ENV 790.30 - Time Series Analysis for Energy Data | Spring 2021

Assignment 2 - Due date 02/05/21

Stefan Chen

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 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. Rename the pdf file such that it includes your first and last name (e.g., "LuanaLima_TSA_A02_Sp21.Rmd"). 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.\

Data set information

Consider the data provided in the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx" on our **Data** folder. The data comes from the US Energy Information and Administration and corresponds to the January 2021 Monthly Energy Review. The spreadsheet is ready to be used. Use the command *read.table*() to import the data in R or *panda.read_excel*() in Python (note that you will need to import pandas package). }

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.

Below is the first six rows generated from head() function as a confirmation of a functional data frame.

```
##   Total Biomass Energy Production Total Renewable Energy Production
## 1                         129.787                             403.981
## 2                         117.338                             360.900
## 3                         129.938                             400.161
## 4                         125.636                             380.470
## 5                         129.834                             392.141
## 6                         125.611                             377.232
##   Hydroelectric Power Consumption
## 1                         272.703
## 2                         242.199
## 3                         268.810
## 4                         253.185
## 5                         260.770
## 6                         249.859
```

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().

The time series is set to start on January of 1973 (start=c(1973,1)) and with a monthly frequency (frequency=12). Below is the first six rows of the time series.

```
##          Total Biomass Energy Production Total Renewable Energy Production
## Jan 1973                         129.787                             403.981
## Feb 1973                         117.338                             360.900
## Mar 1973                         129.938                             400.161
## Apr 1973                         125.636                             380.470
## May 1973                         129.834                             392.141
## Jun 1973                         125.611                             377.232
##          Hydroelectric Power Consumption
## Jan 1973                         272.703
## Feb 1973                         242.199
## Mar 1973                         268.810
## Apr 1973                         253.185
## May 1973                         260.770
## Jun 1973                         249.859
```

Question 3

Compute mean and standard deviation for these three series.

Total Biomass Energy Production

-Mean: 270.70

-Standard Deviation: 87.36

Total Renewable Energy Production

-Mean: 572.73

-Standard Deviation: 168.46

Hydroelectric Power Consumption

-Mean: 236.95

-Standard Deviation: 43.90

```
#c(mean, sd)
round(c(mean(ts_bio),sd(ts_bio)),2)
```

```
## [1] 270.70  87.36
```

```
round(c(mean(ts_re),sd(ts_re)),2)
```

```
## [1] 572.73 168.46
```

```
round(c(mean(ts_hydro),sd(ts_hydro)),2)
```

