

**BUSA 511: Project Assignment cover page**

Assignment for Course:      BUSA 511: Business Analytics for Managers

Submitted to:                 Dr. Syed A. Raza

Submitted by:

|  |  |
| --- | --- |
| **Student Name** | **CWID** |
| **Ashtageeta Ashok Hingmire** | **50333659** |

:

Title of Assignment: **Time Series Forecasting Project**

CERTIFICATION OF AUTHORSHIP: I certify that I am the author of this paper and that any assistance received in its preparation is fully acknowledge and disclosed in the paper. I have also cited any sources from which we used data, ideas of words, whether quoted directly or paraphrased.  I also certify that this paper was prepared by me specifically for this course

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

**Time Series Forecasting Project Instructions**

***Instructions***

This is an individual assignment and therefore must be completed by an individual student without outside assistance of any type. Follow the instructions below in order to complete the assignment:

* **Step 1**: Download and read the project statement for the *on the Forecasting problem* stated in the Microsoft Word file “**project\_description.docx**”. Read carefully this document
* **Step 2**: Download the Excel file “**project\_data.xlsx**”. Use the data in the file to perform timeseries forecasting analyses.
* **Step 3**: Based on your analytics performed and use of the models in Step 2, answer the questions and provide a short (*not exceeding 500 words*) recommendation/explanation

***Submittals***

1. Submit your answers to this assignment using this Microsoft Word document and post to the assignment drop box before the required deadline. Be sure to complete the above first page cover sheet. Enter your answers in the pages below to include all of your answers and results of your interpretations of calculations for the assignment questions. Your answers must be entered directly into this Word document below each question. Use as much space as needed.
2. Submit your Excel spreadsheet(s) with calculations/ visualizations to the assignment drop box before the required deadline. Your Excel model calculations will be used to substantiate your answers to the assignment questions herein.

***Grading***

This is an individual assignment. A total of 100 percentage points is possible for this assignment. This includes the point values which are assigned to each question - point values are noted next to each question below. Use APA 7 writing style as a guide for answers that require a written explanation. Up to 10 points will be deducted from the combination of written explanation answers that are poorly written. The percentage points earned on this assignment will be 20% weight obtain the final assignment grade.

Please answer following Question with reference to the project file: “**Project\_description.docx**”

***Question 1*** (6 points): Define a problem statement which reflects the challenge faced in this prediction problem (*Do not exceed 500 words*)

**Answer**:

For the given ABC Cooperation Dataset have 5 years of sales for every month in the year that is from January to December. The challenge that we are facing is we need to predict the sales of demand for every month of 6th year based on the values of the previous 5 years. ABC Cooperation wants to know the demands of sales of the 6th Year. The following graph shows Time Series in that, time series prediction is the use of a model to predict future events based on known past events.

A time series is a sequence of data points, measured typically at successive times, spaced at (often uniform) time intervals. Time series analysis comprises methods thatattempt to understand such time series to understand the underlying theory of the data points

Where did they come from? what generated them?,

Time series method is used to make forecasts (predictions).

Predicting the development of possible future enterprises is one of the most difficult aspects of producing any form of prediction. As a result, new sales opportunities are often disregarded when conducting top-down revenue forecasting. The demand data that is presently available is sometimes inadequate to undertake a meaningful analysis when looking at the sales history of newer items. This absence of past sales information is usually a severe disadvantage. Because of this, it's difficult to predict how much demand there will be for new items, which may lead to wasteful stock management and increased prices.

On Plotting Demand (Dt) Vs Period (t), lots of highs and lows values can be seen (Trend). Also, from the graph at specific periods that is after almost 6 periods a seasonal cyclic pattern is observed. The data values are consistently increasing or decreasing over time and changing from time period to time period. These are the challenges observed. In time-series data to predict the future data responses, which are based on past data. Generally, in a time series, some unusual effect of seasonality or trends and noise makes the prediction wrong. For better forecasting with time series, we need a stationary time series data set in which the effect of trends or seasons is less. Once the pattern is established, we can interpret

and integrate it with other data (i.e., use it in the

theory of the investigated phenomenon.

***Question 2*** (6 points): Develop a 3-period moving average forecasting model. Report the forecasts for year 6 from months January through December inclusive. Discuss briefly these forecasts (*Discussion NOT to exceed 500 words*)

**Answer**: Moving Average is defined as the average of the most recent *m* observations, where *m* is a fixed integer. The 3-period moving average is formed by averaging the 3 most recent observations.

We compute the out-of-sample forecast for Year 6 January month that is period 61 by averaging the forecast of the most recent/previous 3 consecutive forecasted demands. The lower the interval, the closer your moving average is to your original data set.

Therefore, the forecasted demand is 16666.67 for period 61. All the future forecasts are the same because the moving average cannot be updated since the previous 3 demands are not updated. Hence, the forecast for the 62nd period is also 16666.67. Therefore, the forecasted demand till period 72nd is the same 16666.67.

