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

Assignment 4 - Due date 02/10/26

Loo Si Min

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
library(forecast)
library(tseries)
library(Kendall)
library(readxl)
library(ggplot2)
```

Questions

Consider the same data you used for A3 from the spreadsheet "Table_10.1_Renewable_Energy_Production_and_Consumpt
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_data <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sourc

read_col_names <- read_excel(path="../Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_So

colnames(energy_data) <- read_col_names

df <- data.frame(
  Month = energy_data$`Month`,
  Renewable = energy_data$`Total Renewable Energy Production`
```

```
)

head(df)
```

```
##        Month Renewable
## 1 1973-01-01   219.839
## 2 1973-02-01   197.330
## 3 1973-03-01   218.686
## 4 1973-04-01   209.330
## 5 1973-05-01   215.982
## 6 1973-06-01   208.249
```

Stochastic Trend and Stationarity Tests

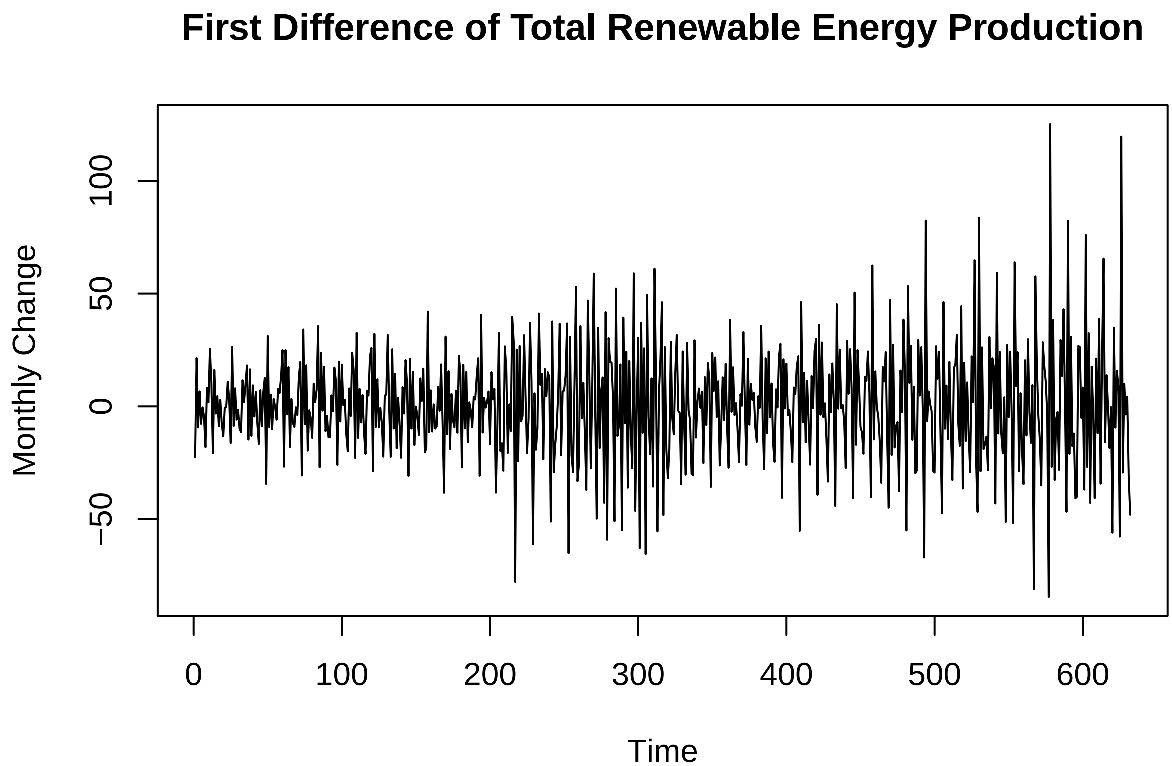
For this part you will work only with the column Total Renewable Energy Production.

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?

```
# First difference at lag 1
renew_diff <- diff(df$Renewable, lag = 1, differences = 1)

# Plot the differenced series
plot(renew_diff,
     type = "l",
     main = "First Difference of Total Renewable Energy Production",
     ylab = "Monthly Change",
     xlab = "Time")
```



After differencing the Total Renewable Energy Production series once at lag 1, the resulting series no longer exhibits a clear trend. The values fluctuate around a constant mean close to zero, indicating that the upward trend present in the original series has been removed. However, noticeable seasonal fluctuations and changing variability remain, suggesting the series is not yet fully stationary.

