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

Assignment 2 - Due date 01/26/22

Colin Lee

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_Sp22.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 on our **Data** folder. The data comes from the US Energy Information and Administration and corresponds to the January 2022 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). }

```
#Importing data set

mydata <- read.xlsx(file = "/Users/colinlee/Documents/Duke/Spring 2022/ENV790/ENV790_TimeSeriesAnalysis_
```

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.

```
mydata <- mydata[,4:6]

colnames(mydata)=c("Total Biomass Energy Production","Total Renewable Energy Production", "Hydroelectric
head(mydata)
```

```
##   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().

```
ts_mydata <- ts(mydata, start = 1, frequency = 1)
```

Question 3

Compute mean and standard deviation for these three series.

```
#summary(mydata)

mean_biomass <- mean(mydata$`Total Biomass Energy Production`)
mean_renewable <- mean(mydata$`Total Renewable Energy Production`)
mean_hydro <- mean(mydata$`Hydroelectric Power Consumption`)

mean_biomass
```

```
## [1] 273.7839
```

```
mean_renewable
```

```
## [1] 581.1708
```

```
mean_hydro
```

```
## [1] 235.9653
```

```
sd_biomass <- sd(mydata$`Total Biomass Energy Production`)
sd_renewable <- sd(mydata$`Total Renewable Energy Production`)
sd_hydro <- sd(mydata$`Hydroelectric Power Consumption`)

sd_biomass
```

```
## [1] 89.42852
```

```
sd_renewable
```

```
## [1] 177.5607
```

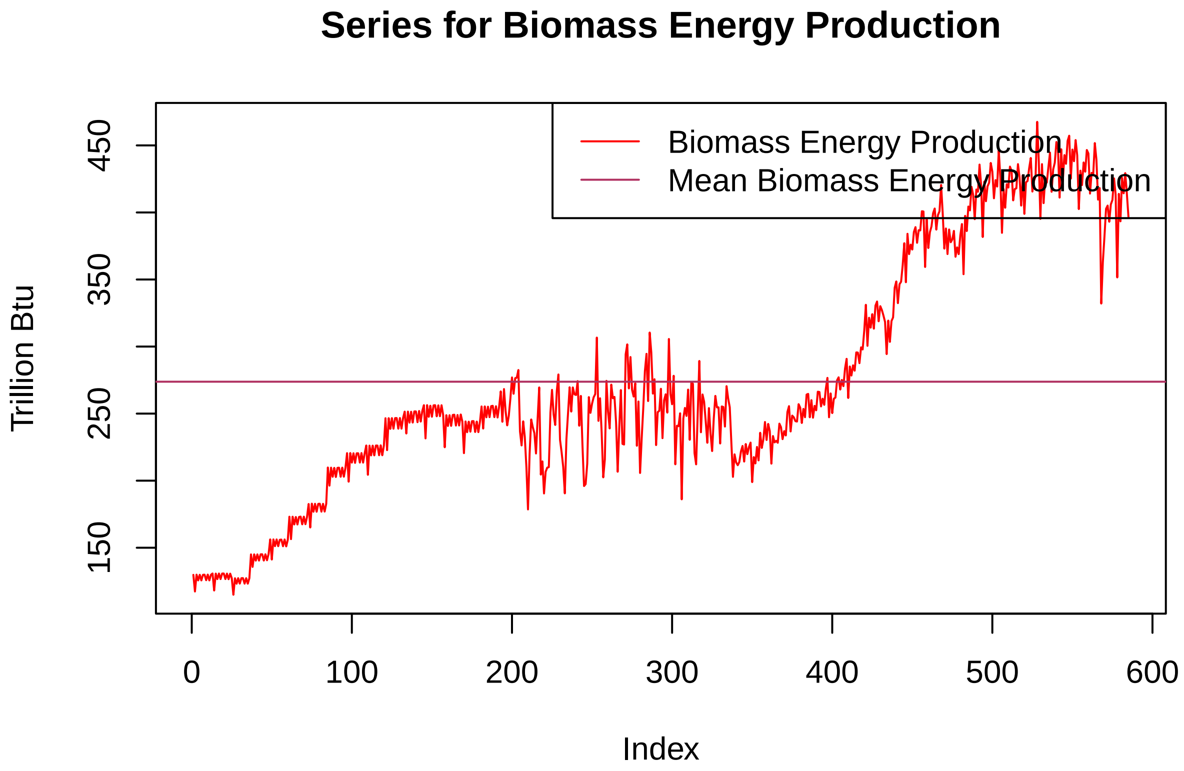
```
sd_hydro
```