The lower the interval, the closer your moving average is to your original data set.

Moving Average model requires history or previous most recent values to calculate the next forecast for the next period. Since, the demands are not updated after 60th period, hence the value after 61th period remains consistent throughout.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 61 |  | January |  | 16666.667 |
| 62 |  | February |  | 16666.667 |
| 63 |  | March |  | 16666.667 |
| 64 |  | April |  | 16666.667 |
| 65 |  | May |  | 16666.667 |
| 66 |  | June |  | 16666.667 |
| 67 |  | July |  | 16666.667 |
| 68 |  | August |  | 16666.667 |
| 69 |  | September |  | 16666.667 |
| 70 |  | October |  | 16666.667 |
| 71 |  | November |  | 16666.667 |
| 72 |  | December |  | 16666.667 |

***Question 3*** (6 points): Compute for the model developed in Question 2, compute the error parameters MAD, MSE, MAPE, and TS. Explain these error computations (*Explanation NOT to exceed 500 words*)

**Answer**:

|  |
| --- |
| Mean Absolute Deviation (MAD*)* of a set of data is the average distance between each data value and the mean. |

The mean absolute deviation is the "average" of the "positive distances" of each point from the mean.

The larger the MAD, the greater variability there is in the data (the data is more spread out).

The MAD helps determine whether the set's mean is a useful indicator of the values within the set.  
The larger the MAD, the less relevant is the mean as an indicator of the values within the set.

Mean squared error (MSE) measures the amount of error in statistical models. It assesses the average squared difference between the observed and [predicted values](https://statisticsbyjim.com/glossary/fitted-values/). When a model has no error, the MSE equals zero. As model error increases, its value increases. The mean squared error is also known as the mean squared deviation (MSD).For example, in [regression](https://statisticsbyjim.com/regression/when-use-regression-analysis/), the mean squared error represents the average squared residual. As the data points fall closer to the regression line, the model has less error, decreasing the MSE. A model with less error produces more [precise predictions](https://statisticsbyjim.com/regression/prediction-precision-applied-regression/).

Mean absolute percentage error (MAPE) is a statistical measure of how accurate a forecast system is. It measures this accuracy as a percentage, and can be calculated as the average absolute percent error for each time period minus actual values divided by actual values.

A MAPE less than 5% is considered as an indication that the forecast is acceptably accurate. A MAPE greater than 10% but less than 25% indicates low, but acceptable accuracy and MAPE greater than 25% very low accuracy, so low that the forecast is not acceptable in terms of its accuracy.

**The preferred model will have the lowest values for MSE, MAD and MAPE. The original time series has a jagged appearance, suggesting the presence of an important random component of the series The moving average proved a smooth picture.**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Reported | Lowest | Highest |
| MSE | 27974658.9 | 111111.1 | 27974658.9 |
| MAD | 4380.11696 | 333.3333 | 4380.11696 |
| MAPE | 74.7721682 | 11.11111 | 78.095927 |
| TSt | -5.4032043 | -8.58509 | 0.24540902 |

***Question 4*** (6 points): Develop a simple exponential smoothing forecasting model, assume, α=0.2. Report the forecasts for year 6 from months January through December inclusive. Discuss briefly these forecasts (*Discussion NOT to exceed 500 words)*

**Answer**:

Simple Exponential Smoothing uses a weighted moving average as the forecast, with the assigned weights decreasing exponentially for periods farther in the past. The formula to calculate exponential smoothening is as follows ŷ(t+1) = α(yt) + (1 – α) ŷ t where yt = ŷt and 0 < α <1. Computed the values using the for values for periods 1 to 60 using the above formula.

To forecast for year 6 for the month of January through December we use the above formula, we have

(yt) = 8000 demand, (ŷ t) = 14040.4106 forecasted demand, α = 0.2. Therefore,

Ŷ61 = 8000\*0.2 + (1 – 0.8)\*14040.4106 = 12832.33. The forecasted demand for ŷ61 is 12832.33. As with the moving average technique, any further out-of-sample also assumes this same value, that is, forecasts for the 62nd,63rd,64th,…. 72nd periods will be 12832.33 forecasted demand as we don’t have the information of required parameter to compute in the formula.

