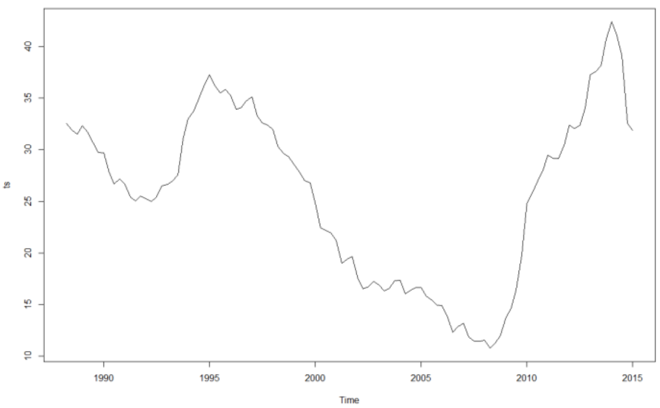
Master Universitario en Ciencia de Datos

Homework 3

Time Series

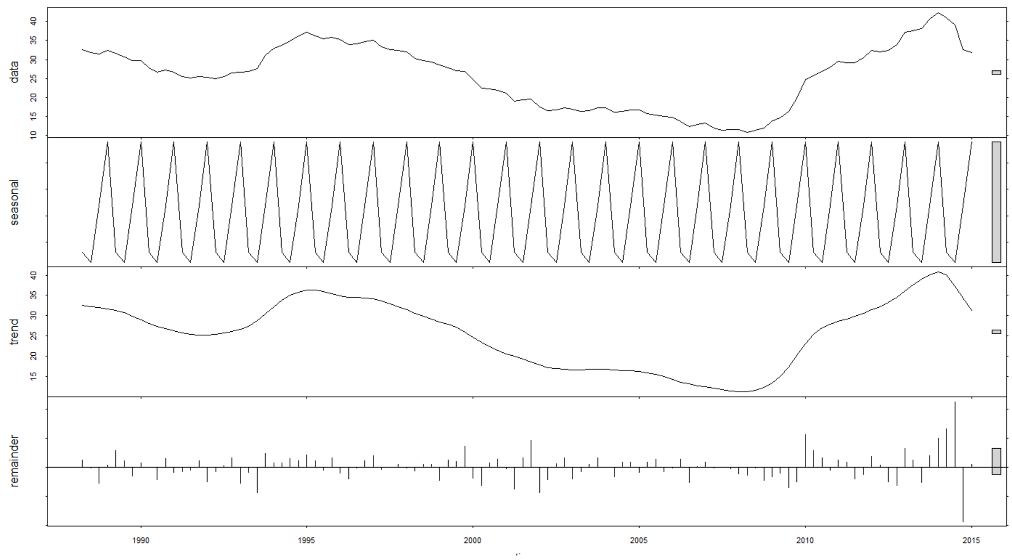
El Abbassi Widad

1. Plot :

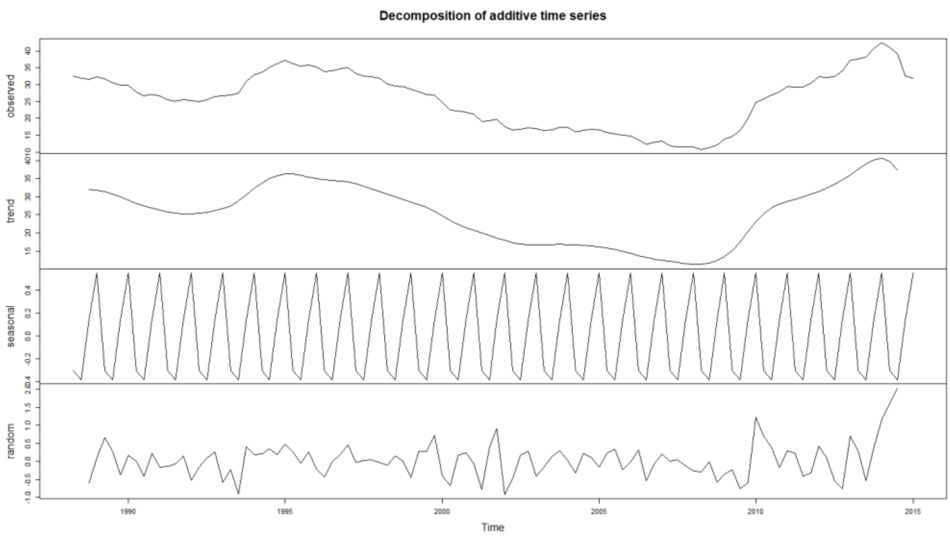


At a first glance, it may seem that this data has a seasonal movement, but actually this behavior can be more described as a cyclic movement that is repeated in the different years of the sample: the data exhibit rises and falls at non-fixed periods. Also, we observe variability in variance through time.

