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

Assignment 2 - Due date 01/28/25

Mazhar Bhuyan

Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sour

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
install.packages("forecast")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
install.packages("tseries")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```
install.packages("dplyr")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```r
install.packages("readxl")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```r
install.packages("openxlsx")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

```r
install.packages("ggplot2")
```

```
## Installing package into '/home/guest/R/x86_64-pc-linux-gnu-library/4.4'
## (as 'lib' is unspecified)
```

Loading packages and initializing

It's useful to designate one code chunk to load packages on the beginning of the file. You can always add to this chunk as needed. But concentrate the packages needed on only one chunk.

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
getwd()
```

```
## [1] "/home/guest/TSA_Mazhar/Time Series_Mazhar"
```

```r
Energy_Data <- read_excel("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx
```

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

```r
head(Energy_Data)
```

```
## # A tibble: 6 x 14
##   ...1                 ...2 ...3        ...4  ...5  ...6  ...7 ...8  ...9  ...10
##   <dttm>              <dbl> <chr>      <dbl> <dbl> <dbl> <dbl> <chr> <chr> <dbl>
## 1 1973-01-01 00:00:00  130. Not Avail~  130.  220.  89.6 0.49  Not ~ Not ~  130.
## 2 1973-02-01 00:00:00  117. Not Avail~  117.  197.  79.5 0.448 Not ~ Not ~  117.
## 3 1973-03-01 00:00:00  130. Not Avail~  130.  219.  88.3 0.464 Not ~ Not ~  130.
## 4 1973-04-01 00:00:00  125. Not Avail~  126.  209.  83.2 0.542 Not ~ Not ~  125.
## 5 1973-05-01 00:00:00  130. Not Avail~  130.  216.  85.6 0.505 Not ~ Not ~  130.
## 6 1973-06-01 00:00:00  125. Not Avail~  126.  208.  82.1 0.579 Not ~ Not ~  125.
## # i 4 more variables: ...11 <dbl>, ...12 <chr>, ...13 <dbl>, ...14 <dbl>
```

```r
read_col_names <- read_excel("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xl
```

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

```r
head(Energy_Data)
```

```
## # A tibble: 6 x 14
##   ...1                 ...2 ...3        ...4  ...5  ...6  ...7 ...8  ...9  ...10
##   <dttm>              <dbl> <chr>      <dbl> <dbl> <dbl> <dbl> <chr> <chr> <dbl>
## 1 1973-01-01 00:00:00  130. Not Avail~  130.  220.  89.6 0.49  Not ~ Not ~  130.
## 2 1973-02-01 00:00:00  117. Not Avail~  117.  197.  79.5 0.448 Not ~ Not ~  117.
## 3 1973-03-01 00:00:00  130. Not Avail~  130.  219.  88.3 0.464 Not ~ Not ~  130.
## 4 1973-04-01 00:00:00  125. Not Avail~  126.  209.  83.2 0.542 Not ~ Not ~  125.
## 5 1973-05-01 00:00:00  130. Not Avail~  130.  216.  85.6 0.505 Not ~ Not ~  130.
## 6 1973-06-01 00:00:00  125. Not Avail~  126.  208.  82.1 0.579 Not ~ Not ~  125.
## # i 4 more variables: ...11 <dbl>, ...12 <chr>, ...13 <dbl>, ...14 <dbl>
```

```r
head(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>
```

```r
colnames(Energy_Data) <- read_col_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
three_data <- Energy_Data[,4:6]
head(three_data)
```

```
## # 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().

```r
Energy_Data_Subset_ts <- ts(three_data, start = c(1973, 1), frequency = 12)
head(Energy_Data_Subset_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.

```
mean_biomass <- mean(Energy_Data_Subset_ts[, 1])
mean_biomass
```

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

```
sd_biomass <- sd(Energy_Data_Subset_ts[,1])
sd_biomass
```

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

```
mean_renewable <- mean(Energy_Data_Subset_ts[,2])
mean_renewable
```

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

```
sd_renewable <- sd(Energy_Data_Subset_ts[,2])
sd_renewable
```

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

```
mean_hydro <- mean(Energy_Data_Subset_ts[,3])
mean_hydro
```

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

