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

Assignment 3 - Due date 02/03/26

Sylvia Hipp

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 2025 Monthly Energy Review. This time 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(dplyr)
library(Kendall)

# additional packages
library(readxl)
library(lubridate)
library(ggplot2)
library(glue)
```

Read in the raw data

```r
filepath <- "./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx"

# load the data
energy_data_raw <- read_excel(filepath, skip = 11)
```

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

```r
# read in column names
colnames <- read_excel(filepath, skip = 10) %>% names()
names(energy_data_raw) <- colnames  # assign colnames to raw dataset
```

Basic data processing

```r
# clean up names and select columns of interest
energy_data <- energy_data_raw %>%
  janitor::clean_names() %>%
  select(month, total_renewable_energy_production, hydroelectric_power_consumption)
```

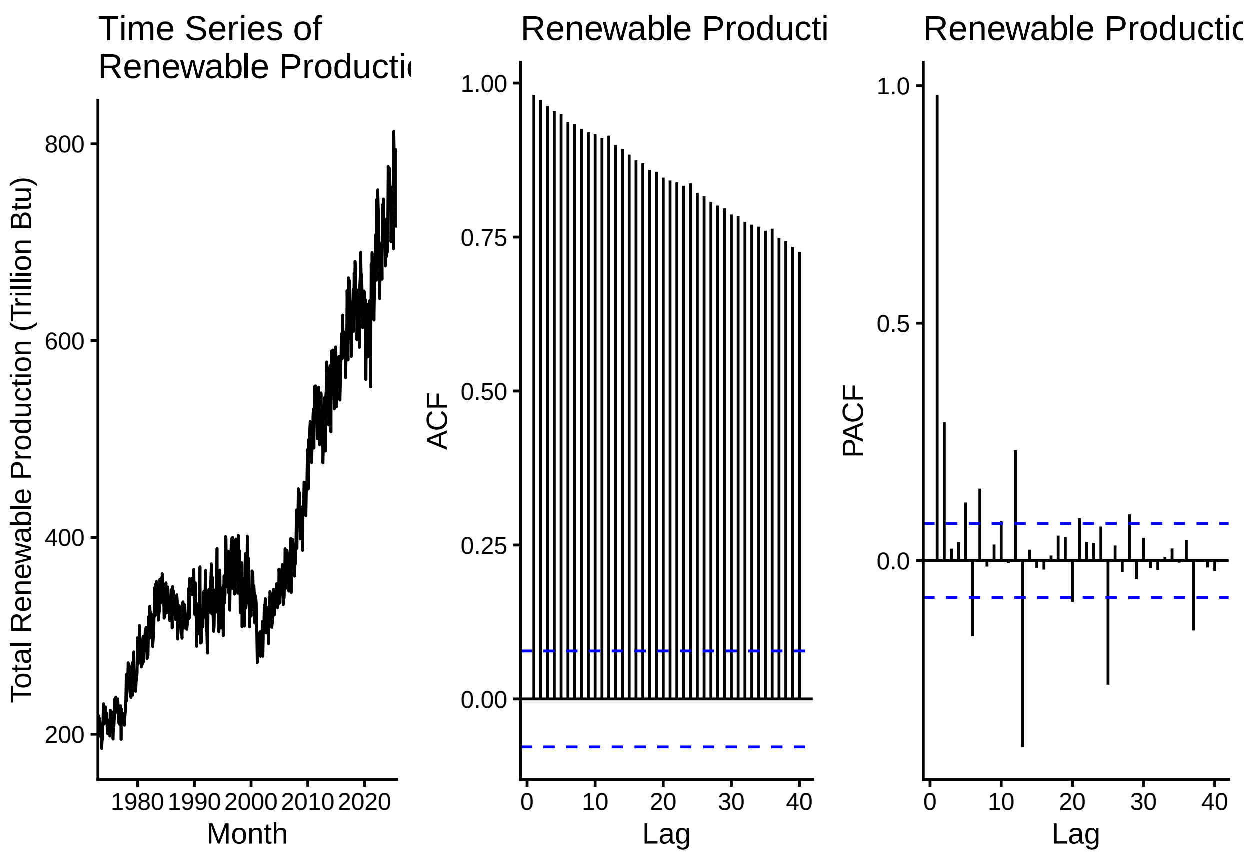
##Trend Component

Q1

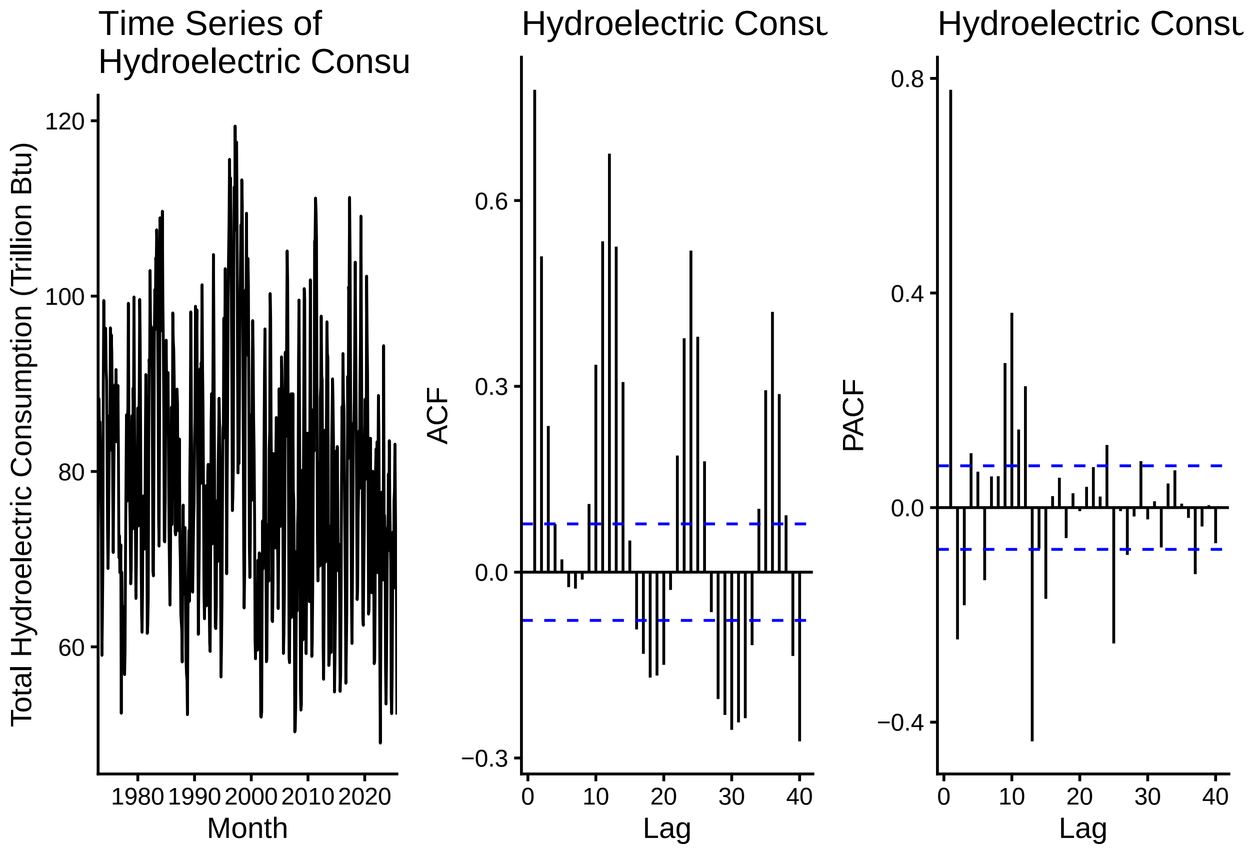
For each series (Total Renewable Production and Hydroelectric Consumption) create three plots arranged in a row (side-by-side): (1) time series plot, (2) ACF, (3) PACF. Use cowplot::plot_grid() to place them in a grid.

