ENV 790.30 - Time Series Analysis for Energy Data | Spring 2025
Assignment 2 - Due date 01/27/26

Si Min (Lucy) Loo

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_Sp26.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(readxl)
library(openxlsx)
library(lubridate)
```

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

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

```r
library(ggplot2)
library(forecast)
```

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

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 2025 Monthly Energy Review. The spreadsheet is ready to be used. Refer to the file "M2_ImportingData_XLSX.Rmd" in our Lessons folder for instructions on how to read *.xlsx* files.

```r
#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`
```

```r
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`
```

```r
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>, ...
```
read_col_names

```
## # A tibble: 1 x 14
##   ...1  ...2   ...3  ...4  ...5  ...6  ...7  ...8  ...9  ...10 ...11 ...12 ...13
##   <chr> <chr>  <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr> <chr>
## 1 Month Wood ~ Biof~ Tota~ Tota~ Hydr~ Geot~ Sola~ Wind~ Wood~ Wast~ Biof~ Tota~
## # i 1 more variable: ...14 <chr>
```

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
df <- data.frame(
  Month = energy_data$`Month`,
  Biomass = energy_data$`Total Biomass Energy Production`,
  Renewable = energy_data$`Total Renewable Energy Production`,
  Hydroelectric = energy_data$`Hydroelectric Power Consumption`
)

head(df)
```

```
##        Month Biomass Renewable Hydroelectric
## 1 1973-01-01 129.787   219.839        89.562
## 2 1973-02-01 117.338   197.330        79.544
## 3 1973-03-01 129.938   218.686        88.284
## 4 1973-04-01 125.636   209.330        83.152
## 5 1973-05-01 129.834   215.982        85.643
## 6 1973-06-01 125.611   208.249        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().

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

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

Question 3

Compute mean and standard deviation for these three series.

```r
# Means
mean_biomass <- mean(ts_df[, "Biomass"], na.rm = TRUE)
mean_renew   <- mean(ts_df[, "Renewable"], na.rm = TRUE)
mean_hydro   <- mean(ts_df[, "Hydroelectric"], na.rm = TRUE)
```

3

```r
# Standard deviations
sd_biomass <- sd(ts_df[, "Biomass"], na.rm = TRUE)
sd_renew   <- sd(ts_df[, "Renewable"], na.rm = TRUE)
sd_hydro   <- sd(ts_df[, "Hydroelectric"], na.rm = TRUE)
```
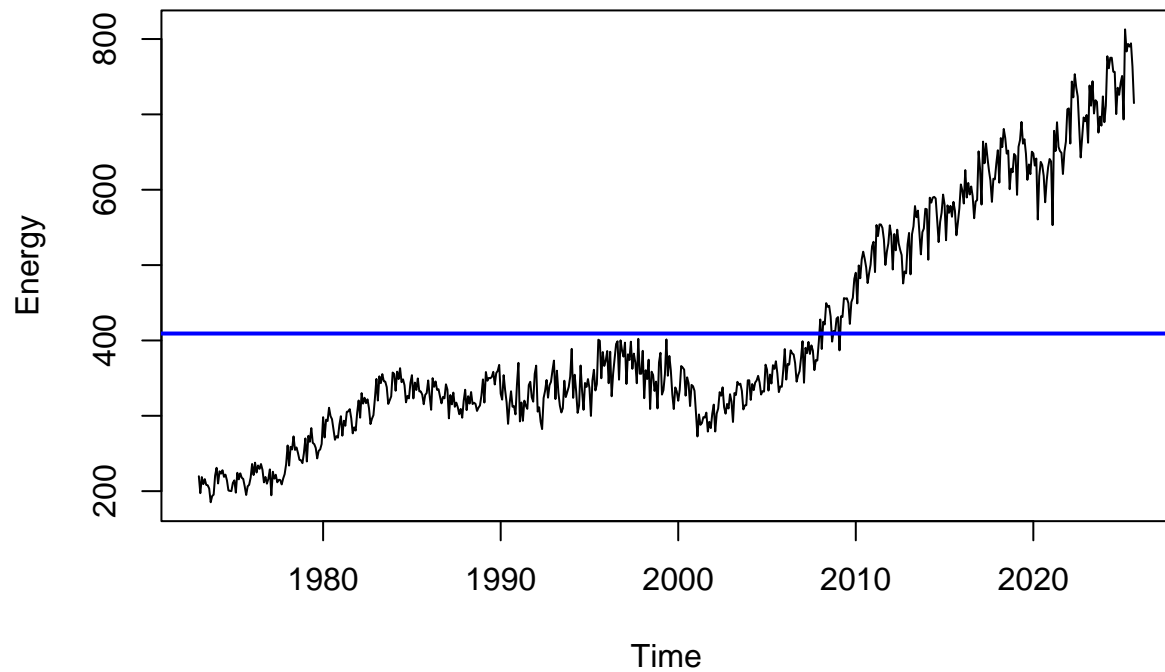
Question 4

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.

