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

Assignment 2 - Due date 01/23/25

Ellie Shang

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)
```

```
## 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)
library(readxl)
```

Data set information

Consider the data provided in the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.x 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
energy_data <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```
read_col_names <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sou
```

```
## New names:
## * `` -> `...1`
## * `` -> `...2`
## * `` -> `...3`
## * `` -> `...4`
## * `` -> `...5`
## * `` -> `...6`
## * `` -> `...7`
## * `` -> `...8`
## * `` -> `...9`
## * `` -> `...10`
## * `` -> `...11`
## * `` -> `...12`
## * `` -> `...13`
## * `` -> `...14`
```

```
colnames(energy_data) <- read_col_names
head(energy_data)
```

```
## # A tibble: 6 x 14
##   Month               `Wood Energy Production` `Biofuels Production`
##   <dttm>                                 <dbl> <chr>
## 1 1973-01-01 00:00:00                     130. Not Available
## 2 1973-02-01 00:00:00                     117. Not Available
## 3 1973-03-01 00:00:00                     130. Not Available
## 4 1973-04-01 00:00:00                     125. Not Available
```

```
## 5 1973-05-01 00:00:00                    130. Not Available
## 6 1973-06-01 00:00:00                    125. Not Available
## # i 11 more variables: `Total Biomass Energy Production` <dbl>,
## #   `Total Renewable Energy Production` <dbl>,
## #   `Hydroelectric Power Consumption` <dbl>,
## #   `Geothermal Energy Consumption` <dbl>, `Solar Energy Consumption` <chr>,
## #   `Wind Energy Consumption` <chr>, `Wood Energy Consumption` <dbl>,
## #   `Waste Energy Consumption` <dbl>, `Biofuels Consumption` <chr>,
## #   `Total Biomass Energy Consumption` <dbl>, ...
```

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.

```
energy_df = select(energy_data, "Total Biomass Energy Production", "Total Renewable Energy Production",

head(energy_df)
```

```
## # A tibble: 6 x 3
##   Total Biomass Energy Productio~1 Total Renewable Ener~2 Hydroelectric Power ~3
##                              <dbl>                  <dbl>                  <dbl>
## 1                             130.                   220.                   89.6
## 2                             117.                   197.                   79.5
## 3                             130.                   219.                   88.3
## 4                             126.                   209.                   83.2
## 5                             130.                   216.                   85.6
## 6                             126.                   208.                   82.1
## # i abbreviated names: 1: `Total Biomass Energy Production`,
## #   2: `Total Renewable Energy Production`,
## #   3: `Hydroelectric Power Consumption`
```

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().

```
energy_ts = ts(energy_df, start=c(1973,1), frequency=12)
```

Question 3

Compute mean and standard deviation for these three series.

```
mean(energy_ts[,"Total Biomass Energy Production"])
```

```
## [1] 282.6779
```

```
sd(energy_ts[,"Total Biomass Energy Production"])
```

```
## [1] 94.05815
```

```
mean(energy_ts[,"Total Renewable Energy Production"])
```

```
## [1] 402.0167
```

```
sd(energy_ts[,"Total Renewable Energy Production"])
```

```
## [1] 143.7927
```

```
mean(energy_ts[,"Hydroelectric Power Consumption"])
```

```
## [1] 79.55371
```

