ENV 797 - Time Series Analysis for Energy and Environment Applications | Spring 2026

Assignment 4 - Due date 02/10/26

Aisha Shen

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_Sp26.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
#install.packages(c("forecast","readxl","tseries","Kendall","cowplot", "ggplot2"))
library(readxl)
library(openxlsx)
library(forecast)
library(Kendall)
library(tseries)
library(ggplot2)
```

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 December 2025 Monthly Energy Review. **For this assignment you will work only with the column "Total Renewable Energy Production"**.

```
#Importing data set - you may copy your code from A3
energy_data1 <- read_excel(path="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_Renewabl

#Now let's extract the column names from row 11
read_col_names <- read_excel(path="/Users/meilishen/Documents/TimeSeries/TSA_Sp26/Data/Table_10.1_Renewa
```

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

timeseries_df <- energy_data1 [, "Total Renewable Energy Production"]
head(timeseries_df)
```

```
## # A tibble: 6 x 1
##   `Total Renewable Energy Production`
##                                 <dbl>
## 1                                220.
## 2                                197.
## 3                                219.
## 4                                209.
## 5                                216.
## 6                                208.
```

```r
energy_ts <- ts(timeseries_df, start=c(1931,1),frequency=12 ) # TAKEN FROM A02
```

Stochastic Trend and Stationarity Tests

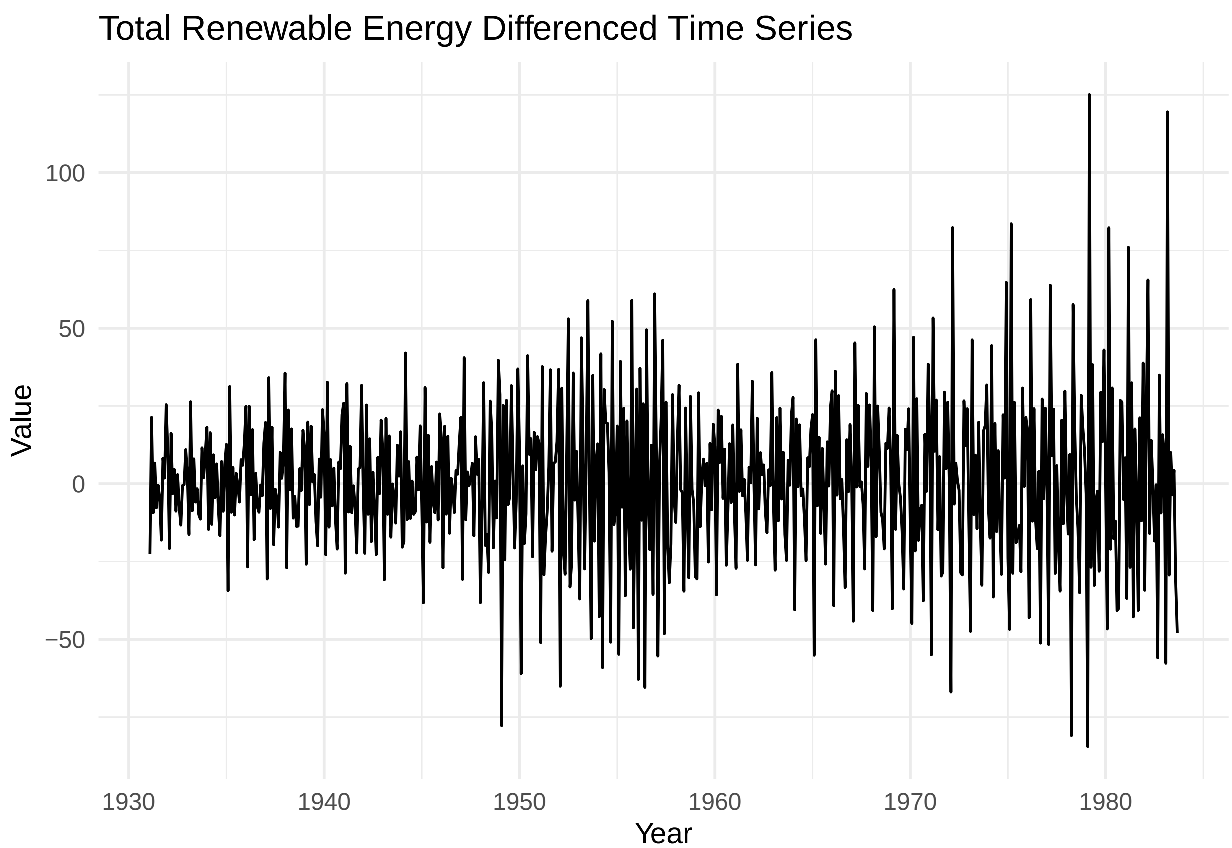
For this part you will work only with the column Total Renewable Energy Production.

