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

Assignment 3 - Due date 02/04/25

Jessalyn Chuang

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

##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)

```r
energy_data <- read_excel(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Source
```

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

```r
#Extract column names from from row 11
read_col_names <- read_excel(path="./Data/Table_10.1_Renewable_Energy_Production_and_Consumption_by_Sou
```

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

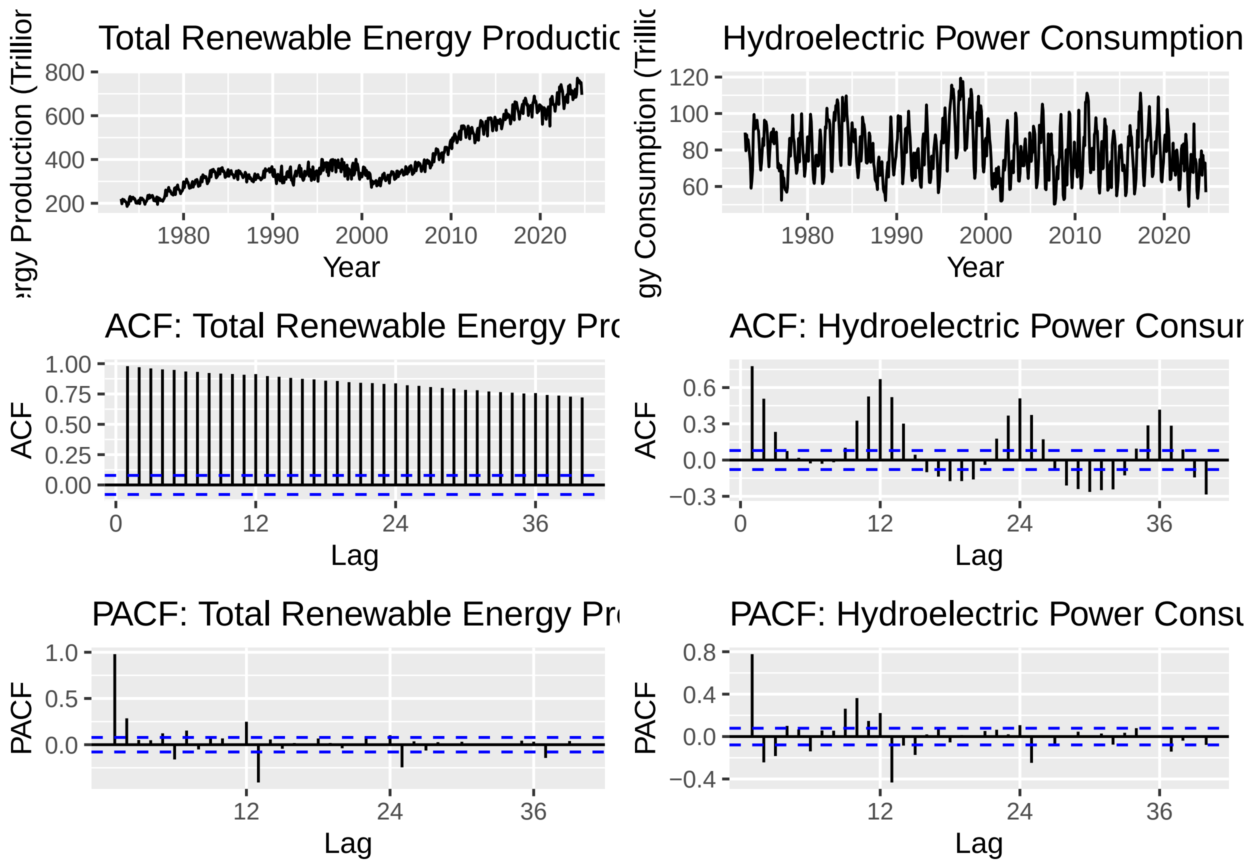
```r
#Assign the column names to the data set
colnames(energy_data) <- read_col_names

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

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

```r
#pulling only Renewable Energy Production and Hydroelectric Consumption
energy_subset <- energy_data[, 5:6]

#turning into a time series
energy_subset_ts <- ts(energy_subset[,1:2],start=c(1973,1),frequency=12)
```

```r
#time series
ts_renewable <- autoplot(energy_subset_ts[,1], main = colnames(energy_subset_ts)[1],
                         xlab = "Year",ylab = "Energy Production (Trillion Btu)")
ts_hydro <- autoplot(energy_subset_ts[,2], main = colnames(energy_subset_ts)[2],
                     xlab = "Year", ylab = "Energy Consumption (Trillion Btu)")