Exponential Smoothening model requires history or previous most recent values to calculate the next forecast for the next period. Since, the demands are not updated after 60th period, hence the value after 61th period remains consistent throughout. There is no increase or decrease throughout the year. We cannot say if there is positive or negative impact on the forecast.

|  |  |  |  |
| --- | --- | --- | --- |
| 61 | January |  | 12832.3285 |
| 62 | February |  | 12832.3285 |
| 63 | March |  | 12832.3285 |
| 64 | April |  | 12832.3285 |
| 65 | May |  | 12832.3285 |
| 66 | June |  | 12832.3285 |
| 67 | July |  | 12832.3285 |
| 68 | August |  | 12832.3285 |
| 69 | September |  | 12832.3285 |
| 70 | October |  | 12832.3285 |
| 71 | November |  | 12832.3285 |
| 72 | December |  | 12832.3285 |

***Question 5*** (6 points): For the model developed in Question 4, compute the error parameters MAD, MSE, MAPE, and TS.

**Answer**:

First calculate the **level** of the series which the average value of a specific period, Growth of the series – the average increase or decrease of the value over a period of time. Here we are considering 3 period time. Average of first 3 consecutive demands will give Level (Lt) of 3rd period. Similarly we find the level upto 60th period. In column E, forecast is calculated by taking Average of Demands of first 3 periods and computing it into 4th period. . Similarly we find the Forecast (Ft) upto 60th period using same formula.

To calculate MSE, MAD, and MAPE, we first compute the forecast error, where Forecast Error = Forecasted Demand - Demand, therefore, et = yt – ŷt. For calculating Forecast Error (et) of period 4th subtract, (Forecasted Demand – Demand) of the 4th period that is, 8216.667 - 2000 = 6216.667. Similarly, we find the Forecast Error (Et) upto 60th period using same formula. To calculate MSE, formula is as follows,

**MSE = \* Σ(et)² where n is the number of observations taken and et is Forecast Error, for calculating MSE of 4th period using above formula we get, (6216.667)²/1 = 38646944.44 in excel compute =SUMSQ($E$3:E3)/A3 (in excel).** Similarly, we find the Forecast MSE upto 60th period using same formula.

**To calculate MAD, formula is as follows,**

**MAD = \* Σ|et| where n is the number of observations taken and et is the Forecast Error for calculating MAD of 4th period using above formula we get |6216.667|\*1/1 = 6216.667 in excel computed = ABS(E3) (in excel).** . Similarly, we find the MAD upto 60th period using same formula.

**MAPE = \* Σ| where n is the number of observations taken and et is the Forecast Error and yt is demand for calculating MAPE of 4th period using above formula we get |6216.667/2000|/1/1 = 310.833 in Excel computed =AVERAGE($I$3:I3) (in excel).** Similarly, we find the MAPE upto 60th period using same formula.

**TSt = \* Σ where n is the number of observations taken and et is the Forecast Error and yt is demand and ŷt is Forecasted Demand for calculating TSt of 4th period using above formula we get, (8216.667 – 2000) / |6216.667| = 1 in Excel computed =SUM($E$3:E3)/H3 (in excel).** Similarly, we find the TSt upto 60th period using same formula.

|  |  |  |  |
| --- | --- | --- | --- |
| α | 0.2 |  |  |
|  | Estimated | Lowest | Highest |
| MSE | 27525821.03 | 9688538 | 38646944.4 |
| MAD | 4364.834986 | 2575.863 | 6216.66667 |
| MAPE | 81.55976758 | 67.24021 | 310.833333 |
| TSt | -5.2873268 | -6.7149 | 5.22342932 |

***Question 6* (6 points):** For the model developed in Question 4, using Excel Solver optimize the value of α with and objective to minimize MSE. Report your results provide a discussion how optimization improved the forecasting error (*Discussion NOT to exceed 500 words*

**Answer:**

The preferred model will have the lowest values for MSE, MAD and MAPE.

On solving into the solver by setting Objective variable to α cell in excel, changing variable to MSE cell and subject to constraints as α ≤ 1, α ≥ 0 we get α = 1 and setting method to GRG nonlinear we get α = 1, MSE value is changed to 17010782.41 and the forecast Error (Et) is Lesser compared to previous model when α =0.2. Lesser the forecast error, the more accurate our model is.

Also, upon plotting Demands Vs Forecasted Demands Line Graph the the fluctuations are almost coinciding. Showing the Actual Demands and Forecasted demands are almost similar.

Hence, on optimization forecasting error improved as the fore casting Error (Et) decreased.

***Question 7*** (6 points): Develop a Holt’s model for forecasting, assume, α=0.3, and β =0.1. Report the forecasts for year 6 from months January through December inclusive. Discuss briefly these forecasts (*Discussion NOT to exceed 500 words)*

**Answer**:

The Holt exponential smoothing method also referred to as the double exponential smoothing method incorporates long-term upward or downward movements of a time series that is trends (Tt). It is appropriate when the time series does not exhibit seasonal variations or the series has been deseasonalized. The forecasts depend on the values of the smoothing parameters α and β, for level and trend, respectively.

Systematic component of demand = level + trend

Obtain initial estimate of level and trend by running a linear regression of the following form:

Dt = at + b

T0 = a = X-Variable

L0 = b = Intercept

Therefore,

To forecast for 6 year from month January through December, we first do the Regression Analysis taking 60 periods and 60 Demands using Data Analysis > Regression Analysis > Specify the data range by highlighting cells A1:A62, B1:B62. We get Intercept = 4800.565, X-Variable = 112.003.

