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

Assignment 2 - Due date 01/23/25

Jingze Dai

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.\

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

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

```r
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
```

```r
library(openxlsx)
```

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*.

```r
energy_data <- read.xlsx(
  xlsxFile="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  sheet = "Monthly Data", startRow = 13, colNames = FALSE)

read_col_names  <- read.xlsx(
  xlsxFile="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
  sheet = "Monthly Data", rows = 11, colNames = FALSE)

colnames(energy_data) <- read_col_names
energy_data$Data <- convertToDate(energy_data$Month)
head(energy_data)
```

```
##   Month Wood Energy Production Biofuels Production
## 1 26665                129.630       Not Available
## 2 26696                117.194       Not Available
## 3 26724                129.763       Not Available
## 4 26755                125.462       Not Available
## 5 26785                129.624       Not Available
## 6 26816                125.435       Not Available
##   Total Biomass Energy Production Total Renewable Energy Production
## 1                         129.787                           219.839
## 2                         117.338                           197.330
## 3                         129.938                           218.686
## 4                         125.636                           209.330
## 5                         129.834                           215.982
## 6                         125.611                           208.249
##   Hydroelectric Power Consumption Geothermal Energy Consumption
## 1                          89.562                         0.490
## 2                          79.544                         0.448
## 3                          88.284                         0.464
## 4                          83.152                         0.542
## 5                          85.643                         0.505
## 6                          82.060                         0.579
##   Solar Energy Consumption Wind Energy Consumption Wood Energy Consumption
## 1            Not Available           Not Available                 129.630
## 2            Not Available           Not Available                 117.194
## 3            Not Available           Not Available                 129.763
## 4            Not Available           Not Available                 125.462
## 5            Not Available           Not Available                 129.624
## 6            Not Available           Not Available                 125.435
```

```
##   Waste Energy Consumption Biofuels Consumption
## 1                    0.157        Not Available
## 2                    0.144        Not Available
## 3                    0.176        Not Available
## 4                    0.174        Not Available
## 5                    0.210        Not Available
## 6                    0.176        Not Available
##   Total Biomass Energy Consumption Total Renewable Energy Consumption
## 1                          129.787                            219.839
## 2                          117.338                            197.330
## 3                          129.938                            218.686
## 4                          125.636                            209.330
## 5                          129.834                            215.982
## 6                          125.611                            208.249
##         Data
## 1 1973-01-01
## 2 1973-02-01
## 3 1973-03-01
## 4 1973-04-01
## 5 1973-05-01
## 6 1973-06-01
```

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
# selecting the columns of interest
power_data <- energy_data %>%
  select("Total Biomass Energy Production",
         "Total Renewable Energy Production",
         "Hydroelectric Power Consumption")

head(power_data)
```

```
##   Total Biomass Energy Production Total Renewable Energy Production
## 1                         129.787                           219.839
## 2                         117.338                           197.330
## 3                         129.938                           218.686
## 4                         125.636                           209.330
## 5                         129.834                           215.982
## 6                         125.611                           208.249
##   Hydroelectric Power Consumption
## 1                          89.562
## 2                          79.544
## 3                          88.284
## 4                          83.152
## 5                          85.643
## 6                          82.060
```

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().

```
ts_power_data <- ts(power_data, start=c(1973,1), frequency=12)
```

Question 3

Compute mean and standard deviation for these three series.

```
mean_ts <- sapply(ts_power_data, mean)
sd_ts <- sapply(ts_power_data, sd)
print("Mean for the three series")
```

```
## [1] "Mean for the three series"
```

```
mean_ts
```

```
##   Total Biomass Energy Production Total Renewable Energy Production
##                         282.67785                         402.01667
##   Hydroelectric Power Consumption
##                          79.55371
```