```
sd_hydro <- sd(Energy_Data_Subset_ts[,3])
sd_hydro
```

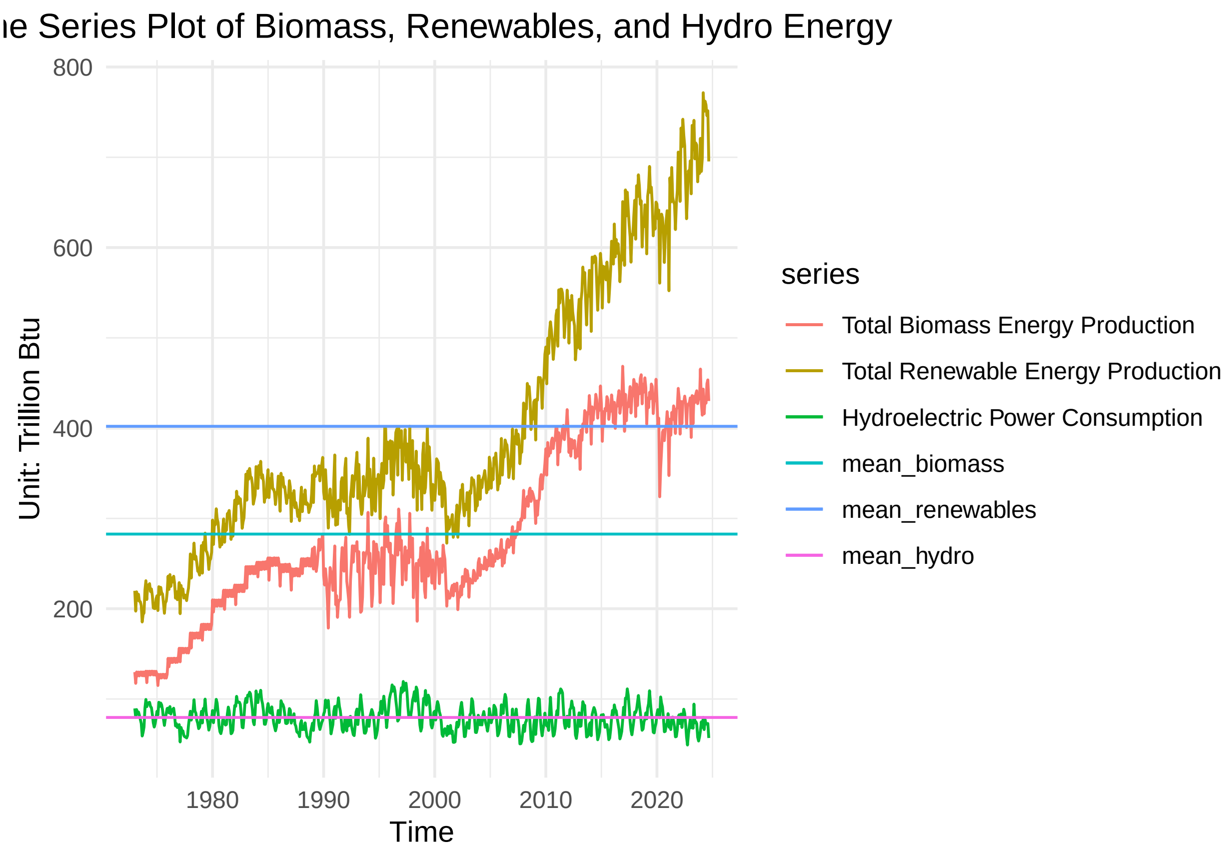
```
## [1] 14.10737
```

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_Data_Subset_ts) +
  xlab("Time") +
  ylab("Unit: Trillion Btu") +
  ggtitle("Time Series Plot of Biomass, Renewables, and Hydro Energy")+
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5)) +
  # Add horizontal lines at the mean of each series
  geom_hline(aes(yintercept = mean(Energy_Data_Subset_ts[, 1]), color = "mean_biomass"),
```

```r
            linetype = "solid", show.legend = TRUE) +
  geom_hline(aes(yintercept = mean(Energy_Data_Subset_ts[, 2]), color = "mean_renewables"),
            linetype = "solid", show.legend = TRUE) +
  geom_hline(aes(yintercept = mean(Energy_Data_Subset_ts[, 3]), color = "mean_hydro"),
            linetype = "solid", show.legend = TRUE)
```



