ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2025

Assignment 4 - Due date 02/11/25

Mazhar Bhuyan

Directions

You should open the .rmd file corresponding to this assignment on RStudio. The file is available on our class repository on Github. And to do so you will need to fork our repository and link it to your RStudio.

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_A04_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).

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

R packages needed for this assignment: "xlsx" or "readxl", "ggplot2", "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(dplyr)
library(forecast)
library(ggplot2)
library(tidyverse)
library(cowplot)
library(stargazer)
library(readxl)
library(trend)
library(tseries)
knitr::opts_chunk$set(echo = TRUE,tidy.opts=list(width.cutoff=80), tidy=FALSE)
knitr::opts_knit$set(root.dir = "/home/guest/TSA_Mazhar/Time_Series_Mazhar")
```

Questions

Consider the same data you used for A3 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 January 2021 Monthly Energy Review. **For this assignment you will work only with the column "Total Renewable Energy Production"**.

```r
#Importing data set - you may copy your code from A3
Energy_Data <- read_excel("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx
  skip = 12,
  sheet = "Monthly Data",
  col_names = FALSE)
#head(Energy_Data)

col_names <- read_excel("./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source.xlsx",
      skip=10,
      n_max = 1,
      sheet="Monthly Data",
      col_names = FALSE)
#head(col_names)

colnames(Energy_Data) <- col_names

data <- Energy_Data[,c(1,5)]
nobs <- nrow(Energy_Data)
#Create vector t - time index
t <- 1:nobs

Renewable_Energy_Data_ts <- ts(data[t,2], frequency=12,start=c(1973,1))
#head(Renewable_Energy_Data_ts)
```

Stochastic Trend and Stationarity Tests

For this part you will work only with the column Total Renewable Energy Production.

Q1

Difference the "Total Renewable Energy Production" series using function diff(). Function diff() is from package base and take three main arguments: * *x* vector containing values to be differenced; * *lag* integer indicating with lag to use; * *differences* integer indicating how many times series should be differenced.

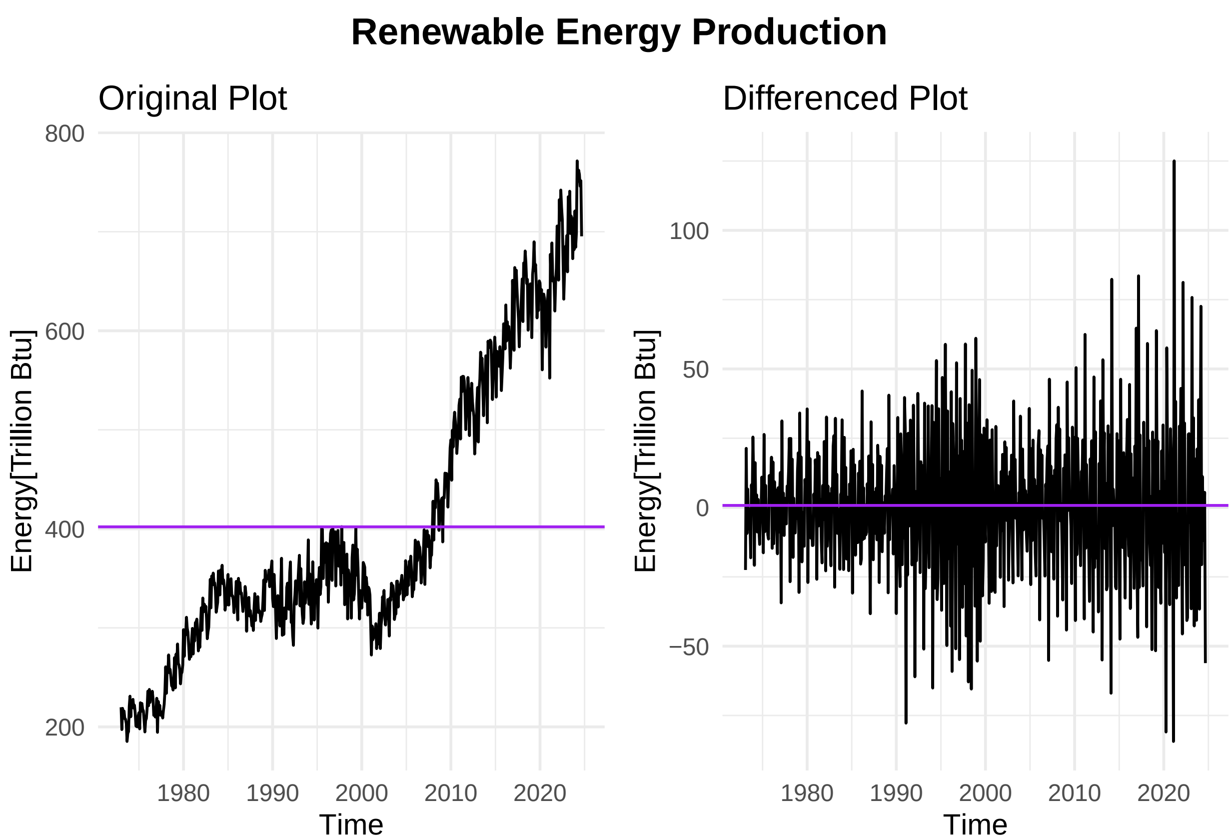
Try differencing at lag 1 only once, i.e., make `lag=1` and `differences=1`. Plot the differenced series. Do the series still seem to have trend?

