ENV 790.30 - Time Series Analysis for Energy Data | Spring 2022

Assignment 3 - Due date 02/08/22

Tatiana Sokolova

### 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 project open the first thing you will do is change “Student Name” on line 3 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. Rename the pdf file such that it includes your first and last name (e.g., “LuanaLima\_TSA\_A03\_Sp22.Rmd”). Submit this pdf using Sakai.

### Questions

Consider the same data you used for A2 from the spreadsheet “Table\_10.1\_Renewable\_Energy\_Production\_and\_Consumption\_by\_Source.xlsx”. The data comes from the US Energy Information and Administration and corresponds to the January 2022 **Monthly** Energy Review. Once again you will work only with the following columns: Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption. Create a data frame structure with these three 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.

### Set Up

#Importing data set  
energy\_data <- read.xlsx(file="../Data/Table\_10.1\_Renewable\_Energy\_Production\_and\_Consumption\_by\_Source.xlsx",header=FALSE, startRow=13, sheetIndex=1)  
#extracting column names from row 11  
read\_col\_names <- read.xlsx(file="../Data/Table\_10.1\_Renewable\_Energy\_Production\_and\_Consumption\_by\_Source.xlsx",header=FALSE,startRow=11,endRow=11,sheetIndex=1)  
colnames(energy\_data) <- read\_col\_names  
head(energy\_data)

## Month Wood Energy Production Biofuels Production  
## 1 1973-01-01 129.630 Not Available  
## 2 1973-02-01 117.194 Not Available  
## 3 1973-03-01 129.763 Not Available  
## 4 1973-04-01 125.462 Not Available  
## 5 1973-05-01 129.624 Not Available  
## 6 1973-06-01 125.435 Not Available  
## Total Biomass Energy Production Total Renewable Energy Production  
## 1 129.787 403.981  
## 2 117.338 360.900  
## 3 129.938 400.161  
## 4 125.636 380.470  
## 5 129.834 392.141  
## 6 125.611 377.232  
## Hydroelectric Power Consumption Geothermal Energy Consumption  
## 1 272.703 1.491  
## 2 242.199 1.363  
## 3 268.810 1.412  
## 4 253.185 1.649  
## 5 260.770 1.537  
## 6 249.859 1.763  
## Solar Energy Consumption Wind Energy Consumption Wood Energy Consumption  
## 1 Not Available Not Available 129.630  
## 2 Not Available Not Available 117.194  
## 3 Not Available Not Available 129.763  
## 4 Not Available Not Available 125.462  
## 5 Not Available Not Available 129.624  
## 6 Not Available Not Available 125.435  
## Waste Energy Consumption Biofuels Consumption  
## 1 0.157 Not Available  
## 2 0.144 Not Available  
## 3 0.176 Not Available  
## 4 0.174 Not Available  
## 5 0.210 Not Available  
## 6 0.176 Not Available  
## Total Biomass Energy Consumption Total Renewable Energy Consumption  
## 1 129.787 403.981  
## 2 117.338 360.900  
## 3 129.938 400.161  
## 4 125.636 380.470  
## 5 129.834 392.141  
## 6 125.611 377.232

#creating df structure for columns of interest  
df <- energy\_data[,c('Month','Total Biomass Energy Production', 'Total Renewable Energy Production','Hydroelectric Power Consumption')]  
head(df)

## Month Total Biomass Energy Production Total Renewable Energy Production  
## 1 1973-01-01 129.787 403.981  
## 2 1973-02-01 117.338 360.900  
## 3 1973-03-01 129.938 400.161  
## 4 1973-04-01 125.636 380.470  
## 5 1973-05-01 129.834 392.141  
## 6 1973-06-01 125.611 377.232  
## Hydroelectric Power Consumption  
## 1 272.703  
## 2 242.199  
## 3 268.810  
## 4 253.185  
## 5 260.770  
## 6 249.859

#transforming data into time series  
ts\_energy\_data <- ts(data=df[,2:4], start=c(1973,1),frequency=12)  
head(ts\_energy\_data)