```r
# Biomass
plot(ts_df[, "Biomass"],
     main = "Total Biomass Energy Production",
     ylab = "Energy",
     xlab = "Time")
abline(h = mean_biomass, col = "red", lwd = 2)
```
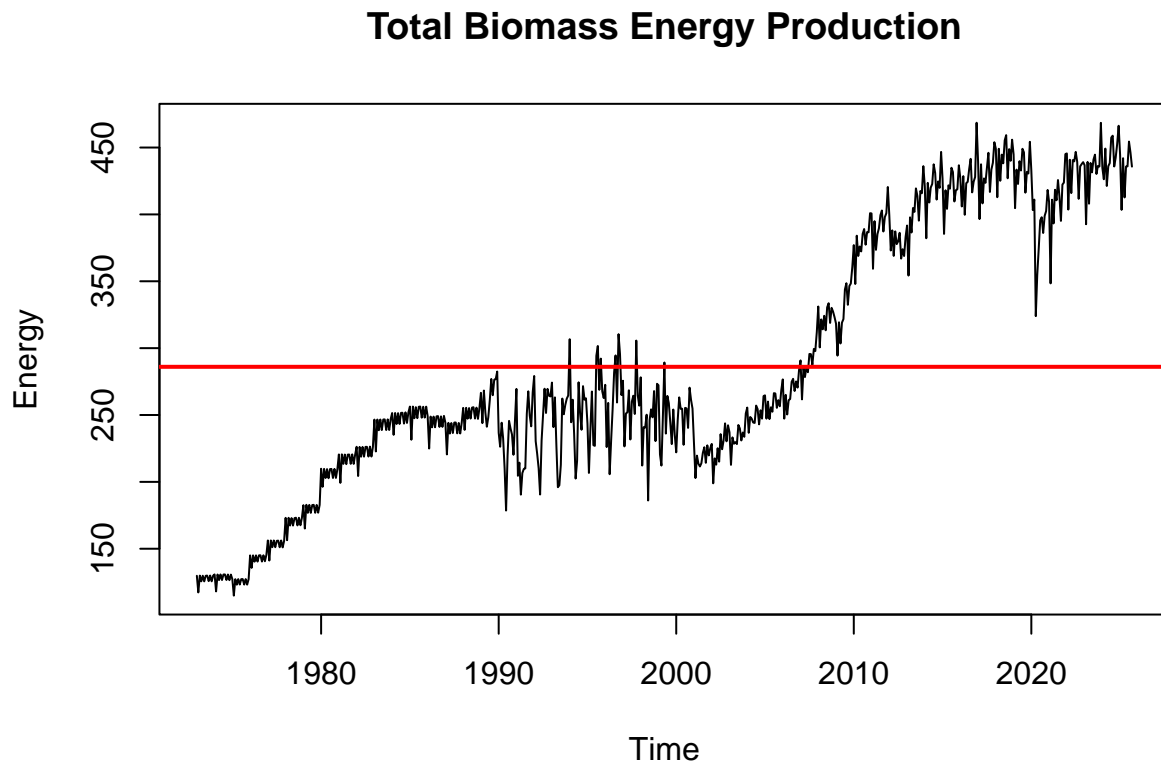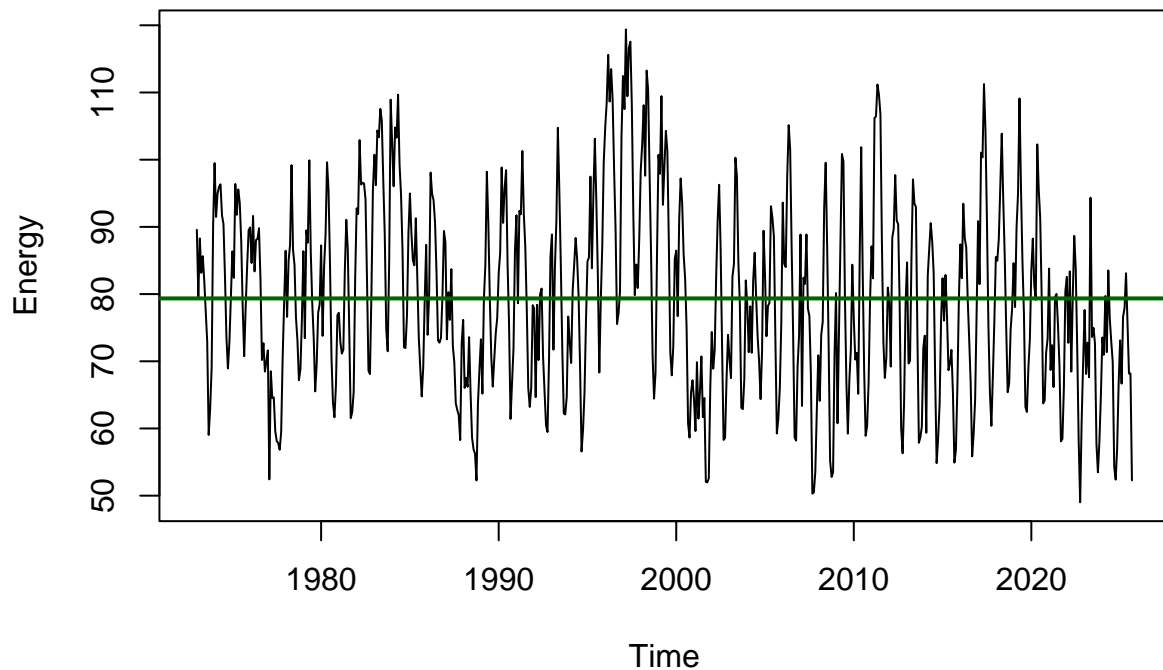