Question 5

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

```r
#Extracting the data to create 3 columns of data

biofuel <- Energy_Data_Subset_ts[, 1]
renewables <- Energy_Data_Subset_ts[, 2]
hydro <- Energy_Data_Subset_ts[, 3]

#Converting those to df
data <- data.frame(Biofuel = biofuel, Renewables = renewables, Hydro = hydro)
#Correlation of data
correlation_matrix <- cor(data)
print("Correlation Matrix:")
```

```
## [1] "Correlation Matrix:"
```

```r
print(correlation_matrix)
```

```
##               Biofuel  Renewables       Hydro
## Biofuel     1.0000000  0.96781371 -0.11429266
## Renewables  0.9678137  1.00000000 -0.02916103
## Hydro      -0.1142927 -0.02916103  1.00000000
```

```r
#Biofuel and Renewables: Correlation = 0.9678 strong positve correlation

#Biofuel and Hydro:Correlation = -0.1143 weak negative correlation

#Renewables and Hydro:Correlation = -0.0292 weak negative correlation


#Significance Test
cat("\nBiofuel vs. Renewables Correlation Test:\n")
```

```
##
## Biofuel vs. Renewables Correlation Test:
```

```r
print(cor.test(biofuel, renewables))
```

```
##
##  Pearson's product-moment correlation
##
## data:  biofuel and renewables
## t = 95.677, df = 619, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.9624198 0.9724443
## sample estimates:
##       cor
## 0.9678137
```

```r
cat("\nBiofuel vs. Hydro Correlation Test:\n")
```

```
##
## Biofuel vs. Hydro Correlation Test:
```

```r
print(cor.test(biofuel, hydro))
```

```
##
##  Pearson's product-moment correlation
##
## data:  biofuel and hydro
## t = -2.8623, df = 619, p-value = 0.004348
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.19125123 -0.03593747
## sample estimates:
##        cor
## -0.1142927
```

```r
cat("\nRenewables vs. Hydro Correlation Test:\n")
```

```
##
## Renewables vs. Hydro Correlation Test:
```

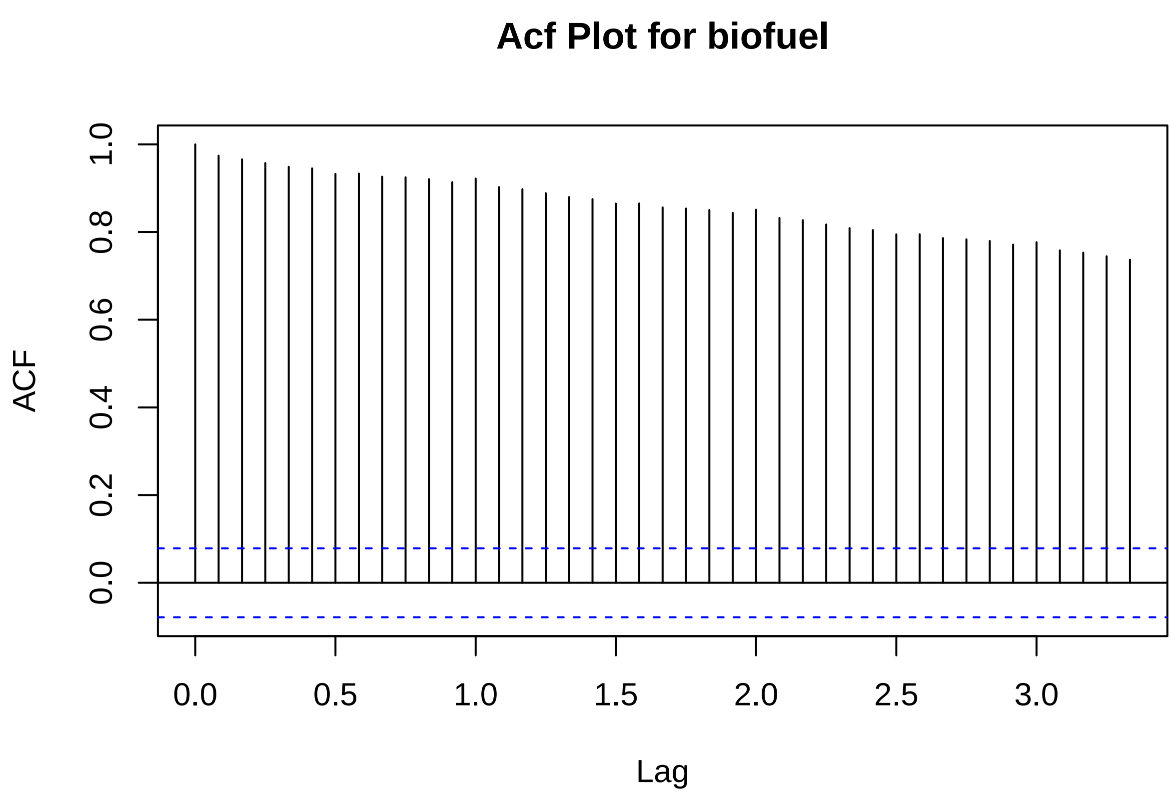
```r
print(cor.test(renewables, hydro))
```

```
##
##  Pearson's product-moment correlation
##
## data:  renewables and hydro
## t = -0.72583, df = 619, p-value = 0.4682
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.1075925  0.0496312
## sample estimates:
##         cor
## -0.02916103
```

