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

Assignment 3 - Due date 02/04/25

Nicole Leines

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_A03_Sp25.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 2024 **Monthly** Energy Review. Once again you will work only with the following columns: Total Renewable Energy Production and Hydroelectric Power Consumption. Create a data frame structure with these two 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)
library(tseries)
library(tidyverse)
library(readxl)
library(openxlsx)
library(ggplot2)
library(Kendall)
library(patchwork)
library(ggplotify)
```

Data

```
#Importing data set
base_dir <- "D:/Geani/Box/Home Folder gnl13/Private/1 Academics/3 Time series/TSA_Sp25"
data_dir <- file.path(base_dir, "Data")
file_name <- "Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx"
file_path <- file.path(data_dir, file_name)

#Importing data set without change the original file using read.xlsx
energy_data1 <- read_excel(path=file_path,
                           skip = 12,
                           sheet="Monthly Data",
                           col_names=FALSE)

#Now let's extract the column names from row 11
read_col_names <- read_excel(path=file_path,
                             skip = 10,n_max = 1,
                             sheet="Monthly Data",
                             col_names=FALSE)

#Assign the column names to the data set
colnames(energy_data1) <- read_col_names

#Visualize the first rows of the data set
head(energy_data1)
```

```
## # 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>, ...
```

Time series

```
#Create a data frame structure with Total Renewable Energy Production and Hydroelectric Power Consumpti

column_names <- colnames(energy_data1)
print(column_names)
```

```
##  [1] "Month"                              "Wood Energy Production"
```

```
##  [3] "Biofuels Production"                "Total Biomass Energy Production"
##  [5] "Total Renewable Energy Production"  "Hydroelectric Power Consumption"
##  [7] "Geothermal Energy Consumption"      "Solar Energy Consumption"
##  [9] "Wind Energy Consumption"            "Wood Energy Consumption"
## [11] "Waste Energy Consumption"           "Biofuels Consumption"
## [13] "Total Biomass Energy Consumption"   "Total Renewable Energy Consumption"
```

```
energy_data2 <- energy_data1[, c("Month",
                                 "Total Renewable Energy Production",
                                 "Hydroelectric Power Consumption")]

energy_data2$Month <- as.Date(energy_data2$Month)

energy_data2<-energy_data2|>
  rename(date=Month)

year1 <- year(energy_data2$date[1])
month1 <- month(energy_data2$date[1])

ts_energy_data2 <- ts(
  data = energy_data2[,2:3],
  start = c(year1,month1),
  frequency = 12
)

