PRN: 22060641026

Name: Sainyukta Harichandra Gujar

Time Series Analysis – Assignment

Defining the Question:

In the question, a dataset is provided and it is expected to work on the columns of Revenue, Sales and Average cost from the dataset. The question can be divided into three sub-questions:

1. To suggest an appropriate smoothing technique among Exponential Smoothing, Holts Smoothing and Holts-Winter Smoothing for all the three columns (Revenue, Sales and Average cost).

2. To compare between Average cost and Revenue.

3. To justify whether both the columns (Average cost and Revenue), use the same technique or not.

Analysing the dataset:

Overview:

A time series is actually a set of observations measured at specified (usually equal) time interval. Here, the interest lies in smoothing out the series so that longer trend patterns can be revealed and short term fluctuations and noise can be smoothened out.

In the dataset there are 64 observations and 3 variables (Revenue, Sales and Average cost) and another column is Period. The observations are measured along a specified and equal time interval and thus the dataset is a time series.

Moving forward, among the three smoothing techniques (Exponential Smoothing, Holts Smoothing and Holts-Winter Smoothing), Exponential Smoothing is a weighted moving average technique and all the past observations are weighted equally. Holts Smoothing is when trend is present in the series. Lastly, Holts-Winter Smoothing is used if trend as well as seasonality is present in the series.

Software/Tools used:

The softwares used for the analysis are MS Excel and R. Graphs for smoothened data and smoothing of the series is done in MS-Excel. R is used to know whether the dataset is time series, to confirm the trend (using Mann Kendall test) and to look out for seasonality and lags using ACF plots.

Outputs and Graphs

```
> is.ts(dataset)
[1] TRUE
```

```
> MannKendall(dataset$Average_cost)
tau = 0.252, 2-sided pvalue =0.0033101
> MannKendall(dataset$Revenue)
tau = 0.617, 2-sided pvalue =< 2.22e-16
> MannKendall(dataset$Sales_quantity)
tau = 0.494, 2-sided pvalue =< 2.22e-16
```

Fig. (1)

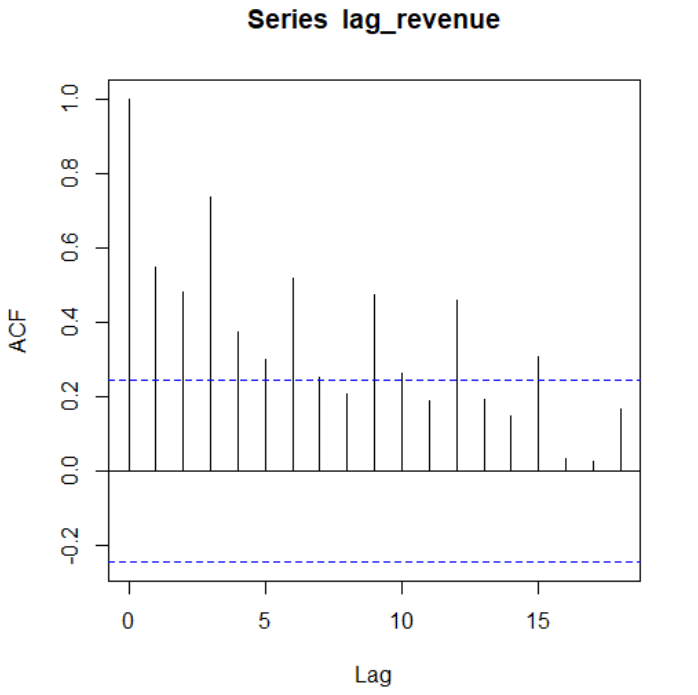

Fig. (2)