2. Decomposition of the series



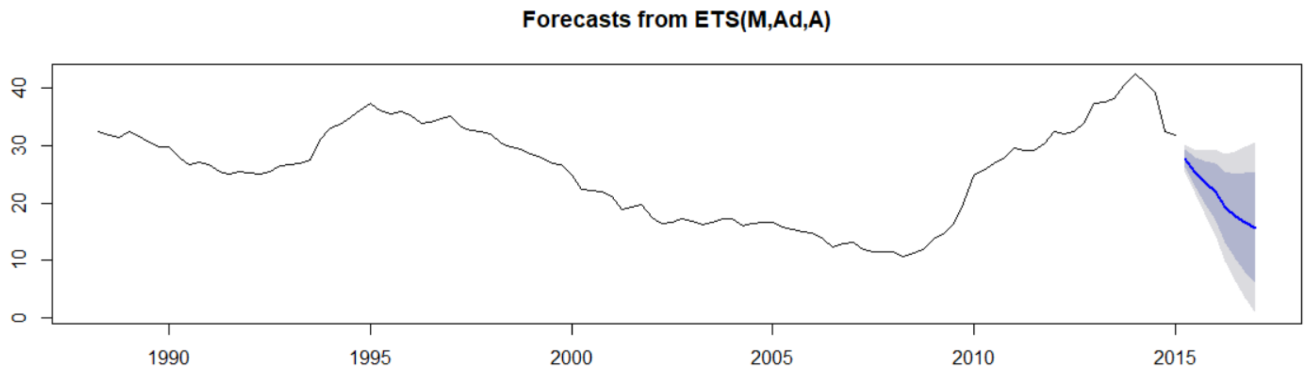
Additive decomposition:



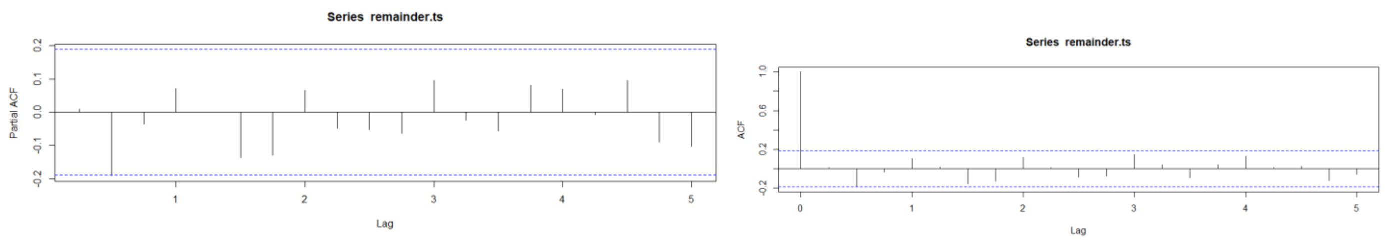
The three components are shown separately in the bottom, we notice that the seasonal component changes very slowly over time, the trend in the other hand is following the cycle change which may be related to the general business cycle, the remainder component shown in the bottom panel is what is left over when the seasonal and trend-cycle components have been subtracted from the data which does not seem very random.

-Multiplicative decomposition: After trying the Log transformation, no visible change has been observed, so there's no need in using the multiplicative decomposition in this case.

-The function forecast() was used as follow :



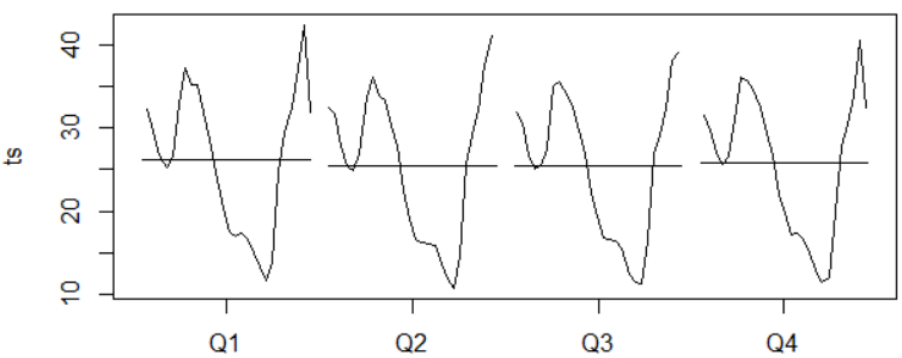
After plotting, the ACf and PACF of the reminder component, it looks like it has the shape of white noise:



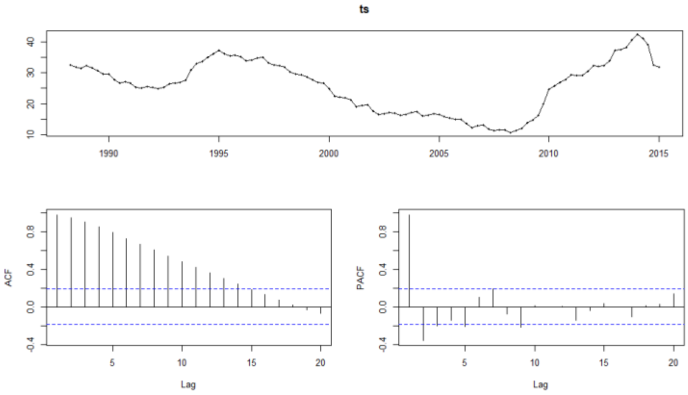
3. Fitting an ARIMA model :