head(ts_energy_data2)
```

```
##          Total Renewable Energy Production Hydroelectric Power Consumption
## Jan 1973                           219.839                           89.562
## Feb 1973                           197.330                           79.544
## Mar 1973                           218.686                           88.284
## Apr 1973                           209.330                           83.152
## May 1973                           215.982                           85.643
## Jun 1973                           208.249                           82.060
```

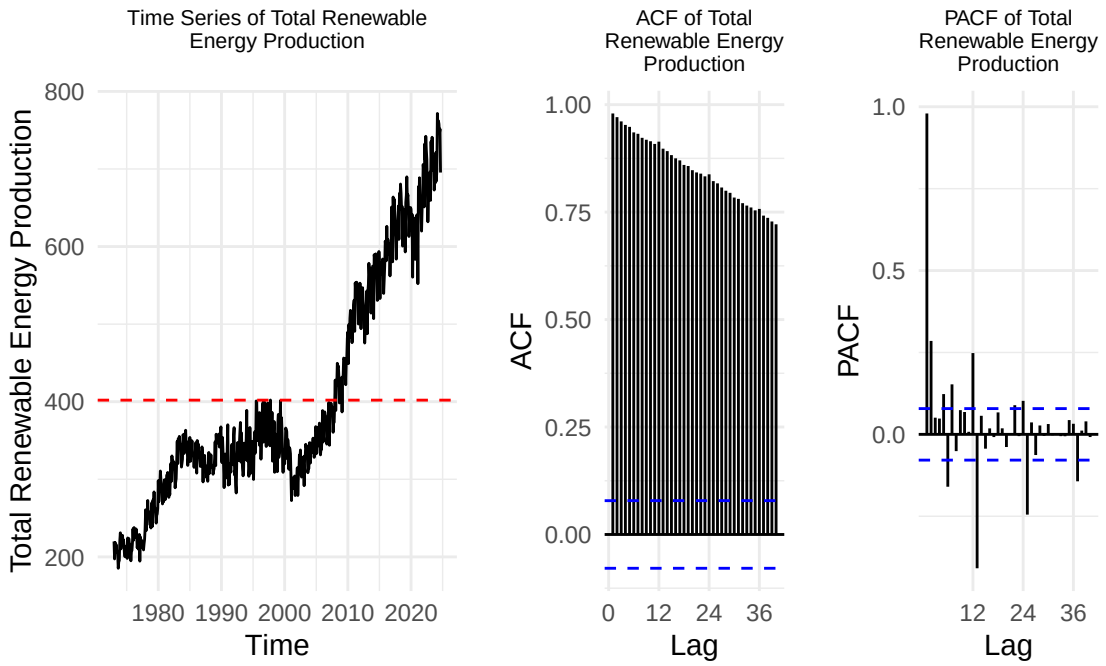
##Trend Component

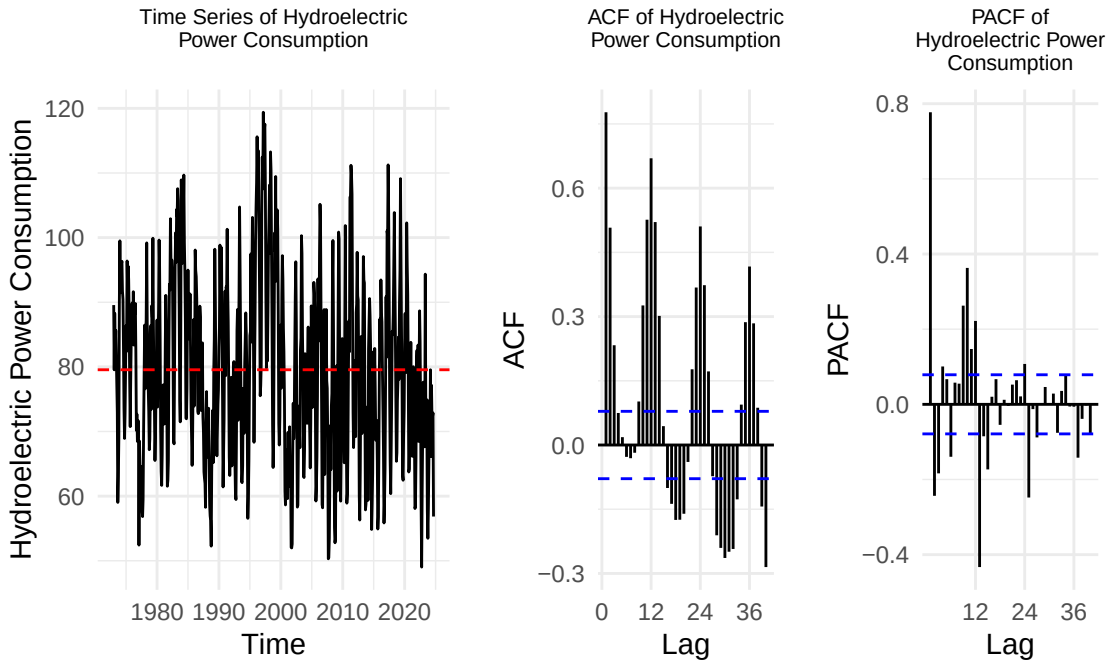
Q1

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)

```
nvars <- ncol(ts_energy_data2)

for (i in 1:nvars) {
  # 1. Time Series Plot
  p1 <- autoplot(ts_energy_data2[, i]) +
    geom_hline(yintercept = mean(ts_energy_data2[, i], na.rm = TRUE),
               color = "red", linetype = "dashed") +
    labs(title = str_wrap(paste("Time Series of", colnames(ts_energy_data2)[i]), width = 30),
         x = "Time", y = colnames(ts_energy_data2)[i]) +
```

```
    theme_minimal() +
    theme(plot.title = element_text(size = 8, hjust = 0.5),
          plot.margin = margin(30, 10, 10, 10))

  # 2. ACF Plot
  p2 <- ggAcf(ts_energy_data2[, i], lag.max = 40) +
    labs(title = str_wrap(paste("ACF of", colnames(ts_energy_data2)[i]), width = 20)) +
    theme_minimal() +
    theme(plot.title = element_text(size = 8, hjust = 0.5),
          plot.margin = margin(30, 10, 10, 10))

  # 3. PACF Plot
  p3 <- ggPacf(ts_energy_data2[, i], lag.max = 40) +
    labs(title = str_wrap(paste("PACF of", colnames(ts_energy_data2)[i]), width = 20)) +
    theme_minimal() +
    theme(plot.title = element_text(size = 8, hjust = 0.5),
          plot.margin = margin(30, 10, 10, 10))

  # Combine plots
  final_plot <- p1 + p2 + p3 +
    plot_layout(ncol = 3, widths = c(2, 1, 1)) +
    plot_annotation(theme = theme(plot.margin = margin(20, 20, 20, 20)))

  print(final_plot)
}
```





Q2

From the plot in Q1, do the series Total Renewable Energy Production and 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 two 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, but not 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, but not mandatory. Did the plots change? How?