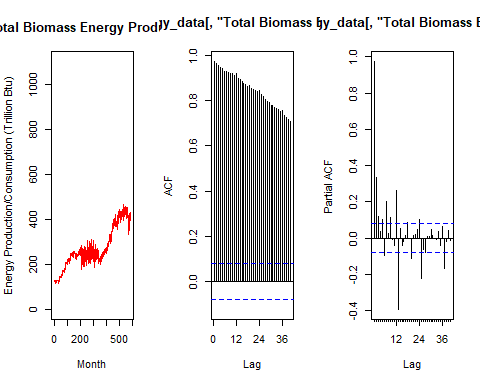
## Total Biomass Energy Production Total Renewable Energy Production  
## Jan 1973 129.787 403.981  
## Feb 1973 117.338 360.900  
## Mar 1973 129.938 400.161  
## Apr 1973 125.636 380.470  
## May 1973 129.834 392.141  
## Jun 1973 125.611 377.232  
## Hydroelectric Power Consumption  
## Jan 1973 272.703  
## Feb 1973 242.199  
## Mar 1973 268.810  
## Apr 1973 253.185  
## May 1973 260.770  
## Jun 1973 249.859

### Trend Component

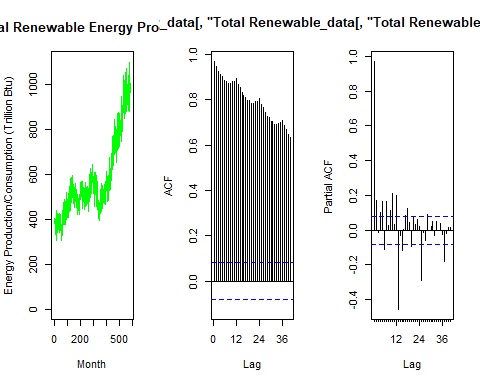
### Q1

Create a plot window that has one row and three columns. And then for each object on your data frame, fill the plot window with time series plot, ACF and PACF. You may use the some code from A2, but I want all three plots on the same window this time. (Hint: use par() function)

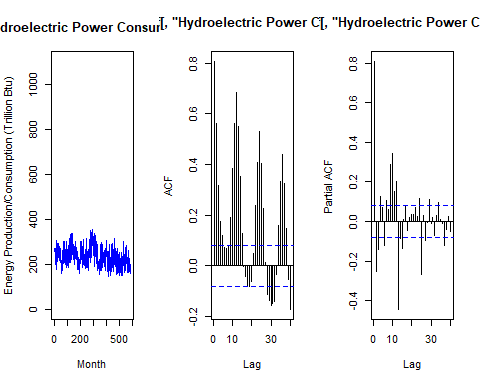
#time series, ACF, and PACF  
  
par(mfrow=c(1,3))  
plot(df[,"Total Biomass Energy Production"],type="l",col="red",ylab="Energy Production/Consumption (Trillion Btu)", xlab="Month", ylim=c(0,1100))   
title(main="Total Biomass Energy Production")  
Acf(ts\_energy\_data[,"Total Biomass Energy Production"],lag.max=40, type="correlation", plot=TRUE)  
Pacf(ts\_energy\_data[,"Total Biomass Energy Production"],lag.max=40, plot=TRUE)



par(mfrow=c(1,3))  
plot(df[,"Total Renewable Energy Production"],type="l",col="green",ylab="Energy Production/Consumption (Trillion Btu)", xlab="Month", ylim=c(0,1100))   
title(main="Total Renewable Energy Production")  
Acf(ts\_energy\_data[,"Total Renewable Energy Production"],lag.max=40, type="correlation", plot=TRUE)  
Pacf(ts\_energy\_data[,"Total Renewable Energy Production"],lag.max=40, plot=TRUE)



par(mfrow=c(1,3)) #place three plots in the same window.  
plot(df[,"Hydroelectric Power Consumption"],type="l",col="blue",ylab="Energy Production/Consumption (Trillion Btu)", xlab="Month", ylim=c(0,1100))   
title(main="Hydroelectric Power Consumption")  
Acf(df[,"Hydroelectric Power Consumption"],lag.max=40, type="correlation", plot=TRUE)  
Pacf(df[,"Hydroelectric Power Consumption"],lag.max=40, plot=TRUE)



### Q2

From the plot in Q1, do the series Total Biomass Energy Production, Total Renewable Energy Production, Hydroelectric Power Consumption appear to have a trend? If yes, what kind of trend?