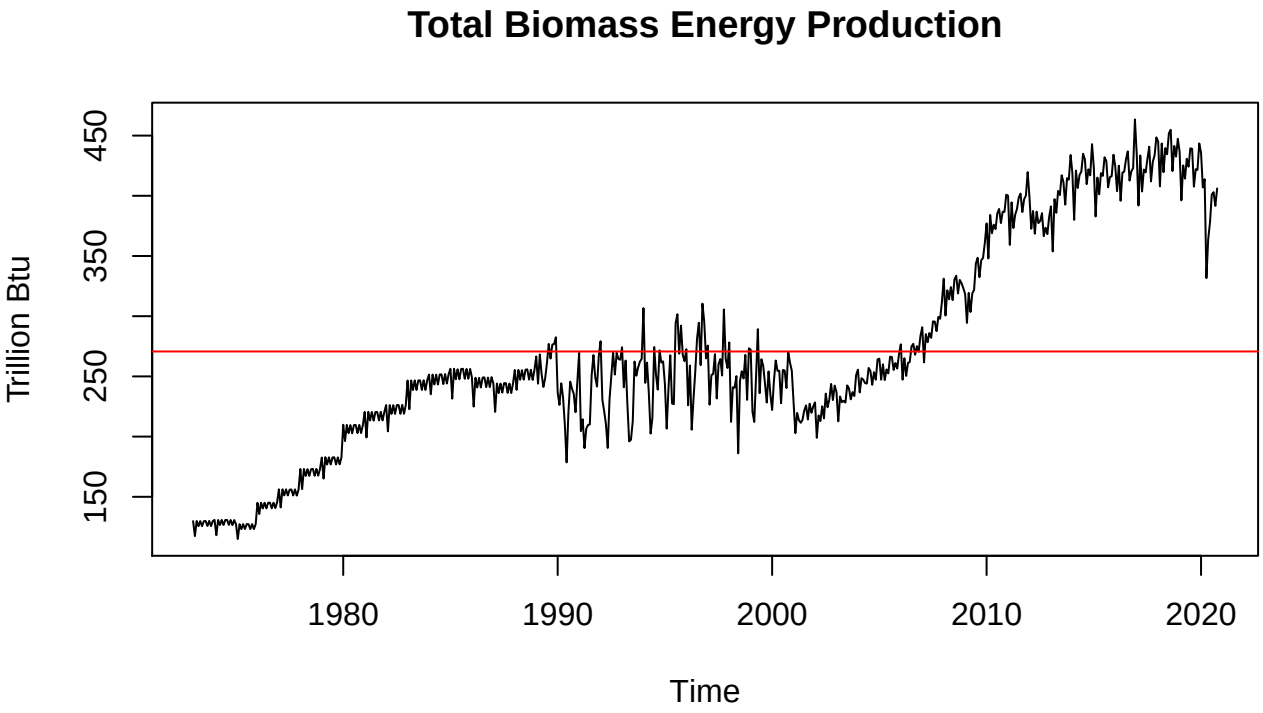
```
## [1] 236.95  43.90
```

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.

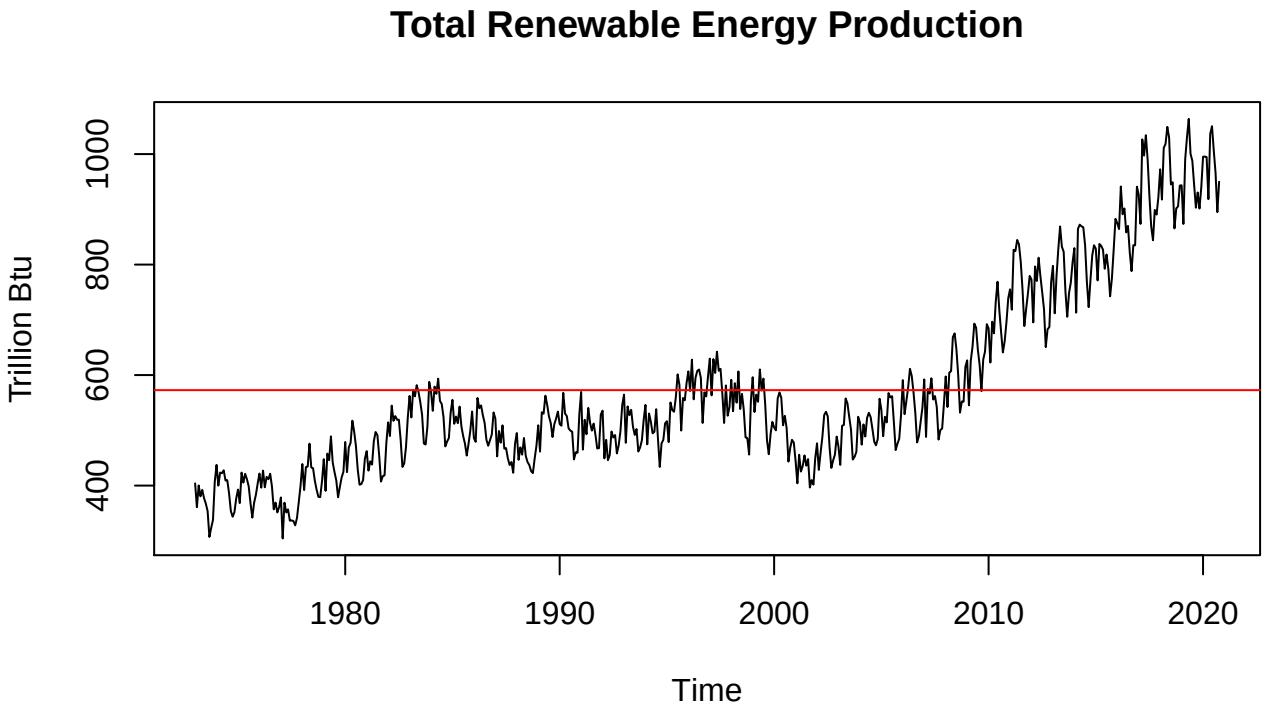
Total Biomass Energy Production

Below is a graph of total biomass energy production from 1973 to 2020. The graph demonstrates a significant growth in biomass energy production between circa 1975 and 1985 as well as between circa 2000 and 2012.