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.

```r
# Time index
nobs <- nrow(df)
t <- 1:nobs

tsdata <- ts(df$Renewable, frequency = 12, start = c(1973,1))

# Linear trend regression
# Linear trend regression (same names as A3)
regmodel_renewable <- lm(tsdata ~ t)

beta0_renewable <- regmodel_renewable$coefficients[1]
beta1_renewable <- regmodel_renewable$coefficients[2]

renewable_detrend <- tsdata - (beta0_renewable + beta1_renewable * t)
renewable_detrend <- ts(renewable_detrend, frequency = 12, start = c(1973,1))
```

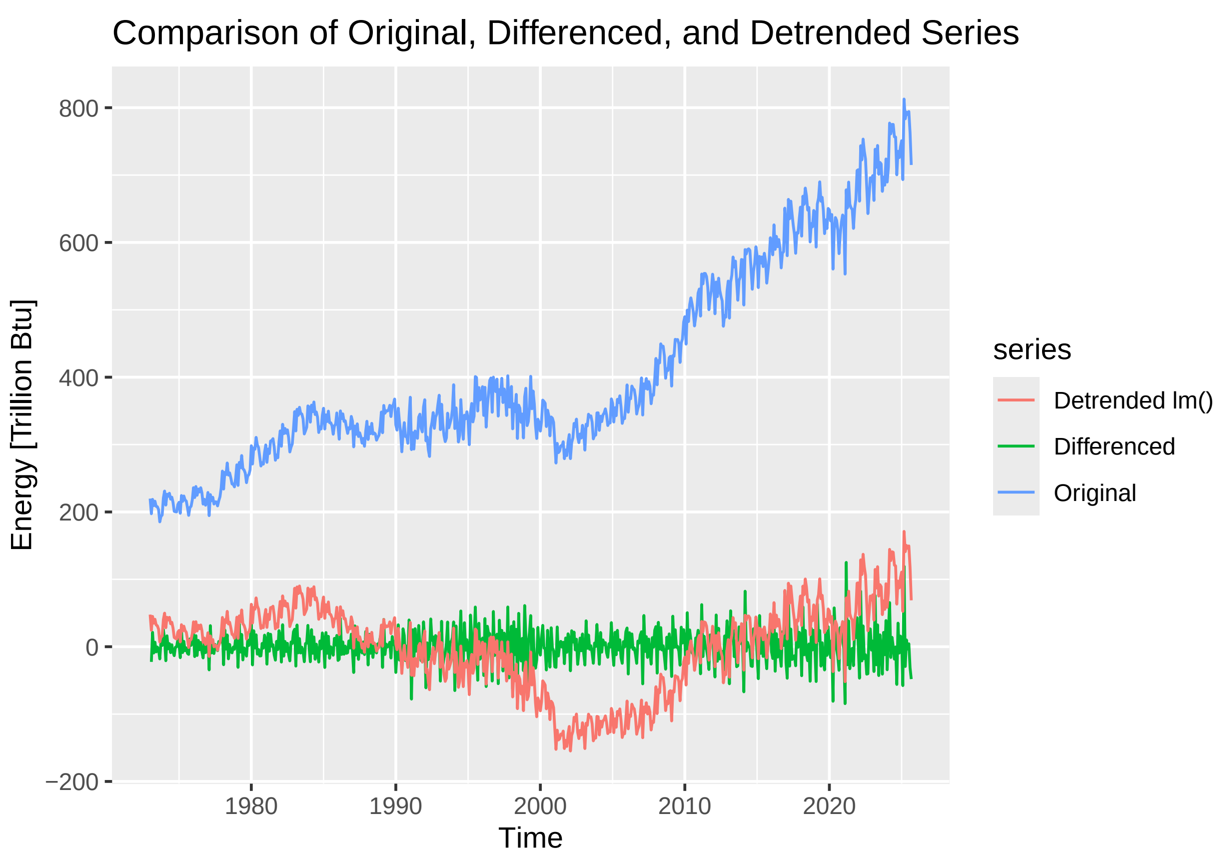
Q3

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
# Convert differenced series to ts (align start)
renew_diff_ts <- ts(renew_diff, frequency = 12, start = c(1973,2))