The formula to calculate Level (Lt) and Trends (Tt) is as follows :-

and

Substitute the Intercept value in Level (Lt) for 0th period be Lt0 and X-Variable in Trend be Tt0. For Calculating 1st period in Level (Lt)

= α\*(yt) + (1-α)\*(Lt0+Tt0)

where yt is demand of period 1, Lt0 is previous period value i.e 0th period and Tt0 is the previous value of period 0. Therefore,

=0.3\*(2000) +(1-0.3)\* (4800.565+112.003) =4038.798. Similarly, we find Level (Lt) upto 60th period using same formula.

**To calculate 1st period of trend, formula is as follows,**

**=β\*(Lt1 – Lt0) + (1 – β)\*Tt0 ,**

**where Lt1 and Lt0 are the level of period 1 and 2, and Trend Tt0 of period 0.** Similarly, we find Trend (Tt) upto 60th period using same formula.

**Therefore,**

**= 0.1\*(**4038.798 - 4800.565) + (1 – 0.1)\* 112.003 = 24.6262

In period t, the forecast for future periods is expressed as follows:

Ft+1 = Lt + Tt

To calculate the forecast for period 1 we take Lt0 + Tt0 , therefore, 4038.798 + 4800.565 = 4912.568. Similarly, we find Forecast Error (Ft) upto 60th period using same formula.

Now, to calculate for period 61th we use the below formula:

**Ft+n = Lt + n\*Tt**

Where, n takes the values as 1,2,3,4…..12.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 61 | January |  |  | 15050.665 |
| 62 | February |  |  | 15531.6475 |
| 63 | March |  |  | 16012.63 |
| 64 | April |  |  | 16493.6125 |
| 65 | May |  |  | 16974.595 |
| 66 | June |  |  | 17455.5775 |
| 67 | July |  |  | 17936.56 |
| 68 | August |  |  | 18417.5426 |
| 69 | September | |  | 18898.5251 |
| 70 | October |  |  | 19379.5076 |
| 71 | November | |  | 19860.4901 |
| 72 | December | |  | 20341.4726 |

If we observe the above forecast values for the year 6, they keep increasing monthly. There is a huge difference between the month January and December. It is a positive growth all along the way. There is increase of 5,291 for 12 months. Also, the very small value of β means that the slope hardly changes over time.

***Question 8*** (6 points): For the model developed in Question 7, compute the error parameters MAD, MSE, MAPE, and TS.

**Answer**:

To forecast for 6 year from month January through December, we first do the Regression Analysis taking 60 periods and 60 Demands using Data Analysis > Regression Analysis > Specify the data range by highlighting cells A1:A62, B1:B62. We get Intercept = 4800.565, X-Variable = 112.003. Substitute the Intercept value in Level (Lt) for 0th period be Lt0 and X-Variable in Trend be Tt0. For Calculating 1st period in Level (Lt)

= α\*(yt) + (1-α)\*(Lt0+Tt0) where yt is demand of period 1, Lt0 is previous period value i.e 0th period and Tt0 is the previous value of period 0. Therefore,

=0.3\*(2000) +(1-0.3)\* (4800.565+112.003) =4038.798. Similarly, we find Level (Lt) upto 60th period using same formula.

**To calculate 1st period of trend, formula is as follows,**

**=β\*(Lt1 – Lt0) + (1 – β)\*Tt0 , where Lt1 and Lt0 are the level of period 1 and 2, and Trend Tt0 of period 0.** Similarly, we find Trend (Tt) upto 60th period using same formula.

**Therefore,**

**= 0.1\*(**4038.798 - 4800.565) + (1 – 0.1)\* 112.003 = 24.6262

To calculate the forecast for period 1 we take Lt0 + Tt0, therefore, 4038.798 + 4800.565 = 4912.568. Similarly, we find Forecast Error (Ft) upto 60th period using same formula.

To calculate Error(Et), substract Demand – Forecasted Error = 2000 - 4912.568 = 2912.568.

**MSE = \* Σ(et)² where n is the number of observations taken and et is Forecast Error, for calculating MSE of 1st period using above formula we get, (4912.5683)²/1 = 8483054.137 in excel compute =SUMSQ($F$3:F3)/A3 (in excel).** Similarly, we find the Forecast MSE upto 60th period using same formula.

**To calculate MAD, formula is as follows,**

**MAD = \* Σ|et| where n is the number of observations taken and et is the Forecast Error for calculating MAD of 4th period using above formula we get|**2912.568**|\*1/1 =** 2912.568 **in excel computed = ABS(F3) (in excel).** Similarly, we find the MAD upto 60th period using same formula.

**MAPE = \* Σ| where n is the number of observations taken and et is the Forecast Error and yt is demand for calculating MAPE of 4th period using above formula we get |**2912.568**/2000|/1/1 = 145.6284in Excel computed =AVERAGE($I$3:I3) (in excel).** Similarly, we find the MAPE upto 60th period using same formula.