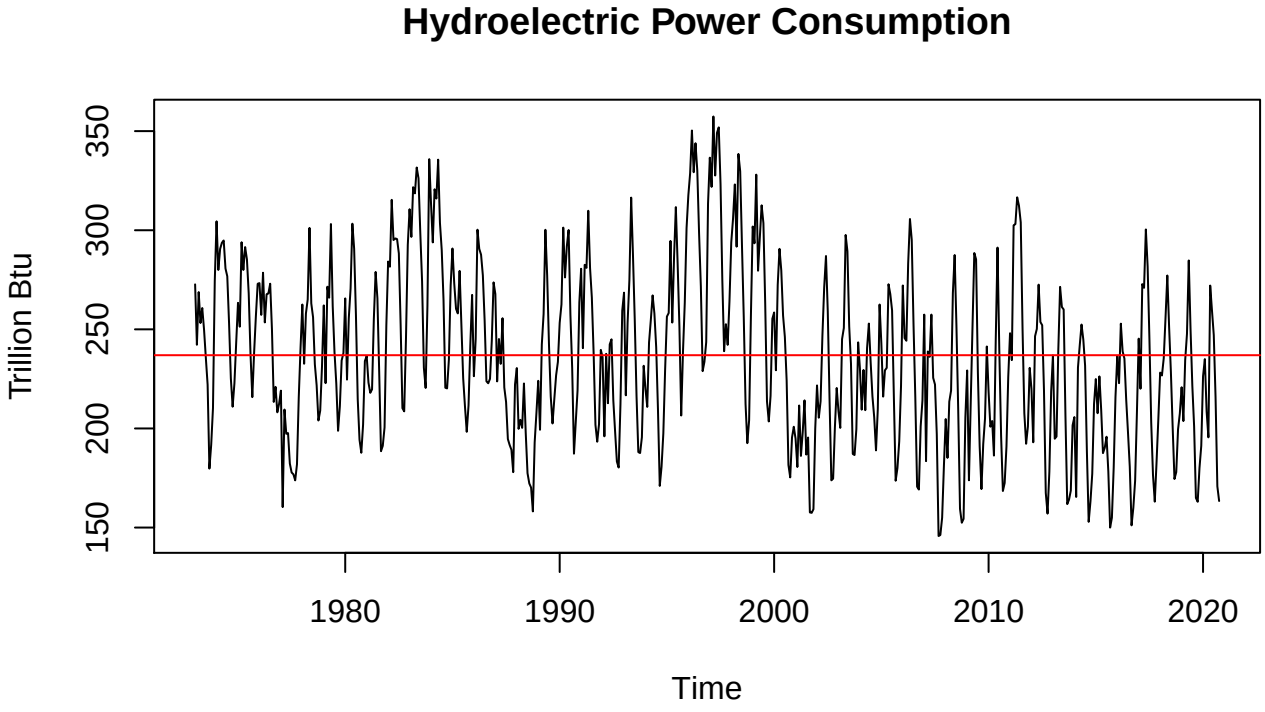
Total Renewable Energy Production

Below is a graph of total renewable energy production from 1973 to 2020. The graph demonstrates a significant growth in renewable energy production since circa 2010 and have since exceed the mean production average of the 47-year period.



Hydroelectric Power Consumption

Below is a graph of hydroelectric power consumption from 1973 to 2020. The graph demonstrates a stable consumption of hydroelectric power in the 47-year period since 1973. Additionally, it appears that variation was significant seasonally.



