STAT 1321/2320: Final Programming

Instructions:

- You should work on this exam on your own. You may ask the instructor any clarifying instructions.
- You may use the R files posted on Canvas for help.
- You should submit one document with all your R code, output, and answers to the problems. This may be an Rmarkdown document or a Word/PDF file with the relevant information.
- You should also submit your R file separately as a backup.
- A corrupt or unreadable file in not a valid reason for re-submission.
- Be sure to include **warning=F, message = F** when you define your R chunks so the output knitted file is concise. Here is an example:

$$\{\text{r, Problem 1, warning = F, message = F}\}$$

Use the retail.csv data for problems 1 and 2. The data available in column 2 is the monthly data on retail book store sales (in millions of dollars) in US from July 2009 to December 2019.

1. (30 points) Fit a harmonic regression model with appropriate trend for the retail sales data. Give a plot of the fitted values (line +markers plot) overlaid on the original series (scatter plot) and comment on the fit of the data. Here are a few things to keep in mind as you design your analysis plan:
    - Start by creating a time series object, a time series plot and describe it.
    - Check the suitability of an additive vs multiplicative model. Use transformations if necessary.
    - Choose the appropriate number of harmonic components to be included in your model.
    - Analyze your residuals from the lm model thoroughly.
    - Fit a model to the residuals if they are autocorrelated. Make sure that you look at the ACF/PACF plots along with built in functions to choose a model for the residuals.
    - Decide wisely on when to stop fitting the model.
    - The plot of the fitted values should be based on all the models fit in the process.
    - Make sure that you explain each decision made with quantitative evidences (plots, hypothesis tests) as well as your expert observations.
    - Clearly specify each model that is being fit. You are free to compare multiple models and then choose your preferred model.

2. (20 points) Fit a SARIMA model to the retail sales data. Forecast and plot the observations for next 12 months. Here are a few things to keep in mind:

   - Consider using a transformed data wherever suitable.
   - You should figure out the order yourself and make recommendations before using built in functions to finalize the model.
   - Make sure that you explain each decision made with quantitative evidences (plots, hypothesis tests) as well as your expert observations.
   - Use trace=F and/or details=F arguments to make sure your model summary is concise.

3. (20 points) **Use the "ausbeer" data from the "fpp" package for this problem. The problem requires a training and a test dataset. The training data should be from 1959 Quarter 2 to 1973 Quarter 4. The test data should be from 1974 Quarter 1 to 1975 Quarter 4.**

   (a) Fit a seasonal means model with appropriate trend to the training data. Pay attention to pointers given in problem 1 as you finalize the model. Use this model to forecast the values for the next 2 years (8 quarters).

   (b) Fit a SARIMA model to the training data. Pay attention to pointers given in problem 2 as you finalize the model. Use this model to forecast the values for the next 2 years.

   (c) Create a plot of observed test series (scatter plot) and overlay forecasts from the two models (regression and SARIMA model). Use different colors for each series and comment on the quality of the predictions. Comment on the quality of the forecasts.

4. (15 points) **Use the euro.csv dataset for this problem. The dataset records US to Euro spot exchange rate. No dates are available for this data.**

   I believe a random walk is a good model for this data. Argue this using a combination of techniques - time series plots, differencing, ACF plots, and/or built in functions to pick orders for ARIMA model.

5. (20 points) This is a simulation exercise. Generate 3 time series data sets each of length $n = 250$ for the following models:

   - AR(1) with $\phi = -0.6$
   - MA(1) with $\theta = 0.8$
   - ARMA(1,1) with $\phi = -0.6, \theta = 0.8$

   (a) For each series plot the simulated time series, create the sample ACF and PACF plots, use the armasubsets function to see which model is picked by the algorithm.

   (b) Do the ACF/PACF plots agree with what we have learnt about these models in class?

   (c) Does the armasubsets function pick the correct models?

   (d) Repeat this exercise for $n = 1000$. You don't need to give the plots or any output; just write a summary of what you observed. Did the results change?