# ACF and PACF using ggAcf() and ggPacf()
acf_renewable <- ggAcf(energy_subset_ts[,1], lag.max = 40) +
  ggtitle(paste("ACF:", colnames(energy_subset_ts)[1]))
acf_hydro <- ggAcf(energy_subset_ts[,2], lag.max = 40) +
  ggtitle(paste("ACF:", colnames(energy_subset_ts)[2]))

pacf_renewable <- ggPacf(energy_subset_ts[,1], lag.max = 40) +
  ggtitle(paste("PACF:", colnames(energy_subset_ts)[1]))
pacf_hydro <- ggPacf(energy_subset_ts[,2], lag.max = 40) +
  ggtitle(paste("PACF:", colnames(energy_subset_ts)[2]))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(ts_renewable, ts_hydro, acf_renewable, acf_hydro, pacf_renewable,
          pacf_hydro, ncol = 2)
```



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?

Total Renewable Energy Production seems to have a clear long-term upward trend. This time series plot shows a steady increase over time, suggesting that renewable energy production has been growing consistently. The ACF shows strong positive autocorrelation with the ACF values being high and decreasing gradually over time. From the PACF, there may be some seasonality with larger PACF values above the blue line at evenly spaced lags.

Hydroelectric power consumption does not have a clear long-term upward or downward trend. It instead fluctuates around a relatively stable mean, suggesting stationarity in the mean but with seasonality present. This is shown in the time series plot, which shows repetition in peaks and troughs over time. The ACF plot supports this since it shows significant autocorrelations at particular lags, indicating a recurring seasonal pattern. The PACF further supports this by showing notable spikes, suggesting that hydroelectric consumption depends on past seasonal patterns.

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.

```r
nobs <- nrow(energy_subset)
col_renewable <- 1
col_hydro <- 2
#Create vector t
t <- c(1:nobs)

#Fit a linear trend to Renewable Energy Production
renewable_linear_model <- lm(energy_subset[[col_renewable]] ~ t)
summary(renewable_linear_model)
```

```
##
## Call:
## lm(formula = energy_subset[[col_renewable]] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -151.11  -37.84   13.53   41.76  149.42
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 176.87293    4.96189   35.65   <2e-16 ***
## t             0.72393    0.01382   52.37   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 61.75 on 619 degrees of freedom
## Multiple R-squared:  0.8159, Adjusted R-squared:  0.8156
## F-statistic:  2743 on 1 and 619 DF,  p-value: < 2.2e-16
```

```r
renewable_beta0 <- as.numeric(renewable_linear_model$coefficients[1])
renewable_beta1 <- as.numeric(renewable_linear_model$coefficients[2])
```

For Renewable Energy Production, the intercept of 176.87 is the estimated renewable energy production at the start of the dataset when t = 0. The slope being 0.7239 indicates that renewable energy is increasing by this many units per time step. Both the intercept and time coefficient are highly significant, meaning that the null hypothesis that there is no linear relationship between time and this seris can be rejected. There is indeed a strong and significant upward trend in renewable energy production over time.

```r
#Fit a linear trend to Hydroelectric Power Consumption
hydro_linear_model <- lm(energy_subset[[col_hydro]] ~ t)
summary(hydro_linear_model)
```