Question 5

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

Correlation(renewable energy production, biomass energy production): 0.9235

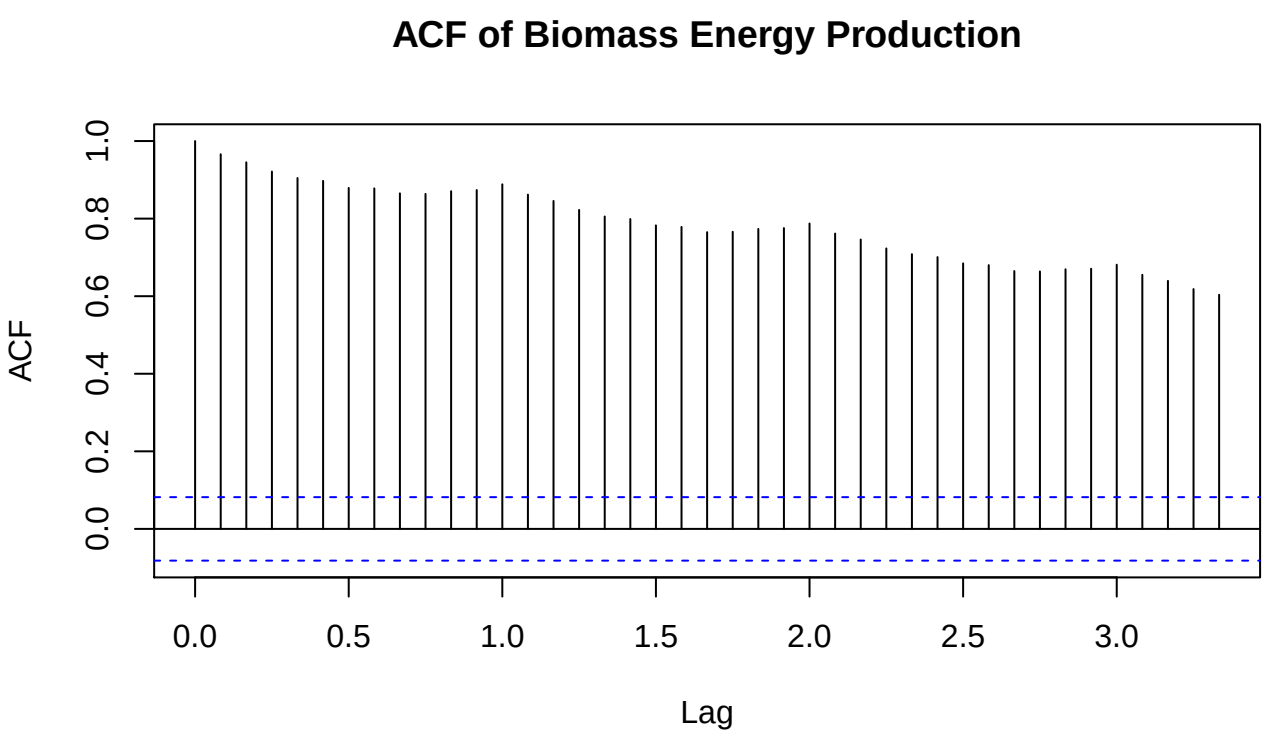
Correlation(biomass energy production, hydropower consumption): -0.2556

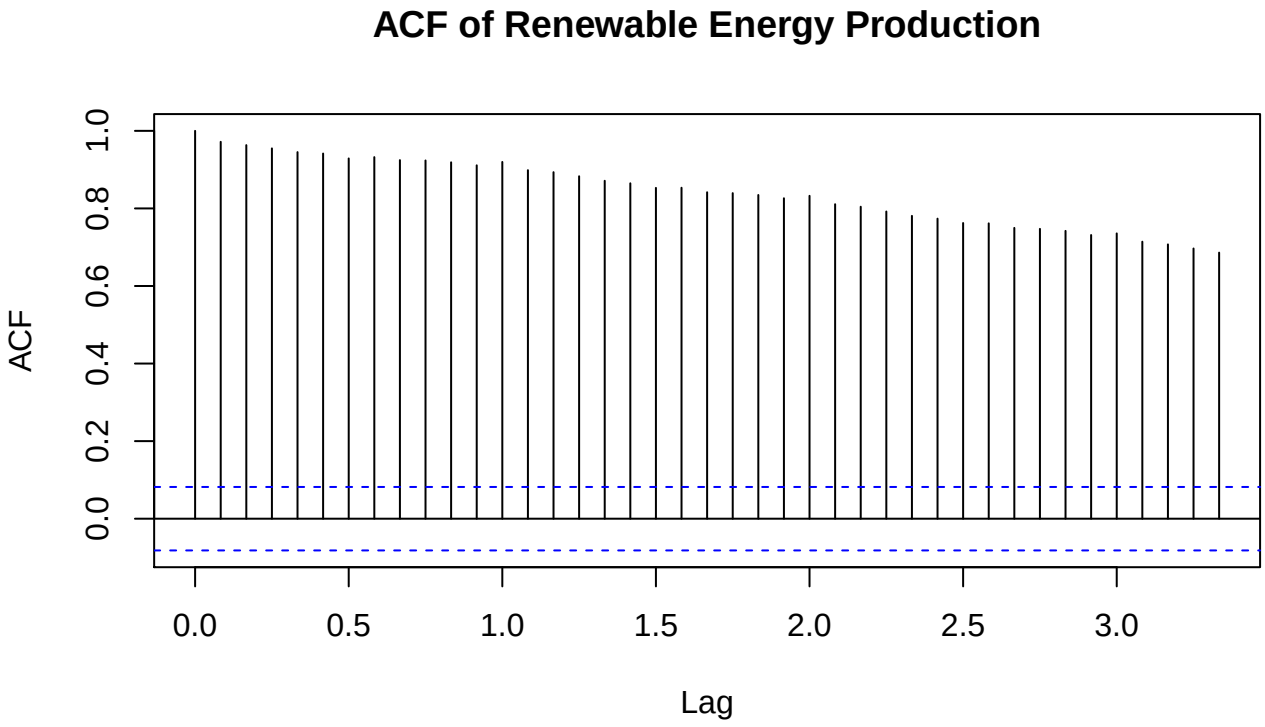
Correlation(hydropower consumption, renewable energy production): -0.0028 Based on the results, it appears that renewable energy production and biomass energy production are significantly positively correlated. On the other hand, hydropower consumption is negatively correlated with both biomass energy production and renewable energy production, but weakly correlated with renewable energy production only.