# Plot original, differenced, and detrended series together
autoplot(tsdata, series = "Original") +
  autolayer(renew_diff_ts, series = "Differenced") +
  autolayer(renewable_detrend, series = "Detrended lm()") +
  ylab("Energy [Trillion Btu]") +
  ggtitle("Comparison of Original, Differenced, and Detrended Series")
```



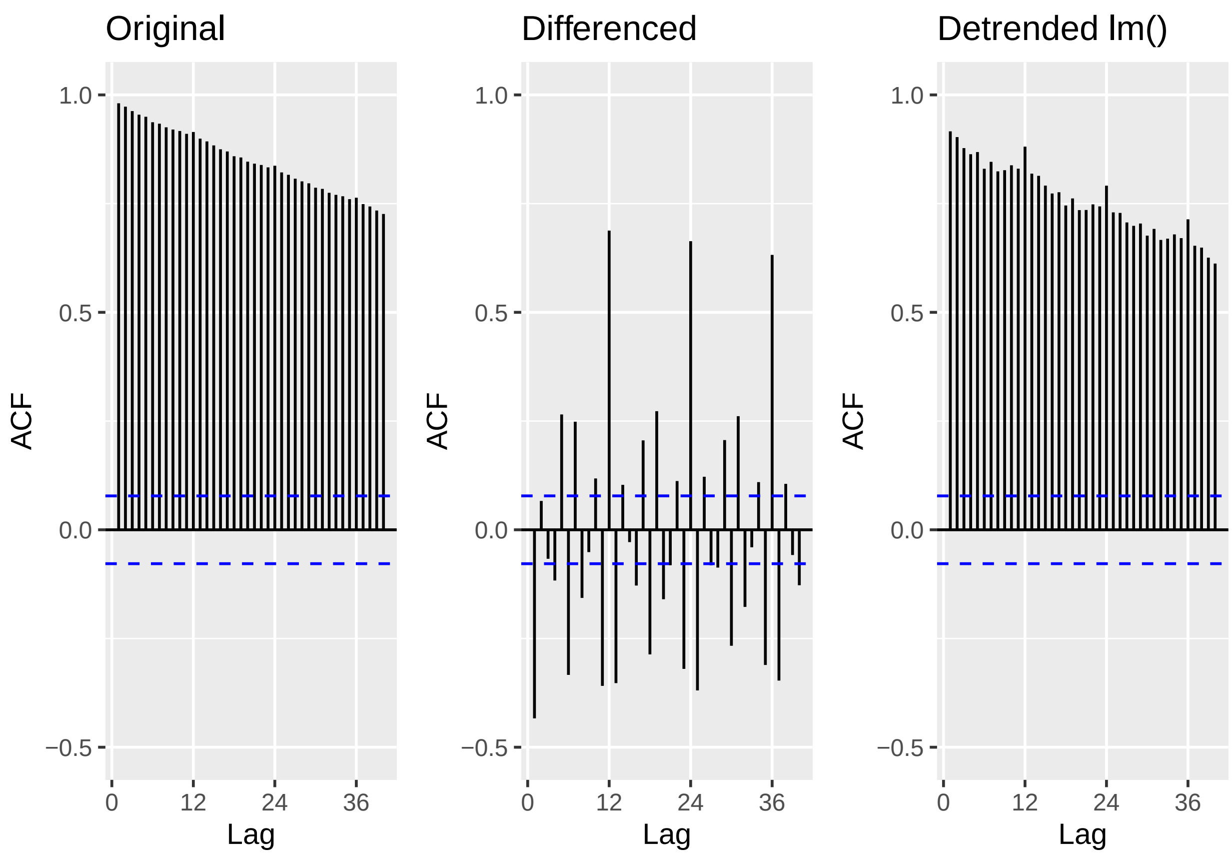
Answer: From the plot, the first-differenced series fluctuates closely around zero with no visible long-term trend, indicating that differencing effectively removes the trend component. In contrast, the linearly detrended series still shows low-frequency movements and systematic deviations over time, suggesting that a simple linear trend is insufficient to capture the underlying growth pattern in renewable energy production. Therefore, first differencing is more effective than linear detrending in removing the trend for this 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?