```r
Renewable_Production_diff <- diff(Renewable_Energy_Data_ts,
                                  lag = 1,
                                  differences = 1
)

ts_plot <- autoplot(Renewable_Energy_Data_ts) +
          xlab("Time") +
          ylab("Energy[Trillion Btu]") +
          geom_hline(yintercept = mean(Renewable_Energy_Data_ts),
                          color = "purple",
                          linetype = "solid") +
          ggtitle("Original Plot")+
          theme_minimal()
#print(ts_plot)
```

```
diff_plot <- autoplot(Renewable_Production_diff) +
          xlab("Time") +
          ylab("Energy[Trillion Btu]")+
          geom_hline(yintercept = mean(Renewable_Production_diff),
                          color = "purple",
                          linetype = "solid") +
          ggtitle("Differenced Plot") +
          theme_minimal()
#print(diff_plot)

plot_comparison <- plot_grid(ts_plot,diff_plot,nrow=1,ncol=2)
#print(plot_comparison)
title <- ggdraw() +
          draw_label("Renewable Energy Production", fontface='bold')
plot_grid(title,plot_comparison,nrow=2,ncol=1,rel_heights = c(0.1,1))
```



Answer: After differencing at lag 1, difference 1, it seems like that the strong upward trend is removed. The values fluctuate around zero indicating a more stationary behavior. But the plot also shows very long spikes at the latest times. It might require more investigation to find out what is going on.

Q2

Copy and paste part of your code for A3 where you run the regression for Total Renewable Energy Production and subtract that from the original series. This should be the code for Q3 and Q4. make sure you use the same name for you time series object that you had in A3, otherwise the code will not work.

```r
## Linear Fit Model
lm_renewable=lm(Renewable_Energy_Data_ts~t)
beta0_renewable=lm_renewable$coefficients[1]
beta1_renewable=lm_renewable$coefficients[2]
#print(summary(lm_renewable))

stargazer(lm_renewable, type = "text",
          title = "Renewable Energy Production Linear Fit Model",
          align = TRUE, star.cutoffs = c(0.1, 0.05, 0.01))
```