```
##
## Call:
## lm(formula = energy_subset[[col_hydro]] ~ t)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -29.995 -10.422  -0.720   9.161  39.624
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) 82.96766    1.12339  73.855  < 2e-16 ***
## t           -0.01098    0.00313  -3.508 0.000485 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 13.98 on 619 degrees of freedom
## Multiple R-squared:  0.01949,    Adjusted R-squared:  0.01791
## F-statistic:  12.3 on 1 and 619 DF,  p-value: 0.0004848
```

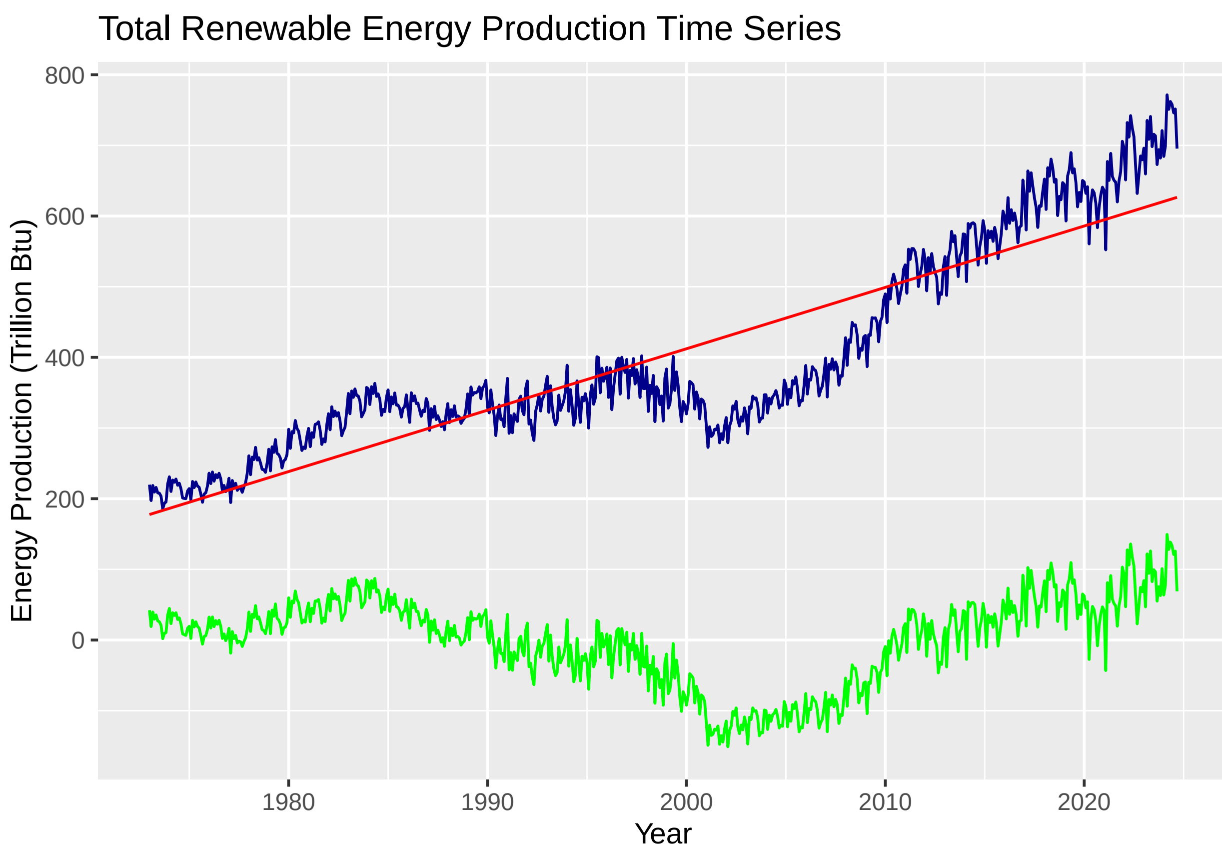
```r
hydro_beta0 <- as.numeric(hydro_linear_model$coefficients[1])
hydro_beta1 <- as.numeric(hydro_linear_model$coefficients[2])
```

The intercept of 82.97 represents the estimated hydroelectric power consumption at the start of the dataset. The slop of -0.01098 being so small suggests that there is just a very small decrease over time. However, since the intercept and slope coefficients have p-values that are much smaller than 0.05, we can still reject the null hypothesis and conclude that time does have a statistically significant effect on hydroelectric power consumption. The low R-squared value (0.01949) suggests that while there is a trend, time alone does not explain much of the variability in hydro consumption.

Q4

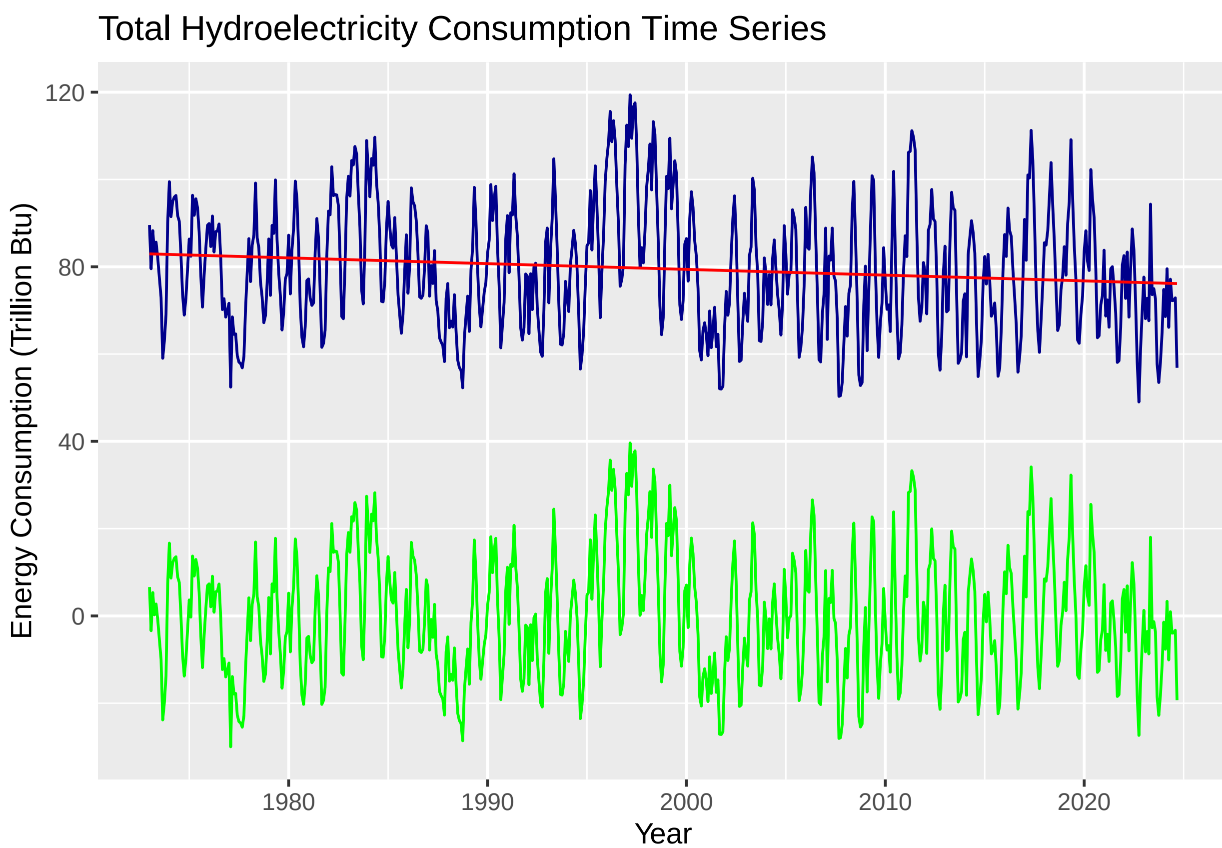
Use the regression coefficients from Q3 to detrend the series. Plot the detrended series and compare with the plots from Q1. What happened? Did anything change?