```
## [1] 44.01749
```

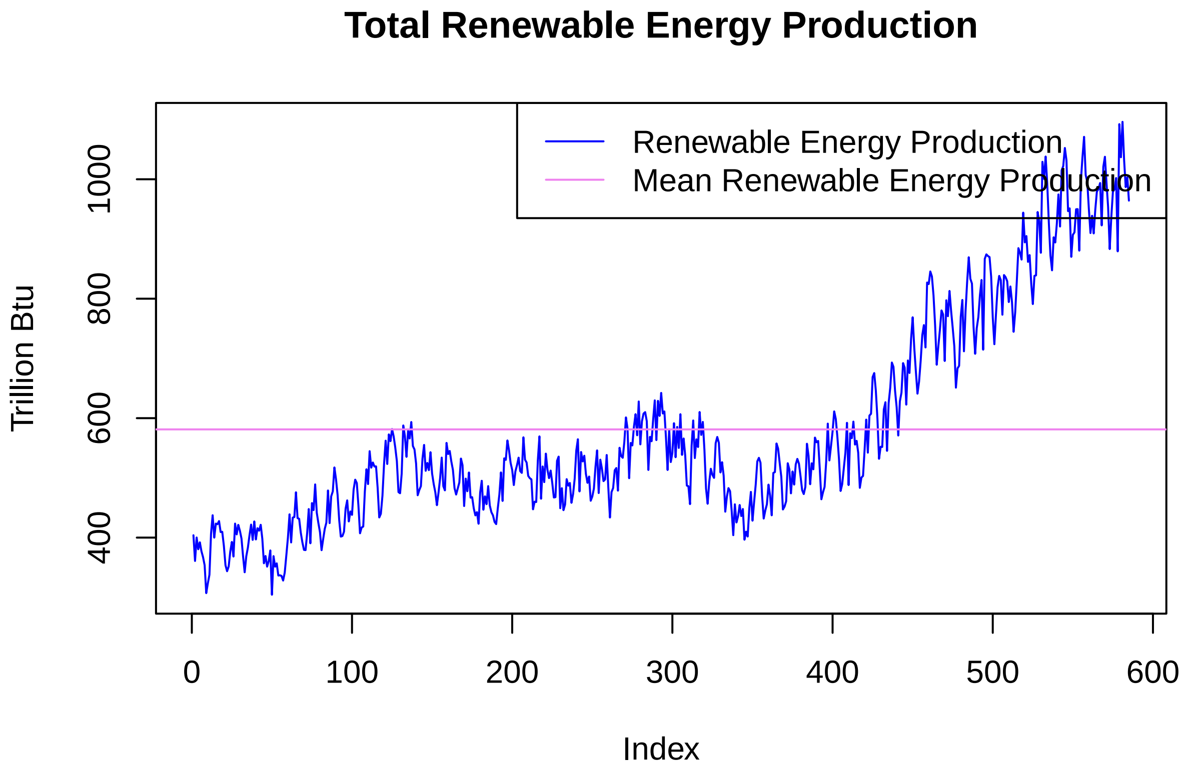
Question 4

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.

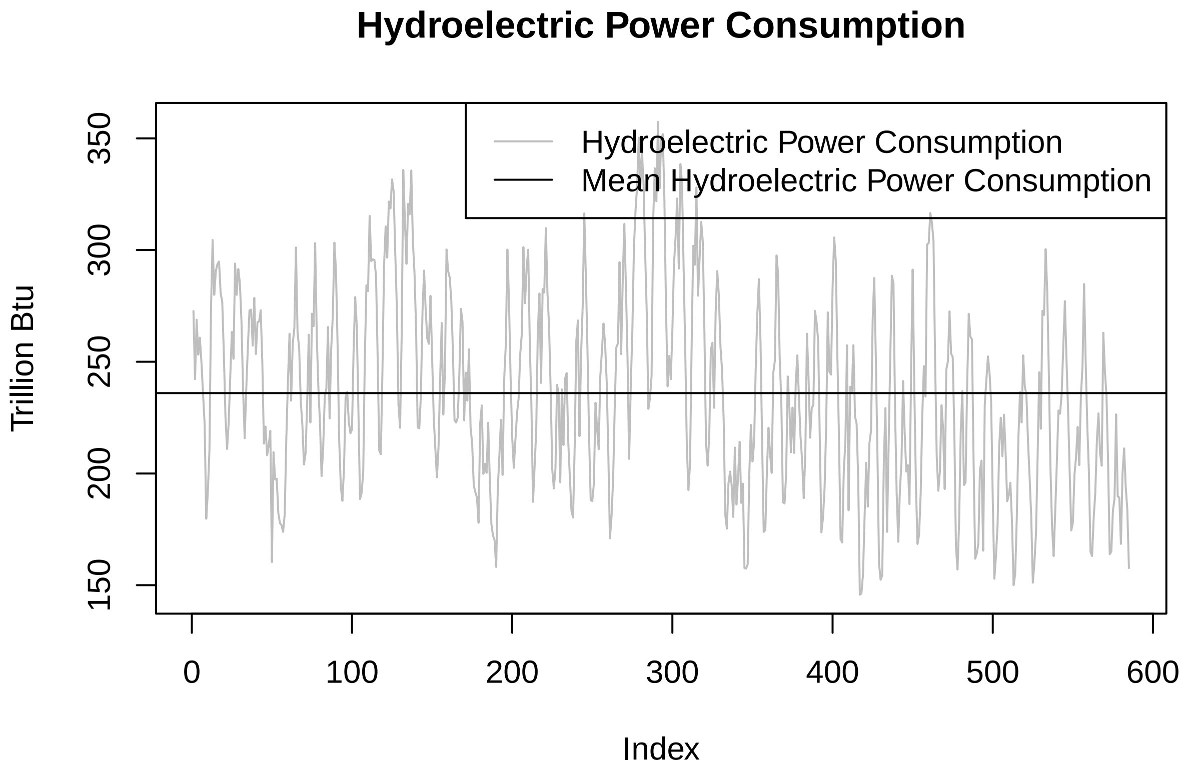
```
plot(mydata[,"Total Biomass Energy Production"],type="l",col="red",ylab="Trillion Btu")
title(main="Series for Biomass Energy Production")
abline(h=mean(mydata[,"Total Biomass Energy Production"]),col="maroon")
legend("topright",legend=c("Biomass Energy Production", "Mean Biomass Energy Production"), lty=c("solid
```