```r
# Renewable
plot(ts_df[, "Renewable"],
     main = "Total Renewable Energy Production",
     ylab = "Energy",
     xlab = "Time")
abline(h = mean_renew, col = "blue", lwd = 2)
```

Total Renewable Energy Production


```r
# Hydroelectric
plot(ts_df[, "Hydroelectric"],
     main = "Hydroelectric Power Consumption",
     ylab = "Energy",
     xlab = "Time")
abline(h = mean_hydro, col = "darkgreen", lwd = 2)
```

Hydroelectric Power Consumption

Question 5

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

```
cor(ts_df)
```

```
##                 Biomass    Renewable Hydroelectric
## Biomass       1.0000000  0.96529851   -0.13473742
## Renewable     0.9652985  1.00000000   -0.05842436
## Hydroelectric -0.1347374 -0.05842436    1.00000000
```

Biomass and Renewable are significantly positively correlated with each other. This is expected since biomass is a major component of total renewable energy. The correlations involving hydroelectric power consumption are relatively weak, suggesting that hydroelectric generation follows different drivers such as weather patterns and reservoir conditions rather than long-term renewable expansion trends.

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?

```
Acf(ts_df[, "Biomass"], lag.max = 40, main = "ACF: Biomass Energy Production")
```
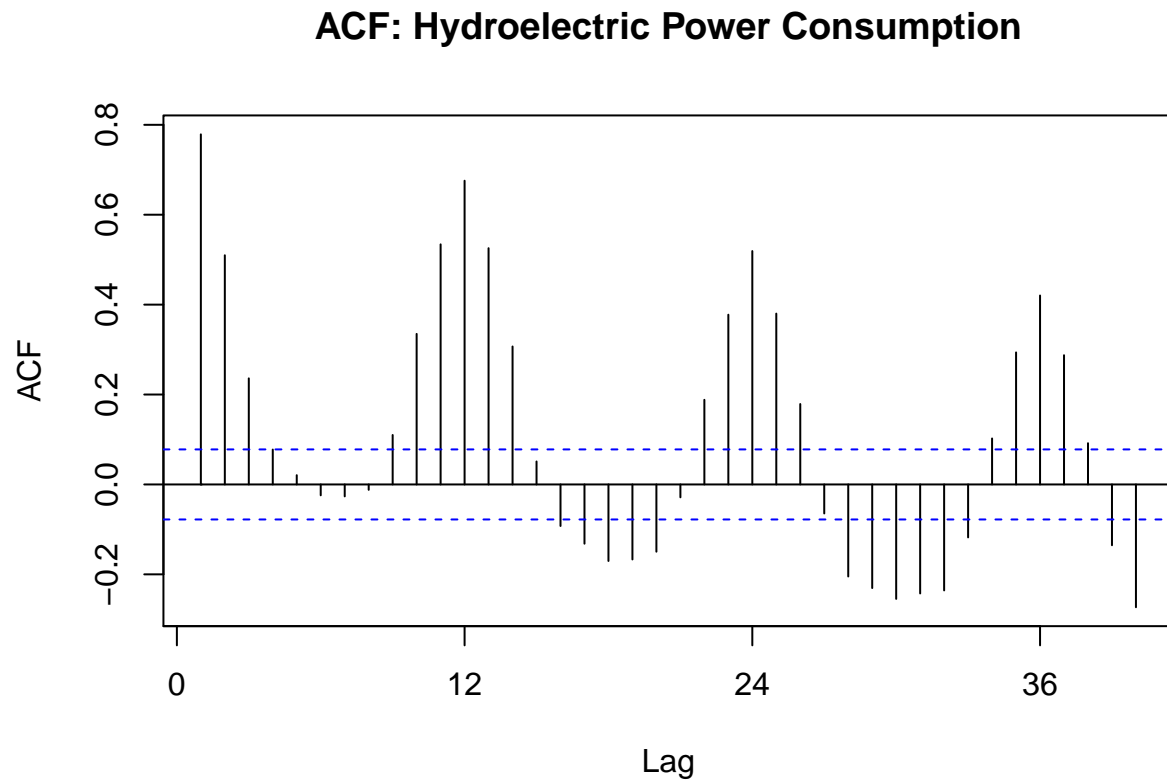
ACF: Biomass Energy Production
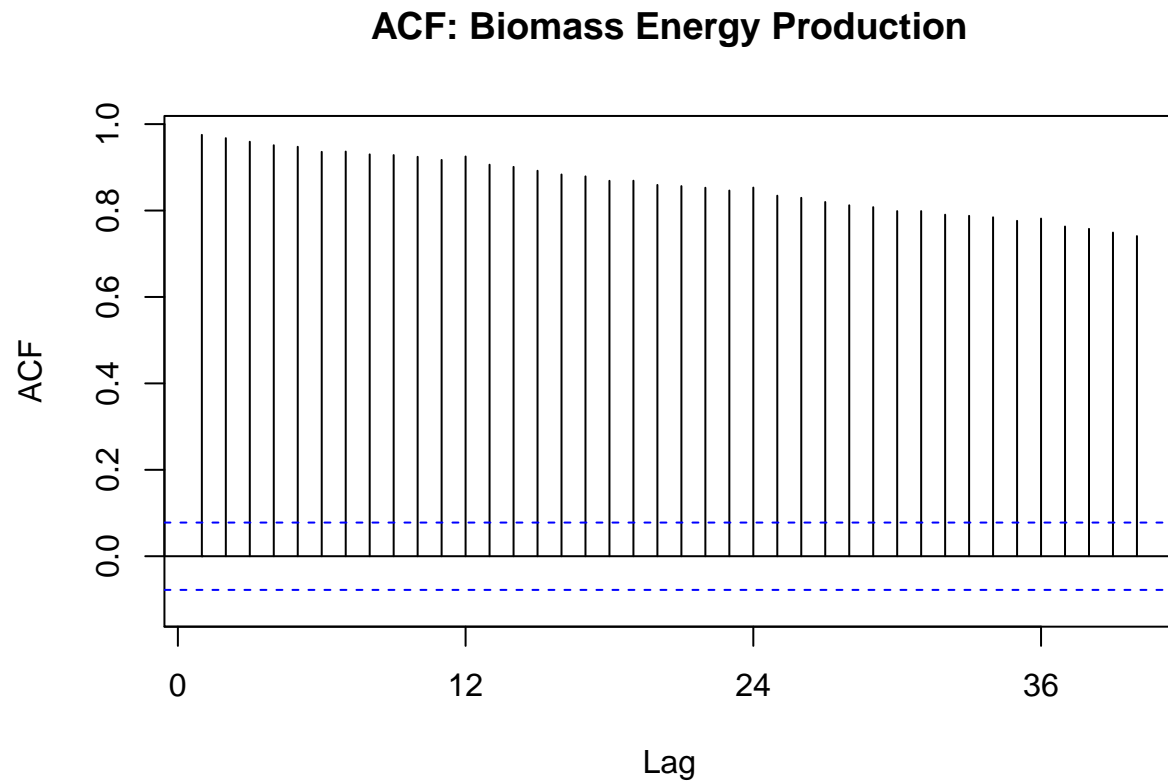