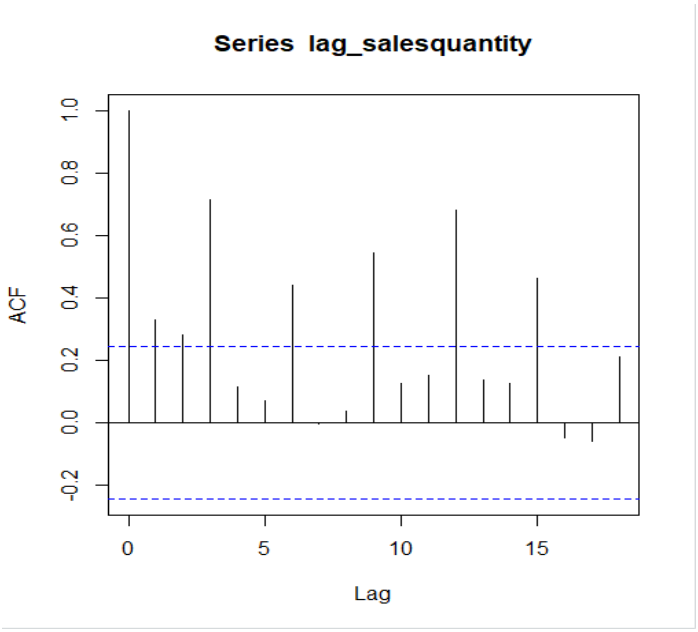

Fig. (3)

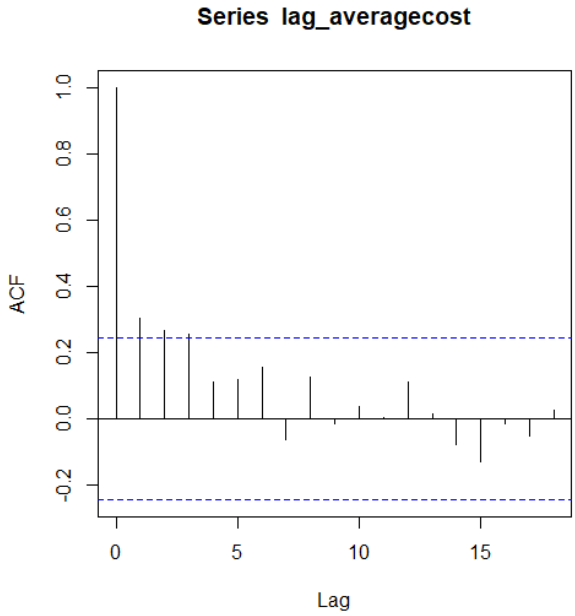
Series lag_averagecost
ACF
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
0
5
10
15
Lag

Fig. (4)

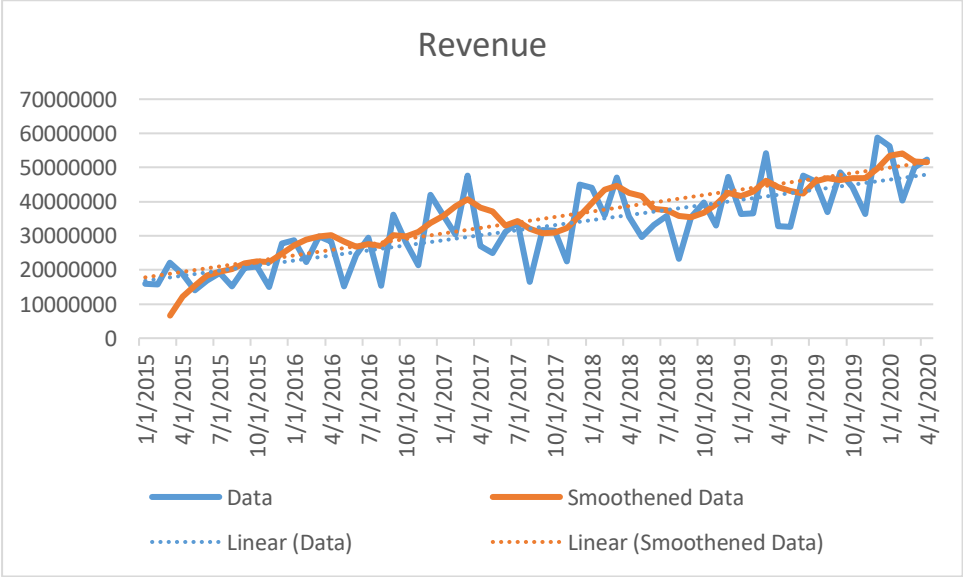
Revenue
70000000
60000000
50000000
40000000
30000000
20000000
10000000
0
1/1/2015
4/1/2015
7/1/2015
10/1/2015
1/1/2016
4/1/2016
7/1/2016
10/1/2016
1/1/2017
4/1/2017
7/1/2017
10/1/2017
1/1/2018
4/1/2018
7/1/2018
10/1/2018
1/1/2019
4/1/2019
7/1/2019
10/1/2019
1/1/2020
4/1/2020
Data
Smoothened Data
Linear (Data)
Linear (Smoothened Data)