```
print("Standard deviation for the three series")
```

```
## [1] "Standard deviation for the three series"
```

```
sd_ts
```

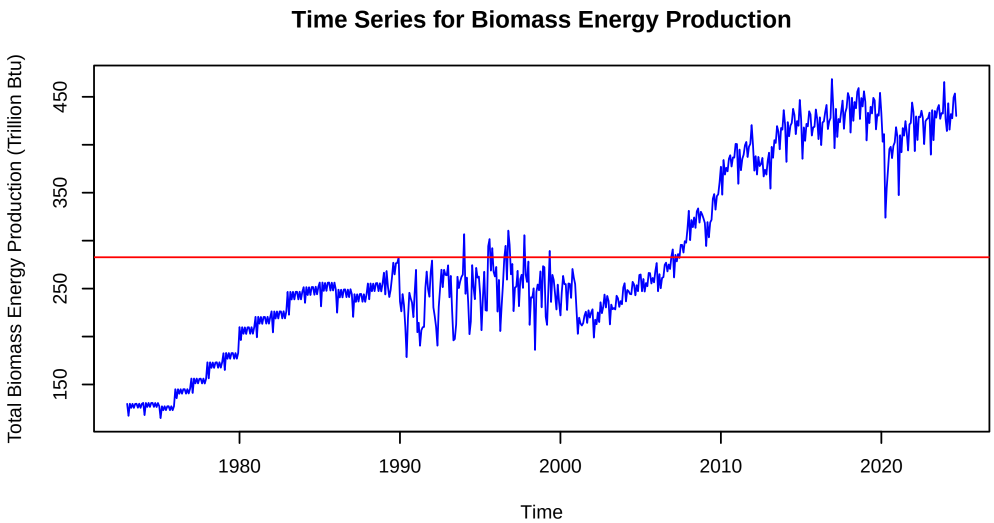
```
##   Total Biomass Energy Production Total Renewable Energy Production
##                          94.05815                         143.79270
##   Hydroelectric Power Consumption
##                          14.10737
```

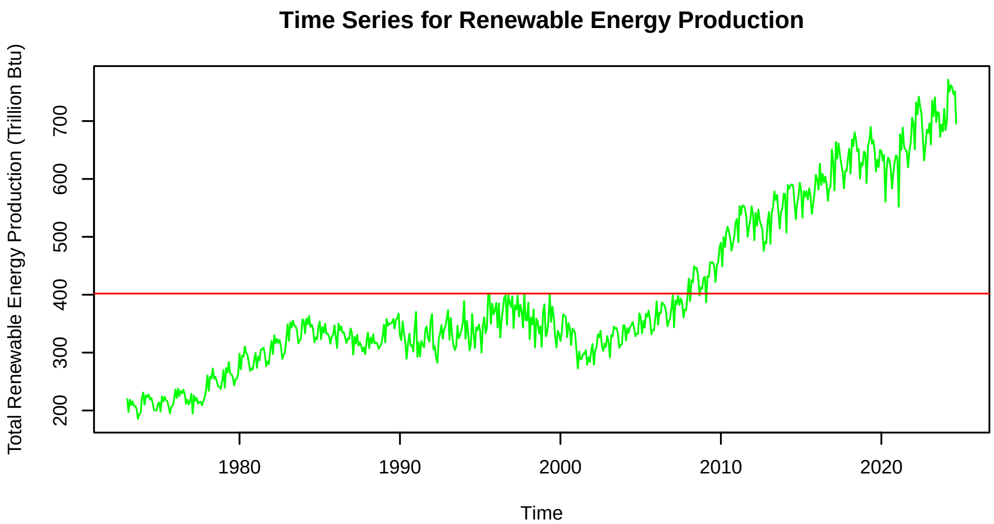
Question 4

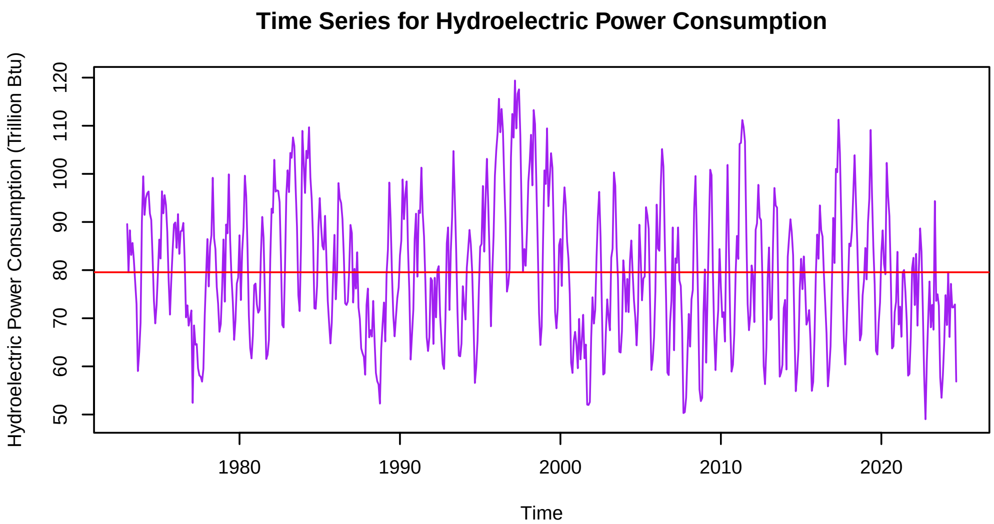
Display and interpret the time series plot for each of these variables. Try to make your plot as informative as possible by writing titles, labels, etc. For each plot add a horizontal line at the mean of each series in a different color.