```r
library(cowplot)

# ACF comparison
plot_grid(
  autoplot(Acf(tsdata, lag.max = 40, plot = FALSE),
           main = "Original",
           ylim = c(-0.5, 1)),

  autoplot(Acf(renew_diff_ts, lag.max = 40, plot = FALSE),
           main = "Differenced",
           ylim = c(-0.5, 1)),

  autoplot(Acf(renewable_detrend, lag.max = 40, plot = FALSE),
           main = "Detrended lm()",
           ylim = c(-0.5, 1)),

  nrow = 1
)
```

```
## Warning in ggplot2::geom_segment(lineend = "butt", ...): Ignoring unknown parameters: `main` and `yl
## Ignoring unknown parameters: `main` and `ylim`
## Ignoring unknown parameters: `main` and `ylim`
```



Answer: The ACF of the original series shows very slow decay with large, persistent autocorrelations, indicating strong trend and non-stationarity. After linear detrending, the ACF remains highly persistent with slow decay, suggesting that the linear regression did not fully remove the trend. In contrast, the ACF of the differenced series drops off rapidly and oscillates around zero, indicating that first differencing is more effective than linear detrending in eliminating the trend.

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.

```
# Convert to ts object (monthly)
renew_ts <- ts(df$Renewable, frequency = 12, start = c(1973,1))

# Seasonal Mann-Kendall test
smk_test <- SeasonalMannKendall(renew_ts)
print(smk_test)
```

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

```
# Augmented Dickey-Fuller test
adf_test <- adf.test(renew_ts)
print(adf_test)
```

```
##
```

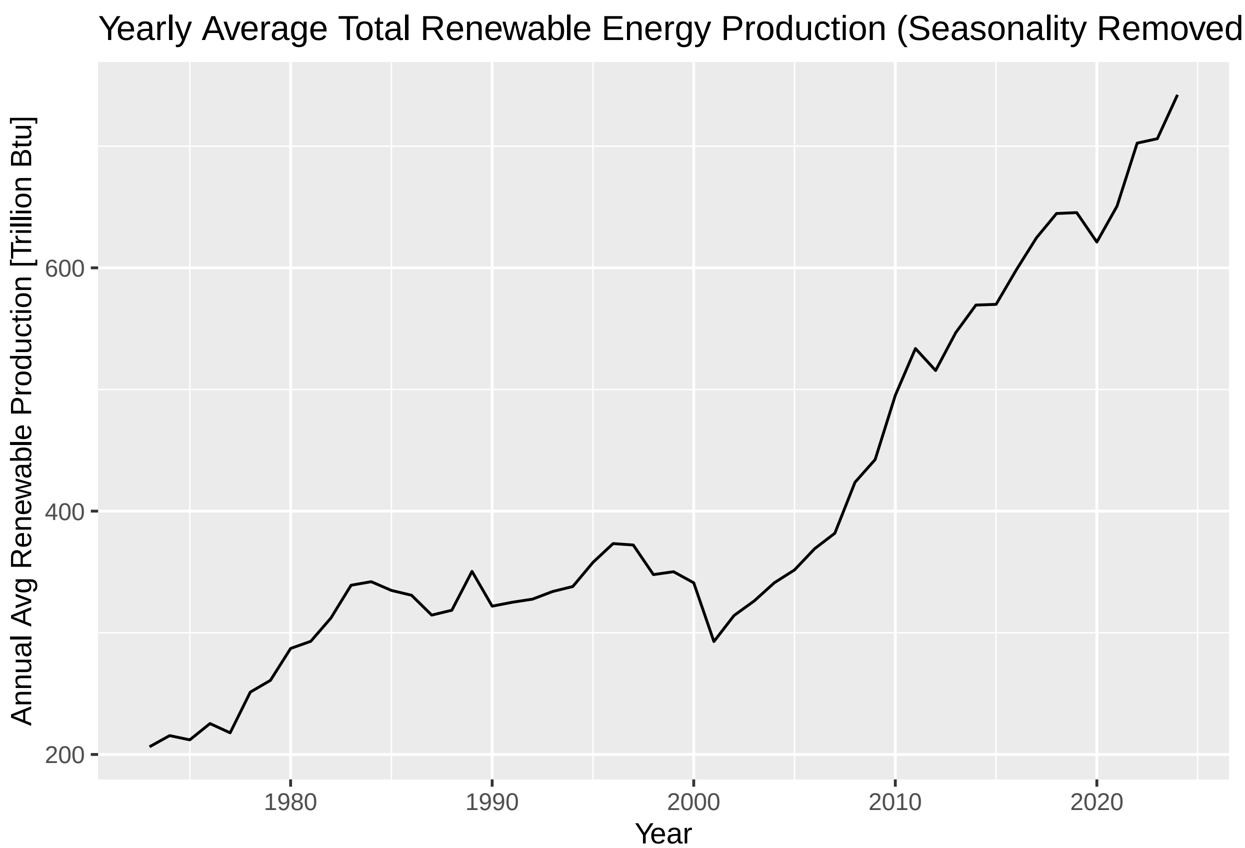
```
##  Augmented Dickey-Fuller Test
##
## data:  renew_ts
## Dickey-Fuller = -1.0247, Lag order = 8, p-value = 0.9347
## alternative hypothesis: stationary
```

> Answer: The Seasonal Mann–Kendall test yields a Kendall's tau = 0.799 with a p-value < 2.22×10^-16, indicating a strong and statistically significant increasing monotonic trend in Total Renewable Energy Production after accounting for seasonality. The Augmented Dickey–Fuller test fails to reject the null hypothesis of a unit root (p-value = 0.9347), indicating that the series is non-stationary and contains a stochastic trend. The results of both tests are consistent and align with the visual evidence from Q3 and Q4. The series exhibits a strong upward trend that is stochastic in nature, implying that first differencing is necessary to remove the trend, rather than relying on linear detrending.

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
renew_ts <- ts(df$Renewable, frequency = 12, start = c(1973,1))

# Aggregate to yearly means (remove seasonality)
y <- as.numeric(renew_ts)
n <- length(y)

# Trim to full years (12 months each)
n_full <- floor(n/12) * 12
y_full <- y[1:n_full]

# Store in matrix: rows = months, cols = years
mat <- matrix(y_full, nrow = 12, byrow = FALSE)

# Yearly average (column means)
renew_yearly <- colMeans(mat, na.rm = TRUE)

# Convert to ts object (yearly frequency) and plot
start_year <- start(renew_ts)[1]
renew_yearly_ts <- ts(renew_yearly, start = start_year, frequency = 1)

autoplot(renew_yearly_ts) +
  ylab("Annual Avg Renewable Production [Trillion Btu]") +
  xlab("Year") +
  ggtitle("Yearly Average Total Renewable Energy Production (Seasonality Removed)")
```