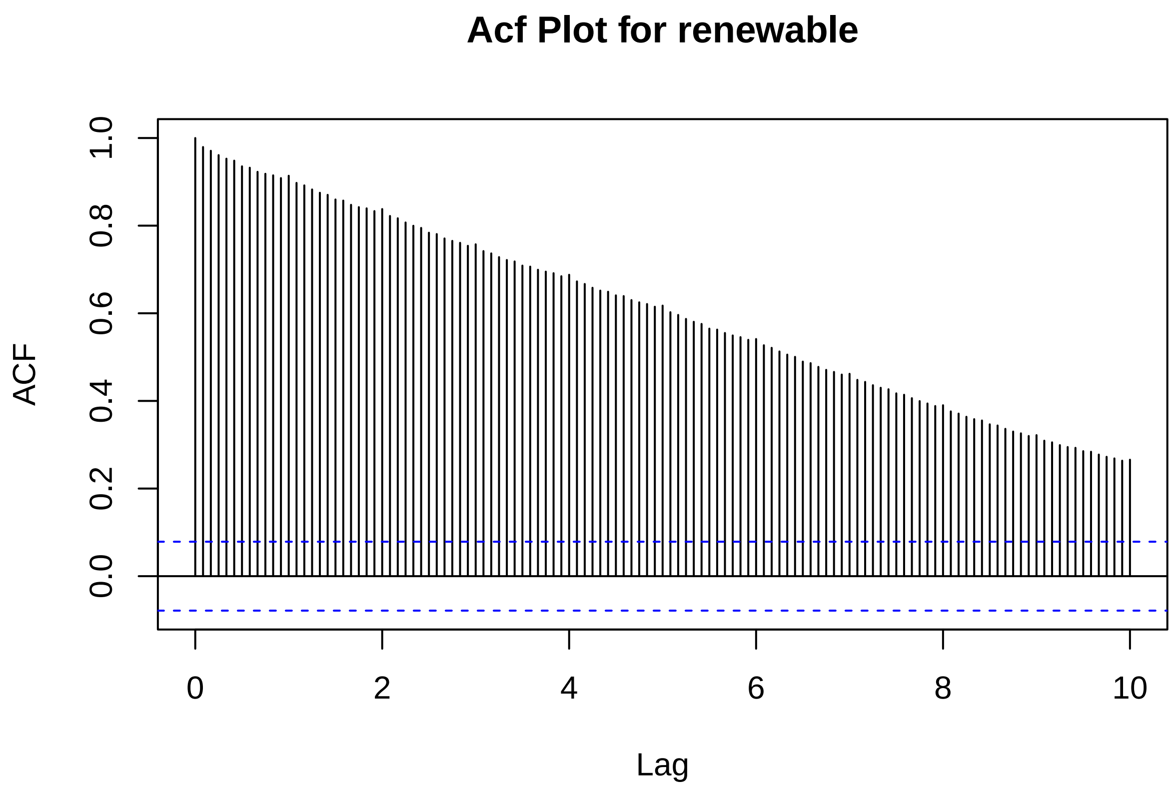
Question 6

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?