```
Acf(ts_df[, "Renewable"], lag.max = 40, main = "ACF: Total Renewable Energy Production")
```

ACF: Total Renewable Energy Production
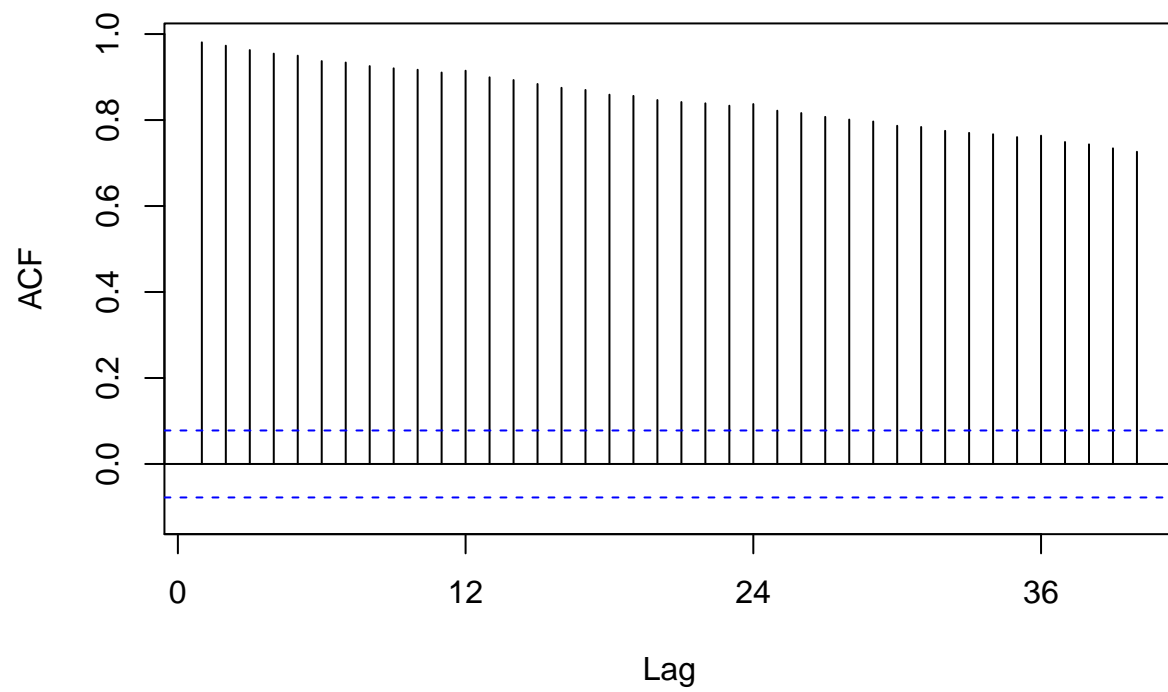