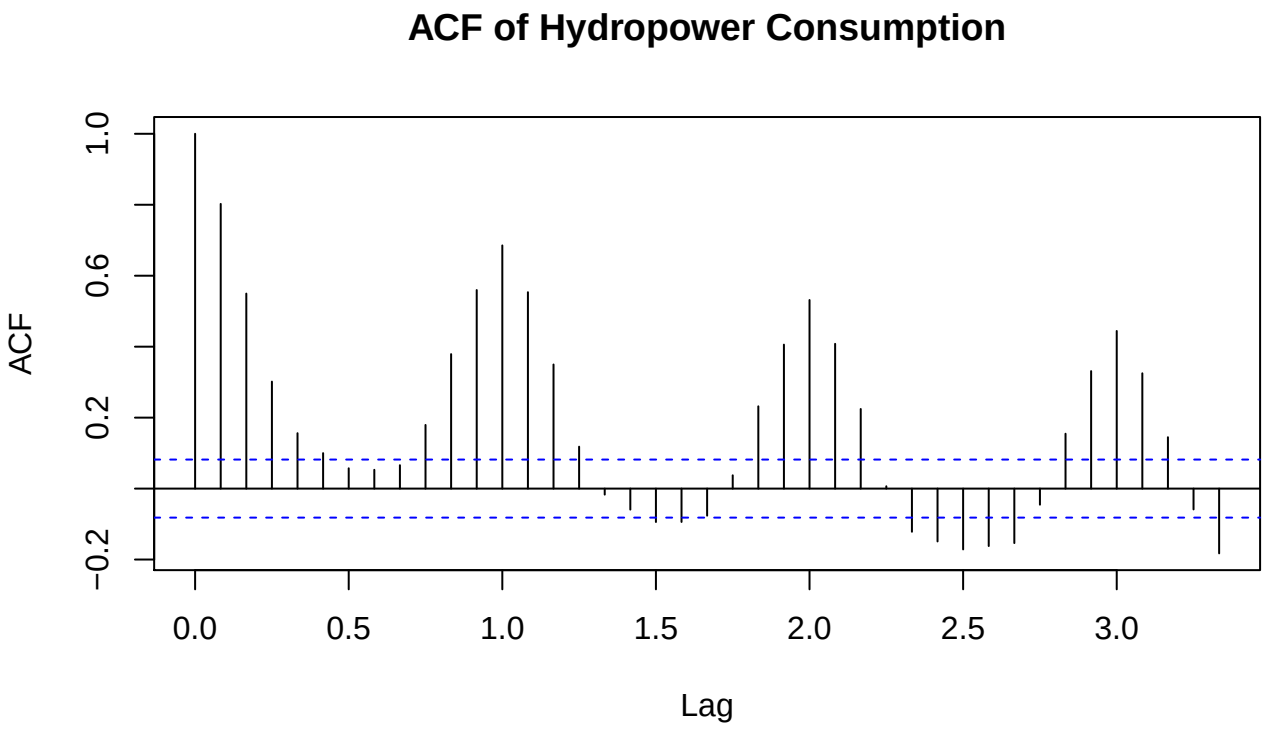
Question 6

Compute the autocorrelation function (ACF) 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 of renewable energy production and biomass energy production show similar behavior of strong autocorrelation with slight decrease over time. Autocorrelation of hydropower consumption, on the other hand, shows a flunctuating trend.