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
diff_ts<-diff(energy_ts,1,1)
autoplot(diff_ts[, 1]) +
labs(title = paste("Total Renewable Energy Differenced", "Time Series"), x = "Year", y = "Value")  +
theme_minimal()
```



The time series appears to be heteroskedastic noise, with variance increasing in some periods.

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 assign same name for the time series object that you had in A3, otherwise the code will not work.

```
time_index <- 1:nrow(energy_ts)
model_renewable <- lm(energy_ts[, 1] ~ time_index) # lm(y given x)
trend_renewable <- fitted(model_renewable)
summary(model_renewable)
```

```
##
## Call:
## lm(formula = energy_ts[, 1] ~ time_index)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -154.81  -39.55   12.52   41.49  171.15
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 171.44868    5.11085   33.55   <2e-16 ***
```

```
## time_index    0.74999    0.01397   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
renew_coeffs<-coefficients(model_renewable)
print(renew_coeffs)
```

```
## (Intercept)  time_index
##  171.448682    0.749989
```

```r
detrended_ts<-energy_ts-trend_renewable
```

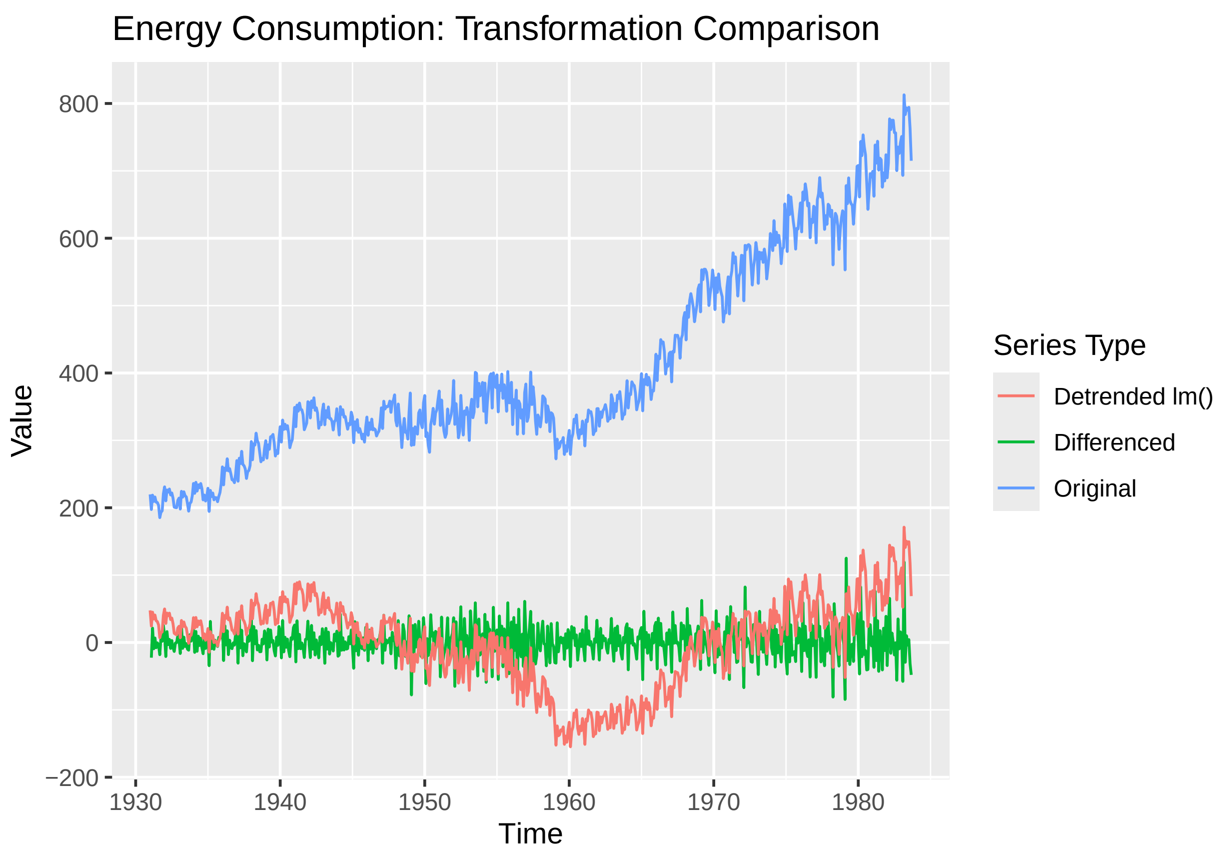
Q3

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 (i.e. "Original", "Differenced", "Detrended lm()"). 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?