Total Biomass Energy Production has an increasing trend.

Total Renewable Energy Production also has an increasing trend.

Hydroelectric Power Consumption’s data appears to have a slight decreasing trend and a much more obvious seasonality.

### Q3

Use the *lm()* function to fit a linear trend to the three 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.

#Create vector t  
nobs=nrow(df)  
t <- c(1:nobs)  
  
#Fit a linear trend to TS of Total Biomass  
linear\_trend\_model\_bio=lm(df[,2]~t)   
summary(linear\_trend\_model\_bio)

##   
## Call:  
## lm(formula = df[, 2] ~ t)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -101.892 -24.306 4.932 33.103 82.292   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.348e+02 3.282e+00 41.07 <2e-16 \*\*\*  
## t 4.744e-01 9.705e-03 48.88 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 39.64 on 583 degrees of freedom  
## Multiple R-squared: 0.8039, Adjusted R-squared: 0.8035   
## F-statistic: 2389 on 1 and 583 DF, p-value: < 2.2e-16

beta0b=as.numeric(linear\_trend\_model\_bio$coefficients[1]) #first coefficient is the intercept term or beta0  
beta1b=as.numeric(linear\_trend\_model\_bio$coefficients[2]) #second coefficient is the slope or beta1

The slope of Total Biomass Energy Production has a slightly positive slope indicating a correlation between time and Biomass Energy Production e.g. a slightly increasing trend. The intercept is a large positive number indicating that at the beginning of the time series (e.g. in 1973), 134.80 trillion Btus of Biomass Energy was being produced.

#Fit a linear trend to TS of Total Renewable Energy Production  
linear\_trend\_model\_renew=lm(df[,3]~t)   
summary(linear\_trend\_model\_renew)

##   
## Call:  
## lm(formula = df[, 3] ~ t)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -230.488 -57.869 5.595 62.090 261.349   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 323.18243 8.02555 40.27 <2e-16 \*\*\*  
## t 0.88051 0.02373 37.10 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 96.93 on 583 degrees of freedom  
## Multiple R-squared: 0.7025, Adjusted R-squared: 0.702   
## F-statistic: 1377 on 1 and 583 DF, p-value: < 2.2e-16

beta0r=as.numeric(linear\_trend\_model\_renew$coefficients[1]) #first coefficient is the intercept term or beta0  
beta1r=as.numeric(linear\_trend\_model\_renew$coefficients[2]) #second coefficient is the slope or beta1

The slope of Total Renewable Energy Production has a slightly positive slope indicating a correlation between time and Renewable Energy Production, e.g. the data has a slightly increasing trend. The intercept is a large positive number indicating that at the beginning of the time series (e.g. in 1973), 323.18 trillion Btus of Renewable Energy was being produced.

#Fit a linear trend to TS of Hydroelectric Power Consumption  
linear\_trend\_model\_hydro=lm(df[,4]~t)   
summary(linear\_trend\_model\_hydro)

##   
## Call:  
## lm(formula = df[, 4] ~ t)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -94.892 -31.300 -2.414 27.876 121.263   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 259.18303 3.47464 74.593 < 2e-16 \*\*\*  
## t -0.07924 0.01027 -7.712 5.36e-14 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 41.97 on 583 degrees of freedom  
## Multiple R-squared: 0.09258, Adjusted R-squared: 0.09103   
## F-statistic: 59.48 on 1 and 583 DF, p-value: 5.364e-14

beta0h=as.numeric(linear\_trend\_model\_hydro$coefficients[1]) #first coefficient is the intercept term or beta0  
beta1h=as.numeric(linear\_trend\_model\_hydro$coefficients[2]) #second coefficient is the slope or beta1

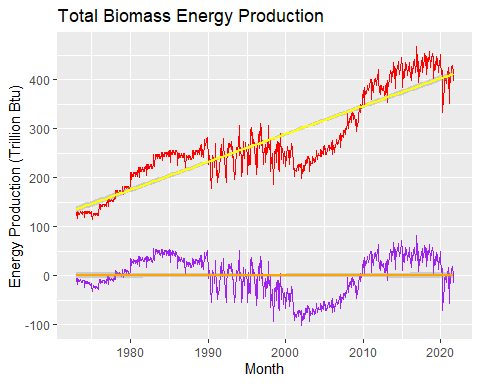
The slope of Hydroelectric Power Consumption has a slightly negative slope indicating that the data has a slightly decreasing trend. The intercept is a large positive number indicating that at the beginning of the time series (e.g. in 1973), 259.18 trillion Btus of Hydroelectric power was being consumed.