```
par(mfrow = c(3,1))
plot(ts_power_data[,1],
     type="l",col="blue",
     ylab="Total Biomass Energy Production (Trillion Btu)",
     main="Time Series for Biomass Energy Production")
abline(h=mean(ts_power_data[,1]),col="red")

plot(ts_power_data[,2],
```

```
     type="l",col="green",
     ylab="Total Renewable Energy Production (Trillion Btu)",
     main="Time Series for Renewable Energy Production")
abline(h=mean(ts_power_data[,2]),col="red")

plot(ts_power_data[,3],
     type="l",col="purple",
     ylab="Hydroelectric Power Consumption (Trillion Btu)",
     main="Time Series for Hydroelectric Power Consumption")
abline(h=mean(ts_power_data[,3]),col="red")
```


Time Series for Biomass Energy Production
Total Biomass Energy Production (Trillion Btu)
150
250
350
450
1980
1990
2000
2010
2020
Time


Time Series for Renewable Energy Production
Total Renewable Energy Production (Trillion Btu)
200
300
400
500
600
700
1980
1990
2000
2010
2020
Time


Time Series for Hydroelectric Power Consumption
Hydroelectric Power Consumption (Trillion Btu)
50
60
70
80
90
100
110
120
1980
1990
2000
2010
2020
Time

Question 5

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

```
cor(ts_power_data)
```

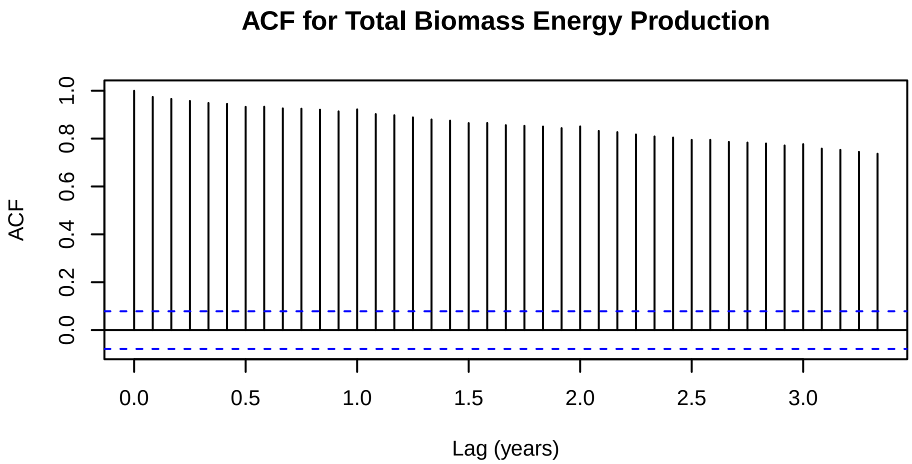
```
##                                   Total Biomass Energy Production
## Total Biomass Energy Production                         1.0000000
## Total Renewable Energy Production                       0.9678137
## Hydroelectric Power Consumption                        -0.1142927
##                  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