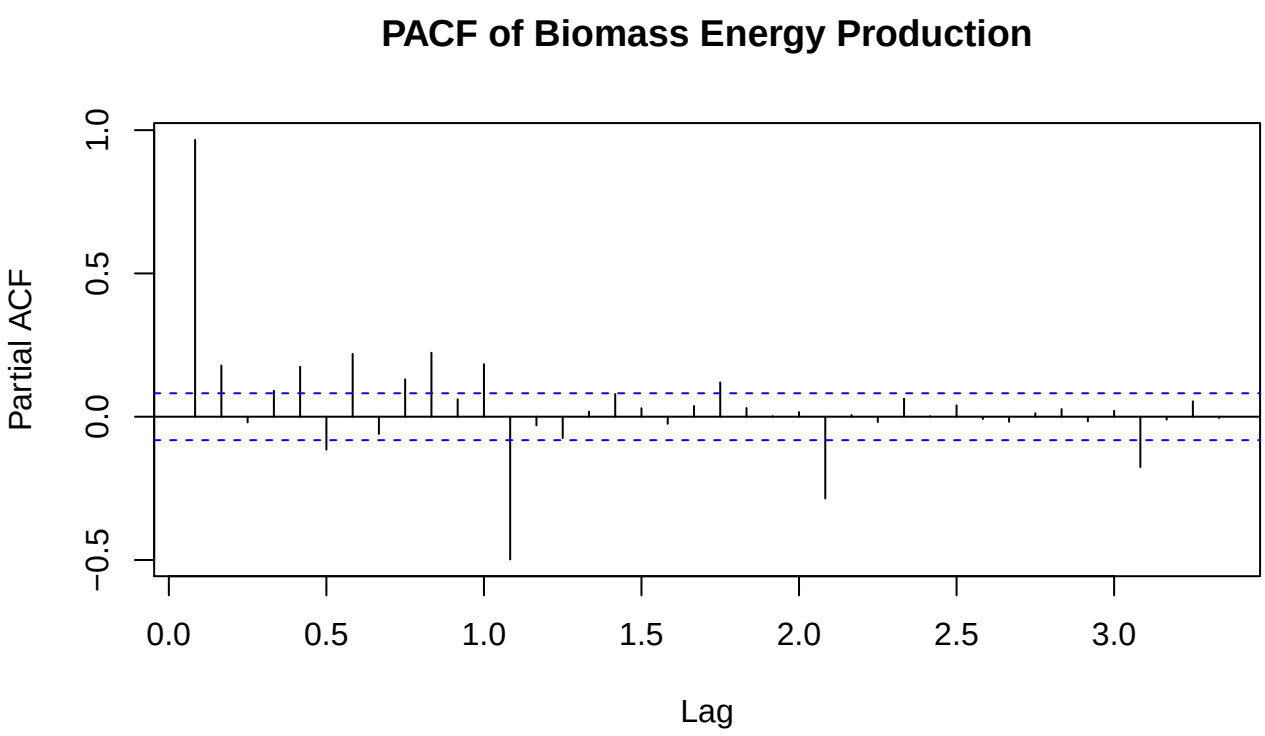


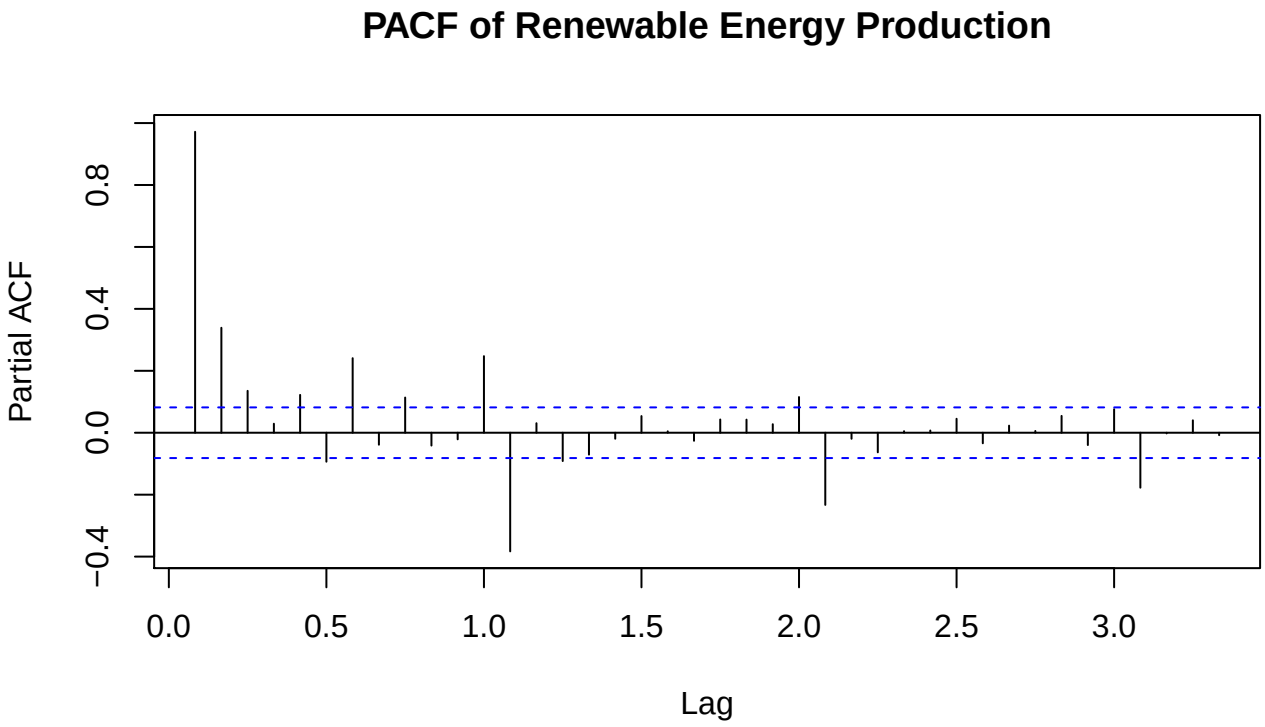


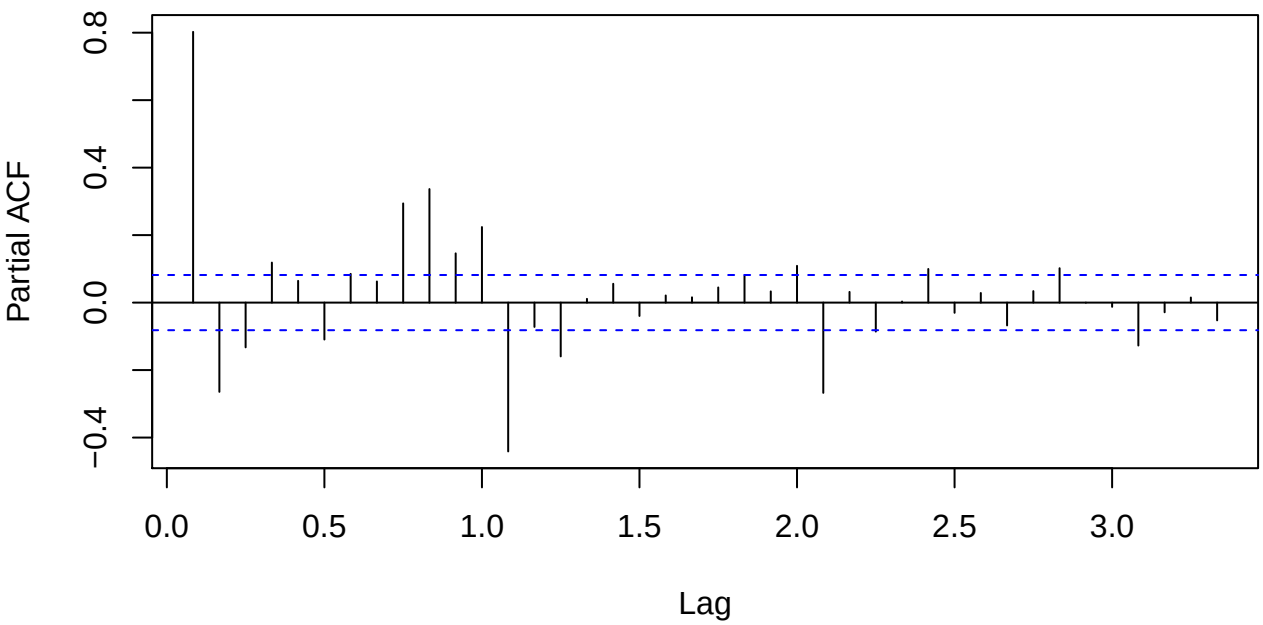
Question 7

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

The PACF plots of all three variables demonstrated similar trend which indicates that intermediate variables in renewable energy production and biomass energy production have a stronger influence on the autocorrelation of the ACF.


PACF of Biomass Energy Production
Partial ACF
1.0
0.5
0.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


PACF of Renewable Energy Production
Partial ACF
0.8
0.4
0.0
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


PACF of Hydropower Consumption
Partial ACF
0.8
0.4
0.0
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag

Appendix

```
#Load/install required package here
library(tseries)
library(forecast)
library(dplyr)
library(readxl)

#setting digit to maximum of 6
options(digits=6)

#Importing initial data set
setwd("/Users/stefanchen/Documents/Duke/Classes/Spring 2021/ENV 790/GitHub/ENV790_30_TSA_S2021/Data")
re.df<-read_xlsx("Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx", skip=10)

#Create data frame with the selected columns