```r
renewable_linear_trend <- renewable_beta0 + renewable_beta1 * t
ts_renewable_linear <- ts(renewable_linear_trend,star=c(1973,1),frequency=12)

detrend_renewable <- energy_subset[,col_renewable] - renewable_linear_trend
ts_renewable_detrend <- ts(detrend_renewable, start = c(1973,1),frequency = 12)

#Plot
autoplot(energy_subset_ts[,col_renewable],color="darkblue")+
  autolayer(ts_renewable_detrend,series="Detrended",color="green")+
  autolayer(ts_renewable_linear,series="Linear Component",color="red")+
  labs(title = "Total Renewable Energy Production Time Series")+
  ylab("Energy Production (Trillion Btu)") +
  xlab("Year")
```



After detrending the renewable production time series, the increasing trend present in the original data (blue) was removed, as shown by the red trend line. The detrended series (green) now fluctuates around a mean of zero, indicating that the long-term upward movement has been eliminated. This transformation improves stationarity in that the mean and variance remain more stable over time. While whatever seasonal trend is still present, the overall trend component has been extracted.

```r
hydro_linear_trend <- hydro_beta0 + hydro_beta1 * t
ts_hydro_linear <- ts(hydro_linear_trend,start=c(1973,1),frequency=12)

detrend_hydro <- energy_subset[,col_hydro] - hydro_linear_trend
ts_hydro_detrend <- ts(detrend_hydro, start = c(1973,1),frequency = 12)

#Plot
autoplot(energy_subset_ts[,col_hydro],color="darkblue")+
  autolayer(ts_hydro_detrend,series="Detrended",color="green")+
  autolayer(ts_hydro_linear,series="Linear Component",color="red")+
  labs(title = "Total Hydroelectricity Consumption Time Series")+
  ylab("Energy Consumption (Trillion Btu)") +
  xlab("Year")
```

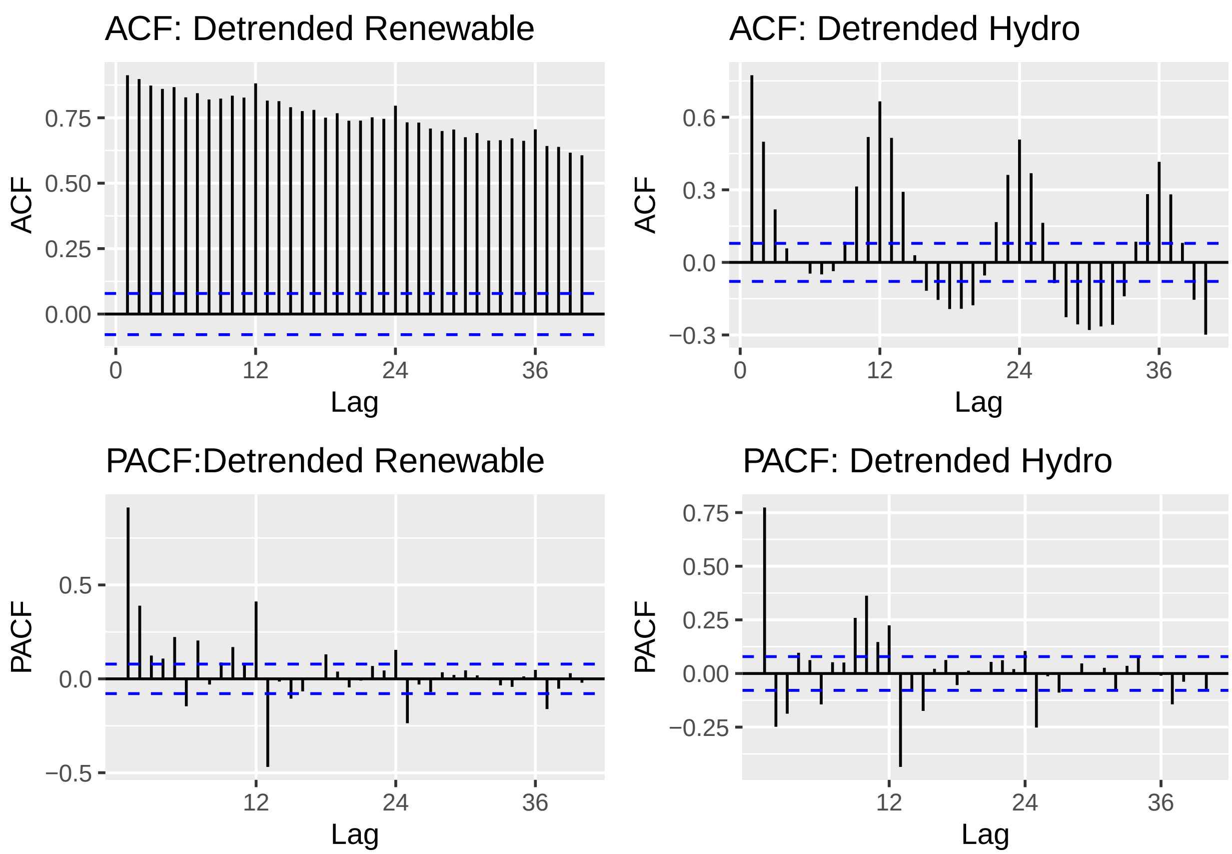


After detrending the hydroelectricity consumption time series, the overall trend present (red) in the original data (blue) has been removed. The detrended series (green) now oscillates around a mean of zero, demonstrating that any long-term upward or downward drift has been eliminated. Unlike the renewable data, the original data does not exhibit a strong trend, meaning the detrended series primarily captures seasonal and cyclical variations.

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?