### 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?

#remove the trend from TS of Total Biomass Energy Production  
detrend\_bio\_data <- df[,2]-(beta0b+beta1b\*t)  
  
#plotting detrended series  
ggplot(df, aes(x=df[,1], y=df[,"Total Biomass Energy Production"])) +  
 geom\_line(color="red") +  
 ggtitle("Total Biomass Energy Production")+  
 ylab(paste0("Energy Production (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_smooth(color="yellow",method="lm") +  
 geom\_line(aes(y=detrend\_bio\_data), col="purple")+  
 geom\_smooth(aes(y=detrend\_bio\_data),color="orange",method="lm")

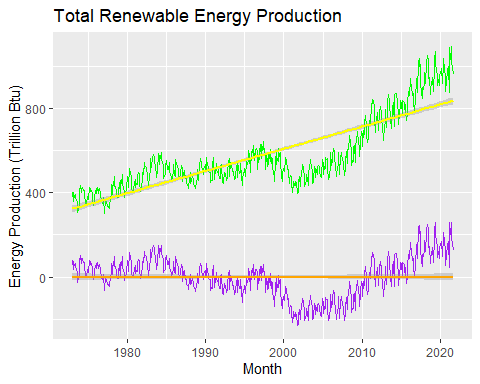
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'



The increasing trend was removed from the Total Biomass Energy Production series.

#remove the trend from TS of Total Renewable Energy Production  
detrend\_renew\_data <- df[,3]-(beta0r+beta1r\*t)  
  
#plotting detrended series  
ggplot(df, aes(x=df[,1], y=df[,"Total Renewable Energy Production"])) +  
 geom\_line(color="green") +  
 ggtitle("Total Renewable Energy Production")+  
 ylab(paste0("Energy Production (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_smooth(color="yellow",method="lm") +  
 geom\_line(aes(y=detrend\_renew\_data), col="purple")+  
 geom\_smooth(aes(y=detrend\_renew\_data),color="orange",method="lm")

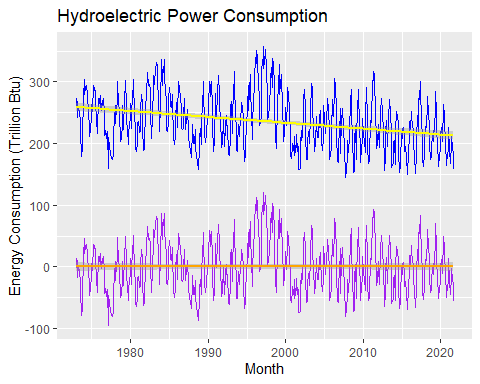
## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'



The increasing trend was removed from the Total Renewable Energy Production series.

#remove the trend from TS of Hydroelectric Power Consumption  
detrend\_hydro\_data <- df[,4]-(beta0h+beta1h\*t)  
  
#plotting detrended series  
ggplot(df, aes(x=df[,1], y=df[,"Hydroelectric Power Consumption"])) +  
 geom\_line(color="blue") +  
 ggtitle("Hydroelectric Power Consumption")+  
 ylab(paste0("Energy Consumption (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_smooth(color="yellow",method="lm") +  
 geom\_line(aes(y=detrend\_hydro\_data), col="purple")+  
 geom\_smooth(aes(y=detrend\_hydro\_data),color="orange",method="lm")

## `geom\_smooth()` using formula 'y ~ x'  
## `geom\_smooth()` using formula 'y ~ x'

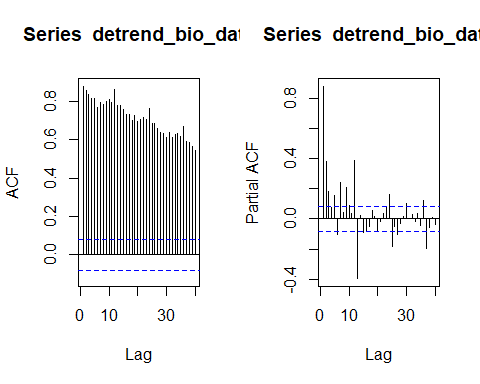


The very slight decreasing trend was removed from the Hydroelectric Power Consumption series.