```
plot(mydata[,"Total Renewable Energy Production"],type="l",col="blue",ylab="Trillion Btu")
title(main="Total Renewable Energy Production")
abline(h=mean(mydata[,"Total Renewable Energy Production"]),col="violet")
legend("topright",legend=c("Renewable Energy Production", "Mean Renewable Energy Production"), lty=c("s
```



```
plot(mydata[,"Hydroelectric Power Consumption"],type="l",col="grey",ylab="Trillion Btu")
title(main="Hydroelectric Power Consumption")
abline(h=mean(mydata[,"Hydroelectric Power Consumption"]),col="black")
legend("topright",legend=c("Hydroelectric Power Consumption", "Mean Hydroelectric Power Consumption"), 
```



Question 5

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

Biomass energy production is significantly correlated with renewable energy production with a value of 0.923 between the variables. This indicates a significant positive correlation between the two variables. Furthermore, neither of those two variables are significantly correlated with hydroelectric power consumption, where the correlation coefficient between Total Biomass Energy Production and Hydroelectric Power Consumption is -0.28 and the correlation coefficient between Total Renewable Energy Production and Hydroelectric Power Consumption is -0.056. These values are very low, indicating low correlation, and they are slightly negatively correlated as well.

```
cor(mydata)
```

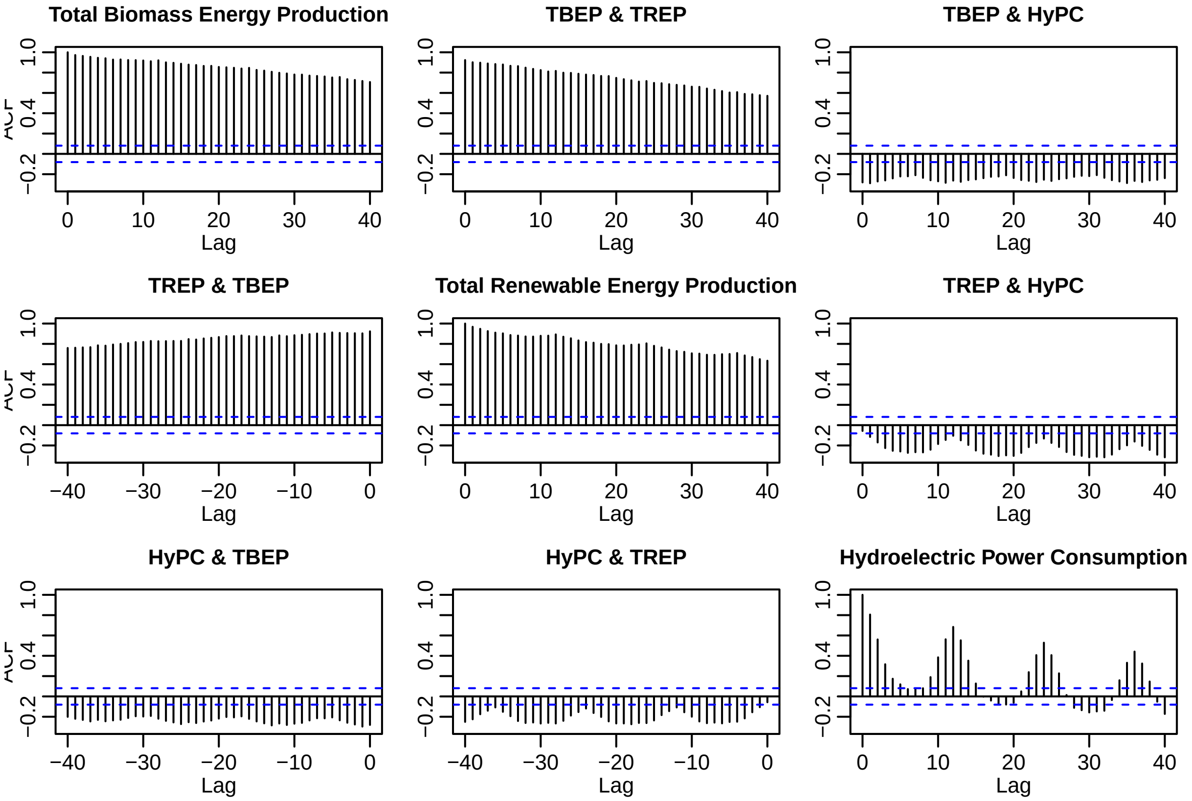
```
##                                   Total Biomass Energy Production
## Total Biomass Energy Production                         1.0000000
## Total Renewable Energy Production                       0.9232838
## Hydroelectric Power Consumption                        -0.2804997
##                                   Total Renewable Energy Production
## Total Biomass Energy Production                          0.92328377
## Total Renewable Energy Production                        1.00000000
## Hydroelectric Power Consumption                         -0.05680651
##                                   Hydroelectric Power Consumption
## Total Biomass Energy Production                       -0.28049970
## Total Renewable Energy Production                     -0.05680651
## Hydroelectric Power Consumption                        1.00000000
```

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?

For the Total Biomass Energy Production and Total Renewable Energy Production plots, we can see that there is high autocorrelation over time (and slightly deteriorates over for greater lags indicating it is non-stationary), indicating that there is high dependence between adjacent values. This is not so much the case with hydroelectric power consumption, where there instead appears to be a seasonal pattern for autocorrelation that peaks every few lags and is also non-stationary.