Fig. (5)

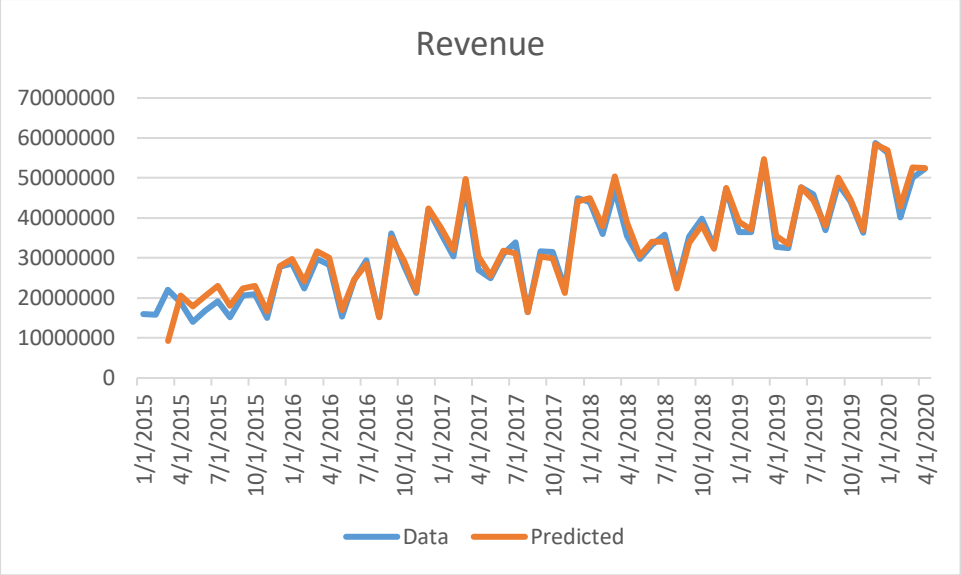
Revenue
70000000
60000000
50000000
40000000
30000000
20000000
10000000
0
1/1/2015
4/1/2015
7/1/2015
10/1/2015
1/1/2016
4/1/2016
7/1/2016
10/1/2016
1/1/2017
4/1/2017
7/1/2017
10/1/2017
1/1/2018
4/1/2018
7/1/2018
10/1/2018
1/1/2019
4/1/2019
7/1/2019
10/1/2019
1/1/2020
4/1/2020
Data
Predicted

Fig. (6)