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?

```r
# Mann-Kendall test
mk_test_yearly <- MannKendall(renew_yearly_ts)
print(mk_test_yearly)
```

```
## tau = 0.817, 2-sided pvalue =< 2.22e-16
```

```r
# Spearman rank correlation test
spearman_test_yearly <- cor.test(
  time(renew_yearly_ts),
  as.numeric(renew_yearly_ts),
  method = "spearman"
)
print(spearman_test_yearly)
```

```
##
##  Spearman's rank correlation rho
##
## data:  time(renew_yearly_ts) and as.numeric(renew_yearly_ts)
## S = 1852, p-value < 2.2e-16
## alternative hypothesis: true rho is not equal to 0
## sample estimates:
##       rho
## 0.9209425
```

```r
# Augmented Dickey-Fuller test
adf_test_yearly <- adf.test(renew_yearly_ts)
print(adf_test_yearly)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  renew_yearly_ts
## Dickey-Fuller = -0.85301, Lag order = 3, p-value = 0.9515
## alternative hypothesis: stationary
```

> Answer: The Mann–Kendall test yields tau = 0.817 with p-value < 2.22×10^-16, indicating a strong and statistically significant increasing monotonic trend in the yearly aggregated Total Renewable Energy Production series. The Spearman test shows a very strong positive rank correlation (= 0.921, p-value < 2.2×10^-16) between time and renewable energy production, further confirming a persistent upward trend. The ADF test fails to reject the null hypothesis of a unit root (p-value = 0.9515), indicating that the yearly series is non-stationary and contains a stochastic trend. The results for the yearly series are consistent with the monthly results in Q5. Both frequencies show a strong increasing trend and evidence of a stochastic trend. While aggregating to yearly data removes seasonality, it does not eliminate non-stationarity, confirming that differencing is still required to remove the trend.