```r
autoplot(energy_ts, series = "Original") +
  autolayer(diff_ts, series = "Differenced") +
  autolayer(detrended_ts, series = "Detrended lm()") +
  labs(title = "Energy Consumption: Transformation Comparison",
       y = "Value",
       colour = "Series Type")
```

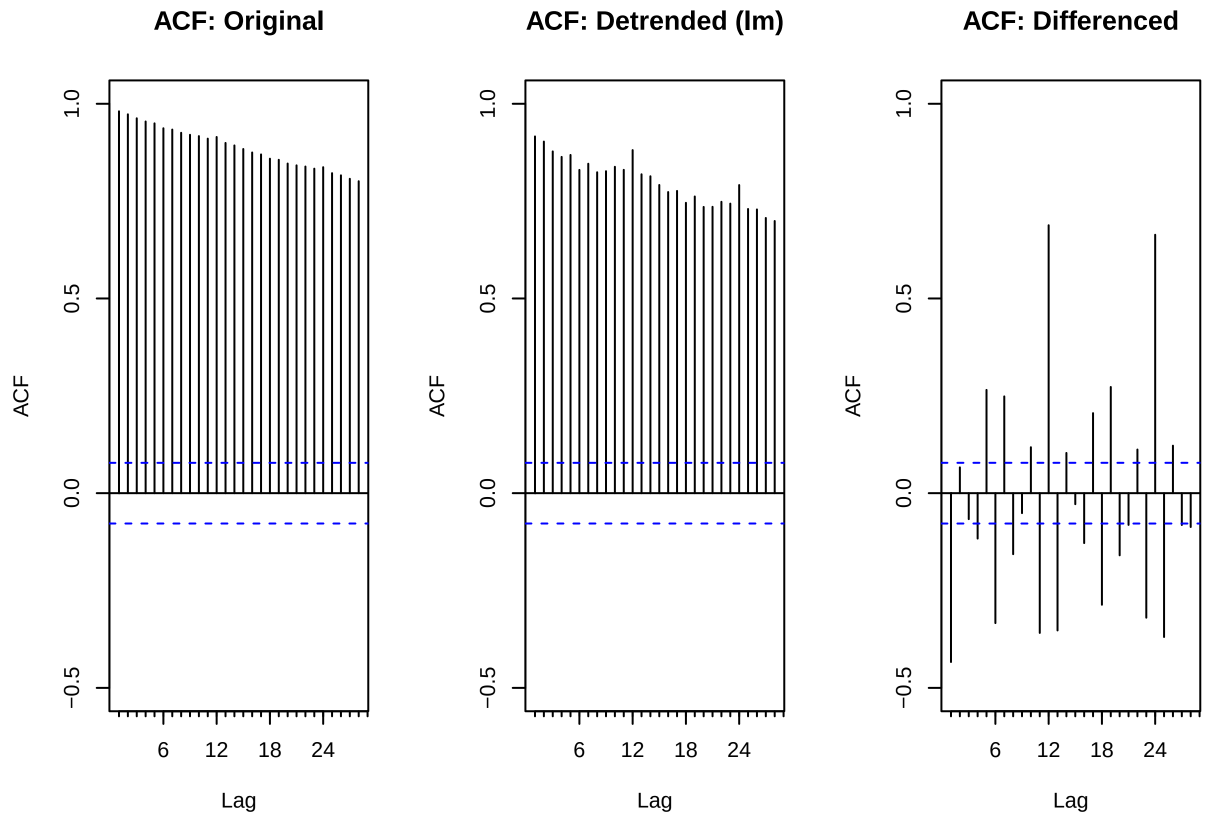


Answer: The Differenced method was significantly more efficient in removing the trend as it results in higher mean stability. This is because there were some non linear trends that the linear model could not remove in the detrended series.

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?