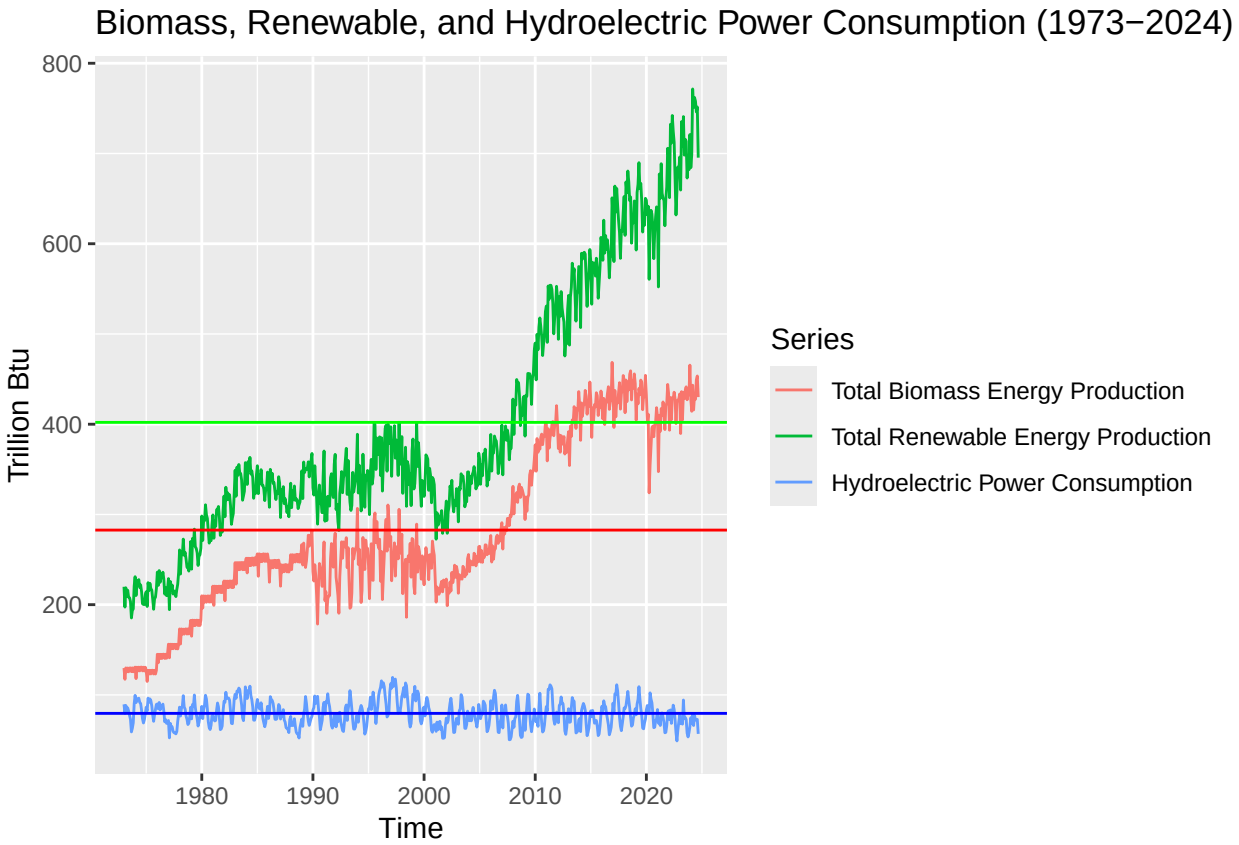
```
sd(energy_ts[,"Hydroelectric Power Consumption"])
```

```
## [1] 14.10737
```

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.

```
autoplot(energy_ts)+
  xlab("Time") +
  ylab("Trillion Btu") +
  labs(color="Series", title="Biomass, Renewable, and Hydroelectric Power Consumption (1973-2024)")+
  geom_hline(yintercept = mean(energy_ts[,"Total Biomass Energy Production"]), color="red")+
  geom_hline(yintercept = mean(energy_ts[,"Total Renewable Energy Production"]), color="green")+
  geom_hline(yintercept = mean(energy_ts[,"Hydroelectric Power Consumption"]), color="blue")
```



> Total Renewable Energy Production appears to have increased slowly, then rapidly after 2000, potentially with a multiplicative pattern. Biomass Energy Production similarly has been increasing, with a rapid increase between 2000 and 2010, though not as quickly as renewable energy overall. Hydroelectric Power Consumption appears to have remained relatively flat over the decades.

