Time series analysis

Name

Institution of affiliation

Date

**Exercise 3.1**

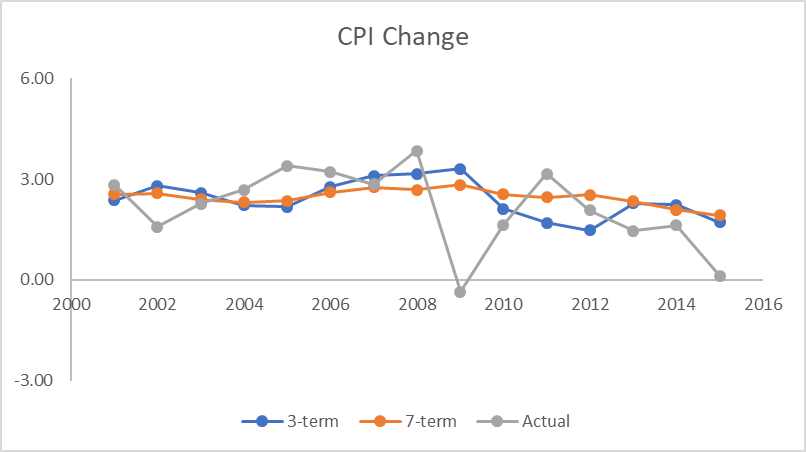
This part required to use the GDP data set to perform the following.

1. To perform three-term and seven-term moving average for the one-step forecast from 2001 to the end of the series. It also requires to visualize the forecast and give a comment on their difference. From the graphs visualized, the seven-term moving average shows a proper fitting than the three-term moving average due to high stationarity.
2. The RMSE of three-term and seven-term were found to be 2.54 and 2.28, respectively, while their RMA was found to be 2.03 and 1.55, respectively. From these residual statistics, the seven-term moving average model has proved to be a better model that the three-term in forecasting. Here MAPE is inappropriate since it involves division technique, and, in the event where an actual value is zero, it produces an undefined result that is difficult for interpretation.

**Exercise 3.2**

This part required to use the CPI time series data set to perform the following tasks.

1. The time-series graph on three and seven moving average for one-step forecast from 2001 to 2015 was drawn to visualize the differences. We can see from the chart below that the seven-term moving average is more stationary and fits the model well compared to the tree-term.

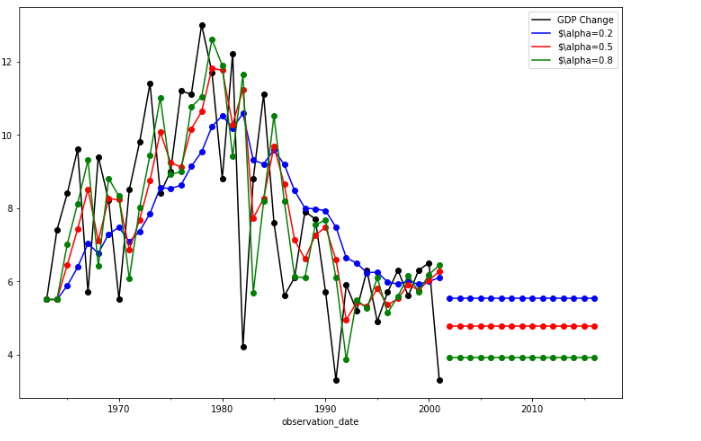


1. The performance metrics of the three and seven-term for one-step forecast and the results were found to be; the RMSE of the three and seven-term moving average were 1.26 and 1.15, respectively. Their MAE were 0.95 and 0.87. According to the residual statistics, the seven-term moving average is a better forecasting model than the three-term. Here MAPE is inappropriate since it involves division technique, and, in the event where an actual value is zero, it produces undefined results difficult for interpretation.

**Exercise 3.3**

The assignment required to perform simple exponential smoothing for the GDP time series data set.

1. The model was built under 0.2, 0.5, 0.8 values of alpha. The trends and forecasts are shown in the graph below.



1. Comparison of the values of alpha using RMSE and MAE

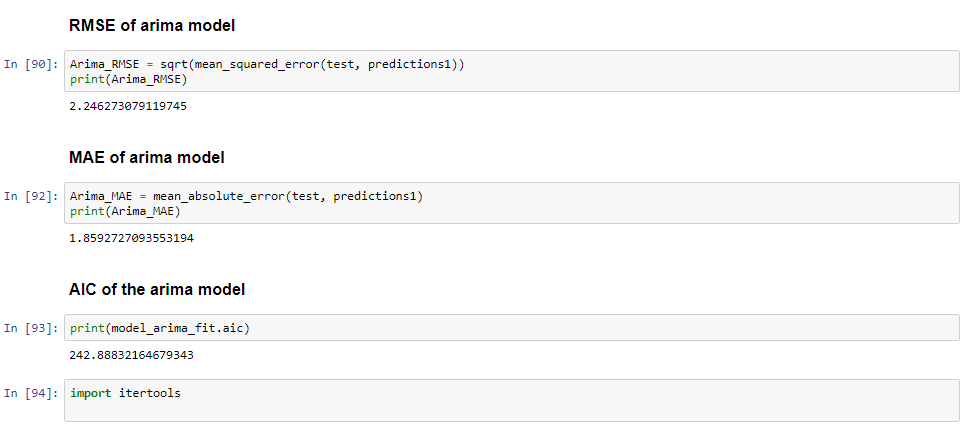
The alpha value of 0.2 produced the least value of 2.894 and 2.210 for RMSE and MAE, respectively. An alpha value of 0.5 provides the value of 2.917 and 2.230 for RMSE and MAE respectively, while the alpha value of 0.8 produced the highest value of 3.025 and 2.323 for RMSE respectively. The screenshot below shows.



