Financial Time Series 26:960:576 - Spring 2019

HOMEWORK 1

Due to February 8, 2019

Question 1.(40 points)
We will perform multiple regression on the Boston housing data. The data contains 506 records with 14 variables. The variable *medv* is the response variable. To assess the data use : library(MASS) and then data(Boston).
(a) First perform a multiple regression with all the variables, what can you say about the significance of the variables based on only the p-values. Next use the "step" function to perform backward selection using the AIC criteria and the BIC criteria then compare the results. (By definition the step function in R performs variable selection based on AIC criteria. You should change the parameters in step function in order to do the selection using BIC criteria.)
(b) Now make a histogram of the response variable (use hist()) to see if it is skewed. Using $log(medv)$ as the response variable, perform the stepwise selection as previously using both AIC and BIC criteria. Compare with the previous results in terms of selected variables and adjusted R^2.

Question 2.(30 points)
The data set fancy (you need to library the fpp package to get the dataset) concerns the monthly sales figures of a shop which opened in January 1987 and sells gifts, souvenirs, and novelties. The sales volume varies with the seasonal population of tourists.
(a) Produce a time plot of the data and describe the patterns in the graph. Identify any unusual or unexpected fluctuations in the time series.
(b) Use R function "tslm" to fit a regression model to the logarithms of these sales data with a linear trend and seasonal component.
(c) Use multiple regression with trend variable and seasonal dummy variables to redo the regression as shown in the lecture example. Check to see that you obtain the same results as tslm.

Question 3.(30 points)
The data set plastics (you need to library the fpp package to get the dataset) represents the monthly sales (in thousands) of product A for a plastics manufacturer for years 1 through 5 (data set plastics).
(a) Plot the time series of sales of product A. Can you identify seasonal fluctuations or a trend?
(b) Perform a classical additive decomposition using "stl" function. Plot out the decomposition for s.window="periodic", t.window=50.
(c) Compute and plot the seasonally adjusted data for s.window="periodic", t.window=50.
(d) Change one observation to be an outlier (pick one data point and add 500 to its value. For instance, if you picked July of the third year, the current value is 1303, then the modified value will be 1803) and recompute the seasonally adjusted data. What is the effect of the outlier. Again, you need to do this for s.window="periodic", t.window=50. Does it make any difference if the outlier is near the end rather than in the middle of the time series? Try it out.

Please type your answers or scan your handwritten answers (photos are not accepted) and upload the files by February 8. Please include the R scripts, abbreviated outputs and answers/explanations to the questions. Name your files as your last name.