Question 5

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

```
cor(energy_ts)
```

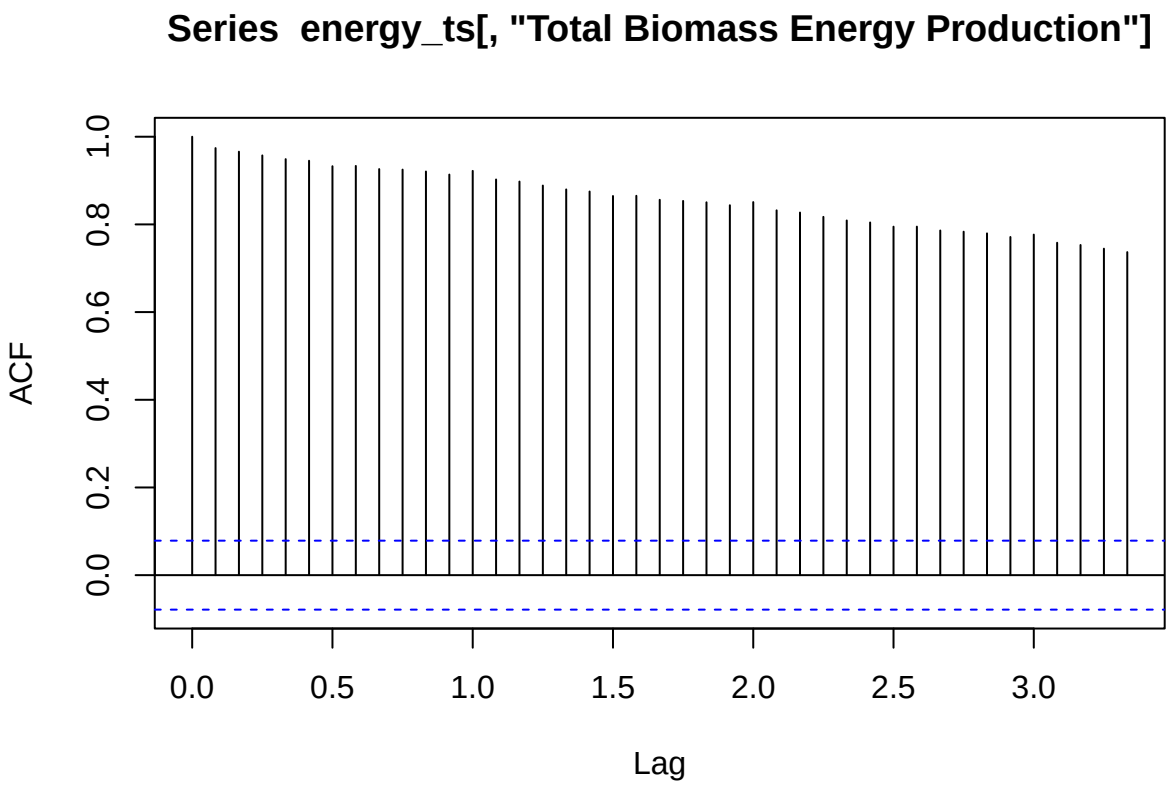
```
##                                  Total Biomass Energy Production
## Total Biomass Energy Production                        1.0000000
## Total Renewable Energy Production                      0.9678137
## Hydroelectric Power Consumption                       -0.1142927
##                                  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
```

Total Biomass Energy Production and Total Renewable Energy Production appear to have a close positive association with r=0.97. Neither are significantly correlated with Hydroelectric Power (r=-0.11 and r=-0.03, respectively).

