

4. **JJ.xls**

This dataset consists of the quarterly sales of Johnson and Johnson over the past couple of years from 1960-1980. A new column is created to accommodate getting each value per year and for each quarter. The time series graph is given below.

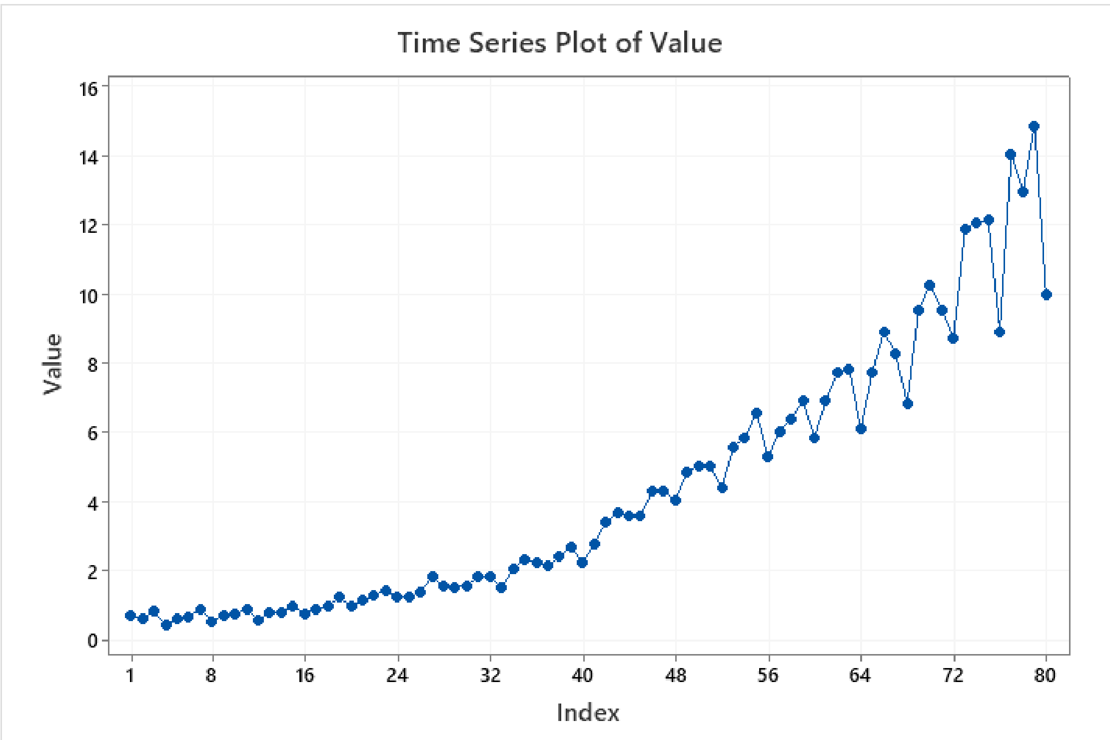


Figure 16 : Time series plot for JJ

From the above figure we can see that the plot has a trend and some seasonality.
Let us take the decomposition for this time series and find out the trend and how are the accuracy measures for the multiplicative and additive model.

