DS 6373 Midterm Fall 2019 Take Home Portion Ver A

For both questions: Consider the half-hourly electricity demand in England and Wales from Monday, 5 June 2000 to Sunday, 16 July 2000.) Data: electricity.csv on GitHub in Unit 8 folder.

Question 1 (10 points):

Below is the realization, acf, periodogram and spectral density of the data.



Is this data stationary or non-stationarity. Discuss and defend your argument addressing all 3 conditions of stationarity. Provide visual evidence when you address condition 3.

Condition 1: The mean appears as if it will vary depending on when the signal is sampled. The time series realization shows a cyclical pattern of increasing and decreasing means across at least two separate but distinct periods. The ACF also indicates a pronounced shift from positive to negative correlation as the lag increases. Thus, there is sufficient evidence to suggest that this signal violates the condition that the mean does not depend on time.

Condition 2: The variance of the signal appears consistent across the entire realization.

Condition 3: The acf of the first half of the time series (1:250) and the second half of the time series (251:500) seems to shift slightly indicating that the correlation between data points depends on how far apart they are and not where they are in the time series.

par(mfrow=c(1, 2))

acf(electricity$x[1:250])

acf(electricity$x[251: 500])

Considering this time series violates the condition 1 that the mean is not dependent on time, and condition 3, stationary covariance, this time series is NOT stationary.

Question 2. (20 points)

Which model do you think is most appropriate/useful for forecasting electricity usage (Model 1, Model 2 or Model 3)? Compare the models based on statistics, properties and characteristics we have learned in class. Answers are graded on completeness, correctness, convincingness and thoroughness.

Do this for two scenarios:

1. Forecasts for the next day of electricity usage.
2. Forecasts for the next week of electricity usage.

Model 1

fore.aruma.wge(electricity$x, phi = c(2.0401,-1.2159, .1751), theta =.9551, s = 336, n.ahead = 4, lastn = T, limits = T, plot = T)

fore.aruma.wge(electricity$x, phi = c(2.0401,-1.2159, .1751), theta =.9551, s = 336, n.ahead = 10, lastn = T, limits = T, plot = T)

ase(m17day$f,electricity$x)

708143

Model 2

m21day = fore.aruma.wge(electricity$x, phi = c(.0866, -.8452, -.1804), theta =c(-.00026,.926), s = 336, d=1, n.ahead = 4, lastn = T, limits = T, plot = T)

m27day = fore.aruma.wge(electricity$x, phi = c(.0866, -.8452, -.1804), theta =c(-.00026,.926), s = 336, d=1, n.ahead = 10, lastn = T, limits = T, plot = T)

ase(m27day$f,electricity$x)

[1] 1083995

Model 3: Custom Model

Use fore.signalplus.wge() to fit a cos model. A good start is to type “?fore.signalplusnoise.wge” in the console to read the R information on the function. You will need to select the most useful frequency and the function will estimate the amplitude, phase shift and mean level (intercept). In addition, pages 260/261 and 266/267 will be of good help as well. As part of your answer, fully specify the model as in equation (6.61) on page 260. Hint: “(Intercept)” = A0 “xmtx1” = A1 “xmtx2” = A2 using attribute *$b* from the fore.signalplusnoise.wge() function call.



The frequency was selected by selecting the prominent frequency in the parzen window.

plotts.sample.wge(electricity$x)

m3=fore.sigplusnoise.wge(electricity$x,linear=T,freq = c(.1), max.p = 5,n.ahead=100)

ase(m3$f,electricity$x)

29490190

The model with the lowest ASE was model 1. Furthermore, model 1 seemed to track the last few known time series measurements. Thus model 1 seems the most appropriate.