Lab 5

Time Series Analysis

Muhammad Ahmad Hesham Sec: 2 BN: 16

Muhammad Alaa Abdel-Khaleq Sec: 2 BN: 22

## Q1) After inspecting the time series

1. What is the time range (start and end) of this time series?

Starts with January 2003 and ends with December 2014.

1. How many values are there in this time series?

There are 144 values.

1. What is the time interval separating between each two consecutive values in the time serious (monthly, weekly, daily, yearly, etc)

Monthly

## Q2) What does the parameter *frequency* mean? Why did we set it to 12?

It is the number of observations per unit time, it is 12 because we have 12 months per year, so we need to take 12 observations per year (which is our unit time).

Q3) After visualizing the time series:

1. Add a neat plot of the generated time series.

Chart, line chart

Description automatically generated

1. Do you think there is a trend in the time series? If yes, then what is the degree of the trend (i.e. is it linear, quadratic, etc.)?

Yes, it seems to be linear from the graph

1. Do you think there is a seasonality in the time series?

Yes, because from the graph also it seems that there is some pattern between years.

Q4) What does the function **stl** do? Add a neat plot of the plot generated.

It decomposes the time serious to its 3 permeative components (trend, seasonal and random)

A picture containing chart

Description automatically generated

Q5) **Back to the original time series**. The time series expects the data to be **stationary** which means the following:

1. Is this time series stationary?

No, it’s not stationary

1. Mention the reasons behind your answer.

We need to do de-trend and seasonally adjust to make its mean, variance and auto correlated constants.

Q6) After **differencing** the time series:

1. Add a neat plot of the time series after differencing.

Chart

Description automatically generated

1. Does the time series become stationary? (Yes/No)

No

1. If no, which of the three conditions mentioned above are still not satisfied?

Variance is not constant and Auto-correlation structure is not constant because the cure is exponential, and the differentiation won’t make the mean constant**.**

1. How does differencing help to make a time series stationary?

It just does a de-trending so it tries to makes the mean constant.

Q7) After applying **logarithm** to the time series, compare visually (5) and (7) and you will know the answer to the following questions:

1. Add a neat plot of the time series after applying logarithm.

Chart, line chart

Description automatically generated

1. Does the time series become stationary? (Yes/No).

No.

1. If no, which of the three conditions mentioned above are still not satisfied?

Mean is not constant Variance is not constant and Auto-correlation structure is not constant because the curve has linear trend.

1. How does applying logarithm help to make a time series stationary? [This is a new piece of information never told in the lecture or the tutorial].

It helps to linearize the trend so we can apply differentiation and get the mean as a constant term

Q8) After applying both **differencing** and **logarithm** to the time series:

1. Add a neat plot of the final time series.

A picture containing line chart

Description automatically generated

1. Does the time series become stationary? (Yes/No).

No

1. If no, which of the three conditions mentioned above are still not satisfied?

Variance is not constant and Auto-correlation structure is not constant.

Q9) **After fitting an ARIMA Model with the logarithm of the time series:**

1. Does the time series passed to the ARIMA model successfully satsify the requirement of a stationary time series? Why?

No, because we passed a non-constant mean, non-constant variance and non-constant auto-correlation structure function.

1. Inspect the summary of the model. The output of the ARIMA model is: (p,d,q) (P,D,Q)[S] What does the (p, d, q) mean?  
   What do you think the (P, D, Q)[S] relate to? (no details are needed).

p: is the number of autoregressive terms.

d: is the number of differences.

q: is the number of moving average terms.

(P, D, Q)[S] is relate to seasonality.

1. Do you think that ARIMA model achieved the requirements of (9-a) internally (i.e. it operated on the stationary version of the time series)? If yes, how did it happen briefly? You don’t need to give any mathematical proofs or so. You just need to observe the ARIMA model output (9-b) and you will get it.

Yes, the output model is ARIMA(0,1,1)(0,1,1)[12] which means it apply a difference 1 and also apply a moving average 1.

1. What do you think will be more suitable for the case of forecasting the tractor sales, an autoregressive (AR) model or a moving average (MA) model? Why?

MA term because it use a short term dependencies.

Q10) After changing **trace = True**:

1. How is the best model selected? What is the information criterion used in selecting the best model? [Mention only the name] [Check the help]

AICc

1. What other information criteria are there that can be used as well? [Mention only the names][Check the help]

AIC and BIC

1. Do we seek to get the minimum value or the maximum value of this criterion?

Minimum value

Q11) What is the meaning of n.head = 36?

The number of steps ahead for which prediction is required.

Q12) After forecasting and plotting the future values:

1. According to your observation, does this forecast work well?

Yes

1. Add a neat plot of the generated time series.

Chart, line chart

Description automatically generated

Q13) After using **TSPred library**:

1. Does this library generate a similar plot to (12)? Why?

No, This function plots ARIMA predictions against its actual values with prediction intervals.

1. Add a neat plot of the generated time series.

Chart

Description automatically generated

Q14) What happened when we tried to forecast the tractor sales for an extended or longer time range? What do you notice?

It starts to have high probability of wrong predictions and the errors start to grow.

Chart

Description automatically generated with low confidence