```r
# ACF and PACF using ggAcf() and ggPacf()
acf_renewable_detrend <- ggAcf(ts_renewable_detrend, lag.max = 40) +
  ggtitle(paste("ACF: Detrended Renewable"))
acf_hydro_detrend <- ggAcf(ts_hydro_detrend, lag.max = 40) +
  ggtitle(paste("ACF: Detrended Hydro"))

pacf_renewable_detrend <- ggPacf(ts_renewable_detrend, lag.max = 40) +
  ggtitle(paste("PACF:Detrended Renewable"))
pacf_hydro_detrend <- ggPacf(ts_hydro_detrend, lag.max = 40) +
  ggtitle(paste("PACF: Detrended Hydro"))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(acf_renewable_detrend, acf_hydro_detrend, pacf_renewable_detrend,
          pacf_hydro_detrend, ncol = 2)
```



After detrending, the ACF and PACF plots changed, indicating the removal of the trend component. Initially, the ACF of total renewable energy production showed a slow decay, suggesting strong autocorrelation, while hydroelectric power consumption displayed noticeable but less persistent correlations. After detrending, long-term dependencies in both series have weakened, particularly in the hydro chart. While the ACF for renewable energy still retains a somewhat similar structure, correlations at lags that are multiples of 12 have become more pronounced, suggesting the presence of some seasonal variation, though not as strongly as in the hydro chart. Additionally, the PACF for renewable energy now reflects some seasonality, with large partial autocorrelations at earlier lags that oscillate. In contrast, the ACF and PACF for hydroelectric power consumption did not change as significantly, indicating that the original trend in this series was relatively small.

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.

Both time series exhibit distinct patterns, but the presence of seasonality differs between them. The total renewable energy production series shows a strong upward trend over time, as seen in the time series plot. Its ACF plot displays high autocorrelation at all lags, characteristic of a trending series, but lacks a clear cyclical pattern, suggesting that any seasonal effects are weak or overshadowed by the dominant trend. In contrast, the hydroelectric power consumption series exhibits noticeable fluctuations that appear to repeat over time, indicating a strong seasonal component. This is further confirmed by the ACF plot, which shows significant spikes at regular interval. Overall, while hydroelectric power consumption clearly exhibits seasonality, total renewable energy production is primarily trend-driven, with any potential seasonal influences being less pronounced.

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?

```r
dummies_renewable <- seasonaldummy(energy_subset_ts[,1])

seas_means_model_renewable <- lm(energy_subset_ts[,1] ~ dummies_renewable)
summary(seas_means_model_renewable)
```

```
##
## Call:
## lm(formula = energy_subset_ts[, 1] ~ dummies_renewable)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -205.65  -91.59  -54.59  117.87  356.19
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)           410.598     20.228  20.299   <2e-16 ***
## dummies_renewableJan    1.973     28.468   0.069    0.945
## dummies_renewableFeb  -34.348     28.468  -1.207    0.228
## dummies_renewableMar    4.721     28.468   0.166    0.868
## dummies_renewableApr   -7.896     28.468  -0.277    0.782
## dummies_renewableMay    7.199     28.468   0.253    0.800
## dummies_renewableJun   -3.394     28.468  -0.119    0.905
## dummies_renewableJul    2.850     28.468   0.100    0.920
## dummies_renewableAug   -4.430     28.468  -0.156    0.876
## dummies_renewableSep  -29.037     28.468  -1.020    0.308
## dummies_renewableOct  -20.205     28.606  -0.706    0.480
## dummies_renewableNov  -20.706     28.606  -0.724    0.469
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 144.5 on 609 degrees of freedom
## Multiple R-squared:  0.008696,   Adjusted R-squared:  -0.009209
## F-statistic: 0.4857 on 11 and 609 DF,  p-value: 0.9126
```

```r
beta_intercept_renewable <- seas_means_model_renewable$coefficients[1]
beta_coeff_renewable <- seas_means_model_renewable$coefficients[2:12]
```

```r
dummies_hydro <- seasonaldummy(energy_subset_ts[,2])

seas_means_model_hydro <- lm(energy_subset_ts[,2] ~ dummies_hydro)
summary(seas_means_model_hydro)
```

```
##
## Call:
## lm(formula = energy_subset_ts[, 2] ~ dummies_hydro)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -31.101  -6.241  -0.444   6.410  32.363
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         79.981      1.452  55.094  < 2e-16 ***
## dummies_hydroJan     4.941      2.043   2.418 0.015880 *
## dummies_hydroFeb    -2.513      2.043  -1.230 0.219219
## dummies_hydroMar     7.053      2.043   3.452 0.000595 ***
## dummies_hydroApr     5.399      2.043   2.642 0.008441 **
## dummies_hydroMay    13.907      2.043   6.807 2.40e-11 ***
## dummies_hydroJun    10.729      2.043   5.251 2.09e-07 ***
## dummies_hydroJul     4.033      2.043   1.974 0.048845 *
## dummies_hydroAug    -5.399      2.043  -2.643 0.008436 **
## dummies_hydroSep   -16.596      2.043  -8.123 2.54e-15 ***
## dummies_hydroOct   -16.370      2.053  -7.973 7.66e-15 ***
## dummies_hydroNov   -10.805      2.053  -5.263 1.97e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.37 on 609 degrees of freedom
## Multiple R-squared:  0.4695, Adjusted R-squared:  0.4599
## F-statistic:    49 on 11 and 609 DF,  p-value: < 2.2e-16
```