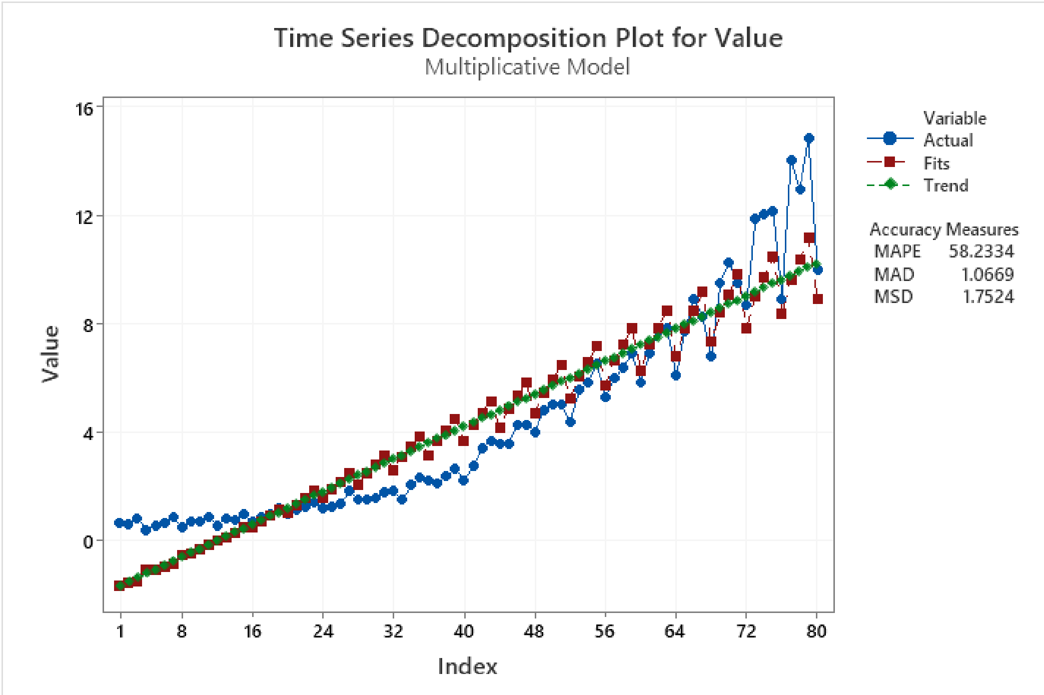
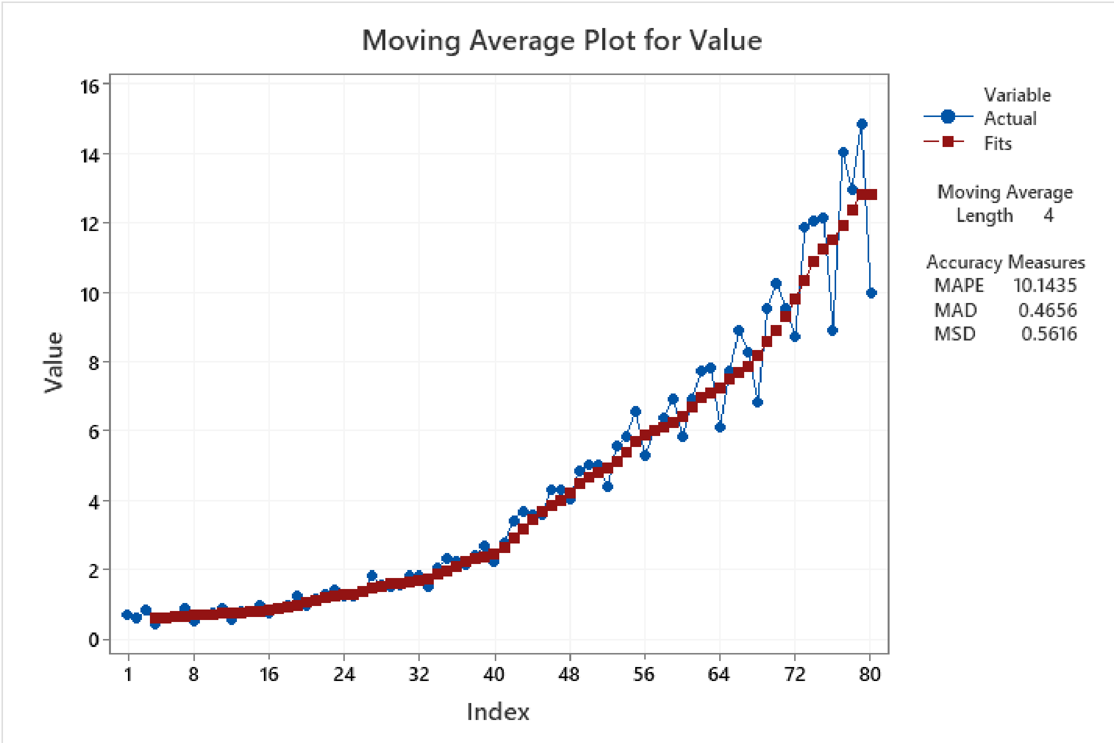


