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

Assignment 3 - Due date 02/12/21

Student Name

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github.

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).

Please keep this R code chunk options for the report. It is easier for us to grade when we can see code and output together. And the tidy.opts will make sure that line breaks on your code chunks are automatically added for better visualization.

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_A01_Sp21.Rmd"). Submit this pdf using Sakai.

Questions

Consider the same data you used for A2 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption
The data comes from the US Energy Information and Administration and corresponds to the January 2021 Monthly Energy Review. Once again you will work only with the following columns: Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption. Create a data frame structure with these three time series only.

R packages needed for this assignment:"forecast","tseries", and "Kendall". Install these packages, if you haven't done yet. Do not forget to load them before running your script, since they are NOT default packages.\

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

```
##
## Attaching package: 'lubridate'
```

```
## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union
```

```
library(readxl)
library(forecast)
```

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

```
library(tseries)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)
```

##Trend Component

```
#Importing data set
raw_data <- read_excel("../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",s

spec_data <- raw_data[,4:6]
head(spec_data)
```

```
## # A tibble: 6 x 3
##   `Total Biomass Energy Pr~ `Total Renewable Energy Pr~ `Hydroelectric Power Co~
##   <chr>                     <chr>                       <chr>
## 1 (Trillion Btu)            (Trillion Btu)              (Trillion Btu)
## 2 129.787                   403.981                     272.703
## 3 117.338                   360.9                       242.199
## 4 129.938                   400.161                     268.81
## 5 125.636                   380.47                      253.185
## 6 129.834                   392.141                     260.77
```

```
my_date <- raw_data[,1]
head(my_date)
```

```
## # A tibble: 6 x 1
##   Month
##   <dttm>
## 1 NA
## 2 1973-01-01 00:00:00
## 3 1973-02-01 00:00:00
## 4 1973-03-01 00:00:00
## 5 1973-04-01 00:00:00
## 6 1973-05-01 00:00:00
```

```
spec_data <- cbind(my_date, spec_data)
colnames(spec_data)=c("Date", "TBEP", "TREP","HPC")
spec_data <- spec_data[-c(1),]
cols.num <- c("TBEP","TREP","HPC")
spec_data[cols.num] <- sapply(spec_data[cols.num],as.numeric)
head(spec_data)
```

```
##         Date    TBEP    TREP     HPC
## 2 1973-01-01 129.787 403.981 272.703
## 3 1973-02-01 117.338 360.900 242.199
## 4 1973-03-01 129.938 400.161 268.810
```

```
## 5 1973-04-01 125.636 380.470 253.185
## 6 1973-05-01 129.834 392.141 260.770
## 7 1973-06-01 125.611 377.232 249.859
```

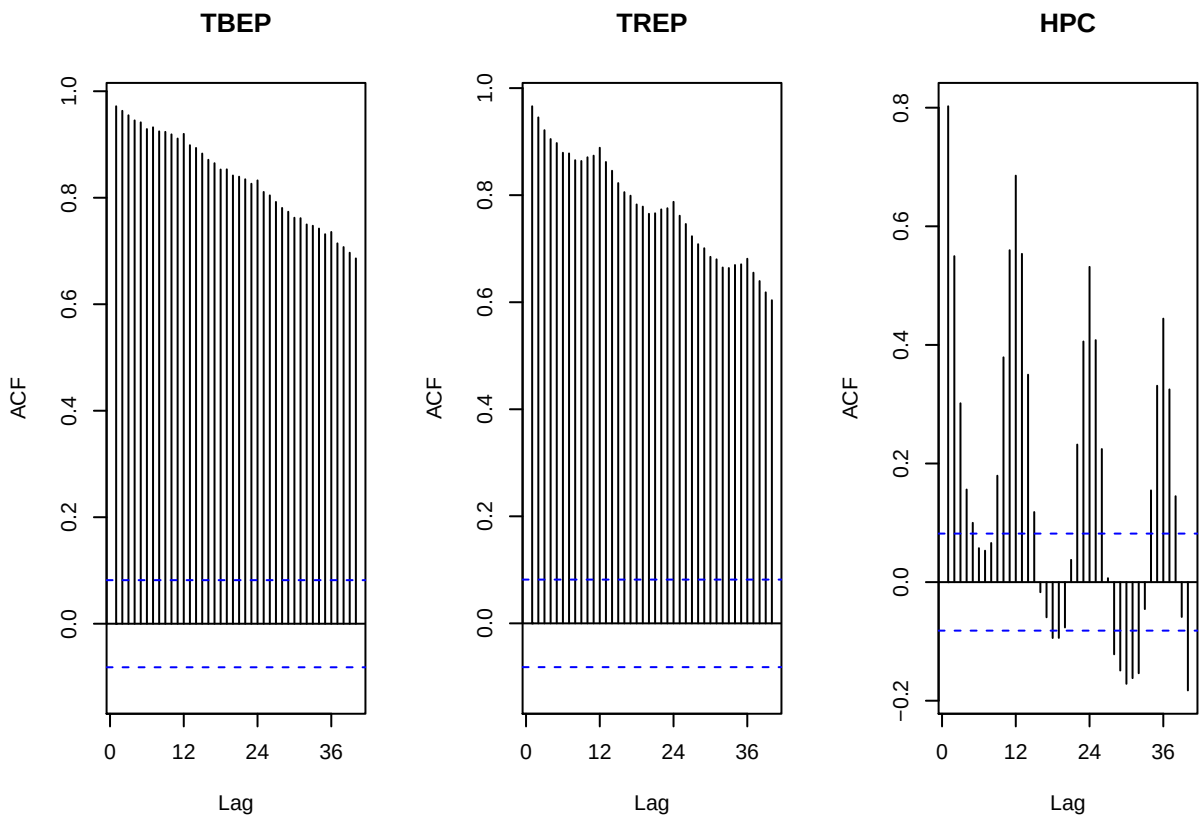
```
ts_data <- ts(spec_data[,2:4],frequency = 12)
head(ts_data,15)
```

```
##          TBEP    TREP     HPC
## Jan 1 129.787 403.981 272.703
## Feb 1 117.338 360.900 242.199
## Mar 1 129.938 400.161 268.810
## Apr 1 125.636 380.470 253.185
## May 1 129.834 392.141 260.770
## Jun 1 125.611 377.232 249.859
## Jul 1 129.787 367.325 235.670
## Aug 1 129.918 353.757 222.077
## Sep 1 125.782 307.006 179.733
## Oct 1 129.970 323.453 191.723
## Nov 1 125.643 337.817 210.285
## Dec 1 129.824 406.694 274.435
## Jan 2 130.807 437.467 304.506
## Feb 2 118.091 399.942 279.950
## Mar 2 130.727 423.474 290.582
```

