# ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 5 - Due date 03/12/21

Yash Doshi

## Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

Once you have the project open the first thing you will do is change "Student Name" on line 3 with your name. Then you will start working through the assignment by **creating code and output** that answer each question. Be sure to use this assignment document. Your report should contain the answer to each question and any plots/tables you obtained (when applicable).

When you have completed the assignment, **Knit** the text and code into a single PDF file. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A05_Sp21.Rmd"). Submit this pdf using Sakai.

## Questions

This assignment has general questions about ARIMA Models.

Packages needed for this assignment: "forecast","tseries". Do not forget to load them before running your script, since they are NOT default packages.\

```
#Load/install required package here
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

```
library(tseries)
library(ggplot2)
```

## Q1

Describe the important characteristics of the sample autocorrelation function (ACF) plot and the partial sample autocorrelation function (PACF) plot for the following models:

(a) AR(2)

AR is an autoregressive model, meaning we perform regression on previous observations. The generalize the idea of regression to represent the linear dependence between a dependent variable and an explanatory variable. If we only have one previous observation, then we say that we are working with 1st order autoregressive process, or AR 1 model. The order is related to how many previous observations you have on your regression, or how many previous observations do we need to have a good representation of Yt (our independent variable).

AR models show a slow decay on ACFs. Slow decay means how long does it take to reach 0 in ACF plot. In simpler terms, if the values go up, they will stay up for sometime. They will not drop from a positive number to a negative number all of a sudden. We can say that AR models have a relatively long memory. Since AR models decay exponentially with time, it means that there is some relationship between the current observation and previous observations.

The PACF plots will help us in identifying the order of the AR model. PACF will cut-off at the order of the AR model. For instance, if the order is 1, then anything after lag two will be insignificant. There is only one significant correlation with previous observations.

(b) MA(1)

MA models have short to no memory. Therefore, its plots will have high fluctuations. For instance, it can go from -3 to 6 to -10 in three consecutive time periods. MA components can have an order higher than 1.

For MA models, ACFs will identify the order of the model. Unlike AR models, there will be no slow decay for MA terms in the ACF. However, PACF of MA models will have a slow, exponential decay. If stationary models have a negative autocorrelation at lag 1, MA terms work the best.

## Q2

Recall that the non-seasonal ARIMA is described by three parameters ARIMA$(p, d, q)$ where $p$ is the order of the autoregressive component, $d$ is the number of times the series need to be differenced to obtain stationarity and $q$ is the order of the moving average component. If we don't need to difference the series, we don't need to specify the "I" part and we can use the short version, i.e., the ARMA$(p, q)$. Consider three models: ARMA(1,0), ARMA(0,1) and ARMA(1,1) with parameters $\phi = 0.6$ and $\theta = 0.9$. The $\phi$ refers to the AR coefficient and the $\theta$ refers to the MA coefficient. Use R to generate $n = 100$ observations from each of these three models

(a) Plot the sample ACF for each of these models in one window to facilitate comparison (Hint: use command $par(mfrow = c(1, 3))$ that divides the plotting window in three columns).

(b) Plot the sample PACF for each of these models in one window to facilitate comparison.

(c) Look at the ACFs and PACFs. Imagine you had these plots for a data set and you were asked to identify the model, i.e., is it AR, MA or ARMA and the order of each component. Would you be identify them correctly? Explain your answer.

> Answer:

(d) Compare the ACF and PACF values R computed with the theoretical values you provided for the coefficients. Do they match? Explain your answer.

> Answer:

(e) Increase number of observations to $n = 1000$ and repeat parts (a)-(d).

## Q3

Consider the ARIMA model $y_t = 0.7 * y_{t-1} - 0.25 * y_{t-12} + a_t - 0.1 * a_{t-1}$

(a) Identify the model using the notation ARIMA$(p, d, q)(P, D, Q)_s$, i.e., identify the integers $p, d, q, P, D, Q, s$ (if possible) from the equation.

p = 1 d = 0 q = 1 P = 1 D = 0 Q = 0

(b) Also from the equation what are the values of the parameters, i.e., model coefficients.

p

## Q4

Plot the ACF and PACF of a seasonal ARIMA$(0, 1) \times (1, 0)_{12}$ model with $\phi = 0.8$ and $\theta = 0.5$ using R. The 12 after the bracket tells you that $s = 12$, i.e., the seasonal lag is 12, suggesting monthly data whose behavior is repeated every 12 months. You can generate as many observations as you like. Note the Integrated part was omitted. It means the series do not need differencing, therefore $d = D = 0$. Plot ACF and PACF for the simulated data. Comment if the plots are well representing the model you simulated, i.e., would you be able to identify the order of both non-seasonal and seasonal components from the plots? Explain.