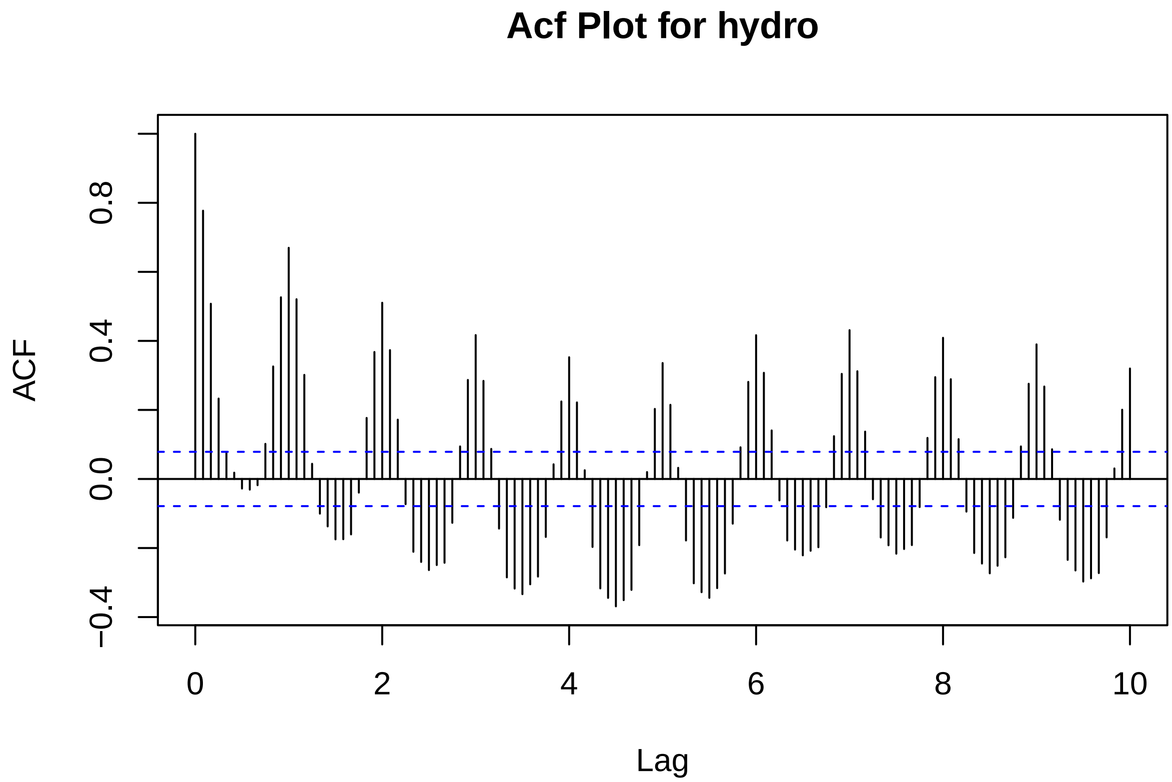
```r
acf_biofuel = acf(Energy_Data_Subset_ts[,1], lag=40, main = "Acf Plot for biofuel")
```



```r
acf_renewable = acf(Energy_Data_Subset_ts[,2], lag=120, main = "Acf Plot for renewable")
```




```r
acf_hydro = acf(Energy_Data_Subset_ts[,3], lag=120, main = "Acf Plot for hydro")
```



#Acf Plots: #Biofuel: The plot shows strong autocorrealtion and a decaying over time. The result is significant since all plots are quite larger than the p-value