```
par(mfrow = c(1, 3))
p1 <- Acf(energy_ts, main="ACF: Original", ylim=c(-0.5, 1))
p2 <- Acf(detrended_ts, main="ACF: Detrended (lm)", ylim=c(-0.5, 1))
p3 <- Acf(diff_ts, main="ACF: Differenced", ylim=c(-0.5, 1))
```



Answer: The differenced approach was more efficient and resulted in an ACF that looks like "white noise" with uncorrelated observations over time.

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_result <- SeasonalMannKendall(energy_ts[,1])
adf_result <- adf.test(energy_ts[,1])
print(smk_result)
```

```
## tau = 0.799, 2-sided pvalue =< 2.22e-16
```

```
print(adf_result)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  energy_ts[, 1]
## Dickey-Fuller = -1.0247, Lag order = 8, p-value = 0.9347
## alternative hypothesis: stationary
```

Answer: In my understanding the Seasonal Mann-Kendall Test is testing the hypothesis that there is a monotonic trend, while the null hypothesis is that there is no such trend. Since the p value for the Seasonal Mann-Kendall test is extremely small <0.05 we can reject the null and therefore there is a statistically significant trend in the "Total Renewable Energy Production" over time. Conversely, in the ADF test the null hypothesis is that the series has unit root and is not stationary while the alternate hypothesis is that it is stationary. Since we get a high p value we can not reject the null hypothesis. These observations are consistent with what we see, since a clear visual trend is visible in the time series. It is also consistent with the ACF plot where we see a gradual decay over time, with significant correlation between subsequent time steps.

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().

```r
iHP <- 1

data_matrix <- matrix(energy_ts[, iHP], nrow = 12, byrow = FALSE)
```