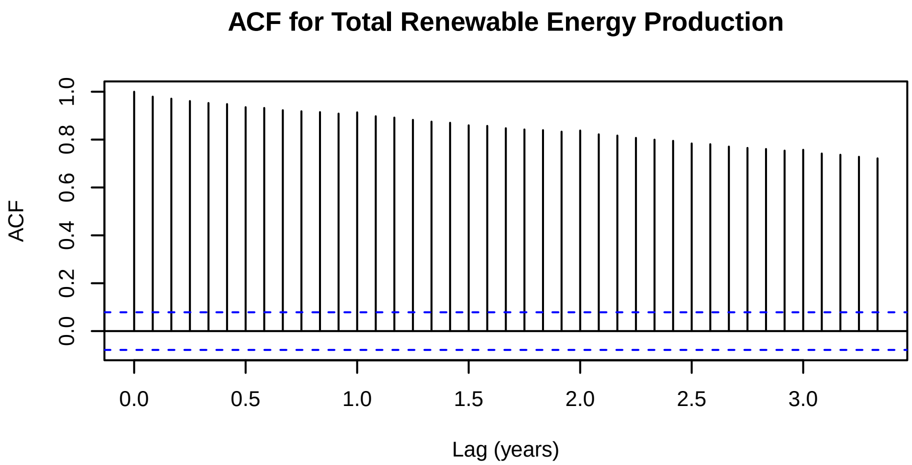
Only total biomass energy and renewable energy productions are positively correlated, as their correlation coefficient is 0.968, close to 1. Hydroelectric power consumption is not correlated to any of the other two variables since the correlation coefficients are close to 0, at -0.114 and -0.029 respectively.

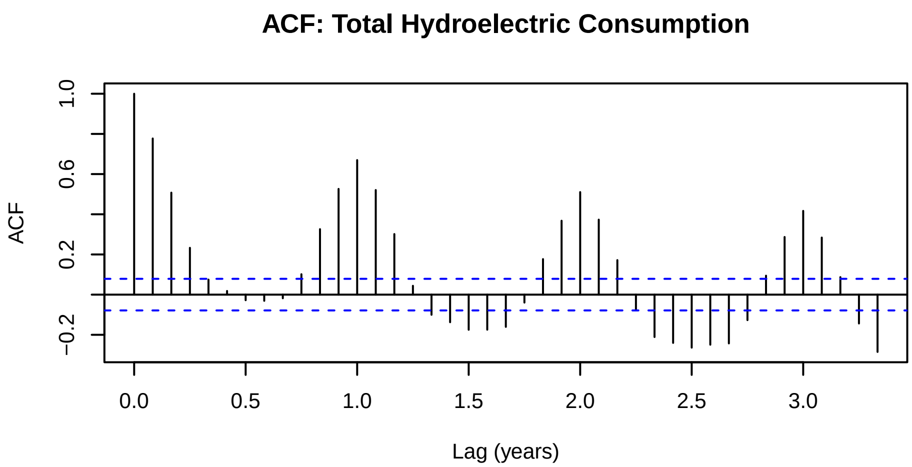
Question 6

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?

```
par(mfrow = c(3,1))
biomass_acf<-acf(ts_power_data[,1],lag.max=40,
                main = "ACF for Total Biomass Energy Production",
                type="correlation", plot=TRUE,
                xlab="Lag (years)")
renewable_acf<-acf(ts_power_data[,2],lag.max=40,
                  main = "ACF for Total Renewable Energy Production",
                  type="correlation", plot=TRUE,
                  xlab="Lag (years)")
hydro_acf<-acf(ts_power_data[,3],lag.max=40,
              main = "ACF: Total Hydroelectric Consumption",
              type="correlation", plot=TRUE,
              xlab="Lag (years)")
```







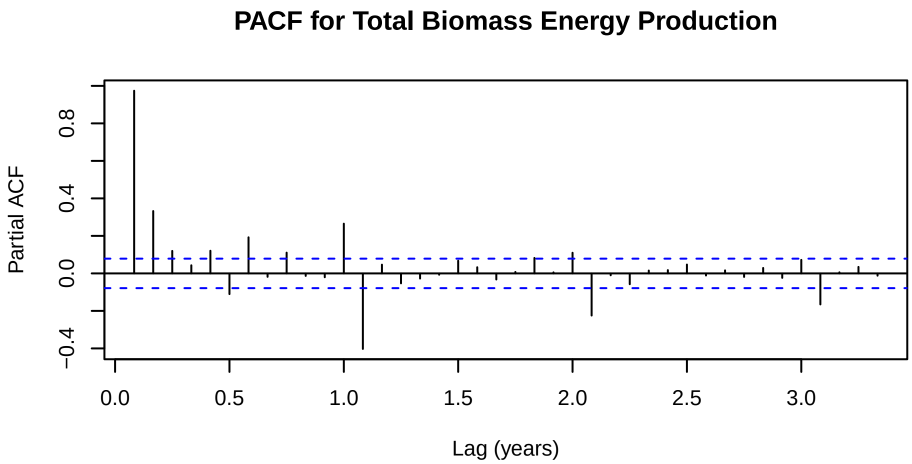
For both the biomass and renewable energy consumption, the ACF values decrease gradually as lags increase, indicating a strong trend. However, it does not have the sinusoidal pattern, suggesting that no significant seasonality is present.