**TSt = \* Σ where n is the number of observations taken and et is the Forecast Error and yt is demand and ŷt is Forecasted Demand for calculating TSt of 4th period using above formula we get, (**2912.568 **– 2000) / |**2912.568**| = 1 in Excel computed =SUM($E$3:E3)/H3 (in excel).** Similarly, we find the TSt upto 60th period using same formula.

|  |  |  |  |
| --- | --- | --- | --- |
| α | 0.3 |  |  |
| β | 0.1 |  |  |
|  |  | Lowest | Highest |
| MSE | 28910358.5 | 2177755 | 28910358.5 |
| MAD | 4531.92837 | 1180.616 | 4531.92837 |
| MAPE | 83.0688097 | 40.8882 | 145.628415 |
| TSt | -2.7139232 | -6.59217 | 4.03527752 |

***Question 9*** (6 points): For the model developed in Question 7, using Excel Solver optimize the values of α and β with an objective to minimize MAD. Report your results provide a Discussion how optimization improved the forecasting error (*Discussion NOT to exceed 500 words)*

**Answer**:

On solving into the solver by setting Objective variable - α, β cell in excel, changing variable to MAD cell and subject to constraints as α ≤ 1, α ≥ 0 and β ≤ 1, β ≥ 0 and setting method to GRG nonlinear we get α = 0.9886 and β = 0, MAD = 4135.92 value is changed 2937.88 and the forecast Error (Et) is Lesser compared to previous model when α =0.3. The very small value of β means that the slope hardly changes over time Lesser the forecast error, the more accurate our model is. Level (Lt) is almost similar to demand and Trend changes to constant value throughout 60th period ie is 112.003.

Also, upon plotting Demands Vs Forecasted Demands Line Graph the the fluctuations seemed less and almost coinciding Showing the Actual Demands and Forecasted demands are almost similar.

Hence, on optimization forecasting error improved as the forecasting Error (Et) decreased.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 61 | January |  |  | 8271.25847 |
| 62 | February |  |  | 8383.2618 |
| 63 | March |  |  | 8495.26514 |
| 64 | April |  |  | 8607.26847 |
| 65 | May |  |  | 8719.2718 |
| 66 | June |  |  | 8831.27514 |
| 67 | July |  |  | 8943.27847 |
| 68 | August |  |  | 9055.28181 |
| 69 | September |  |  | 9167.28514 |
| 70 | October |  |  | 9279.28848 |
| 71 | November |  |  | 9391.29181 |
| 72 | December |  |  | 9503.29514 |
|  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| α | 0.98869783 |  |  |
| β | 0 |  |  |
|  |  | Lowest | Highest |
| MSE | 16526363.1 | 1958391 | 16526363.1 |
| MAD | 2937.88779 | 973.9727 | 2937.88779 |
| MAPE | 51.1109558 | 30.45141 | 145.628415 |
| TSt | 1.15727244 | -5.35899 | 3.22006463 |

MAD indicates how far each data point is from the mean, “on average.” A “large” MAD indicates that the information is spread far out from the mean. A “small” MAD means that the information is more clustered and therefore more predictable.

After the minimization of the MAD, There is still a positive impact on the forecast. The values in the 6th year are seen more clustered and showed positive linear trend on ploting scatter plot. Which means MAD seems small. There is still a huge gap from January and December. There is a 1,232 increase for 12 months.

***Question 10*** (6 points): Develop a Winter’s model for forecasting, assume, α=0.2, β =0.3, and γ =0.1. Report the forecasts for year 6 from months January through December inclusive. Discuss briefly these forecasts (*Discussion NOT to exceed 500 words)*

**Answer**:

Holt Winters exponential smoothing method or the triple exponential smoothing method incorporates the long-term upward or downward movements of the time series as well as seasonality. Holt-Winters exponential smoothing method uses three smoothing parameters α, β and γ indicate the speed of decline. Alpha specifies the coefficient for the level smoothing. Beta specifies the coefficient for the trend smoothing. Gamma specifies the coefficient for the seasonal smoothing.

The Holt-Winters exponential smoothing method is further divided into additive and multiplicative structures depending on the type of seasonality exhibited by the series. The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are proportional to the level of the series.

To calculate the forecast for period 1 we take (Lt0 + Tt0)\*St , therefore, (5997.261+ 70.245) \* 0.4266 = 2588.45. Similarly, we find Forecast Error (Ft) upto 60th period using same formula.

Now, to calculate for period 61th we use the below formula:

**Ft+1 = (Lt + Tt \* n) \* Value of the Previous Years (5th year) of same month forecast**

Where n is 1,2,3,4….12. Here we used Mixed forcasting Method.

