ASSIGNMENT 7

STAT 485/685 E100/G100: Applied Time Series Analysis

Fall 2020

Simon Fraser University

This week's assignment will be based on some further topics in model specification. The topics here have been covered in the Week 10 video, as well as § 6.4 - 6.5 of the textbook *Time Series Analysis with Applications in R (2nd ed.)* by Cryer & Chan.[1]

Due date: **Sunday, Nov. 22nd at 11:59 pm (end of day)** (Pacific Time).

Marks: 10.

Please include your R code for each of questions that requires it below. Some ideas for how you can most easily do this:

- Copy-paste your code/plots into a Word document along with your responses to the questions, and save as a PDF.
- Take images of your code/plots and upload to Crowdmark, along with your responses to the questions.
- Save your code and responses together in an RMarkdown document and save as PDF (if you've worked with RMarkdown before).

Other important policies on assignment submissions:

- Please write each question on a **separate page**!
- Please **show all your code and work**, in order to get full marks.
- Upload your complete answers as PDF files or high-resolution images.
- If you're hand-writing answers, please make sure they are **neat and clearly readable**, and that the photo is high resolution.
- Please **clearly label the question numbers**.

[1] Cryer, J. D., & Chan, K. S. (2008). *Time series analysis: with applications in R.* Springer Science and Business Media.

For this assignment, we will be using some datasets and functions in the `TSA` package in R. For instructions on how to install and load the package, please see the Week 1 module on Canvas.

1. *(9 marks)* The `airmiles` dataset in the `TSA` package gives the monthly U.S. airline passenger-miles, from 1996 to 2005.

   (a) Load in and plot the time series dataset. Does this data appear to come from a stationary process? Why or why not?

   (b) Plot the first difference of the time series. What improvements do you see here? Is there still something in this data that needs to be accounted for?

   (c) Create the sample ACF plot, sample PACF plot and sample EACF table for the *first difference* of the time series. Explain any conclusions you can make from each of these visualizations. If you find that it is difficult to make a single overall conclusion about the underlying model, explain why it is difficult.

   (d) Sometimes this type of data is seasonal. So, perhaps the difficulty in choosing an appropriate ARMA model is due to the fact that this is a subset-ARMA model, i.e. some of the coefficients are zero. Create the "best subset-ARMA selection" plot.
   (Hint: An example of how to create this plot is shown in Video 28. You can ignore the Warning message that R gives you.)

   (e) Interpret the best subset-ARMA selection plot you created in part (d). What does it generally tell you about the "most important" terms in the model?

   (f) Write out the top choice of model based on the best subset-ARMA selection plot in part (d).
   (Hint: Remember that this ARMA model is for the first difference of the series, $\{W_t\}$, not for the original series $\{Y_t\}$.)

   (g) Does this choice of model make sense to you? Why do you think these lags were chosen to be the most important?

   (h) Since this ARMA model was constructed for the *first difference* of the time series, write out the equation for the original series, $\{Y_t\}$.

2. *(1 mark)* Describe the importance of the "$2k$" term in Akaike's Information Criterion (AIC). What does it achieve? What would happen if we were to not include this term in the criterion?
   (This answer only needs to be about 2 sentences long.)