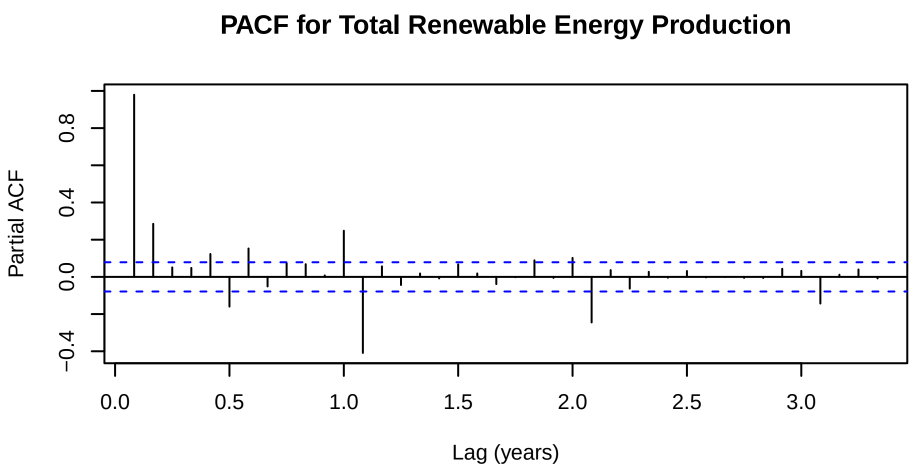
For the hydroelectric consumption, the ACF plot shows a sinusoidal pattern with a periodicity of 12 lags, indicating a seasonal behavior. However, the positive sinusoidal wave has stronger autocorrelation than the negative side.

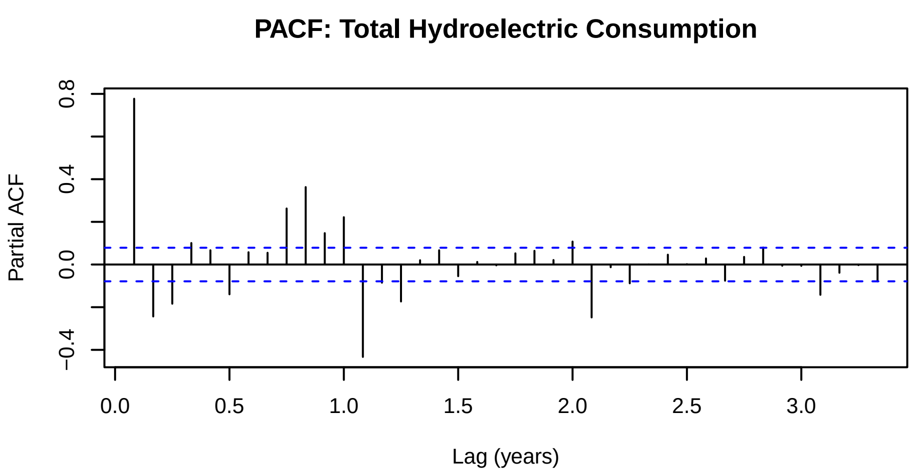
Question 7

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

```
par(mfrow = c(3,1))
biomass_pacf<-pacf(ts_power_data[,1],lag.max=40,
                main = "PACF for Total Biomass Energy Production",
                plot=TRUE,xlab="Lag (years)")
renewable_pacf<-pacf(ts_power_data[,2],lag.max=40,
                  main = "PACF for Total Renewable Energy Production",
                  plot=TRUE,xlab="Lag (years)")
hydro_pacf<-pacf(ts_power_data[,3],lag.max=40,
              main = "PACF: Total Hydroelectric Consumption",
              plot=TRUE,xlab="Lag (years)")
```







All PACF plots shows seasonality. For total hydroelectric consumption, the seasonality is expected as the shape of the PACF plot is similar to its ACF plot. However, for the other two time series, after removing the effects of other lags, there is significant ACF values at the lag value of 12 months (or 1 year), which is different from the ACF plot.