Calculate till period 72th using the above formula

|  |  |  |  |
| --- | --- | --- | --- |
| January |  |  | 4021.986 |
| February |  |  | 4837.599 |
| March |  |  | 4678.643 |
| April |  |  | 4044.109 |
| May |  |  | 6207.374 |
| June |  |  | 8737.546 |
| July |  |  | 9018.126 |
| August |  |  | 12389.93 |
| September |  |  | 18441.22 |
| October |  |  | 19289.3 |
| November |  |  | 23646.49 |
| December |  |  | 12650.38 |

The winter model where we assume α=0.2, β =0.3, and γ =0.1, there is a positive impact on the forecast for the year 6. The growth from January to December is very high. There is huge positive impact of 8,629.

Also, upon plotting Demands Vs Forecasted Demands Line Graph the fluctuations seemed less and almost coinciding Showing the Actual Demands and Forecasted demands are almost similar.

Hence, on optimization forecasting error improved as the forecasting Error (Et) decreased.

The forecasts generated by the method with the multiplicative seasonality display larger and increasing seasonal variation as the level of the forecasts increases.

***Question 11*** (6 points): For the model developed in Question 10, compute the error parameters MAD, MSE, MAPE, and TS.

**Answer**:

To calculate Error(Et), substract Demand – Forecasted Error = 2000 – 2588.45 = 588.4499

To calculate Absolute Error (At), Absolute of Error(Et) ie = |et|

**MSE = \* Σ(et)² where n is the number of observations taken and et is Forecast Error, for calculating MSE of 1st period using above formula we get, (4912.5683**

**588.4499)²/1 = 346273.3 in excel compute =SUMSQ($J$3:J3)/A3 (in excel).** Similarly, we find the Forecast MSE upto 60th period using same formula.

**To calculate MAD, formula is as follows,**

**MAD = \* Σ|et| where n is the number of observations taken and et is the Forecast Error for calculating MAD of 4th period using above formula we get|**588.4499**|\*1/1 =** 588.4499 **in excel computed = ABS(F3) (in excel).** Similarly, we find the MAD upto 60th period using same formula.

**MAPE = \* Σ| where n is the number of observations taken and et is the Forecast Error and yt is demand for calculating MAPE of 4th period using above formula we get |**588.4499**/2000|/1/1 = 29.4225 in Excel computed =AVERAGE($I$3:I3) (in excel).** Similarly, we find the MAPE upto 60th period using same formula.

**TSt = \* Σ where n is the number of observations taken and et is the Forecast Error and yt is demand and ŷt is Forecasted Demand for calculating TSt of 4th period using above formula we get, (**2588.45 **– 2000) / |**588.4499**| = 1 in Excel computed =SUM($E$3:E3)/H3 (in excel).** Similarly, we find the TSt upto 60th period using same formula.

|  |  |  |  |
| --- | --- | --- | --- |
| P | 12 |  |  |
| α | 0.2 |  |  |
| β | 0.3 |  |  |
| γ | 0.1 |  |  |
|  |  | Lowest | Highest |
| MSE | 3844548 | 161289.9 | 3844548 |
| MAD | 1423.752 | 347.6382 | 1423.752 |
| MAPE | 24.72796 | 12.51875 | 29.4225 |
| TSt | -3.77003 | -6.04958 | 9.71978 |
| r | 5 |  |  |

***Question 12*** (6 points): For the model developed in Question 10, using Excel Solver optimize the values of α, β, and γ with an objective to minimize MAPE. Report your results provide a Discussion how optimization improved the forecasting error (*Discussion NOT to exceed 500 words)*

**Answer**:

Alpha specifies the coefficient for the level smoothing. Beta specifies the coefficient for the trend smoothing. Gamma specifies the coefficient for the seasonal smoothing.

On solving into the solver by setting Objective variable - α, β, γ cell in excel, changing variable to MAPE cell and subject to constraints as

α ≤ 1, α ≥ 0,

β ≤ 1, β ≥ 0 and

γ≤1, γ≥0 and setting method to GRG nonlinear we get

α = 0, β = 0.021924, γ = 0. MAPE = 19.87971 that is 20% and lowest value for MAPE = 10.66316 i.e 10.66% value is changed and the forecast Error (Et) is Lesser compared to previous model when α =0.2, β=0.3 and γ = 0.1.

Optimization improved the forecasting error as forecast error after optimization analysis became less. Lesser the forecast error, the more accurate our model is. The very small value of β = 0.021924 means that the slope hardly changes over time. The γ parameter must be in the interval of 0-1. A small value means that older values in X direction are weighted more heavily. Values near 1.0 mean that the latest value has more weight. Level (Lt) is almost similar to Demand(Dt) and Trend(T) changes to constant value throughout 60th period i.e is 70.24578. Which means the same series of events are not happening happening over and over indicating a good model.

Also, upon plotting Demands Vs Forecasted Demands Line Graph values of Demands Vs Forecasted Demands are almost coinciding Showing the Actual Demands and Forecasted demands are almost similar.

