ENV 790.30 - Time Series Analysis for Energy Data | Spring 2025

Assignment 2 - Due date 01/23/25

Justin Maynard

Submission Instructions

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 file open on your local machine the first thing you will do is rename the file such that it includes your first and last name (e.g., "LuanaLima_TSA_A02_Sp24.Rmd"). Then change "Student Name" on line 4 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. Submit this pdf using Sakai.

R packages

R packages needed for this assignment:"forecast","tseries", and "dplyr". 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(forecast)
library(tseries)
library(dplyr)
library(here)
library(readxl)
library(ggplot2)
```

Data set information

Consider the data provided in the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source on our **Data** folder. The data comes from the US Energy Information and Administration and corresponds to the December 2023 Monthly Energy Review. The spreadsheet is ready to be used. You will also find a *.csv* version of the data "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source-Edit.csv". You may use the function *read.table*() to import the *.csv* data in R. Or refer to the file "M2_ImportingData_CSV_XLSX.Rmd" in our Lessons folder for functions that are better suited for importing the *.xlsx*.

```
#Importing data set
data <- read_excel("Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
                   skip = 12, col_names = FALSE)
```

```r
names <- read_excel("Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
                    skip = 10, n_max = 1, col_names = F)
colnames(data) <- names
```

Question 1

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. Use the command head() to verify your data.

```r
data <- data[,c("Month", "Total Biomass Energy Production",
                "Total Renewable Energy Production", "Hydroelectric Power Consumption")]
head(data)
```

```
## # A tibble: 6 x 4
##   Month               `Total Biomass Energy Production` Total Renewable Energy~1
##   <dttm>                                          <dbl>                    <dbl>
## 1 1973-01-01 00:00:00                              130.                     220.
## 2 1973-02-01 00:00:00                              117.                     197.
## 3 1973-03-01 00:00:00                              130.                     219.
## 4 1973-04-01 00:00:00                              126.                     209.
## 5 1973-05-01 00:00:00                              130.                     216.
## 6 1973-06-01 00:00:00                              126.                     208.
## # i abbreviated name: 1: `Total Renewable Energy Production`
## # i 1 more variable: `Hydroelectric Power Consumption` <dbl>
```

Question 2

Transform your data frame in a time series object and specify the starting point and frequency of the time series using the function ts().

```r
ts <- ts(data[,2:4], start = c(1973,1,1), frequency = 12)
head(ts)
```

```
##          Total Biomass Energy Production Total Renewable Energy Production
## Jan 1973                         129.787                           219.839
## Feb 1973                         117.338                           197.330
## Mar 1973                         129.938                           218.686
## Apr 1973                         125.636                           209.330
## May 1973                         129.834                           215.982
## Jun 1973                         125.611                           208.249
##          Hydroelectric Power Consumption
## Jan 1973                          89.562
## Feb 1973                          79.544
## Mar 1973                          88.284
## Apr 1973                          83.152
## May 1973                          85.643
## Jun 1973                          82.060
```

Question 3

Compute mean and standard deviation for these three series.

```
print(paste("Mean of",colnames(ts)[1], "is",round(mean(ts[,1]),2)))
```

```
## [1] "Mean of Total Biomass Energy Production is 282.68"
```

```
print(paste("Mean of",colnames(ts)[2], "is",round(mean(ts[,2]),2)))
```

```
## [1] "Mean of Total Renewable Energy Production is 402.02"
```

```
print(paste("Mean of",colnames(ts)[3], "is",round(mean(ts[,3]),2)))
```

```
## [1] "Mean of Hydroelectric Power Consumption is 79.55"
```

```
print(paste("Standard deviation of",colnames(ts)[1], "is",round(sd(ts[,1]),2)))
```

```
## [1] "Standard deviation of Total Biomass Energy Production is 94.06"
```

```
print(paste("Standard deviation of",colnames(ts)[2], "is",round(sd(ts[,2]),2)))
```

```
## [1] "Standard deviation of Total Renewable Energy Production is 143.79"
```