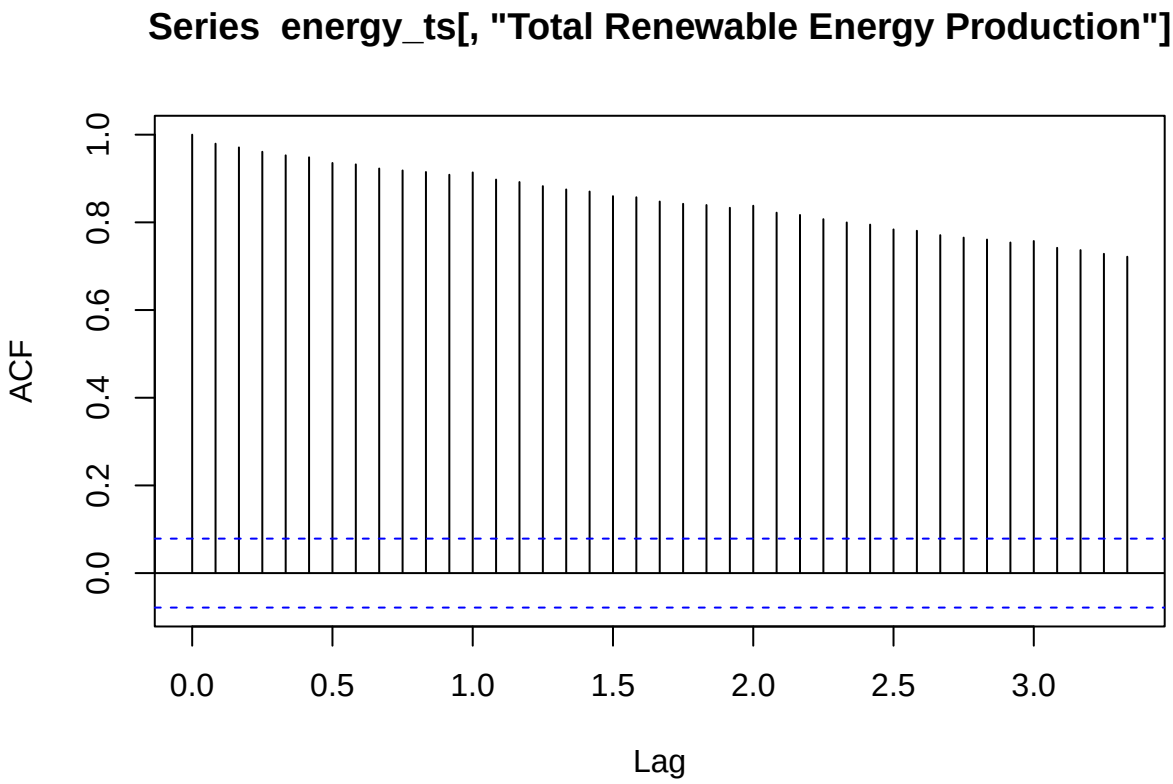
Question 6

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?