```
##
## Renewable Energy Production Linear Fit Model
## ================================================
##                          Dependent variable:
##                      ---------------------------
##                       Renewable_Energy_Data_ts
## ------------------------------------------------
## t                              0.724***
##                                 (0.014)
##
## Constant                      176.873***
##                                 (4.962)
##
## ------------------------------------------------
## Observations                     621
## R2                              0.816
## Adjusted R2                     0.816
## Residual Std. Error      61.750 (df = 619)
## F Statistic         2,742.930*** (df = 1; 619)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```r
## Detrended Series

detrended_renewable <- Renewable_Energy_Data_ts - (beta0_renewable +
                                                     beta1_renewable*t)
detrended_renewable <- ts(detrended_renewable,
                          frequency = 12,
                          start = c(1973,1))

detrend_plot <- autoplot(detrended_renewable) +
                xlab("Time") +
                ylab("Energy[Trillion Btu]")+
                geom_hline(yintercept = mean(detrended_renewable),
                color = "purple",
                linetype = "solid") +
                ggtitle("Detrend Plot") +
                theme_minimal()
#print(detrend_plot)
```

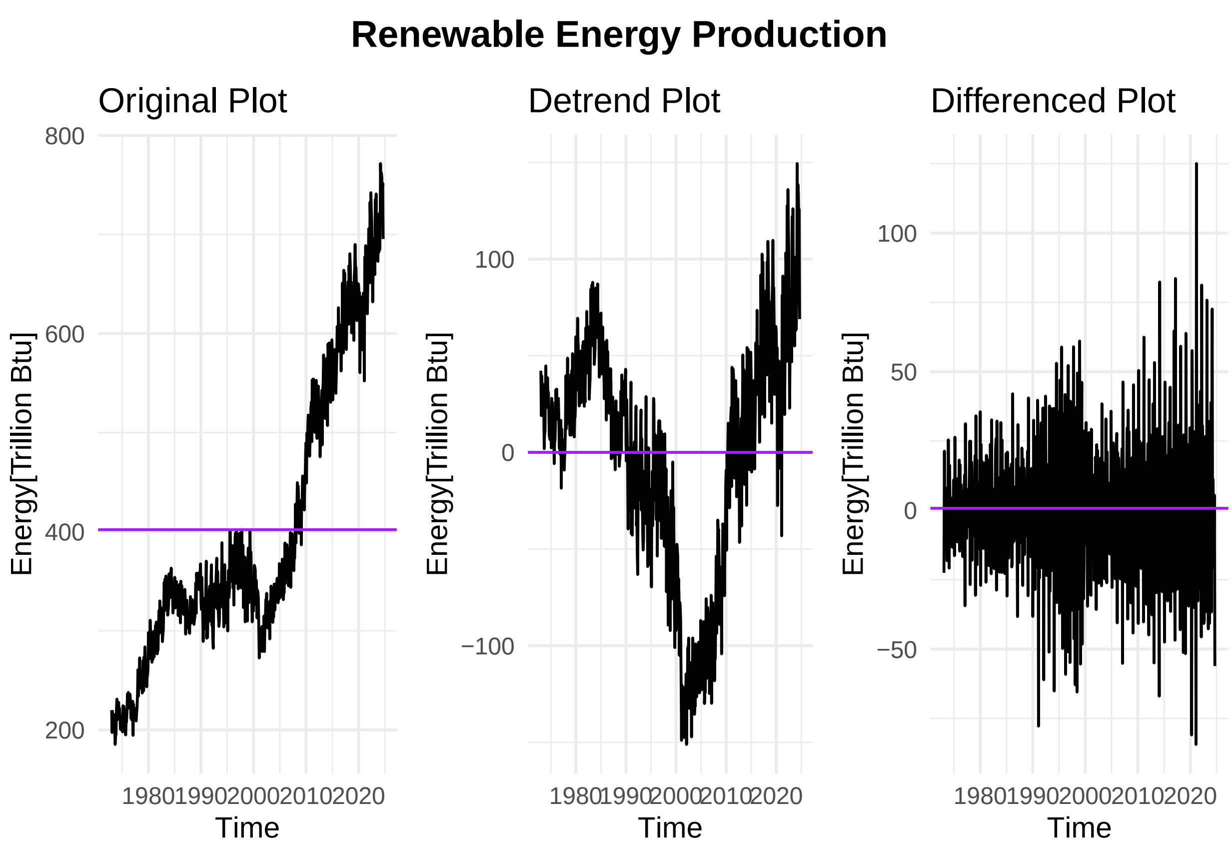
Q3

Now let's compare the differenced series with the detrended series you calculated on A3. In other words, for the "Total Renewable Energy Production" compare the differenced series from Q1 with the series you detrended in Q2 using linear regression.

Using autoplot() + autolayer() create a plot that shows the three series together. Make sure your plot has a legend. The easiest way to do it is by adding the `series=` argument to each autoplot and autolayer function. Look at the key for A03 for an example on how to use autoplot() and autolayer().

What can you tell from this plot? Which method seems to have been more efficient in removing the trend?