```r
beta_intercept_hydro <- seas_means_model_hydro$coefficients[1]
beta_coeff_hydro <- seas_means_model_hydro$coefficients[2:12]
```

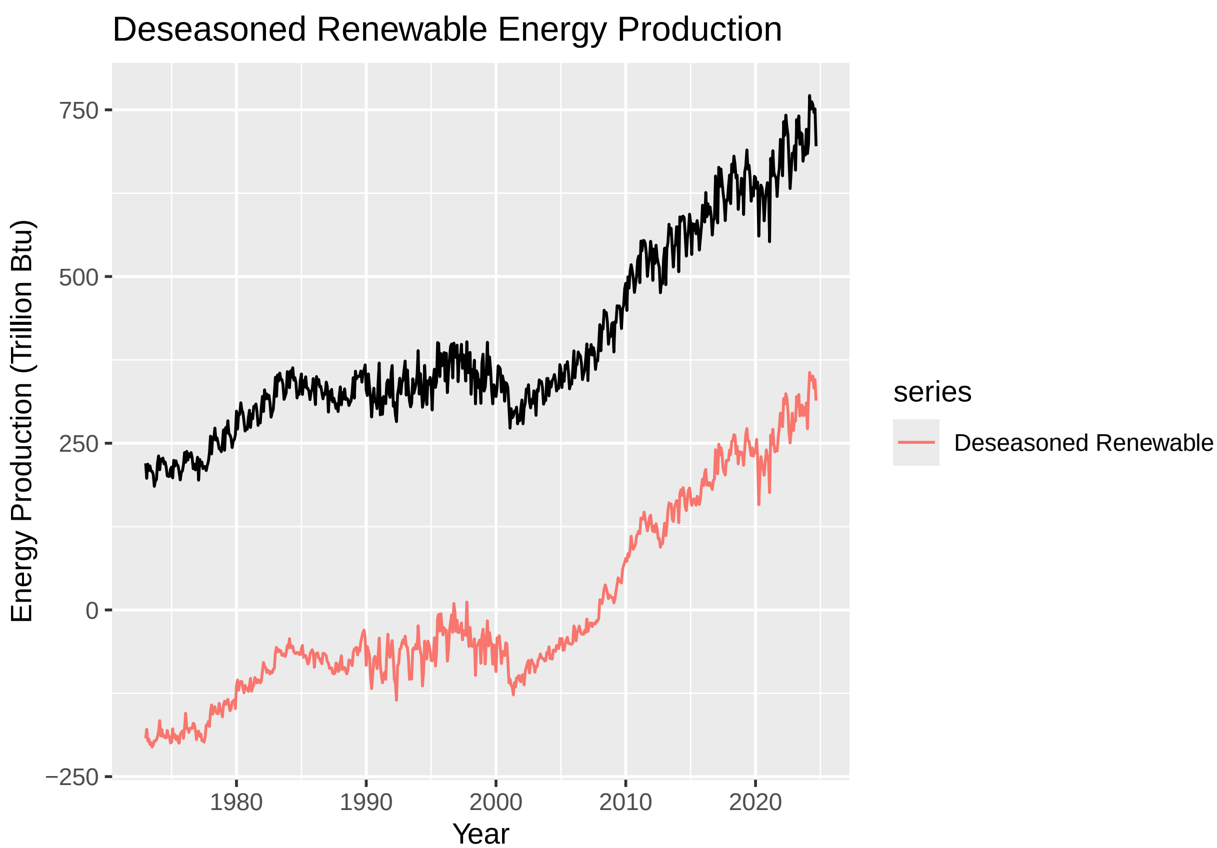
The regression results confirm that hydroelectric power consumption exhibits a strong seasonal trend, while total renewable energy production shows no significant seasonality. The hydroelectric power consumption data demonstrates seasonality, as indicated by p-values being less than 0.05 across most months, meaning monthly variations explain a substantial portion of the data's fluctuations. The model for hydroelectric power consumption has a high R-squared value (0.4695), with many significant monthly coefficients and a highly significant overall model (p-value $<$ 2.2e-16), further reinforcing