```
Acf(ts_df[, "Hydroelectric"], lag.max = 40, main = "ACF: Hydroelectric Power Consumption")
```

ACF: Hydroelectric Power Consumption

Biomass and Total Renewable Energy Production show very slow decay in the ACF, with significant autocorrelations persisting even at high lags. This suggests strong trend components and likely non-stationarity, meaning that current biomass levels are very dependent on past values.

Hydroelectric power consumption also shows autocorrelation, but with a more pronounced seasonal pattern at around 12, 24 and 36 months and alternating positive and negative correlations at intermediate lags.

The three series do not have identical behavior: biomass and renewables are dominated by long-term growth trends, while hydroelectric consumption is more strongly influenced by seasonal dynamics.

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?
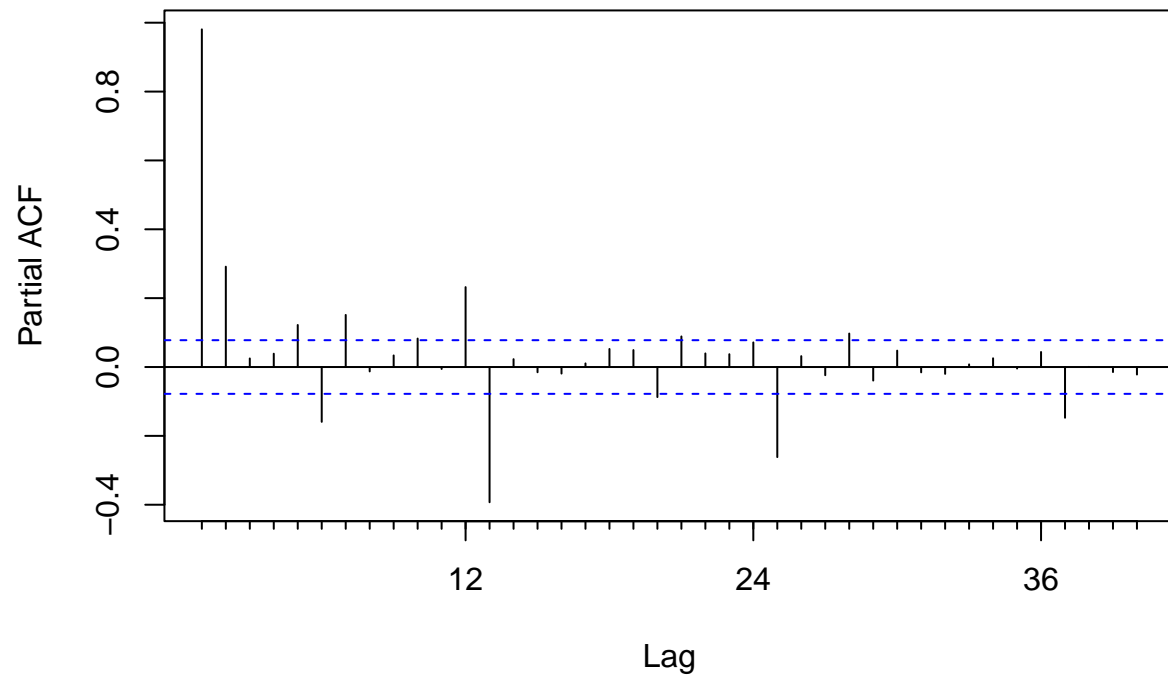
```
Pacf(ts_df[, "Biomass"], lag.max = 40, main = "PACF: Biomass Energy Production")
```
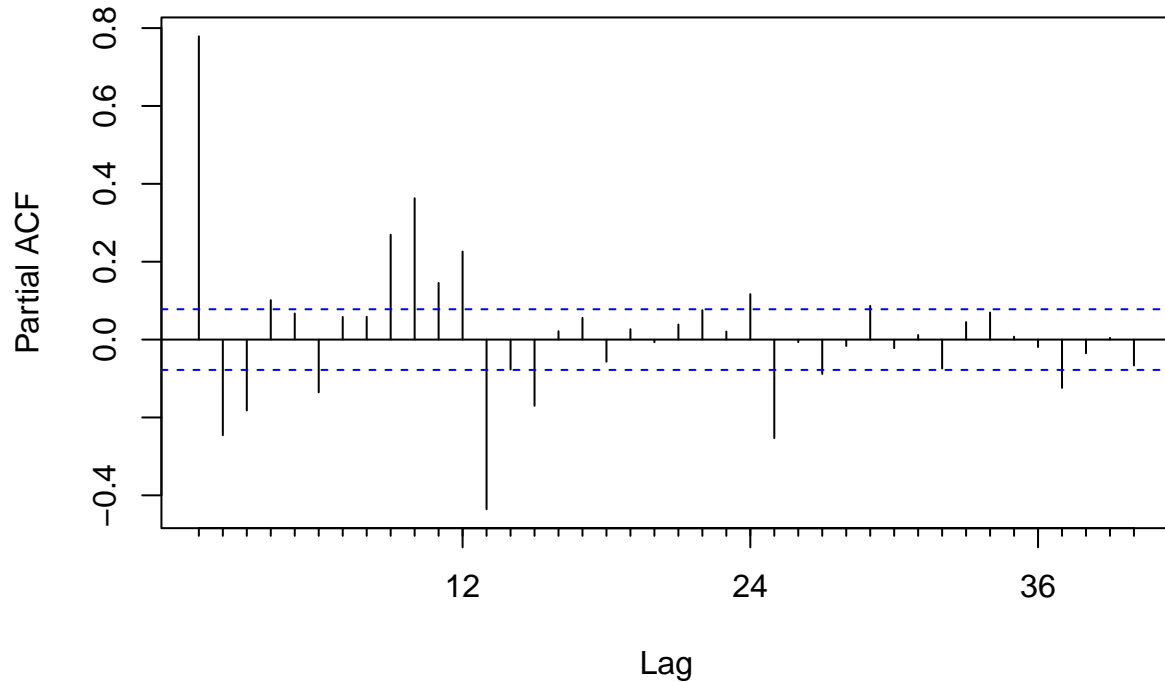
PACF: Biomass Energy Production
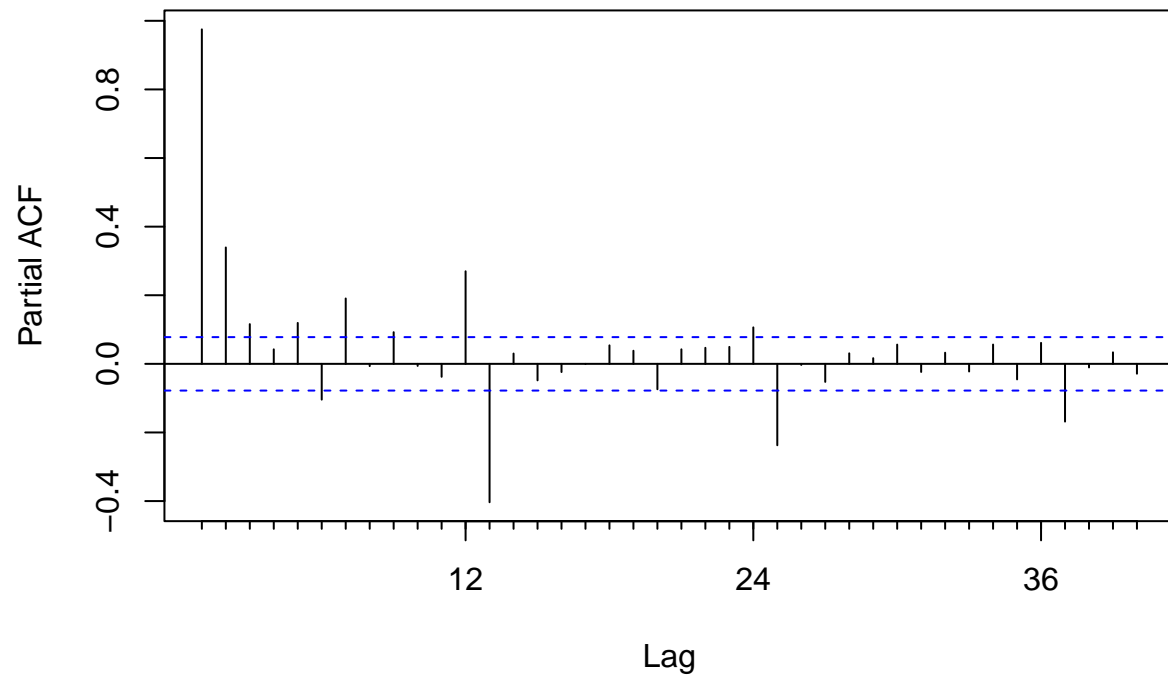