```
plot_comparison <- plot_grid(ts_plot,detrend_plot, diff_plot,
                             nrow=1,ncol=3)
#print(plot_comparison)
title <- ggdraw() +
         draw_label("Renewable Energy Production", fontface='bold')
plot_grid(title,plot_comparison,nrow=2,ncol=1,rel_heights = c(0.1,1))
```



Answer: The Original Series showed a strong upward trend, which was partially removed by detrending. However, the detrended series still exhibits significant variation and does not fluctuate around zero. This indicates that the detrending process only partially removed the trend. From the detrended plot, we can observe that the series began trending upward again around the year 2000.

In contrast, first-order differencing successfully removed the trend, as the series now fluctuates around zero, indicating stationarity.

Q4

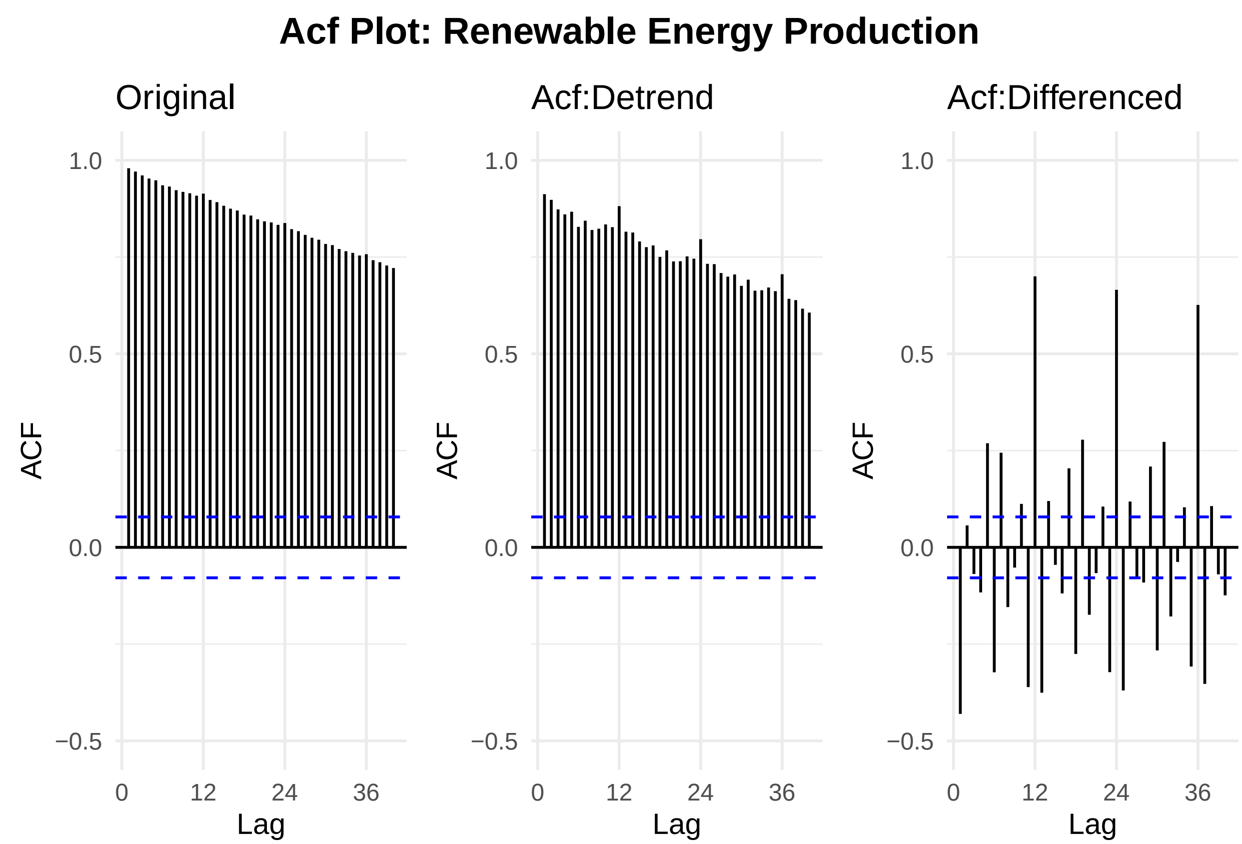
Plot the ACF for the three series and compare the plots. Add the argument `ylim=c(-0.5,1)` to the autoplot() or Acf() function - whichever you are using to generate the plots - to make sure all three y axis have the same limits. Looking at the ACF which method do you think was more efficient in eliminating the trend? The linear regression or differencing?