```
print(paste("Standard deviation of",colnames(ts)[3], "is",round(sd(ts[,3]),2)))
```

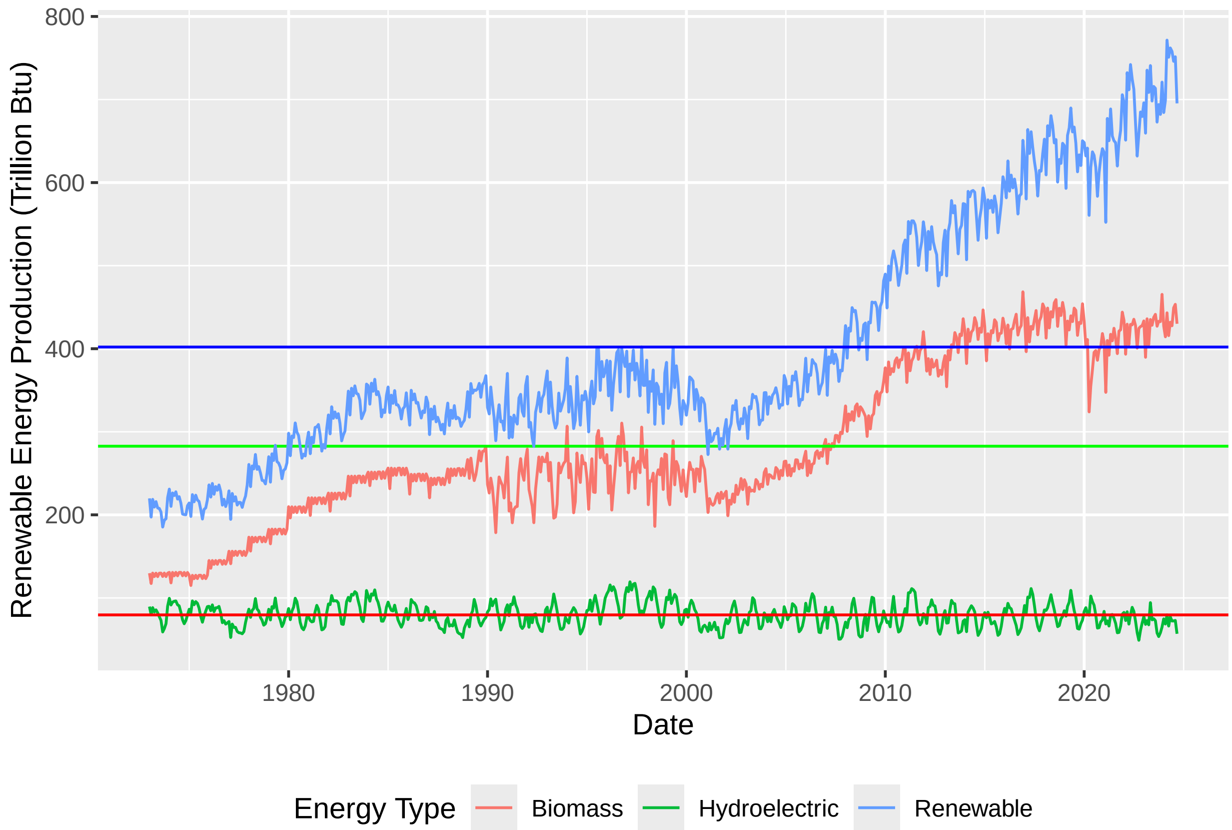
```
## [1] "Standard deviation of Hydroelectric Power Consumption is 14.11"
```

Question 4

Display and interpret the time series plot for each of these variables. Try to make your plot as informative as possible by writing titles, labels, etc. For each plot add a horizontal line at the mean of each series in a different color.

Based on the following plot, hydroelectric power consumption varies mostly seasonally, and stays near it's average with little trend. Biomass energy production and renewable energy production also vary seasonally, but have an increasing trend.

```
ggplot(data, aes(x = Month)) +
  geom_line(aes(y = `Total Biomass Energy Production`, col = "Biomass")) +
  geom_hline(yintercept = mean(data$`Total Biomass Energy Production`), color = 'green') +
  geom_line(aes(y = `Total Renewable Energy Production`, col = "Renewable")) +
  geom_hline(yintercept = mean(data$`Total Renewable Energy Production`), color = 'blue') +
  geom_line(aes(y = `Hydroelectric Power Consumption`, col = "Hydroelectric")) +
  geom_hline(yintercept = mean(data$`Hydroelectric Power Consumption`), color = 'red') +
  xlab("Date") +
  ylab("Renewable Energy Production (Trillion Btu)") +
  labs(color = "Energy Type") +
  theme(legend.position = "bottom")
```



Question 5

Compute the correlation between these three series. Are they significantly correlated? Explain your answer.

Renewable energy production is highly correlated to biomass energy production (0.97), which may be due to the similar increasing trend that biomass and renewable production follow, as well as similar seasonal cycles. Hydroelectric power consumption is not strongly correlated with biomass production (-0.11) or renewable energy production (-0.03) as hydroelectric power consumption does not show the same increasing trend, as new hydroelectric plants are rarely built.

```r
print(cor(ts))
```

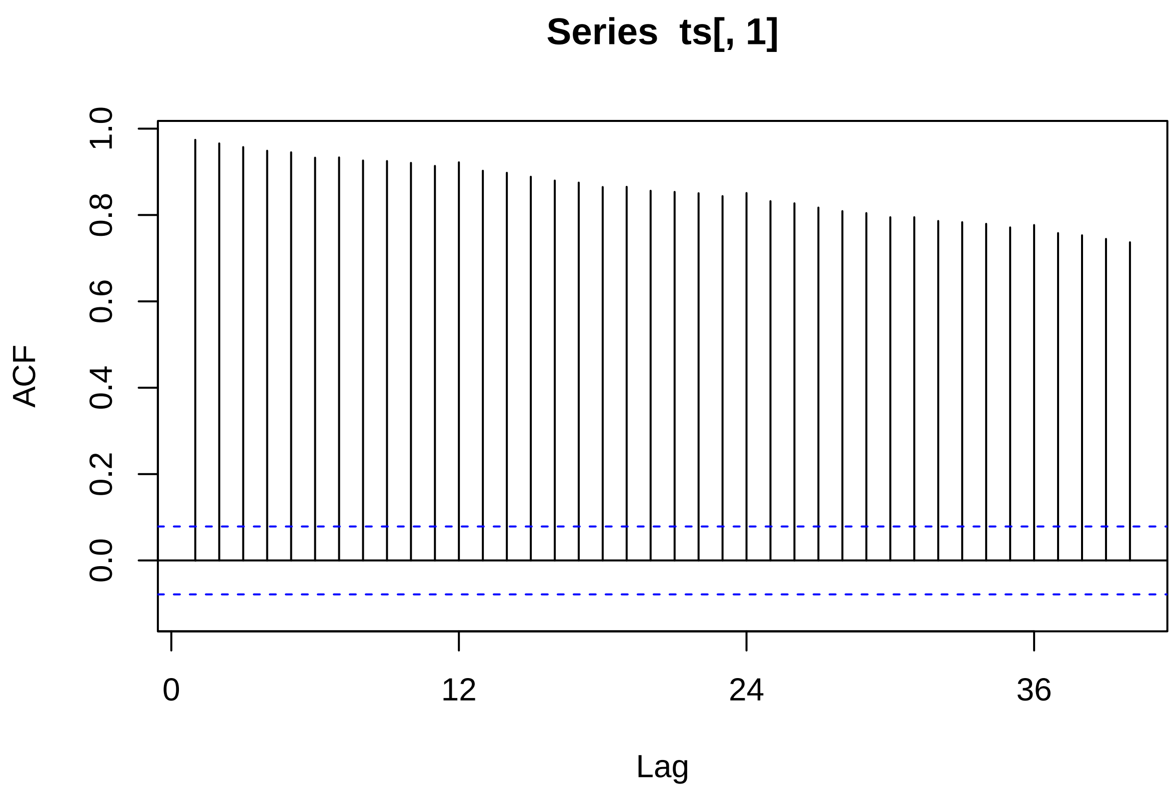
```
##                                  Total Biomass Energy Production
## Total Biomass Energy Production                        1.0000000
## Total Renewable Energy Production                      0.9678137
## Hydroelectric Power Consumption                       -0.1142927
##                                  Total Renewable Energy Production
## Total Biomass Energy Production                          0.96781371
## Total Renewable Energy Production                        1.00000000
## Hydroelectric Power Consumption                         -0.02916103
##                                  Hydroelectric Power Consumption
## Total Biomass Energy Production                      -0.11429266
## Total Renewable Energy Production                    -0.02916103
## Hydroelectric Power Consumption                       1.00000000
```

Question 6

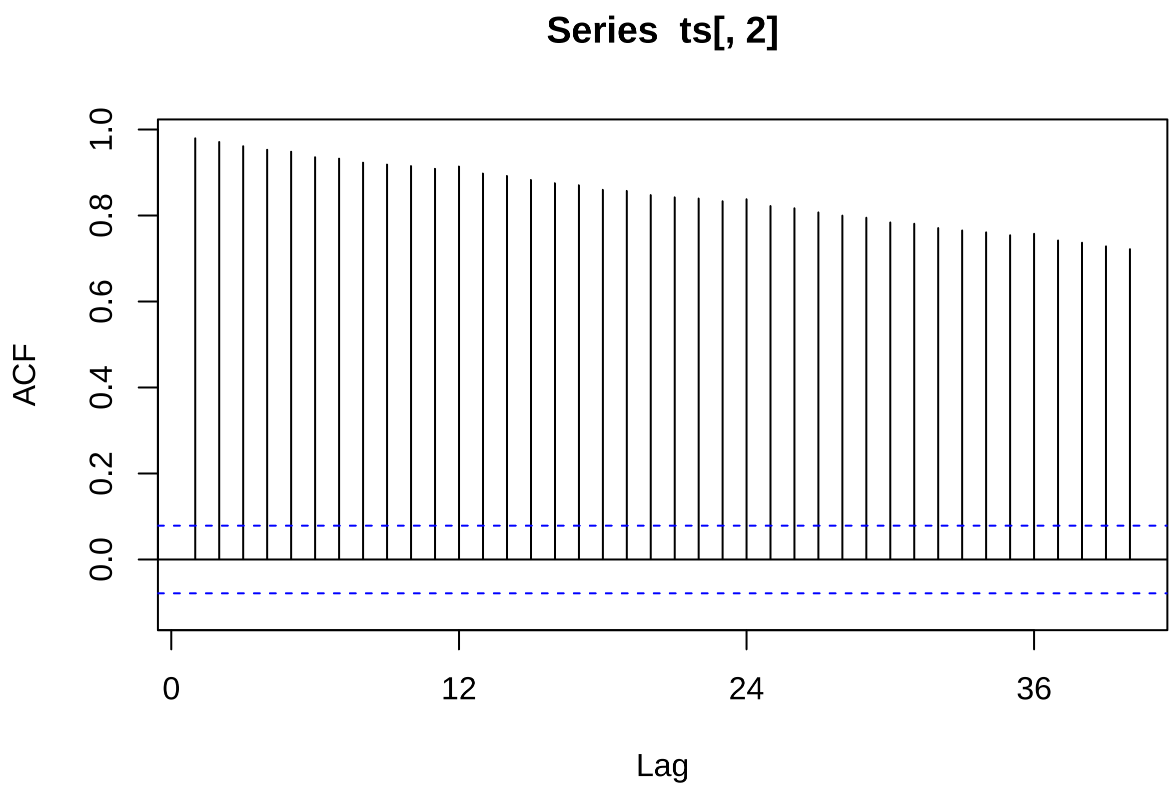
Compute the autocorrelation function from lag 1 up to lag 40 for these three variables. What can you say about these plots? Do the three of them have the same behavior?

The autocorrelation plots for biomass and renewable energy have the same behavior of slight decrease in autocorrelation from lag 1 to lag 40. Hydroelectric power production follows a different trend, and oscillates between positive and negative autocorrelation following a wave shape. This is because there is little trend and high seasonality.

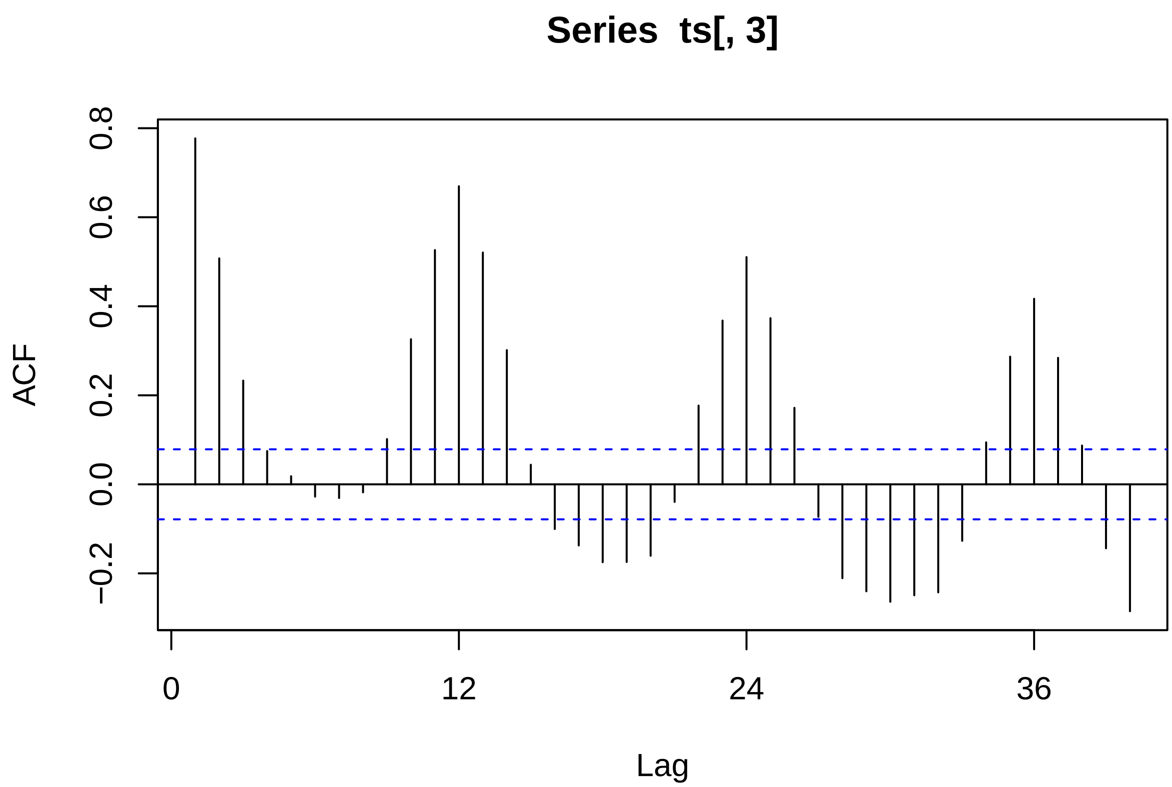
```r
#Biomass
Acf_1 <- Acf(ts[,1],lag=40)
```




```r
#Renewable
Acf_2 <- Acf(ts[,2],lag=40)
```




```r
#Hydro
Acf_3 <- Acf(ts[,3],lag=40)
```

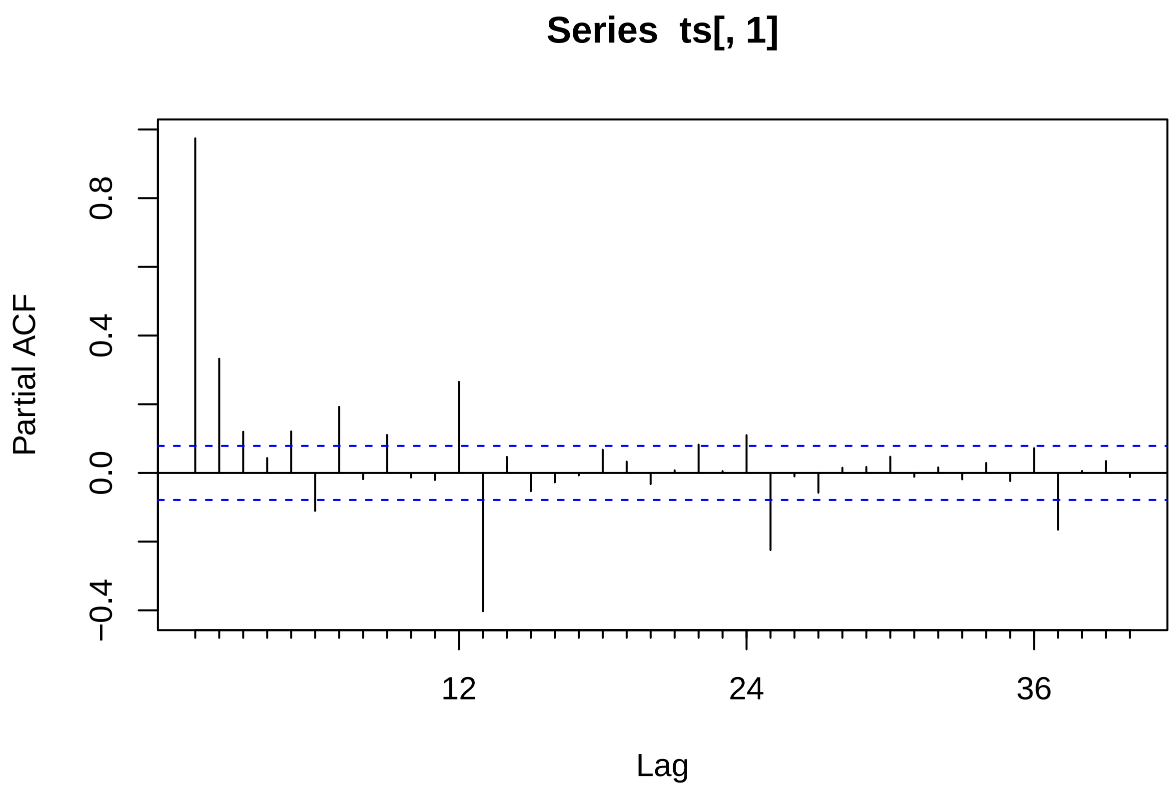


Question 7

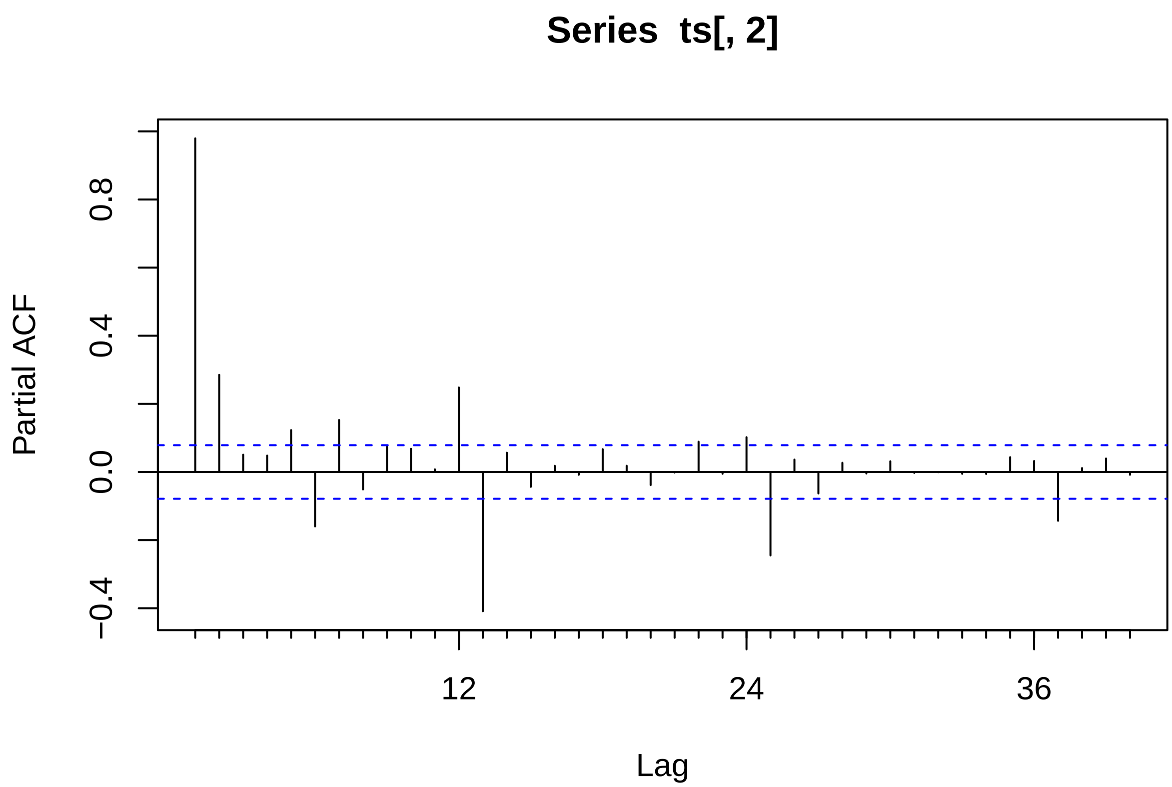
Compute the partial autocorrelation function from lag 1 to lag 40 for these three variables. How these plots differ from the ones in Q6?

These plots differ in that they all start with a high PACF and then exhibit the "wave" pattern found in the hydroelectric plot in A6, except that the wave pattern is decreasing in magnitude as lag increases.

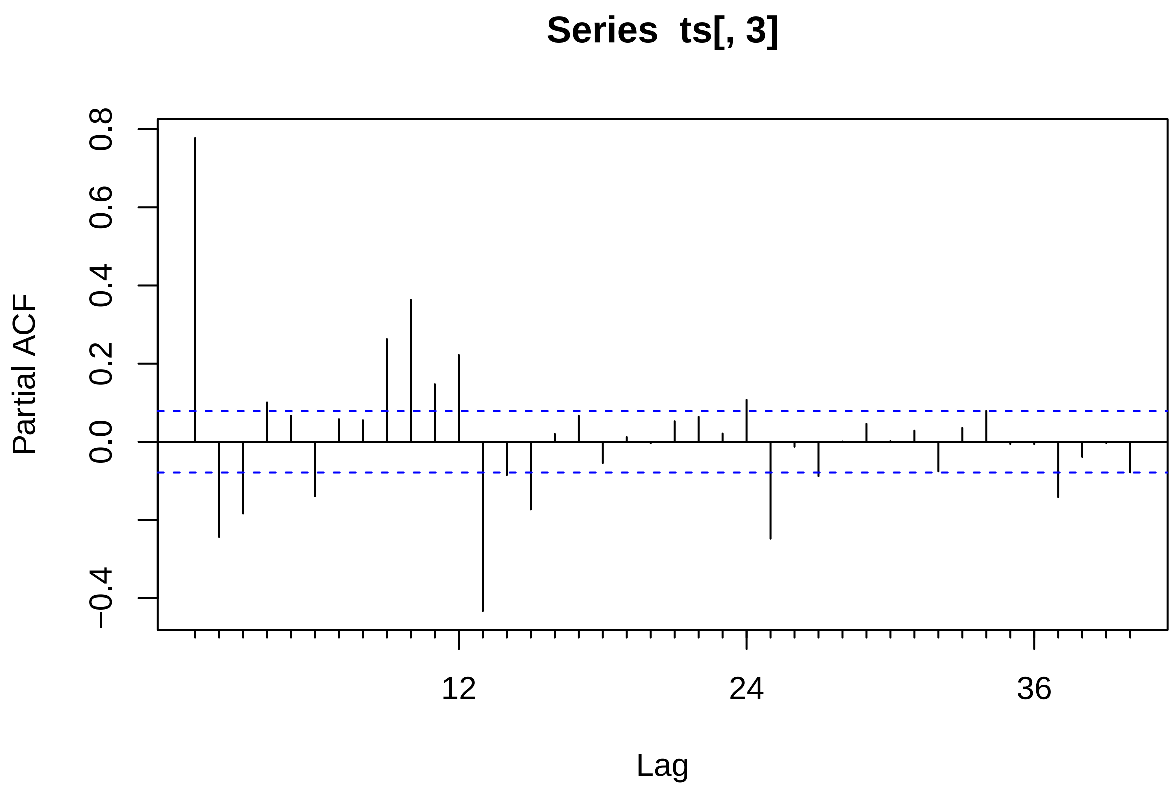
```
Pacf_1 <- Pacf(ts[,1],lag=40)
```




```
Pacf_2 <- Pacf(ts[,2],lag=40)
```




```
Pacf_3 <- Pacf(ts[,3],lag=40)
```


Series ts[, 3]
Partial ACF
-0.4
0.0
0.2
0.4
0.6
0.8
12
24
36
Lag