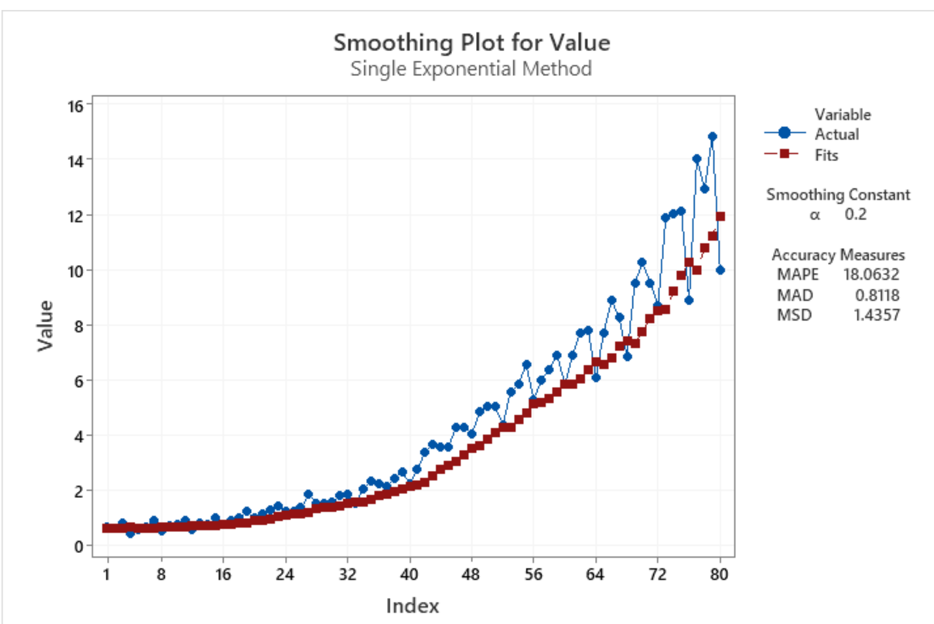
Figure 17 : Multiplicative model for JJ

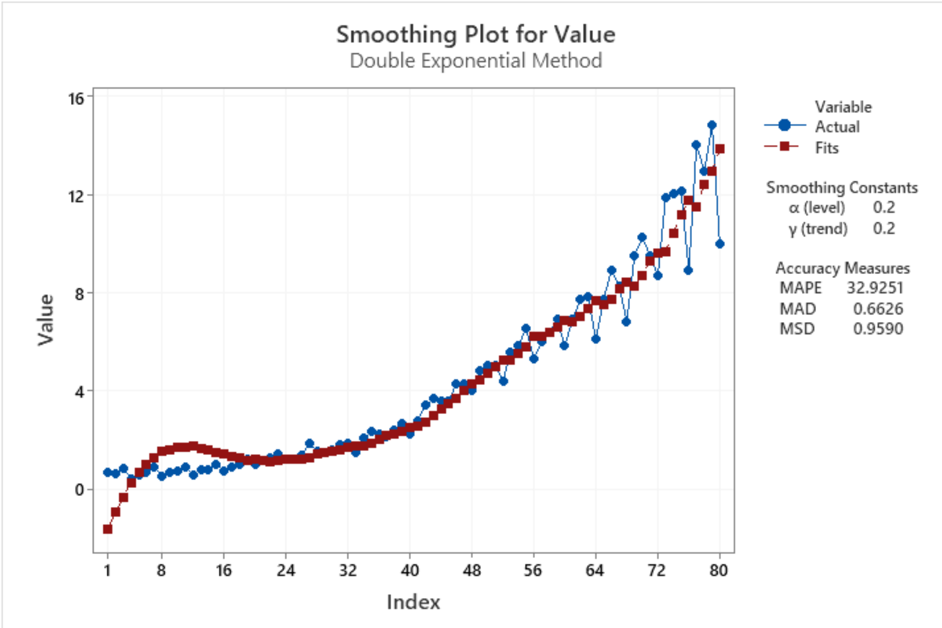
Multiplicative model does better.

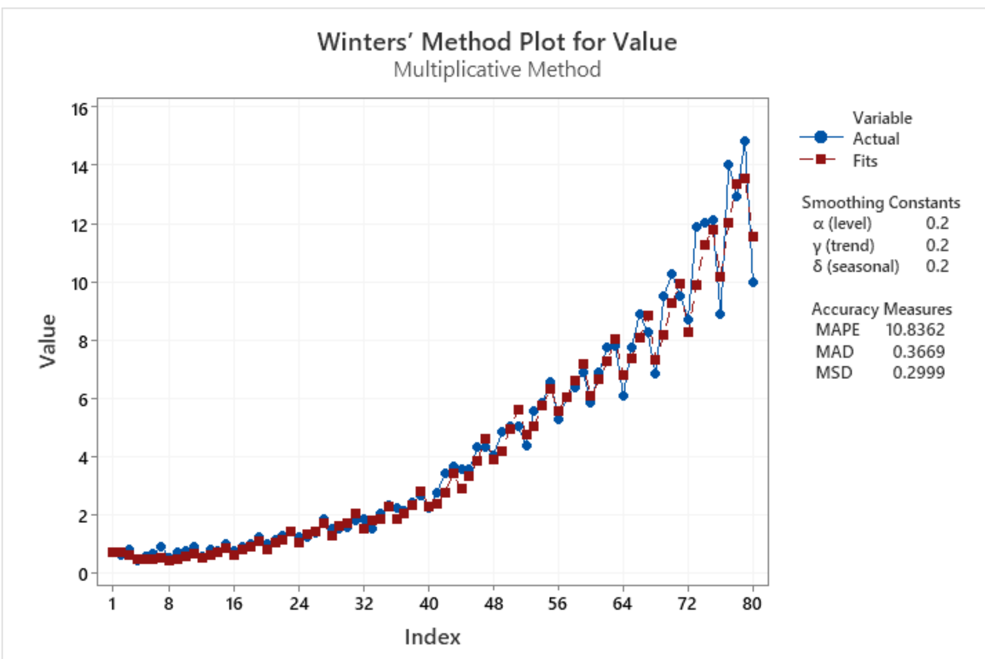
Moving averages plot is given below. We can see that as the index increases there is a dip in the accuracy.