### Q5

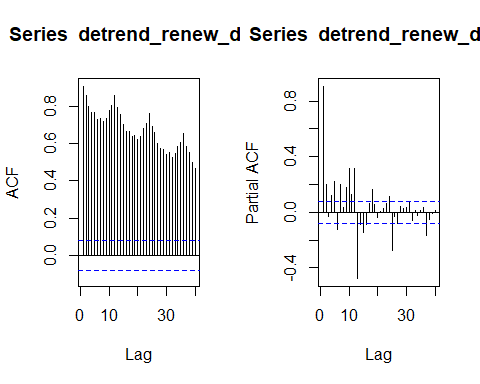
Plot ACF and PACF for the detrended series and compare with the plots from Q1. Did the plots change? How?

#new ACF and PACF of Total Biomass Energy Production  
par(mfrow=c(1,2))  
Acf(detrend\_bio\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(detrend\_bio\_data,lag.max=40, plot=TRUE)



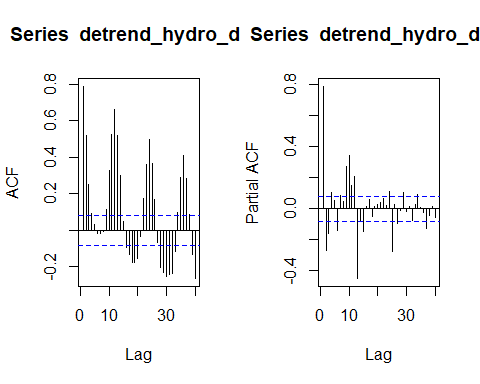
The ACF shows that the values of the detrended series are less related with its past values than that of the orignal series. The PACF shows that there is a greater correlation between the lags of the series when it is detrended than that of the original series.

#new ACF and PACF of Total Renewable Energy Production  
par(mfrow=c(1,2))  
Acf(detrend\_renew\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(detrend\_renew\_data,lag.max=40, plot=TRUE)



The ACF shows that the values of the detrended series are less related with its past values than that of the orignal series and that there might be greater seasonality. The PACF looks pretty similar to that of the original series possibly indicating that the lags in the detrended series have a similar correlation to that of the original series.

#new ACF and PACF of Hydroelectric Power Consumption  
par(mfrow=c(1,2))  
Acf(detrend\_hydro\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(detrend\_hydro\_data,lag.max=40, plot=TRUE)



The ACF and PACF of the detrended series look similar to that of the original series. The ACF of the detrended series looks like it has slightly greater correlation than that of the PACF but I am not sure it is significant. Overall, we can deduce that the minor decreasing trend in this series has minimal impact upon the correlation of the series.

## Seasonal Component

Set aside the detrended series and consider the original series again from Q1 to answer Q6 to Q8.

### Q6

Do the series seem to have a seasonal trend? Which serie/series? Use function *lm()* to fit a seasonal means model (i.e. using the seasonal dummies) to this/these time series. Ask R to print the summary of the regression. Interpret the regression output. Save the regression coefficients for further analysis.

The Hydroelectric Power Consumption series appears to have the most obvious seasonal trend.

#Using seasonal means model  
#Creating the seasonal dummies  
dummies <- seasonaldummy(ts\_energy\_data[,3]) #bc this function only accepts ts object  
  
#Fitting a linear model to the seasonal dummies  
seas\_means\_model=lm(df[,4]~dummies)  
summary(seas\_means\_model)