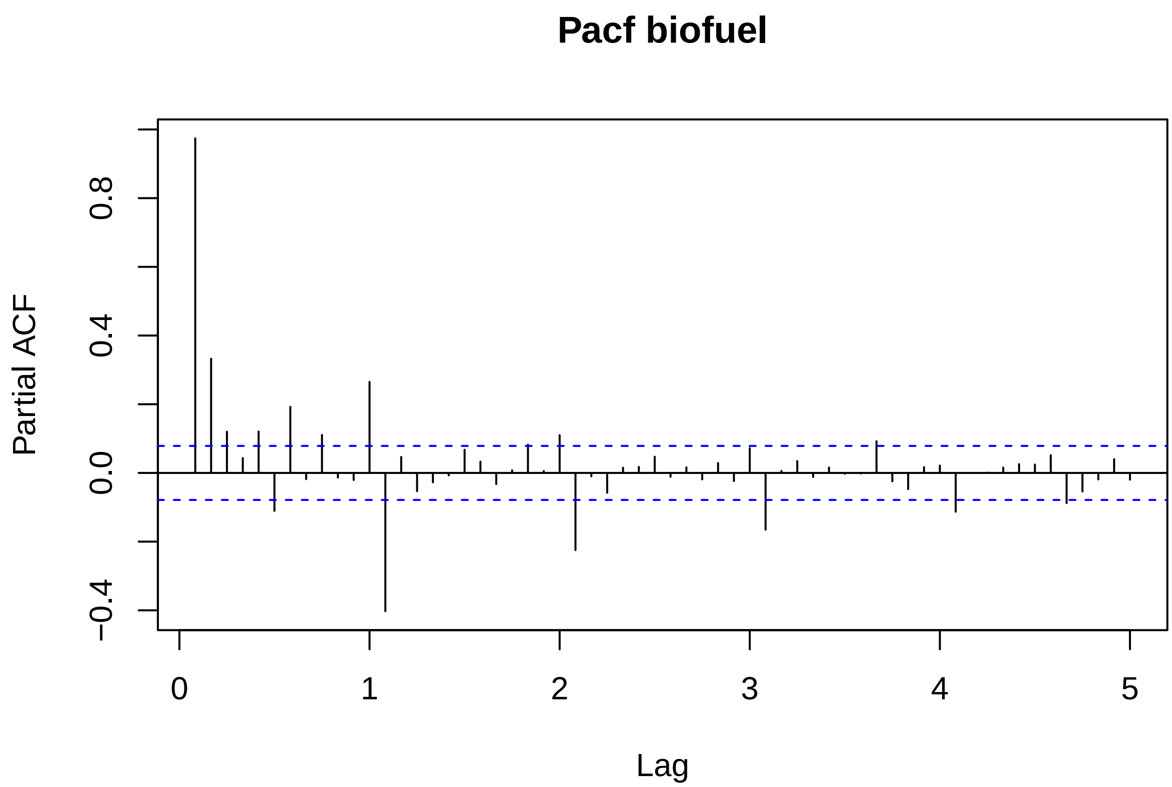
```
#Renewable: Similar to biofuel this plot also shows strong autocorrelation that is         rapidly decre
```

```
#Hydro: The plot shows seasonality. It is expected because this related to rainfall,       drought
```