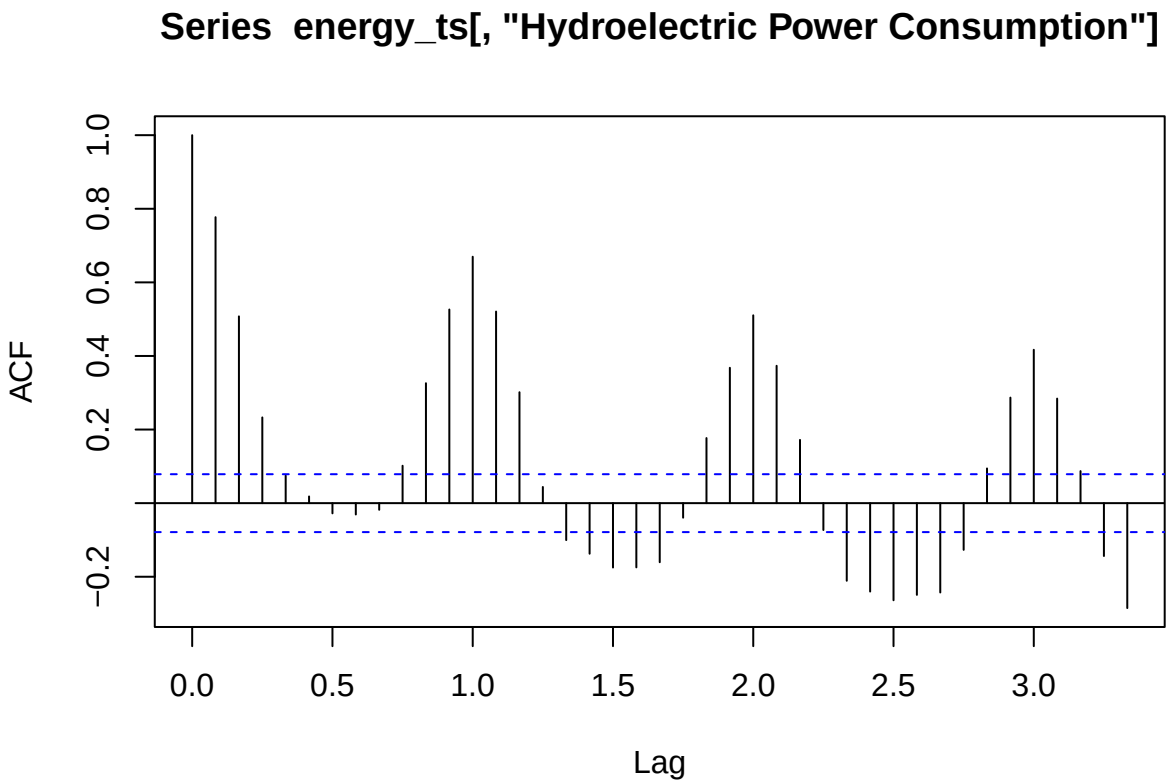
```
biomass_acf=acf(energy_ts[,"Total Biomass Energy Production"],lag=40)
```




```
renewable_acf=acf(energy_ts[,"Total Renewable Energy Production"],lag=40)
```




```
hydro_acf=acf(energy_ts[,"Hydroelectric Power Consumption"],lag=40)
```



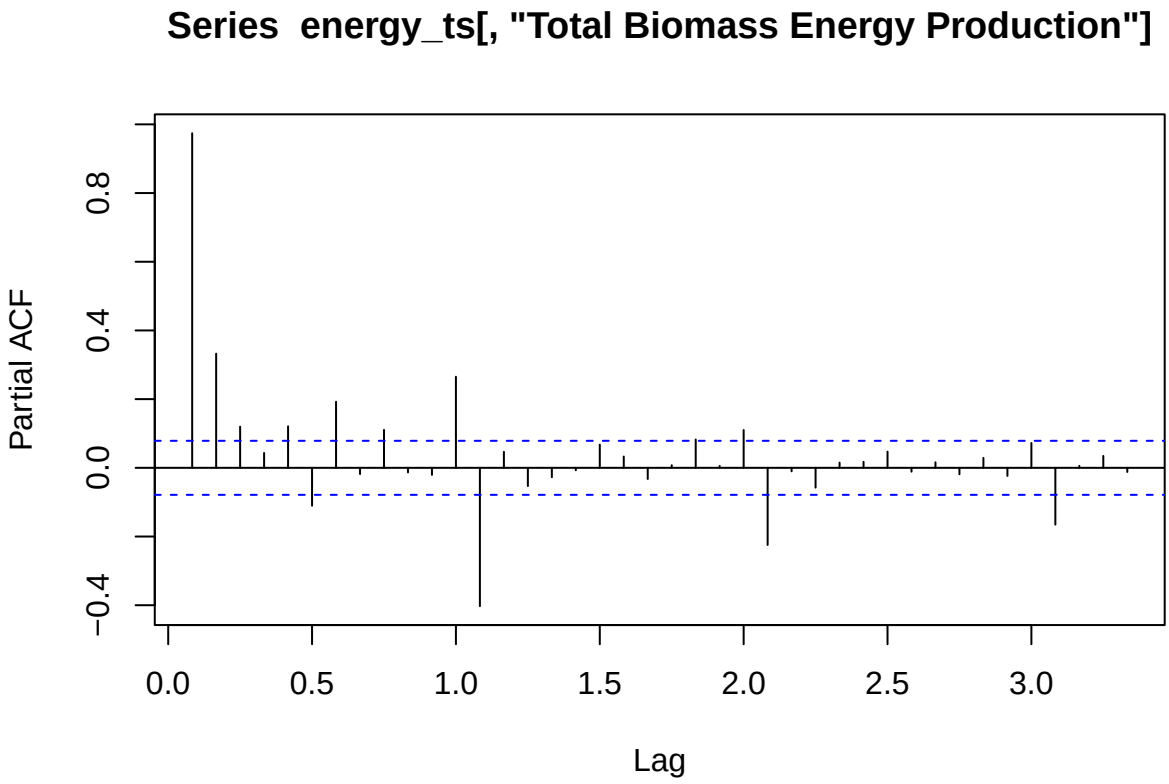
> The ACF for biomass and renewable energy production starts close to 1 and slowly declines as the lag increases, suggesting that values are highly correlated with previous values, particularly those closer in time. The ACF appears to be linearly decreasing. For hydroelectric power, the ACF appears somewhat sinusoidal, suggesting