##   
## Call:  
## lm(formula = df[, 4] ~ dummies)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -90.253 -23.017 -3.042 21.487 99.478   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 237.841 4.892 48.616 < 2e-16 \*\*\*  
## dummiesJan 13.558 6.883 1.970 0.04936 \*   
## dummiesFeb -8.090 6.883 -1.175 0.24037   
## dummiesMar 20.067 6.883 2.915 0.00369 \*\*   
## dummiesApr 16.619 6.883 2.414 0.01607 \*   
## dummiesMay 39.961 6.883 5.805 1.06e-08 \*\*\*  
## dummiesJun 31.315 6.883 4.549 6.57e-06 \*\*\*  
## dummiesJul 10.511 6.883 1.527 0.12732   
## dummiesAug -17.853 6.883 -2.594 0.00974 \*\*   
## dummiesSep -49.852 6.883 -7.242 1.43e-12 \*\*\*  
## dummiesOct -48.086 6.919 -6.950 9.96e-12 \*\*\*  
## dummiesNov -32.187 6.919 -4.652 4.08e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 33.89 on 573 degrees of freedom  
## Multiple R-squared: 0.4182, Adjusted R-squared: 0.4071   
## F-statistic: 37.45 on 11 and 573 DF, p-value: < 2.2e-16

March through July appear to be more positively correlated, possibly indicating a wet period where there is more hydroelectric power available to produce or a greater demand for hydroelectric power. Since August through November are negative, we can assume that is a either a more dry period or less demand for hydroelectric power (or both).

The Total Renewable Energy Production series appears to have a slight seasonal trend as well.

#Using seasonal means model  
#Creating the seasonal dummies  
dummiesr <- seasonaldummy(ts\_energy\_data[,2]) #bc this function only accepts ts object  
  
#Fitting a linear model to the seasonal dummies  
seas\_means\_model\_renew=lm(df[,3]~dummiesr)  
summary(seas\_means\_model\_renew)

##   
## Call:  
## lm(formula = df[, 3] ~ dummiesr)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -272.95 -111.55 -59.35 65.68 480.41   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 589.971 25.464 23.169 <2e-16 \*\*\*  
## dummiesrJan 11.793 35.828 0.329 0.7422   
## dummiesrFeb -40.992 35.828 -1.144 0.2530   
## dummiesrMar 21.892 35.828 0.611 0.5414   
## dummiesrApr 8.908 35.828 0.249 0.8037   
## dummiesrMay 37.500 35.828 1.047 0.2957   
## dummiesrJun 19.465 35.828 0.543 0.5871   
## dummiesrJul 8.115 35.828 0.227 0.8209   
## dummiesrAug -18.359 35.828 -0.512 0.6086   
## dummiesrSep -62.115 35.828 -1.734 0.0835 .   
## dummiesrOct -51.377 36.012 -1.427 0.1542   
## dummiesrNov -41.789 36.012 -1.160 0.2464   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 176.4 on 573 degrees of freedom  
## Multiple R-squared: 0.03139, Adjusted R-squared: 0.0128   
## F-statistic: 1.688 on 11 and 573 DF, p-value: 0.07235

Indeed, there appears to be a slight seasonality in Renewables production with greater correlation in the summer (March through July) and a more negative correlation in the fall/winter (August - November). I would presume this is due to solar consumption, but more inspection of the sources of data would be needed.

To be exhaustive, I looked at the seasonality of the Total Biomass Energy Production series as well.

#Using seasonal means model  
#Creating the seasonal dummies  
dummiesb <- seasonaldummy(ts\_energy\_data[,1]) #bc this function only accepts ts object  
  
#Fitting a linear model to the seasonal dummies  
seas\_means\_model\_bio=lm(df[,2]~dummiesb)  
summary(seas\_means\_model\_bio)

##   
## Call:  
## lm(formula = df[, 2] ~ dummiesb)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -156.96 -51.40 -22.15 60.65 183.31   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 284.241 12.962 21.928 <2e-16 \*\*\*  
## dummiesbJan -1.498 18.238 -0.082 0.9346   
## dummiesbFeb -30.582 18.238 -1.677 0.0941 .   
## dummiesbMar -8.873 18.238 -0.486 0.6268   
## dummiesbApr -21.009 18.238 -1.152 0.2498   
## dummiesbMay -14.065 18.238 -0.771 0.4409   
## dummiesbJun -19.601 18.238 -1.075 0.2829   
## dummiesbJul -3.499 18.238 -0.192 0.8479   
## dummiesbAug -0.252 18.238 -0.014 0.9890   
## dummiesbSep -12.518 18.238 -0.686 0.4928   
## dummiesbOct -3.629 18.331 -0.198 0.8432   
## dummiesbNov -9.592 18.331 -0.523 0.6010   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 89.81 on 573 degrees of freedom  
## Multiple R-squared: 0.01056, Adjusted R-squared: -0.008439   
## F-statistic: 0.5557 on 11 and 573 DF, p-value: 0.8647