the presence of a seasonal pattern. In contrast, total renewable energy production exhibits no significance, as shown by the lack of statistically significant monthly coefficients and an extremely low R-squared value (0.0087), suggesting no meaningful seasonal trend. This aligns with my answer for Q6, where the time series plot and ACF suggested a clear seasonal trend in hydro data, while renewable energy production showed a much less pronounced or no seasonal pattern, being primarily driven by a long-term trend.

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?

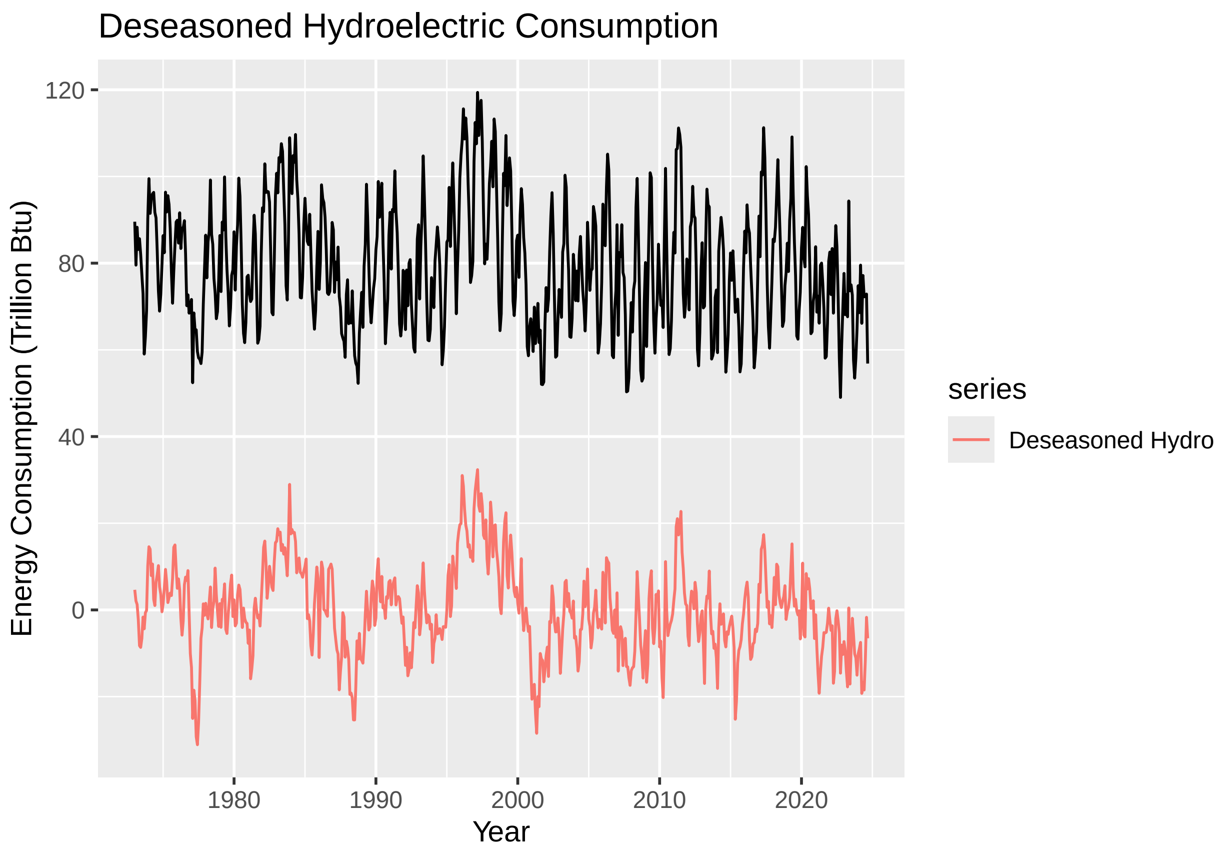
```r
renewable_seas_comp <- array(0, nobs)
i <- 1
for(i in 1:nobs){
  renewable_seas_comp[i] <- beta_intercept_renewable + beta_coeff_renewable %*%
```