Exponential smoothing is applied for this dataset. Let us check with single, double and winter exponential smoothing. After this we can decide which model works best.

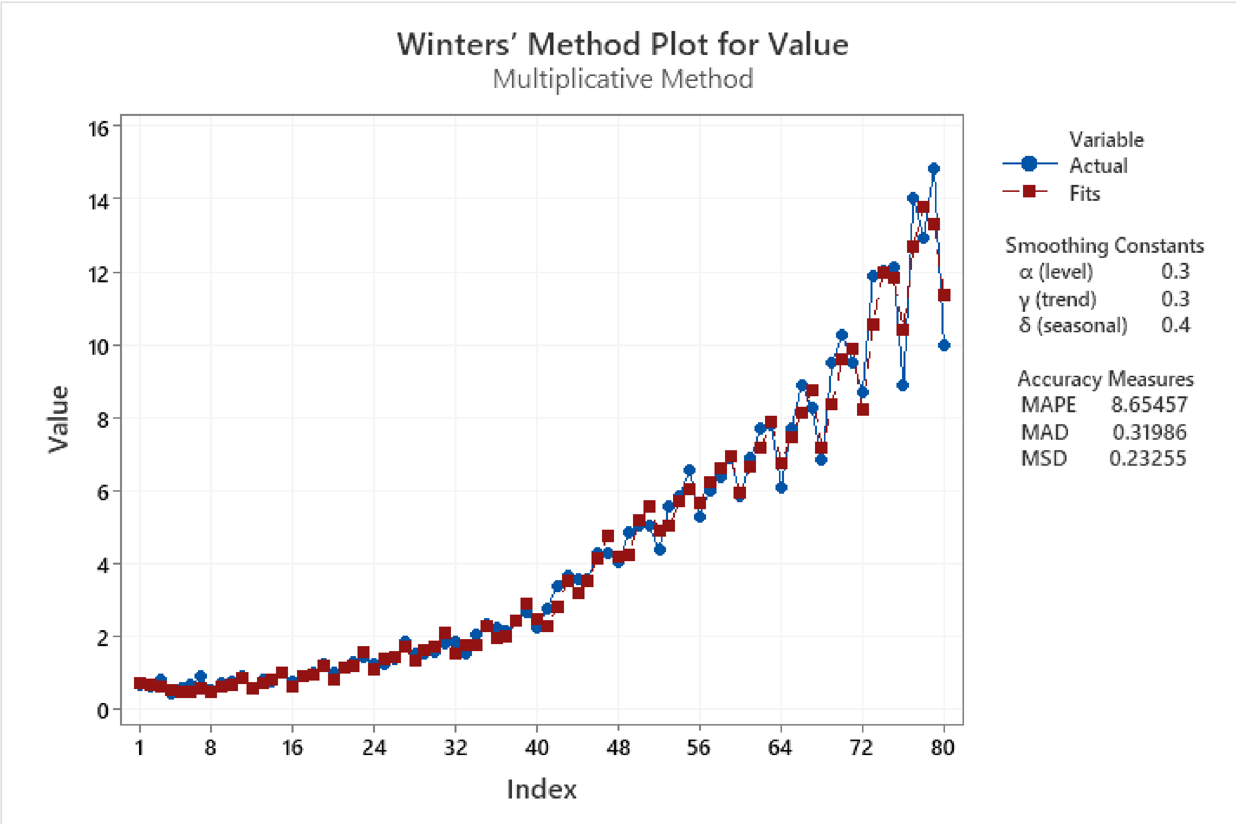






Holt-winters method seems to fit the model the best. The Alpha, beta and gamma values were taken as 0.2. The accuracy measures are lower than single and double exponential methods.

Let us tweak the parameters to get a better fitted model to forecast the values.



The above plot shows us a decent enough model using winters method. The accuracy measures are lower than when compared with lower values of alpha, beta and gamma. There may be other values that can be used to come up with a better model for forecasting.

Given below are the forecasts for JJ dataset.

Forecasts

Period	Forecast	Lower	Upper
81	14.9863	14.2027	15.7700
82	14.9305	14.0963	15.7648
83	15.3155	14.4212	16.2097
84	11.6898	10.7279	12.6517

Appendix :

1. Smoothing constant is taken as 0.9 and we can see that the Accuracy measures are lower than when the smoothing constant is in the range 0.2 to 0.8. The realistic smoothing constant should be taken as 0.4 but there are a few cases where smoothing constants from 0 to 1 can be considered. I checked for all values in the previous mentioned range and accuracy measures only differ slightly.