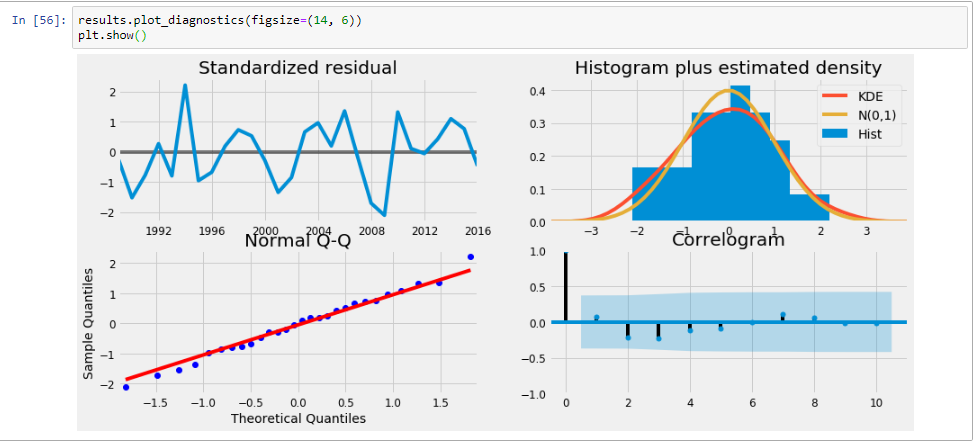
1. The simple exponential smoothing is better than the three-term and seven-term moving average used in part 3.1 above since it performs well where a time series data don’t have clear trend such as the GDP time series dataset.

**Exercise 6.1.**

This part required the development of Arima model to forecast the GDP time series data set. ARIMA is the acronym for autoregressive integrated moving average. The assignment required to compare this model with the simple exponential smoothing used in part 3.3 above. It was found that the Arima model performed better forecasting with MAE value of 1.859, which is lower than the one produced the given alpha values, which show the predicted values were not far away from the mean. The RMSE value was 2.246 which is also low than RMSE values of all the alpha values of the smoothing model and the Akaike information criteria value of 242.888. The diagnostic plots below further show that the results of the arima model are relatively reasonable and so the model is not bad. The screenshot below shows.



**Residual plots of Arima (1, 1, 0)**

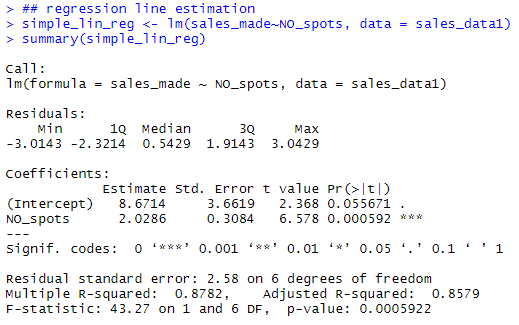


**Exercise 7.1**

1. Regression line

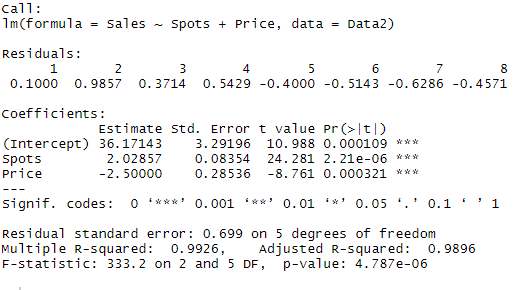
From the simple regression model performed using RStudio the regression was found to be where Y is sales and is the number of spots

1. The slop was found to be significant at the alpha value of 0.1 with t statistic value of 2.368. The screenshot of the RStudio output below shows.
2. The R-squared value for the model was found to be 0. 8782 and the S, standard error of 2.58. This means the model performs 87.82% accurate estimation for the data, and the standard error shows the data is linear with a minimal number of outliers.
3. The value of sales from 20 spots in week nine can be estimated easily from the regression line as as the value of is 20 spots. This gives this is rounded off to 41 sales in week 9.
4. The accuracy of the model is just the level of the R-squared of the model, which 87.82%; hence, the model is adequate.



**Exercise 8.1**

1. Multiple regression performed using RStudio. Both coefficients and the model were found to be significant. Both Spots and Price were found to be significant at alpha 0.0 alpha level with t statistic value of 24.281 and -8.761, respectively. The models F statistic was 333.2 for the p-value of 4.787e-06 hence its significance.
2. The multiple regression model has an R-squared value of 0.9926, which translates to 99.26% accuracy and the standard residual error of 0.699, which means the model has less variability compared to simple linear model. Due to this finding, multiple regression model is better. The screenshot from RStudio output shows.



**Exercise 8.3**

On training both Personal Disposable Income (PDI) and Real Retail Sales (RRS), RRS seemed very nonstationary, and it can be shown by the RMSE results in the screenshot below. Therefore I would prefer unleaded price model for 2008.