Hence, on optimization forecasting error improved as the forecasting Error (Et) decreased.

|  |  |  |  |
| --- | --- | --- | --- |
| P | 12 |  |  |
| α | 0 |  |  |
| β | 0.021924 |  |  |
| γ | 0 |  |  |
|  |  | Lowest | Highest |
| MSE | 1850424 | 123432.6 | 1850424 |
| MAD | 1032.765 | 267.9474 | 1032.765 |
| MAPE | 19.87971 | 10.66316 | 29.4225 |
| TSt | 0.356109 | -7.35752 | 2.646792 |
| r | 5 |  |  |

A MAPE less than 5% is considered as an indication that the forecast is acceptably accurate. A MAPE greater than 10% but less than 25% indicates low, but acceptable accuracy and MAPE greater than 25% very low accuracy, so low that the forecast is not acceptable in terms of its accuracy.

Optimization improved

In this scenario, MAPE can be used as a good accuracy measure since it is scale-independent and can be used to compare different series or forecast scenarios.

The Scatter plot graph Demand Vs Forecasted Demand shows a similar sales demand as previous years, Although, there is slight increase in the forecasted demand in the 6 th year.

***Question 13*** (6 points): Discuss the model developed in Question 10 is how different from model developed in Question 7, (*Discussion NOT to exceed 500 words)*

**Answer**:

**Holt's Smoothing method:** Holt's smoothing technique, also known as linear exponential smoothing, is a widely known smoothing model for forecasting data that has a trend.

the Holt’ model forecasting method applies a double exponential smoothing for level and trend. It is appropriate when the time series does not exhibit seasonal variations or the series has been deseasonalized. The forecasts depend on the values of the smoothing parameters α and β, for level and trend, respectively.

Systematic component of demand = level + trend

The formula to calculate Level (Lt) and Trends (Tt) is as follows :-

and

In period t, the forecast for future periods is expressed as follows:

Ft+1 = Lt + Tt

Ft+n = Lt + nTt

Holts Smoothening calculations and methods of solving is simple as comapred to Holt-Winter Forecasting Method as it is shown below

While, the **Holt-Winters forecasting** method applies a triple exponential smoothing for level, trend and seasonal components. Alpha(α) specifies the coefficient for the level smoothing. Beta(β) specifies the coefficient for the trend smoothing. Gamma(γ) specifies the coefficient for the seasonal smoothing. There is also a parameter for the type of seasonality: Additive seasonality, where each season changes by a constant number. Multiplicative seasonality, where each season changes by a factor.

Additive (default) indicates that X is modeled as level + trend + seasonal.

Multiplicative indicates the model is (level + trend) \* seasonal.

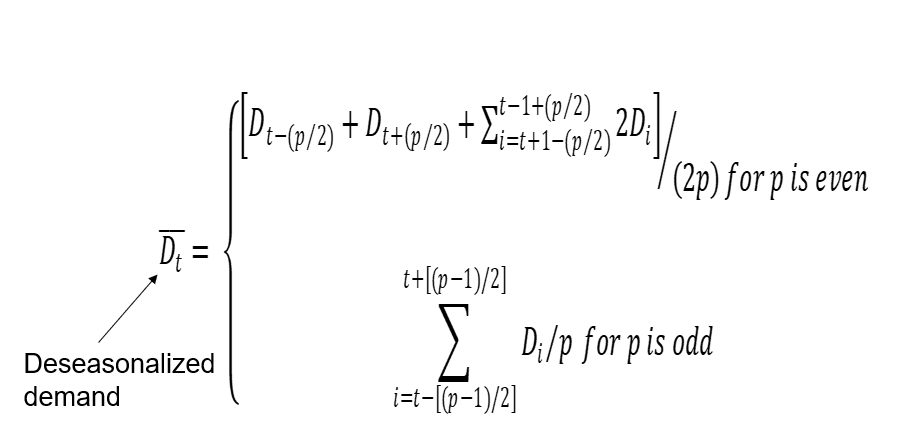
Systematic component = (level+trend)(seasonal factor).

In period t, the forecast for future periods is given by:

Ft+1 = (Lt+Tt)(St+1) and

Ft+n = (Lt + nTt)St+n

Before estimating level and trend, demand data must be deseasonalized. By deseasonalizing data, we're removing seasonal fluctuations, or patterns in the data, to predict or approximate future data values. Deseasonalized demand = demand that would have been observed in the absence of seasonal fluctuations (each season is given equal weight). Formula for calculating Deseasonalizating Demand is as follows :



Where i = t+1-(p/2) to t-1+(p/2))

Where ,

P = the number of periods after which the seasonal cycle repeats

T = period

Dt = actual demand in period t

After removal of seasonality from time series, we can consider it as a seasonal stationary time series.

We calculate Deseasonlizing Demand from period 7 because Seasonal fluctuations are seen in the interval of 6 months, Hence calculating from 7th period.

t - (p/2) = 7 – (12/2) = 1

t + (p/2) = 7 +(12/2) = 13

i = t+1-(p/2) = 2

i = t-1+(p/2) = 14

Seasonal factor (Initialisation) is calculated as,

St = Dt / Dt dash = seasonal factor for period t = Demand of period 1/ Deseasonalised Demand after 1st period.

And the Level and Trend are calculated

Forecast of following year:

Ft+1 = (Lt + Tt \* n) \* Value of the Previous Years (5th year) of same month forecast

**Holts Smoothening Model :-**

If we observe the above forecast values for the year 6, they keep increasing monthly. There is a huge difference between the month January and December. It is a positive growth all along the way. There is increase of 5,291 for 12 months. Also, the very small value of β means that the slope hardly changes over time.

**Winter’ s Holts Model :-**

The Scatter plot graph Demand Vs Forecasted Demand shows a similar sales demand as previous years, Although, there is slight increase in the forecasted demand in the 6 th year.

***Question 14*** (22 points): Which model you will select and Why and how does the optimization improve the forecasting performance of the methods? (*1 to 2-page discussion recommend*)

**Answer :**

We will prefer Holts-Winter Model as it is considering the 3 variables. It takes the seasonality into consideration which is present in the time series data given to us. So below are the forecasting values when we optimize the values of α, β, and γ with an objective to minimize MAPE, MAD, MSE values.

Triple exponential smoothing or Holts-Winter Model applies exponential smoothing three times, which is commonly used when there are three high frequency signals to be removed from a time series under study. There are different types of seasonality: 'multiplicative' and 'additive' in nature, much like addition and multiplication are basic operations.

The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series

Performing Optimization to improve our forecasting we get :

Set Objective as : **MSE**

|  |  |  |  |
| --- | --- | --- | --- |
| P | 12 |  |  |
| α | 0 |  |  |
| β | 0.021924 |  |  |
| γ | 0 |  |  |
|  |  | Lowest | Highest |
| MSE | 1850424 | 123432.6 | 1850424 |
| MAD | 1032.765 | 267.9474 | 1032.765 |
| MAPE | 19.87971 | 10.66316 | 29.4225 |
| TSt | 0.356109 | -7.35752 | 2.646792 |
| r | 5 |  |  |

Set Objective as : **MAD**

|  |  |  |  |
| --- | --- | --- | --- |
| P | 12 |  |  |
| α | 0.002084 |  |  |
| β | 0.021926 |  |  |
| γ | 0 |  |  |
|  |  | Lowest | Highest |
| MSE | 1852149 | 123617.7 | 1852149 |
| MAD | 1031.817 | 268.8132 | 1031.817 |
| MAPE | 19.93767 | 10.67776 | 29.4225 |
| TSt | 1.157811 | -6.55927 | 3.334762 |
| r | 5 |  |  |

Set Objective as : **MAPE**

|  |  |  |  |
| --- | --- | --- | --- |
| P | 12 |  |  |
| α | 0 |  |  |
| β | 0.015953 |  |  |
| γ | 0 |  |  |
|  |  | Lowest | Highest |
| MSE | 1850424 | 123432.6 | 1850424 |
| MAD | 1032.765 | 267.9474 | 1032.765 |
| MAPE | 19.87971 | 10.66316 | 29.4225 |
| TSt | 0.356109 | -7.35752 | 2.646792 |
| r | 5 |  |  |

As seen from above tables the optimization analysis is similar for all. Hence there forecast error must also be similar.

Optimization improved the forecasting error as forecast error after optimization analysis became less. Lesser the forecast error, the more accurate our model is. The very small value of β means that the slope hardly changes over time. The γ parameter must be in the interval of 0-1. A small value means that older values in X direction are weighted more heavily. Values near 1.0 mean that the latest value has more weight. Level (Lt) is almost similar to Demand(Dt) and Trend(T) changes to constant value throughout 60th period i.e is 70.24578. Which means the same series of events are not happening over and over indicating a good model.

Also, upon plotting Demands Vs Forecasted Demands Line Graph values of Demands Vs Forecasted Demands are almost coinciding Showing the Actual Demands and Forecasted demands are almost similar.

Hence, on optimization forecasting error improved as the forecasting Error (Et) decreased.

The forecasts generated by the method with the multiplicative seasonality display larger and increasing seasonal variation as the level of the forecasts increases.

A MAPE less than 5% is considered as an indication that the forecast is acceptably accurate. A MAPE greater than 10% but less than 25% indicates low, but acceptable accuracy and MAPE greater than 25% very low accuracy, so low that the forecast is not acceptable in terms of its accuracy.

Optimization improved

In this scenario, MAPE can be used as a good accuracy measure since it is scale-independent and can be used to compare different series or forecast scenarios.

Moving average, Time Series Dataset which has seasonal cycles.

Hence in comparison to other forecasting model Winter Holts is preferable.