re_select<-re.df %>%
  select("Total Biomass Energy Production",
         "Total Renewable Energy Production",
         "Hydroelectric Power Consumption")

#Remove rows of units by selecting [numeric rows which excludes characters, and all columns]
re_select<-re_select[!is.na(as.numeric(re_select$'Total Biomass Energy Production')),]

#Transforming characters to numeric value
re_select<-sapply(re_select,as.numeric) %>%
  as.data.frame()

#Confirm data type
str(re_select)

#Confirm the first 6 values
head(re_select)

#Transform data frame to time series
ts_select<-ts(re_select, start=c(1973,1), frequency=12)

#Transform each columns in data frame to time series
ts_bio<-ts(re_select[,1], start=1973, frequency=12)
ts_re<-ts(re_select[,2], start=1973, frequency=12)
ts_hydro<-ts(re_select[,3], start=1973, frequency=12)

head(ts_select)  #show first six results

#Calculate mean and standard deviation and display c(mean, sd)
round(c(mean(ts_bio),sd(ts_bio)),2)
round(c(mean(ts_re),sd(ts_re)),2)
round(c(mean(ts_hydro),sd(ts_hydro)),2)

#Plot times series for Total Biomass Energy Production
plot(ts_bio,
     main="Total Biomass Energy Production",
     ylab="Trillion Btu",