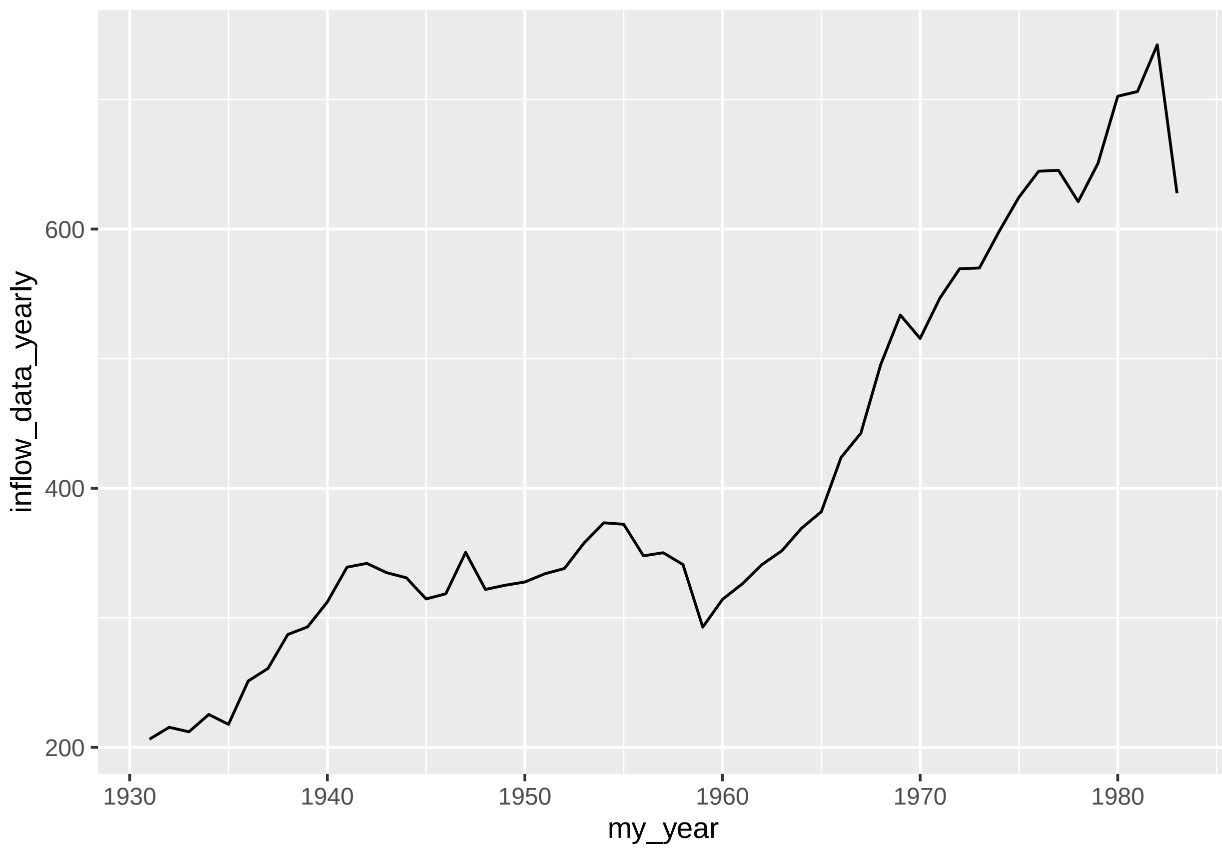
```
## Warning in matrix(energy_ts[, iHP], nrow = 12, byrow = FALSE): data length
## [633] is not a sub-multiple or multiple of the number of rows [12]
```

```r
inflow_data_yearly <- colMeans(data_matrix)

start_year <- start(energy_ts)[1]
my_year <- start_year:(start_year + length(inflow_data_yearly) - 1)

inflow_data_new_yearly <- data.frame(
  my_year = my_year,
  inflow_data_yearly = inflow_data_yearly
)

library(ggplot2)
energy_ts_yearly<-ts(inflow_data_new_yearly)
ggplot(inflow_data_new_yearly, aes(x = my_year, y = inflow_data_yearly)) +
  geom_line()
```



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 Q5?

```
smk_result <- SeasonalMannKendall(energy_ts_yearly[,1])
adf_result <- adf.test(energy_ts_yearly[,1])
spearman_result<-cor.test(energy_ts_yearly[,1],my_year,method="spearman")
print(smk_result)
```

```
## tau = 1, 2-sided pvalue =< 2.22e-16
```

```
print(adf_result)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  energy_ts_yearly[, 1]
## Dickey-Fuller = -1.7321, Lag order = 3, p-value = 0.6823
## alternative hypothesis: stationary
```

```
print(spearman_result)
```

```
##
##  Spearman's rank correlation rho
##
## data:  energy_ts_yearly[, 1] and my_year
## S = 2.7538e-12, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
## rho
##   1
```

> Answer: It is in the agreement with the monthly result. Since the p value for the Seasonal Mann-Kendall test is extremely small <0.05 we can reject the null and therefore there is a statistically significant trend in the "Total Renewable Energy Production" over time. In the ADF test, since we get a high p value we can not reject the null hypothesis. For the Spearman's correlation test, since the p value is close to 0, and we see a sample estimate of rho to be 1, this shows a clear monotonic relationship.