PORTFOLIO TASK 4 – TIME SERIES DATA – DECOMPOSITION TECHNIQUE

Introduction

The goal of this task is to perform analysis using Time series forecasting with decomposition technique on a dataset provided. A common starter time series data set which shows the number of airline passengers there were per month in the United States in the fifties is provided.

Forecasting allows organizations to plan effectively and make better decisions. In the time series forecasting technique historical data is used to project meaningful patterns for the future. Therefore, after analysis of dataset the next step will be to apply an appropriate model.

Analysis of given data

Before analysis, it is always a good practice to re-arrange the data based on the usage. From the excel file, it can be seen that the months column is rearranged and Passengers has a label Y(t). Y(t) indicates the number of passengers at time 't', which is the value to be forecasted or predicted.

The first two steps of analysis, prior to forecasting include finding the type of model (additive or multiplicative) and calculating the seasonal factors. In this case there are 12 seasonal components. It is vital to estimate the trend to find out the type of model, which can be done by plotting time period against the Y(t). The respective plot is shown below.

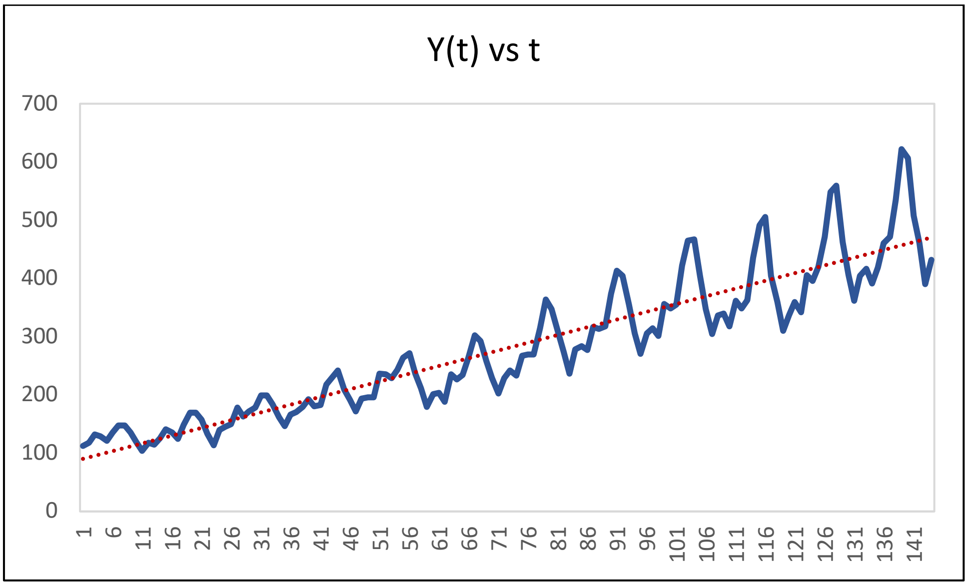

Fig: Chart showing Number of passengers over time

From the plot, it can be observed that the seasonal variation varies over time. Hence, this model is best described by 'Multiplicative Model'.

Estimating the trend

Using Moving Averages, the fluctuations caused by the seasonal component of the time series can be smoothened and the trend can be estimated as well. As there are even number of seasonal components, 'Centered moving averages' ought to be used.

Calculation of centered moving averages.

In this case, centered moving averages of size 12 are calculated. The values of these averages will be represented as 12_m(t). After this, the seasonal factors s(t) will be calculated.

Calculation of seasonal factors

In Multiplicative Model, the time series variable Y(t) is the product of trends, seasonal factors and noise. It is mathematically represented by the following equation.

$Y(t) = m(t) * s(t) * e(t)$

Where y(t) is the time series variable, s(t) is the seasonal factor and e(t) is the noise.

Thus, the seasonal factors can be calculated by using the below equation.

$s(t) = Y(t)/m(t)$

It can be observed in the excel sheet that the seasonal factors obtained for different seasons are different for every year. These values should be corrected.

Correction Table

A correction table has to be created by keeping the seasons in rows and the years in columns.

- Tabulate the moving averages according to the seasons for every year
- After tabulating the values, calculate the averages of moving averages in every season. The values obtained for all the 12 seasons are shown in the image below.
- Finding the sum of the calculated averages (value obtained is 11.974979)
- Finding the correction factor (12/ (sum of averages)) = 12/11.974979 = 1.0020894
- Multiplying the individual seasonal averages with the correction factor to get 'Typical seasonal factor'. The typical seasonal factors can be seen in the image below.

Years (Column), Season(Rows)	1	2	3	4	5	6	7	8	9	10	11	12		
1							1.16726914	1.16306483	1.06284598	0.92546986	0.80620155	0.90944123		
2	0.87619048	0.9467752	1.04508956	0.98961515	0.90964221	1.07387387	1.20638675	1.18742724	1.08435802	0.89612577	0.75226835	0.90492863		
3	0.92283214	0.94019326	1.09989701	0.99314547	1.032	1.05273534	1.1620438	1.14642343	1.04868202	0.91611687	0.82003276	0.9213691		
4	0.9337884	0.96665921	1.02093895	0.94619909	0.94532932	1.11319149	1.16137177	1.21151439	1.03358747	0.92606061	0.81742574	0.90919742		
5	0.90810811	0.89702517	1.06827612	1.05420561	1.02194124	1.08140182	1.17159763	1.20710059	1.05352843	0.93951763	0.80193057	0.89118788		
6	0.89473684	0.81576568	1.01184069	0.97043107	0.99310345	1.11041009	1.25571726	1.20102477	1.04787593	0.91508492	0.80078895	0.8906174		
7	0.92425207	0.87375	0.98478562	0.97744133	0.96947935	1.11718635	1.27384077	1.19930876	1.06393862	0.9220415	0.78737542	0.91010776		
8	0.91625218	0.88099655	0.99489996	0.97280497	0.97996918	1.14343949	1.25325578	1.22049221	1.06141772	0.90655475	0.79579102	0.88931945		
9	0.90452261	0.85269122	0.99545613	0.96298859	0.9739369	1.14934181	1.2585993	1.25805365	1.08553515	0.93175207	0.81824279	0.89929742		
10	0.90606262	0.84145535	0.95388669	0.91578947	0.95348583	1.14185716	1.28590135	1.31624674	1.04527814	0.91972673	0.78539006	0.84540608		
11	0.89431736	0.83995088	0.98573596	0.95116093	0.99881094	1.1092832							SUM	
Average	0.90810628	0.88552625	1.01608067	0.97337817	0.97776984	1.10927206	1.21959836	1.21106566	1.05870475	0.91984507	0.79854472	0.89708724	11.9749791	
													1.00208943	Correction Factor
Typical seasonal factor	0.91000371	0.8873765	1.0182037	0.97541198	0.97981283	1.11158981	1.22214663	1.2135961	1.06091684	0.92176703	0.80021323	0.89896164	12	

Fig: Image showing the correction table

- Add a column with typical seasonal values (TSF) obtained from the above step.

Obtaining De-seasonalised data

The de-seasonalised value (D) can be calculated by dividing the variable Y(t) by the corresponding seasonal factor.

D = Y(t)/(TSF)

Once the forecast for Deseasonalised data is estimated, it can be re-seasonalised and the future value of the original time series can be predicted.

Forecasting

Fitting the trend

The next step is to estimate a regression line into the deseasonalised data.

The line equation used here is Y = a + bt + e

Here a = intercept of the line and b = slope of the line.

In the excel sheet, by using formulae, the slope and intercept values are obtained.

a = 92.49410917

b = 2.553885933

Therefore, the line equation obtained will be **Y = 92.49410917 + 2.553885933t + e**

Using this equation, the values can be forecasted to obtain deseasonalized data; By substituting time period ‘t’ of the required case. To get the actual prediction, the deseasonalised data has to be reseasonalized.

Reseasonalising the data

The product of deseasonalised values (D) and the corresponding typical seasonal factor (TSF) will give the forecast values.

Forecast (F) = D * TSF

Forecasting Errors

Even a good model contains some errors. Errors help determine how good a model is. The two most important measures to check that determine the adequacy of a forecast are Mean Absolute error and Mean Squared error.

Mean Absolute error (MAE):

It is the average of absolute values of errors.

$MAE = \frac{\sum_{t=1}^{n} |e_t|}{n}$

The errors (e) can be calculated by subtracting the Forecast value (F) from Actual value (A).

Errors (e) = A - F

To calculate the mean absolute error in the excel we use the formula ABS (error).

Mean Squared error (MSE):

It is the average of the squares of the errors.

$$MSE = \frac{\sum_{t=1}^{n} e_t^2}{n}$$

Conclusion

The data had an exponential trend with 12 seasonal components. Using time series analysis, a model was built to forecast number of airline passenger using historical data followed by calculations of mean square error and mean absolute error.