     type="l")
abline(h=270.70, col="red")   #adding a horizontal line
```

```
#Plot times series for Total Renewable Energy Production
plot(ts_re,
     main="Total Renewable Energy Production",
```

```
     ylab="Trillion Btu",
     type="l")
abline(h=572.73, col="red")   #adding a horizontal line
```

```
#Plot times series for Hydroelectric Power Consumption
plot(ts_hydro,
     main="Hydroelectric Power Consumption",
     ylab="Trillion Btu",
     type="l")
abline(h=236.95, col="red")   #adding a horizontal line
```

```
#Correlation between the variables
cor(ts_re, ts_bio)
cor(ts_bio, ts_hydro)
cor(ts_hydro, ts_re)

#Autocorrelation between the variables
acf(ts_re, lag.max = 40, main="ACF of Biomass Energy Production")
```

```
acf(ts_bio, lag.max = 40, main="ACF of Renewable Energy Production")
```

```
acf(ts_hydro, lag.max = 40, main="ACF of Hydropower Consumption")
```

```
#Partial autocorrelation between the variables
pacf(ts_re, lag.max = 40, main="PACF of Biomass Energy Production")
```

```
pacf(ts_bio, lag.max = 40, main="PACF of Renewable Energy Production")
```

```
pacf(ts_hydro, lag.max = 40, main="PACF of Hydropower Consumption")
```