```
Pacf(ts_df[, "Renewable"], lag.max = 40, main = "PACF: Total Renewable Energy Production")
```

PACF: Total Renewable Energy Production


```
Pacf(ts_df[, "Hydroelectric"], lag.max = 40, main = "PACF: Hydroelectric Power Consumption")
```

PACF: Hydroelectric Power Consumption

The PACF for biomass shows a very large and significant spike at lag 1, followed by mostly insignificant partial autocorrelations at higher lags (with a few isolated exceptions). This pattern suggests that, once the immediate past value is accounted for, additional lags contribute little explanatory power. The PACF for total renewable energy production closely mirrors that of biomass. For both biomass and total renewable energy production, the ACF shows very high autocorrelation across all lags with slow decay, indicating strong persistence and non-stationarity. In contrast, the PACF shows a dominant spike at lag 1, with most higher-order lags becoming insignificant.

The hydroelectric power consumption shows multiple significant spikes, particularly around seasonal lags.This indicates that hydroelectric consumption depends not only on the immediate past but also on seasonal cycles. The ACF exhibits an oscillating, repeating pattern, with significant spikes at seasonal lags (12, 24, 36), indicating strong annual seasonality, and the PACF does not cut off cleanly, instead showing multiple significant spikes at both short and seasonal lags.This contrast suggests that hydroelectric consumption depends on both recent past values and seasonal cycles.