```r
acf_renewable_ts <- ggAcf(Renewable_Energy_Data_ts, lag.max = 40)+
                    ggtitle("Original") +
                    ylim(c(-0.5,1)) +
                    theme_minimal()
#print(acf_renewable_ts)

acf_detrended_renewable <- ggAcf(detrended_renewable,lag.max = 40) +
                            ggtitle("Acf:Detrend") +
                            ylim(c(-0.5,1)) +
                            theme_minimal()
#print(acf_detrended_renewable)

acf_diff_renewable <-ggAcf(Renewable_Production_diff, lag.max = 40) +
                      ggtitle("Acf:Differenced") +
                      ylim(c(-0.5,1)) +
                      theme_minimal()
#print(acf_diff_renewable)

acf_plot_comparison <- plot_grid(acf_renewable_ts,acf_detrended_renewable,
                                 acf_diff_renewable,
                                nrow=1,ncol=3)
#print(acf_plot_comparison)
title <- ggdraw() +
          draw_label("Acf Plot: Renewable Energy Production", fontface='bold')
plot_grid(title,acf_plot_comparison,nrow=2,ncol=1,rel_heights = c(0.1,1))
```



Answer:

Q5

Compute the Seasonal Mann-Kendall and ADF Test for the original "Total Renewable Energy Production" series. Ask R to print the results. Interpret the results for both test. What is the conclusion from the Seasonal Mann Kendall test? What's the conclusion for the ADF test? Do they match what you observed in Q3 plot? Recall that having a unit root means the series has a stochastic trend. And when a series has stochastic trend we need to use differencing to remove the trend.

```
smk_test <- smk.test(Renewable_Energy_Data_ts)
print(smk_test)
```

```
##
##  Seasonal Mann-Kendall trend test (Hirsch-Slack test)
##
## data:  Renewable_Energy_Data_ts
## z = 28.601, p-value < 2.2e-16
## alternative hypothesis: true S is not equal to 0
## sample estimates:
##      S   varS
##  12468 190008
```

```r
adf_test <- adf.test(Renewable_Energy_Data_ts)
print(adf_test)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  Renewable_Energy_Data_ts
## Dickey-Fuller = -1.0898, Lag order = 8, p-value = 0.9242
## alternative hypothesis: stationary
```

Answer:

Q6

Aggregate the original "Total Renewable Energy Production" series by year. You can use the same procedure we used in class. Store series in a matrix where rows represent months and columns represent years. And then take the columns mean using function colMeans(). Recall the goal is the remove the seasonal variation from the series to check for trend. Convert the accumulates yearly series into a time series object and plot the series using autoplot().

Q7

Apply the Mann Kendall, Spearman correlation rank test and ADF. Are the results from the test in agreement with the test results for the monthly series, i.e., results for Q6?

Answer: