ENV 790.30 - Time Series Analysis for Energy Data | Spring 2024

Assignment 3 - Due date 02/01/24

Cara Kuuskvere

Directions

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).

Please keep this R code chunk options for the report. It is easier for us to grade when we can see code and output together. And the tidy.opts will make sure that line breaks on your code chunks are automatically added for better visualization.

When you have completed the assignment, **Knit** the text and code into a single PDF file. Submit this pdf using Sakai.

Questions

Consider the same data you used for A2 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumpti
The data comes from the US Energy Information and Administration and corresponds to the December 2022 **Monthly** Energy Review. Once again 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.

R packages needed for this assignment:"forecast","tseries", and "Kendall". 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
library(forecast)
```

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

```
library(tseries)
library(Kendall)
library(readr)
library(ggplot2)
library(cowplot)
```

```r
#Importing data set
raw_energy_data <- read.csv(
  "Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.csv",
  header=FALSE,skip=12)
```

```r
#trim the table
energy_data <- raw_energy_data[,4:6] #want all rows
                                     #all columns from 4 to 6
n_energy_sources <- ncol(energy_data) #number of variables
n_obs <- nrow(energy_data) #number observations

#Adding column names
colnames(energy_data)=c("Total Biomass Energy Production (Trillion Btu)",
                        "Total Renewable Energy Production (Trillion Btu)",
                        "Hydroelectric Power Consumption (Trillion Btu)")

head(energy_data)
```

```
##   Total Biomass Energy Production (Trillion Btu)
## 1                                         129.824
## 2                                         130.807
## 3                                         118.091
## 4                                         130.727
## 5                                         126.583
## 6                                         130.789
##   Total Renewable Energy Production (Trillion Btu)
## 1                                          220.755
## 2                                          231.010
## 3                                          210.188
## 4                                          226.384
## 5                                          223.218
## 6                                          227.793
##   Hydroelectric Power Consumption (Trillion Btu)
## 1                                         90.131
## 2                                         99.500
## 3                                         91.476
## 4                                         94.950
## 5                                         95.969
## 6                                         96.337
```

```r
ts_energy_data <- ts(energy_data,start=c(1973,1),frequency = 12)
# frequency is 12 because it's monthly data, and months repeat every 12 entries!
# starts in year 1973
head(ts_energy_data)
```

```
##          Total Biomass Energy Production (Trillion Btu)
## Jan 1973                                        129.824
## Feb 1973                                        130.807
## Mar 1973                                        118.091
## Apr 1973                                        130.727
## May 1973                                        126.583
## Jun 1973                                        130.789
```

```
##          Total Renewable Energy Production (Trillion Btu)
## Jan 1973                                          220.755
## Feb 1973                                          231.010
## Mar 1973                                          210.188
## Apr 1973                                          226.384
## May 1973                                          223.218
## Jun 1973                                          227.793
##          Hydroelectric Power Consumption (Trillion Btu)
## Jan 1973                                         90.131
## Feb 1973                                         99.500
## Mar 1973                                         91.476
## Apr 1973                                         94.950
## May 1973                                         95.969
## Jun 1973                                         96.337
```

##Trend Component

Q1

For each time series, i.e., Renewable Energy Production and Hydroelectric Consumption create three plots: one with time series, one with the ACF and with the PACF. You may use the some code form A2, but I want all the three plots side by side as in a grid. (Hint: use function `plot_grid()` from the `cowplot` package)

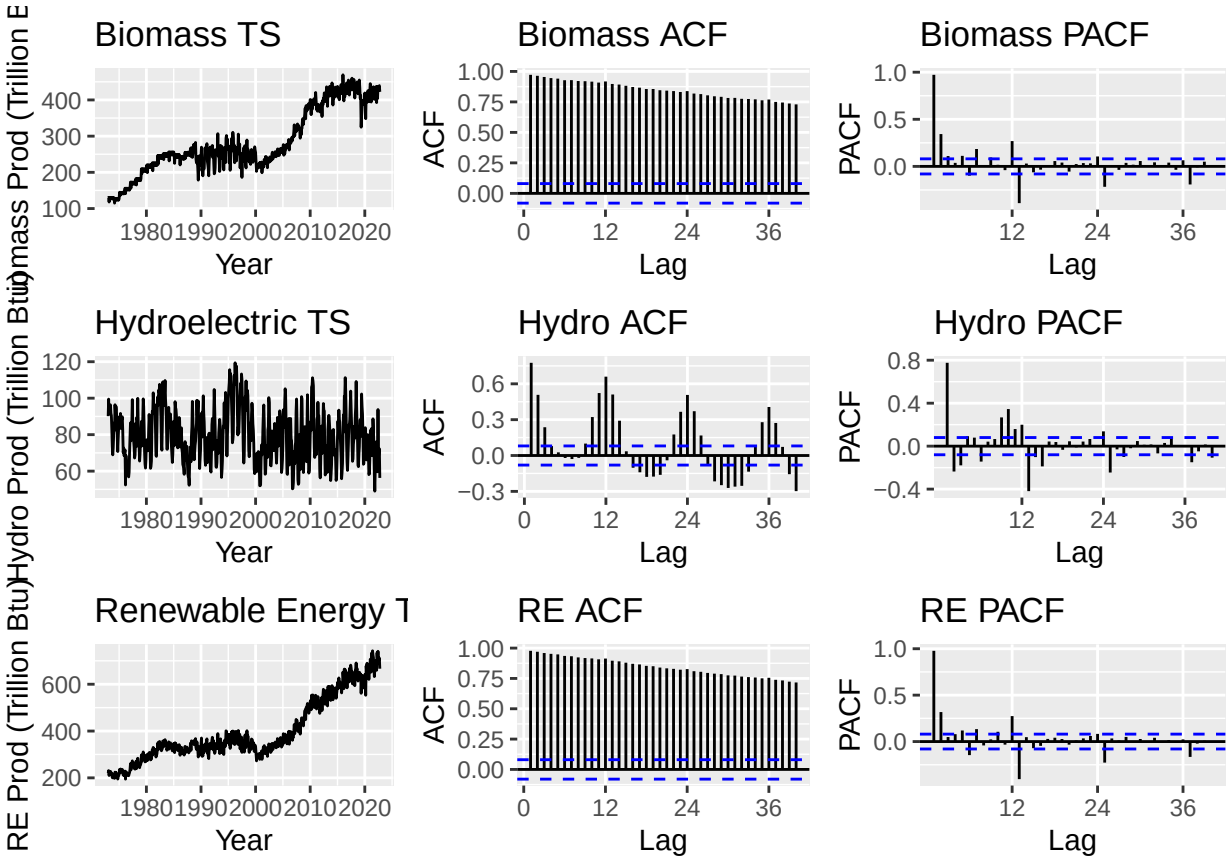
```
biomass_ts_plt <- autoplot(ts_energy_data[,1]) +
  labs(x="Year",y= "Biomass Prod (Trillion Btu)",
    title = "Biomass TS")
#biomass_ts_plt