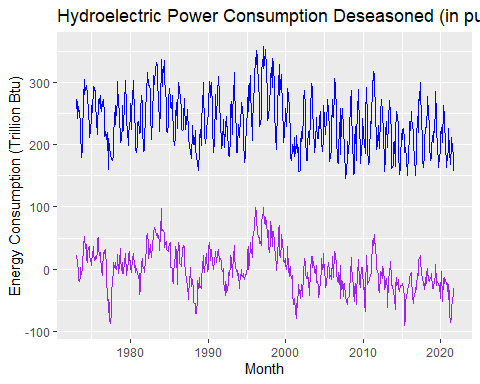
I am not seeing any seasonality trend from the regression coefficients for Biomass Energy Production.

#Storing regression coefficients  
beta\_inth=seas\_means\_model$coefficients[1]  
beta\_coeffh=seas\_means\_model$coefficients[2:12]  
  
beta\_intr=seas\_means\_model\_renew$coefficients[1]  
beta\_coeffr=seas\_means\_model\_renew$coefficients[2:12]  
  
beta\_intb=seas\_means\_model\_bio$coefficients[1]  
beta\_coeffb=seas\_means\_model\_bio$coefficients[2:12]

### Q7

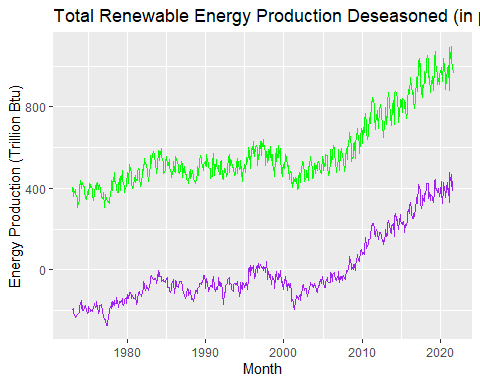
Use the regression coefficients from Q6 to deseason the series. Plot the deseason series and compare with the plots from part Q1. Did anything change?

#Computing seasonal component for Hydro  
hydro\_seas\_comp=array(0,nobs)  
for(i in 1:nobs){  
 hydro\_seas\_comp[i]=(beta\_inth+beta\_coeffh%\*%dummies[i,])  
}  
  
#Removing seasonal component  
deseason\_hydro\_data <- df[,4]-hydro\_seas\_comp  
  
  
#Graphing  
ggplot(df, aes(x=df[,1], y=df[,"Hydroelectric Power Consumption"])) +  
 ggtitle("Hydroelectric Power Consumption Deseasoned (in purple)")+  
 geom\_line(color="blue") +  
 ylab(paste0("Energy Consumption (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_line(aes(y=deseason\_hydro\_data), col="purple")



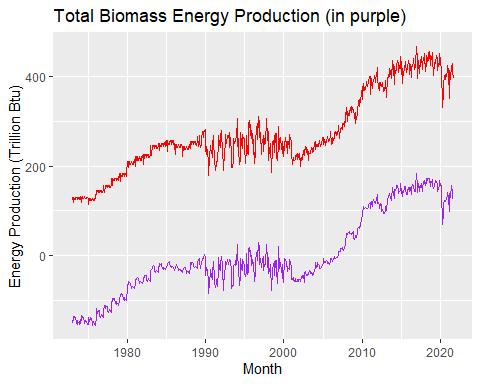
Yes, the plot changed to where the slope of the data appears to be more closely centered around 0 than the original series.

#Computing seasonal component for Renewables Production  
renew\_seas\_comp=array(0,nobs)  
for(i in 1:nobs){  
 renew\_seas\_comp[i]=(beta\_intr+beta\_coeffr%\*%dummies[i,])  
}  
  
#Removing seasonal component  
deseason\_renew\_data <- df[,3]-renew\_seas\_comp  
  
  
#Graphing  
ggplot(df, aes(x=df[,1], y=df[,"Total Renewable Energy Production"])) +  
 ggtitle("Total Renewable Energy Production Deseasoned (in purple)")+  
 geom\_line(color="green") +  
 ylab(paste0("Energy Production (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_line(aes(y=deseason\_renew\_data), col="purple")



There does not appear to be much of a change between the two graphs regarding trends with various time points. Therefore, it could be concluded that deseasoning is not the best way to analyze this series.