some seasonality. All three plots seem to indicate that the ACF values are statistically significant - for all lags for biomass and renewable energy, and for most for hydro consumption.

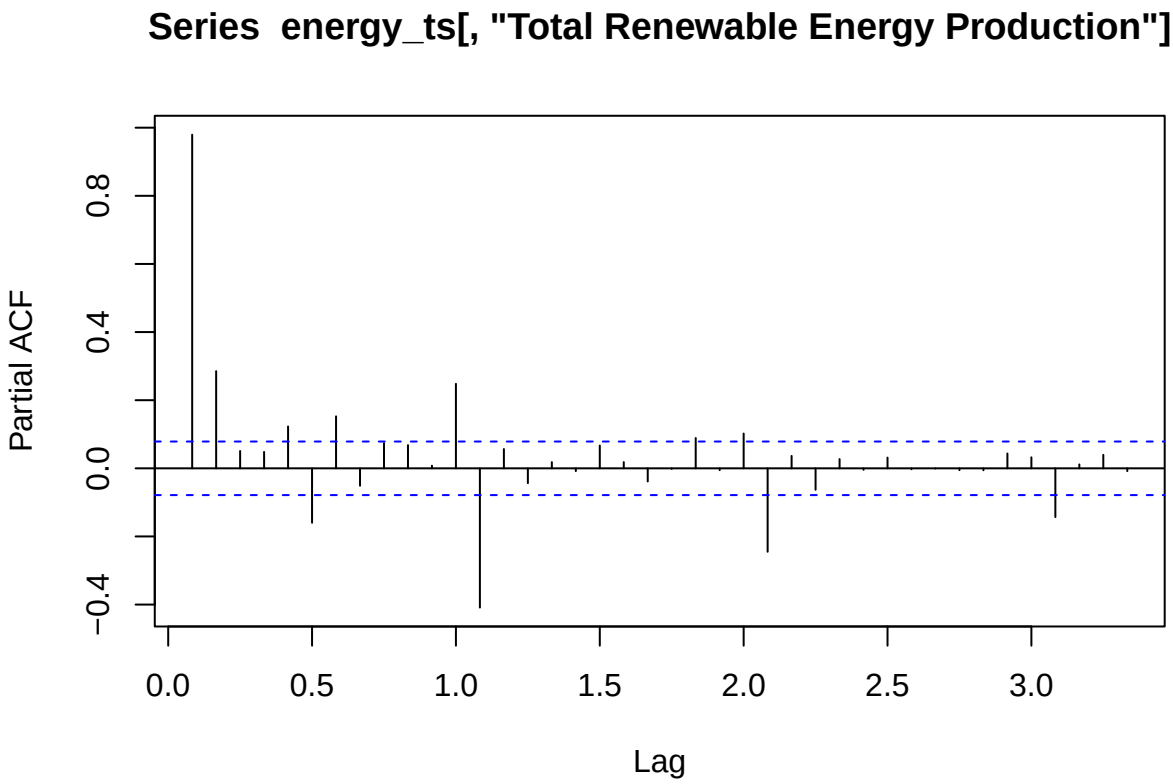
Question 7

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

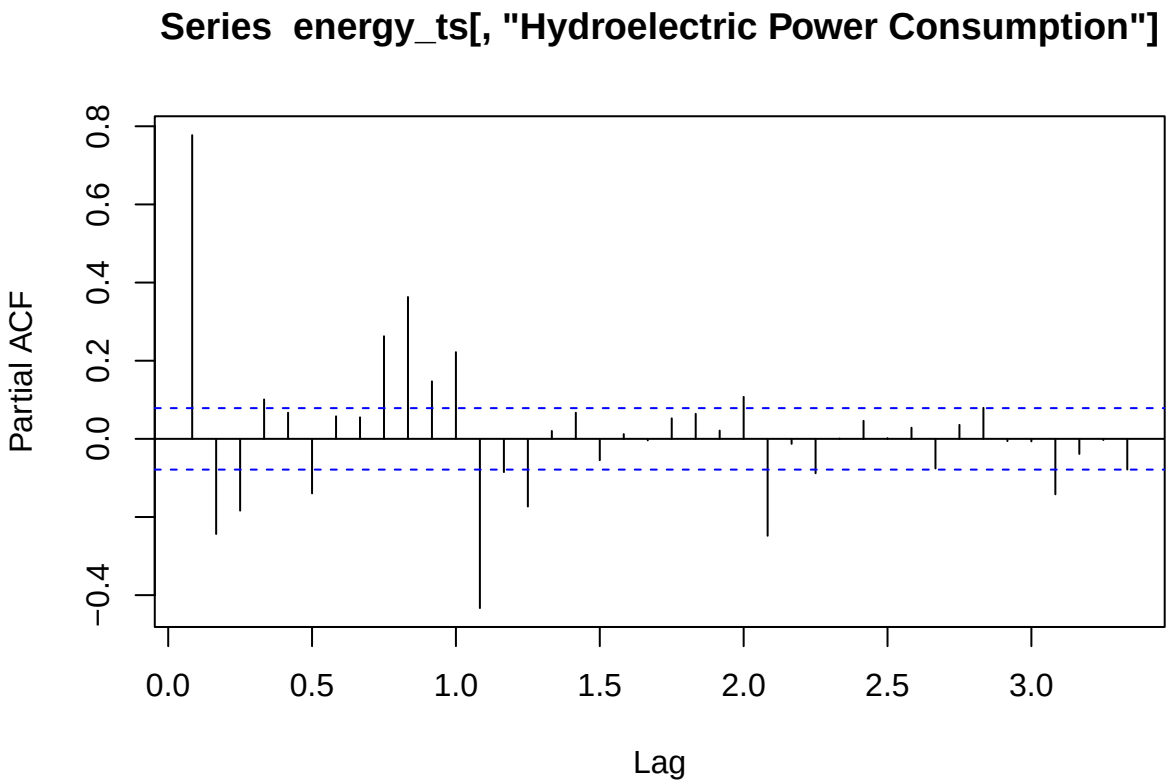
```
biomass_pacf=pacf(energy_ts[,"Total Biomass Energy Production"],lag=40)
```




```
renewable_pacf=pacf(energy_ts[,"Total Renewable Energy Production"],lag=40)
```




```
hydro_pacf=pacf(energy_ts[,"Hydroelectric Power Consumption"],lag=40)
```



>

These plots show much higher correlation at the first lag, and lower values everywhere else, with some significant values at the second lag and occasionally throughout, but mostly concentrated on the first lag. The hydro PACF may still show some mild sinusoidal behavior.