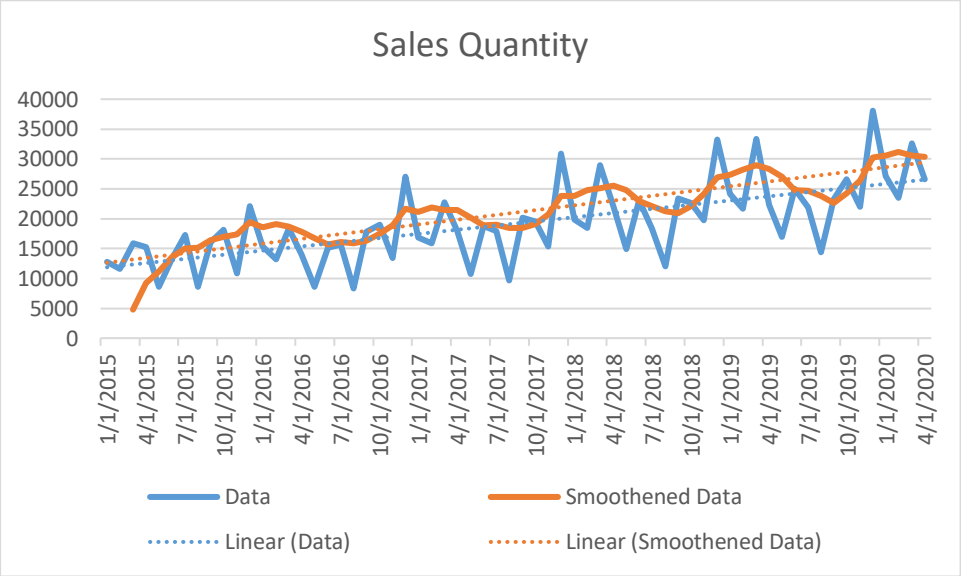
Sales Quantity
40000
35000
30000
25000
20000
15000
10000
5000
0
1/1/2015
4/1/2015
7/1/2015
10/1/2015
1/1/2016
4/1/2016
7/1/2016
10/1/2016
1/1/2017
4/1/2017
7/1/2017
10/1/2017
1/1/2018
4/1/2018
7/1/2018
10/1/2018
1/1/2019
4/1/2019
7/1/2019
10/1/2019
1/1/2020
4/1/2020
Data
Smoothened Data
Linear (Data)
Linear (Smoothened Data)

Fig. (7)

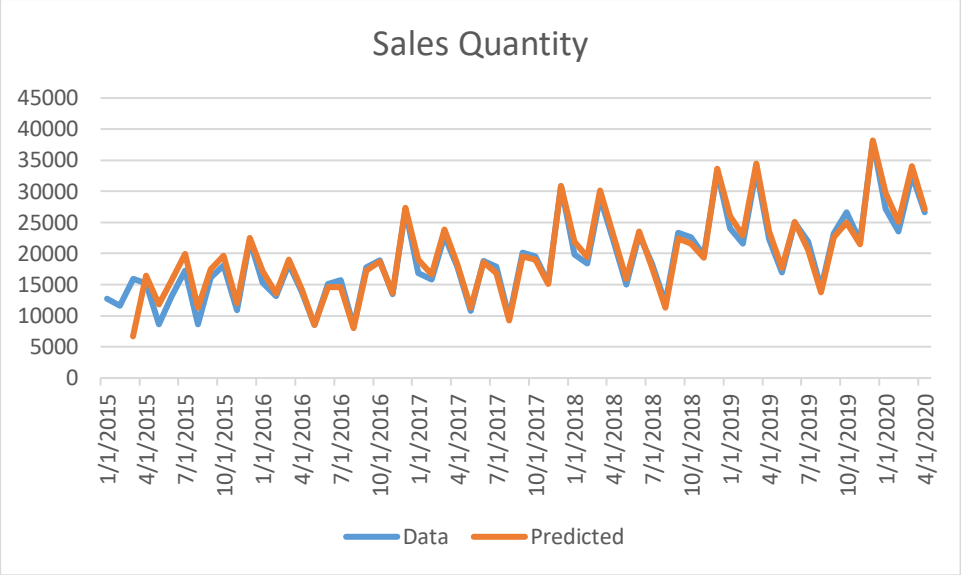

Fig. (8)