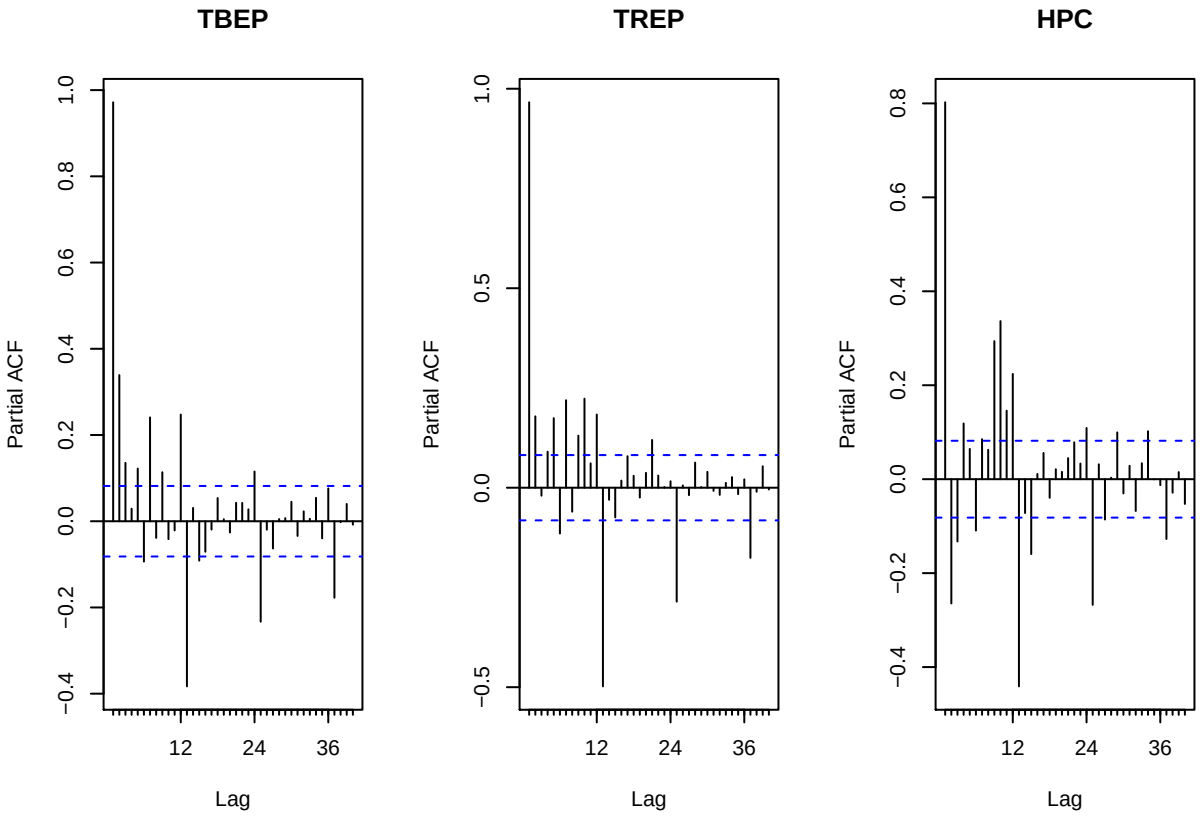
Q1

Create a plot window that has one row and three columns. And then for each object on your data frame, fill the plot window with time series plot, ACF and PACF. You may use the some code form A2, but I want all three plots on the same window this time. (Hint: watch videos for M4)

```
par(mfrow=c(1,3)) #place plot side by side
Acf(ts_data[,1],lag.max=40,main=paste("TBEP"))
Acf(ts_data[,2],lag.max=40,main=paste("TREP"))
Acf(ts_data[,3],lag.max=40,main=paste("HPC"))
```




```
par(mfrow=c(1,3)) #place plot side by side
Pacf(ts_data[,1],lag.max=40,main=paste("TBEP"))
Pacf(ts_data[,2],lag.max=40,main=paste("TREP"))
Pacf(ts_data[,3],lag.max=40,main=paste("HPC"))
```



Q2

From the plot in Q1, do the series Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Q3

Use the *lm()* function to fit a linear trend to the three time series. Ask R to print the summary of the regression. Interpret the regression output, i.e., slope and intercept. Save the regression coefficients for further analysis.

Q4

Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

Q5

Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How?

Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

Q6

Do the series seem to have a seasonal trend? Which serie/series? Use function *lm()* to fit a seasonal means model to this/these time series. Ask R to print the summary of the regression. Interpret the regression

output. Save the regression coefficients for further analysis.

Q7

Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

Q8

Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How?