```r
# function to create all three plots for a given series
create_plots <- function(series, description){
  time_series_plot <- ggplot(energy_data,
                             aes(x = month, y = series)) +
    geom_line() +
    scale_x_date(expand = c(0,0)) +   # compress x axis
    theme_classic() +
    labs(x = "Month",
         y = glue("Total {description} (Trillion Btu)"),
         title = glue("Time Series of\n{description}"))
```

```r
  acf_plot <- ggAcf(series, lag.max = 40) +
    theme_classic() +
    labs(title = glue("{description} ACF"))

  pacf_plot <- ggPacf(series, lag.max = 40) +
    theme_classic() +
    labs(title = glue("{description} PACF"))

  cowplot::plot_grid(time_series_plot, acf_plot, pacf_plot,
                     ncol = 3)
}

# Series 1: Renewable Production
create_plots(energy_data$total_renewable_energy_production,
             "Renewable Production")
```




```r
# Series 2: Hydroelectric Consumption
create_plots(energy_data$hydroelectric_power_consumption,
             "Hydroelectric Consumption")
```



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.

```
# Total Renewable Energy Production
linear_model_renewable <- lm(total_renewable_energy_production~month, energy_data)
summary(linear_model_renewable)
```

```
##
## Call:
## lm(formula = total_renewable_energy_production ~ month, data = energy_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -154.84  -39.52   12.50   41.50  171.17
##
## Coefficients:
```

```
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.452e+02  5.540e+00   26.21   <2e-16 ***
## month       2.852e-07  5.312e-09   53.69   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64.22 on 631 degrees of freedom
## Multiple R-squared:  0.8204, Adjusted R-squared:  0.8201
## F-statistic:  2883 on 1 and 631 DF,  p-value: < 2.2e-16
```

```r
# Hydroelectric Power Consumption
linear_model_hydro <- lm(hydroelectric_power_consumption~month, energy_data)
summary(linear_model_hydro)
```

```
##
## Call:
## lm(formula = hydroelectric_power_consumption ~ month, data = energy_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -30.190 -10.214  -0.714   8.907  39.723
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.365e+01  1.204e+00   69.50  < 2e-16 ***
## month       -4.640e-09  1.154e-09   -4.02 6.52e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.95 on 631 degrees of freedom
## Multiple R-squared:  0.02497,    Adjusted R-squared:  0.02343
## F-statistic: 16.16 on 1 and 631 DF,  p-value: 6.515e-05
```

Q4

Use the regression coefficients to detrend each series (subtract fitted linear trend). Plot detrended series and compare with the original time series from Q1. Describe what changed.

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 to make it easier to compare. 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.

Answer:

Q7

Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to 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. Did the plots change? How?