```r
#acf(mydata, lag=40, pl=FALSE)
#acf(mydata, lag=40)
acf(ts_mydata, lag.max = 40)
```



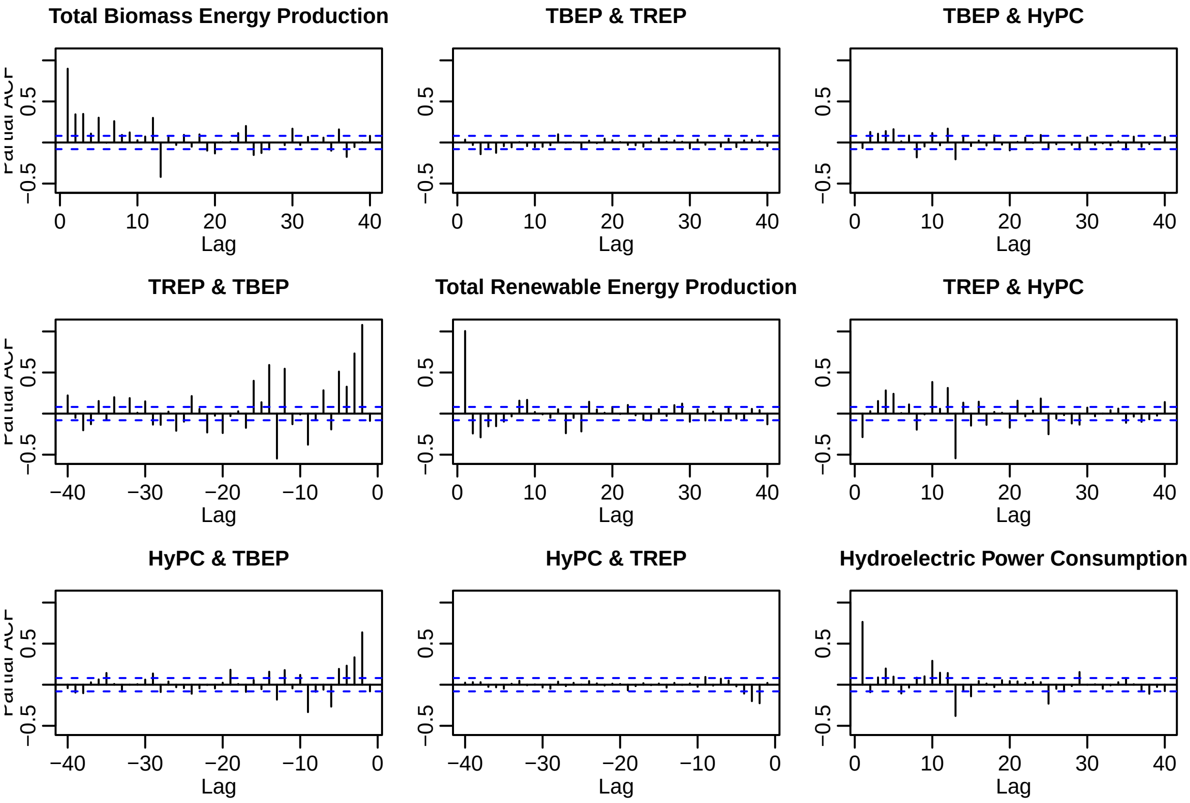
Question 7

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

The PACF plots differ greatly from the ACF plots in that the values are not nearly as high or close to 1.0, and the plots for Total Biomass Energy Production and Total Renewable Energy Production are not as strong with consistently high correlation values compared to the ACF plots. Here, all the charts look slightly similar with low PACF function values and no discernible pattern but possibly some seasonality.

```
#pacf(mydata, lag=40, pl=FALSE)

#pacf(mydata, lag=40)
pacf(ts_mydata, lag.max = 40)
```


Total Biomass Energy Production
TBEP & TREP
TBEP & HyPC
TREP & TBEP
Total Renewable Energy Production
TREP & HyPC
HyPC & TBEP
HyPC & TREP
Hydroelectric Power Consumption
Partial ACF
Lag