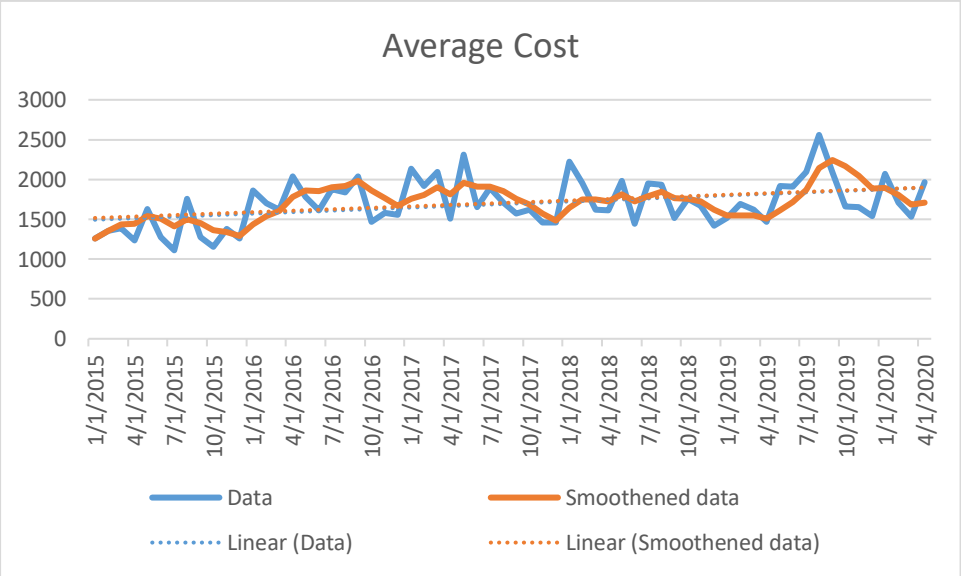

Fig. (9)

Exploratory Data Analysis (EDA):

There are no missing values present in the given dataset. Also, no need arises in data integration (since there is only one dataset and no need merge many datasets together), normalisation (since the values of variables already lie in the predetermined range) and transforming the data into graphs (since work is done individually with each variable column of the dataset). Through data mining seasonality and trend can be observed in Revenue and Sales Quantity and trend is observed in Average cost.

Interpretations:

From Fig. (1), the is.ts() command confirms that the dataset is a time series indeed. Further using the Mann-Kendall test we can interpret that trend is present in the data,

since the p-values of Revenue, Sales quantity and Average cost are less than 0.05 Also, since the tau values of Revenue, Sales quantity and Average cost are greater than 0, the trend is positive(increasing) for all the three variables.

From Fig. (2), it can be clearly understood that seasonality is present in the series (for Revenue). Also from the ACF plot it is pretty clear that a lag of 3 is present.

Similarly, from Fig. (3), it can be clearly understood that seasonality is present in the series (for Sales quantity). Also from the ACF plot it is pretty clear that a lag of 3 is present.

From Fig. (4), however seasonality is not present in the series (Average Cost) and it can be very well understood using the ACF plot since some of the plot lines are below the x-axis while some are above.

Since seasonality is present for Revenue and Sales quantity, as explained earlier Holts Winter Smoothing Technique is used for both the series taking values of α, β and γ as 0.3, 0.4 and 0.5 respectively. Fig. (5) and Fig. (7) show us the graphs of Holts Winter Smoothing Technique used for Revenue and Sales quantity respectively.

As observed from Fig. (6) and Fig. (8) which are graphs plotted for actual values for Revenue and Sales quantity against the predicted values, we can say that the predicted values are closer to the actual values which suggests that the series is actually smoothened out.

Lastly, since only trend and no seasonality is present in Average cost, Holts Smoothing Technique is used. The values of α and β are taken as 0.3, and 0.4 respectively.

Conclusions:

Concluding, it is now known that smoothing is used to reveal long term trends in time series and get rid of short term noises and fluctuations. Also, it can be shown that as value of α tend to 0, the series gets smoothened better. The appropriate smoothing techniques are Holts Winter Smoothing Technique for Revenue and Sales quantity and Holts Smoothing Technique for Average cost. Further, we can compare Revenue and Average cost by taking into account the techniques used to smoothen them which are different since seasonality is present in Revenue but absent in Average cost.