#Computing seasonal component for Biomass Energy Production  
bio\_seas\_comp=array(0,nobs)  
for(i in 1:nobs){  
 bio\_seas\_comp[i]=(beta\_intb+beta\_coeffb%\*%dummies[i,])  
}  
  
#Removing seasonal component  
deseason\_bio\_data <- df[,2]-bio\_seas\_comp  
  
  
#Graphing  
ggplot(df, aes(x=df[,1], y=df[,"Total Biomass Energy Production"])) +  
 ggtitle("Total Biomass Energy Production (in purple)")+  
 geom\_line(color="red") +  
 ylab(paste0("Energy Production (Trillion Btu)")) +  
 xlab(paste0("Month"))+  
 geom\_line(aes(y=deseason\_bio\_data), col="purple")

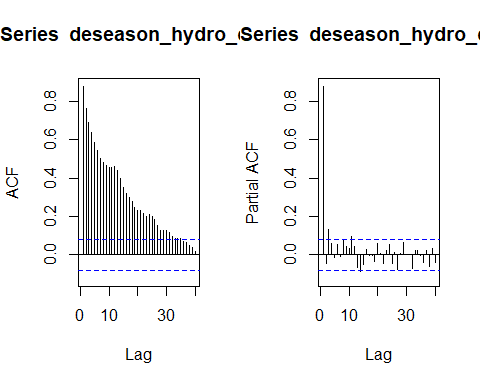


There does not appear any change between the two graphs regarding trends with various time points. Therefore, it could be concluded that deseasoning is not recommended for analyzing this series.

### Q8

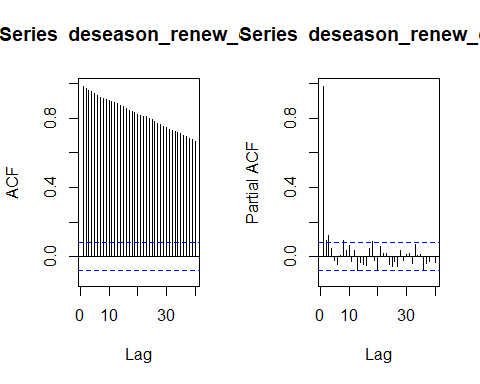
Plot ACF and PACF for the deseason series and compare with the plots from Q1. Did the plots change? How?

#new ACF and PACF of Hydroelectric Power Consumption  
par(mfrow=c(1,2))  
Acf(deseason\_hydro\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(deseason\_hydro\_data,lag.max=40, plot=TRUE)



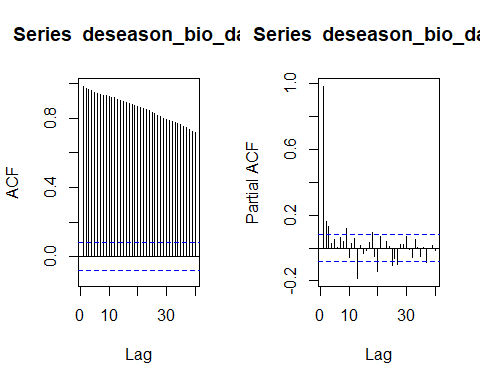
The ACF has no seasonal correlation unlike the original series. The PACF shows no significant correlation between the lags of the deseasoned data.

#new ACF and PACF of Total Renewable Energy Production  
par(mfrow=c(1,2))  
Acf(deseason\_renew\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(deseason\_renew\_data,lag.max=40, plot=TRUE)



The ACF has less of a seasonal correlation unlike the original series, but still has a few spikes that could indicate a different trend. The PACF looks very similar to the plot from Q1 perhaps indicating that the lags are still correlated to one another in what appears to be a seasonal manner (once a year).

#new ACF and PACF of Biomass Energy Production  
par(mfrow=c(1,2))  
Acf(deseason\_bio\_data,lag.max=40, type="correlation", plot=TRUE)  
Pacf(deseason\_bio\_data,lag.max=40, plot=TRUE)



There is very little change with this deseasoned ACF compared to that of the original series. The PACF does appear to to show that lags are less correlated than that of the original series.