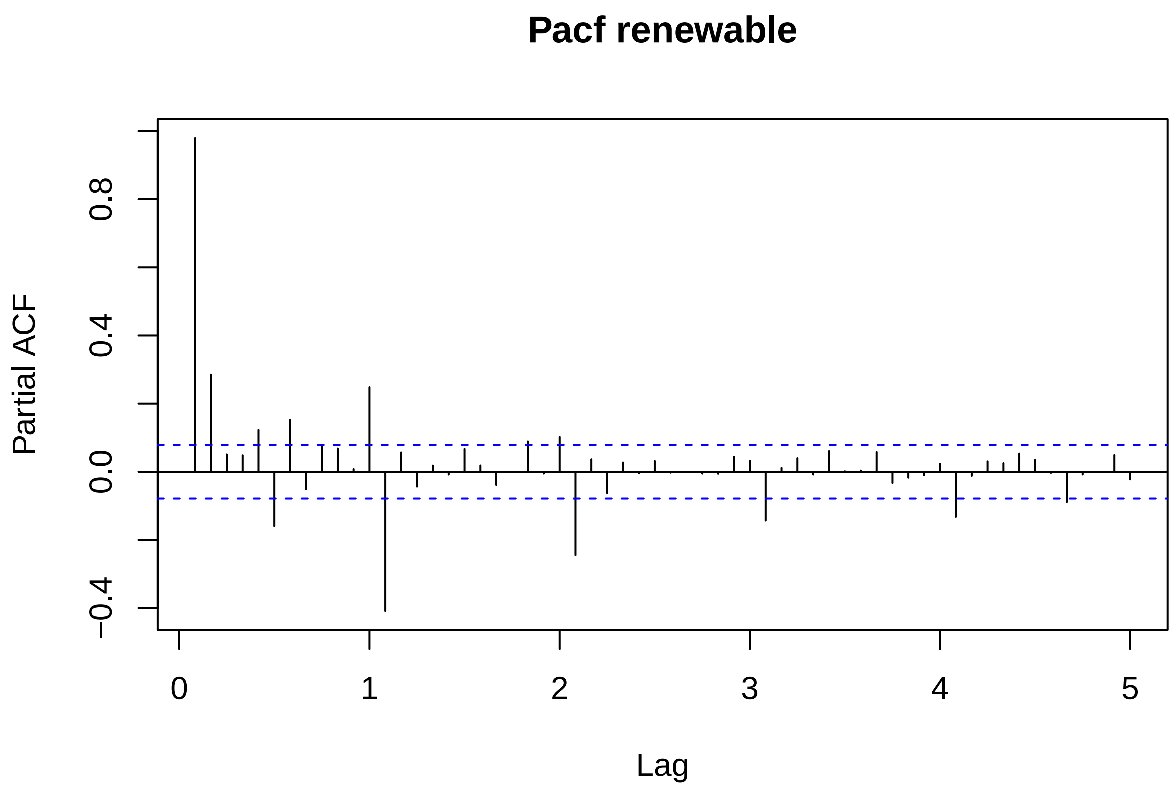
Question 7

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

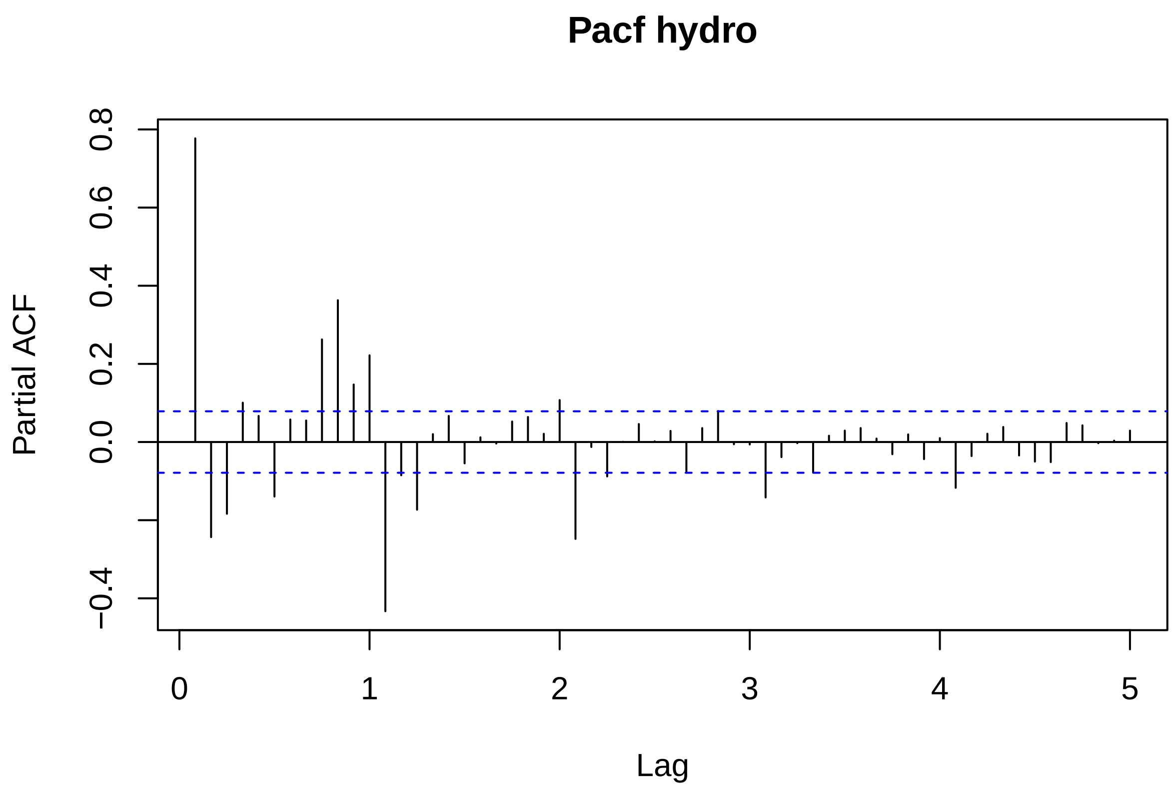
```
pacf_biofuel = pacf(Energy_Data_Subset_ts[,1], lag=60, main = "Pacf biofuel")
```




```
pacf_renewable = pacf(Energy_Data_Subset_ts[,2], lag=60, main = "Pacf renewable")
```




```
pacf_hydor = pacf(Energy_Data_Subset_ts[,3], lag=60, main = "Pacf hydro")
```



#Pacf Plots:

#Biofuel: The plot shows significant partial auto correlation at lag 1 and then kept declining. It also suggests that the time series is influenced by its past value. The plot also shows that most of the spikes are not significant.

#Renewable: Large spikes at lag 1. Most of the values are not significant.

#Hydro: Pacf at lag 1 is significant but continues to decay over time.