F70TS 2016-17: Assessed Project

Feedback

Feedback on individual projects is available through Turnitin and from the course lecturer. Some general feedback on the project is given below.

Question 1

In Question 1, a good approach is to investigate the distribution of your samples of the number of turning points using (for example) QQ plots. Running the turning points test on data you already know to be white noise is not particularly informative. If you did want to use an approach based on constructing confidence intervals using the normal approximation, it would give you much more information if you investigated a variety of confidence levels (50%, 80%, 90%, ...) and looked at the proportion of your sample in each of these confidence intervals. The approach using QQ plots, however, is still more straightforward and easier to interpret.

The 'staircase' shape you see on QQ plots constructed in this question (particularly obvious for small values of n) is because the data are discrete.

Question 2

Linearity of expectation (i.e., the fact that $E[\sum_t X_t] = \sum_t E[X_t]$) does not need independence of the random variables X_i, though the corresponding formula for variance does. In fact, the X_t are not independent here (otherwise all the covariances in the formula given would be zero).

In the question, you are asked to give a proof of the formulas for expectation and variance for a general k. It is not enough to show that when $k = 2$ we get the required expressions. You need to consider the general case, and show how the mean and variance can be calculated here.

Question 3

In a setting with monthly data (as we have here), significant autocorrelation at lag 12 may be related to seasonality in the data - i.e., the fact that sales this March could be related to sales last March through some seasonal effect. The model we have used for the sales data does not include any seasonal effect, and the significant autocorrelation at lag 12 may be some evidence that seasonality should be accounted for in the model.

It is not necessary for data to be normally distributed to be white noise.