```
    dummies_renewable[i,]
}

deseason_renewable <- energy_subset[1] - renewable_seas_comp

ts_deseason_renewable_data <- ts(deseason_renewable, start = c(1973,1), frequency = 12)

autoplot(energy_subset_ts[,col_renewable]) +
  autolayer(ts_deseason_renewable_data, series = "Deseasoned Renewable") +
  ggtitle("Deseasoned Renewable Energy Production") +
  ylab("Energy Production (Trillion Btu)") +
  xlab("Year")
```




```
hydro_seas_comp <- array(0,nobs)
i <- 1
for(i in 1:nobs){
  hydro_seas_comp[i] <- beta_intercept_hydro + beta_coeff_hydro %*%
    dummies_hydro[i,]
}

deseason_hydro <- energy_subset[2] - hydro_seas_comp

ts_deseason_hydro_data <- ts(deseason_hydro, start = c(1973,1), frequency = 12)

autoplot(energy_subset_ts[,col_hydro])+
```

```
  autolayer(ts_deseason_hydro_data, series = "Deseasoned Hydro")+
  ggtitle("Deseasoned Hydroelectric Consumption") +
  ylab("Energy Consumption (Trillion Btu)") +
  xlab("Year")
```



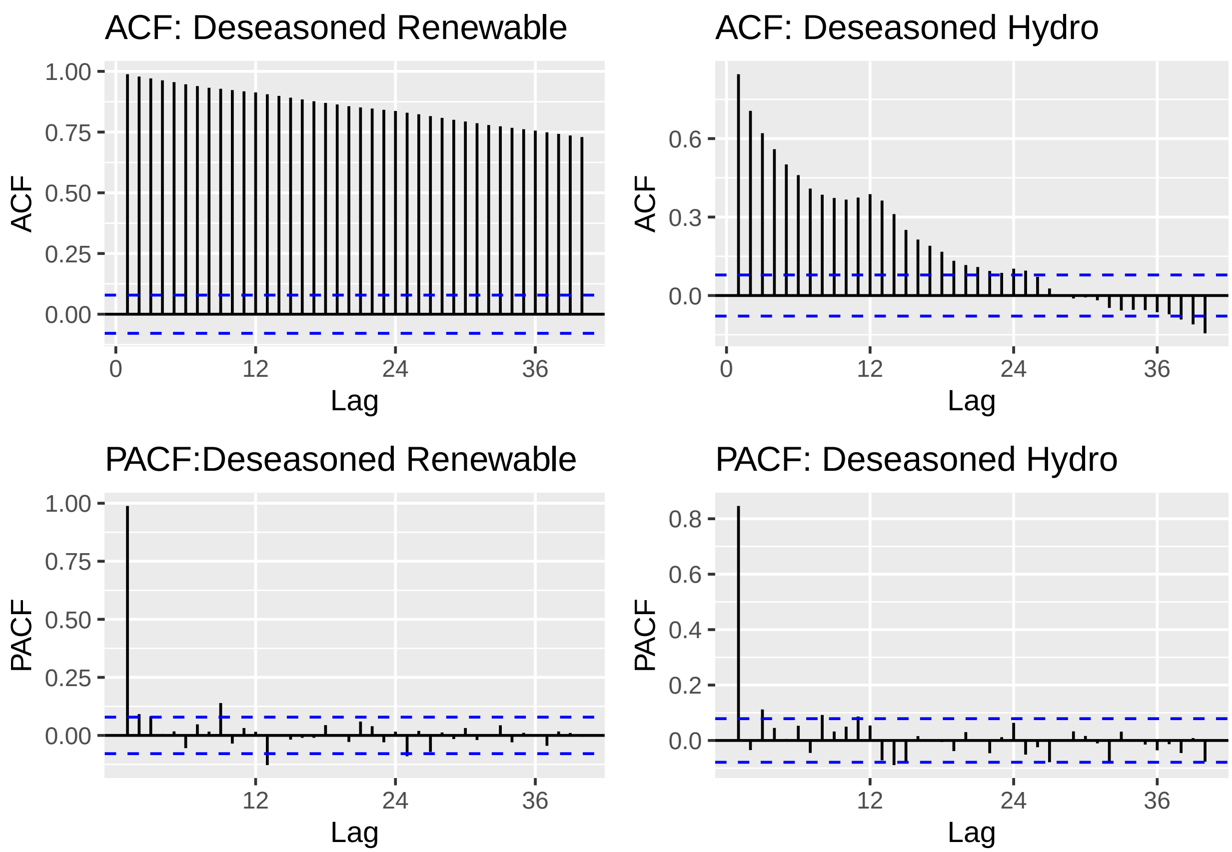
After deseasoning the series, noticeable changes can be observed compared to the original time series plots from Q1 (black lines). The overall structure and long-term trends remain intact, but seasonal fluctuations have been removed, particularly in the hydroelectric power consumption series. In the case of total renewable energy production, the original series exhibited an upward trend with some fluctuations but no strong visible seasonality. After deseasoning, the series (red line) follows a similar trajectory, suggesting that seasonal effects were minimal. In contrast, hydroelectric power consumption initially displayed strong seasonal patterns with clear repeating peaks and troughs. After deseasoning, the red line appears smoother, with the periodic seasonal spikes removed, leaving behind residual noise and long-term trends. This confirms that hydro was heavily influenced by seasonal variations, whereas renewable energy production was primarily driven by long-term trends. These findings align with the previous regression and ACF analysis, which indicated strong seasonality in hydro but a less pronounced seasonal effect in renewable energy production.

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?

```
# ACF and PACF using ggAcf() and ggPacf()
acf_renewable_deseason <- ggAcf(ts_deseason_renewable_data, lag.max = 40) +
  ggtitle(paste("ACF: Deseasoned Renewable"))
```

```r
acf_hydro_deseason <- ggAcf(ts_deseason_hydro_data, lag.max = 40) +
  ggtitle(paste("ACF: Deseasoned Hydro"))

pacf_renewable_deseason <- ggPacf(ts_deseason_renewable_data, lag.max = 40) +
  ggtitle(paste("PACF:Deseasoned Renewable"))
pacf_hydro_deseason <- ggPacf(ts_deseason_hydro_data, lag.max = 40) +
  ggtitle(paste("PACF: Deseasoned Hydro"))

# Using plot_grid to arrange plots (all are now ggplot objects)
plot_grid(acf_renewable_deseason, acf_hydro_deseason, pacf_renewable_deseason,
          pacf_hydro_deseason, ncol = 2)
```



After deseasoning, the ACF and PACF plots show noticeable changes, particularly for the hydroelectric power consumption data. The ACF and PACF for the renewable dataset remain largely unchanged, indicating that seasonality was not a significant factor in its correlations, and the series is still primarily driven by a long-term trend. However, for the hydro data, the ACF and PACF change by a considerable amount, becoming less cyclical. Previously, the hydro series exhibited strong seasonal patterns, with clear spikes in the ACF at regular lags, reflecting its seasonal dependencies. After deseasoning, these cyclical patterns are no longer strongly present, confirming that seasonality played a major role in the original correlations and has now been successfully removed. This aligns with previous findings that hydroelectric power consumption exhibited a strong seasonal trend, whereas total renewable energy production was more influenced by long-term growth.