RE_ts_plt <- autoplot(ts_energy_data[,2]) +
  labs(x="Year",y= "RE Prod (Trillion Btu)",
  title = "Renewable Energy TS")
#RE_ts_plt

hydro_ts_plt <- autoplot(ts_energy_data[,3]) +
  labs(x="Year",y= "Hydro Prod (Trillion Btu)",
    title = "Hydroelectric TS")
#hydro_ts_plt
```

```
biomass_acf <- ggAcf(ts_energy_data[,1],lag.max=40)+
  labs(title="Biomass ACF")
#biomass_acf
hydro_acf <- ggAcf(ts_energy_data[,3],lag.max=40)+
  labs(title="Hydro ACF")
#hydro_acf
RE_acf <- ggAcf(ts_energy_data[,2],lag.max=40)+
  labs(title="RE ACF")
#RE_acf
```

```
biomass_pacf <- ggPacf(ts_energy_data[,1],lag.max=40)+
  labs(title="Biomass PACF")
#biomass_pacf
hydro_pacf <- ggPacf(ts_energy_data[,3],lag.max=40)+
```

```
  labs(title="Hydro PACF")
#hydro_pacf
RE_pacf <- ggPacf(ts_energy_data[,2],lag.max=40)+
  labs(title="RE PACF")
#RE_pacf
```

```
#plot_grid(biomass_ts_plt,biomass_acf,biomass_pacf, align="h",label_size=8)
#plot_grid(hydro_ts_plt,hydro_acf,hydro_pacf, align = "h", label_size=8)
#plot_grid(RE_ts_plt,RE_acf,RE_pacf,align="h",label_size=8)
plot_grid(biomass_ts_plt,biomass_acf,biomass_pacf,
          hydro_ts_plt,hydro_acf,hydro_pacf,
          RE_ts_plt,RE_acf,RE_pacf,
          ncol=3,align="h")
```



Q2

From the plot in Q1, do the series Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Q3

Use the *lm()* function to fit a linear trend to the three time series. Ask R to print the summary of the regression. Interpret the regression output, i.e., slope and intercept. Save the regression coefficients for further analysis.

Q4

Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

Q5

Plot ACF and PACF for the detrended series and compare with the plots from Q1. You may use plot_grid() again to get them side by side. nut mot mandatory. Did the plots change? How?

Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

Q6

Just by looking at the time series and the acf plots, do the series seem to have a seasonal trend? No need to run any code to answer your question. Just type in you answer below.

Q7

Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) the two time series. Ask R to print the summary of the regression. Interpret the regression output. From the results which series have a seasonal trend? Do the results match you answer to Q6?

Q8

Use the regression coefficients from Q7 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

Q9

Plot ACF and PACF for the deseason series and compare with the plots from Q1. You may use plot_grid() again to get them side by side. nut mot mandatory. Did the plots change? How?