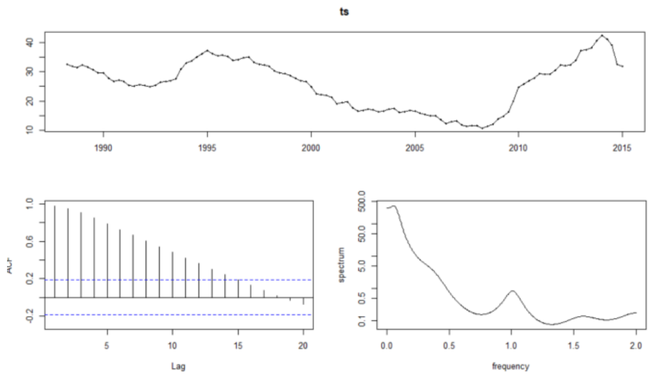
a) After plotting the Log transformation of our data, we don't notice any radical change in the form in the 2 plots.

At the beginning, we declared our dataset with f = 4, now plotting with monthplot() , allows us to see that the data vary in a regular pattern.



b.)





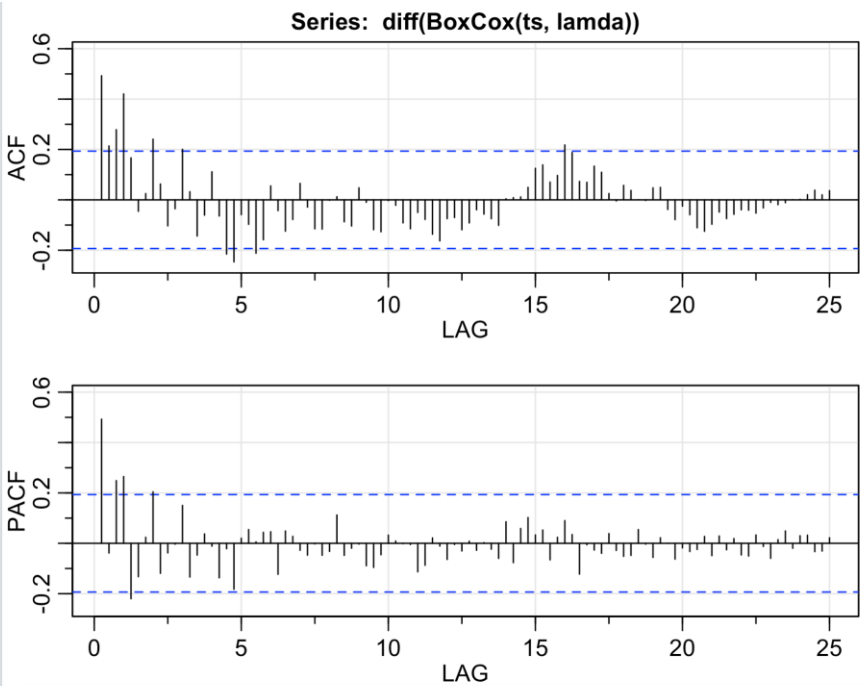
`tsdisplay`   `tsdisplay(ts, plot.type = "spectrum")`

The results of these plots may show the need for differencing. The Time Plot displays short periods of upward and downward trends in the data.

Seasonality and changing variance are not very obvious either .We can say for sure that there is some cyclical behavior, but since we don't dispose of a large dataset or enough information, it will be unpredictable.

The ACF Plot shows spikes that are dropping to zero slowly and PACF Plot show only one significant spike.

c.)



The spikes in ACF of differenced data drops quickly. The data should be differenced again to apply ARMA model (right).

After this we validate our work by applying a stationary test kpss,test:

The null hypothesis for the KPSS test is that the data are stationary, which mean we need the p-value to be greater than 0.05 not less than 0.05.

By standard deviation, we select the lowest value:

```
> sd(ts)  #standard deviations
[1] 8.37697
> sd(diff(ts)) #this one has smallest value
[1] 1.394642
> sd(diff(diff(ts)))
[1] 1.421934
```

In conclusion, d=1.

d.) in this section, we tried different parameters to fit the models :

```
> summary(model.5)
Series: ts
ARIMA(1,1,5) with drift

Coefficients:
          ar1     ma1     ma2     ma3     ma4     ma5    drift
      -0.2544  0.6688  0.4502  0.4001  0.6760  0.3850  -0.0656
s.e.   0.3040  0.2877  0.1887  0.1687  0.2011  0.2049   0.3001

sigma^2 estimated as 1.308:  log likelihood=-163.54
AIC=343.08   AICc=344.55   BIC=364.47

Training set error measures:
                      ME     RMSE       MAE        MPE     MAPE      MASE
Training set 0.002442793 1.100679 0.7409537 0.09048596 3.089521 0.2488349
                  ACF1
Training set -0.011385
```

We get result minimum AIC (343.08) with ARIMA(1,1,5) , but we observe high correlations between coefficients(-0.942974692).

e)

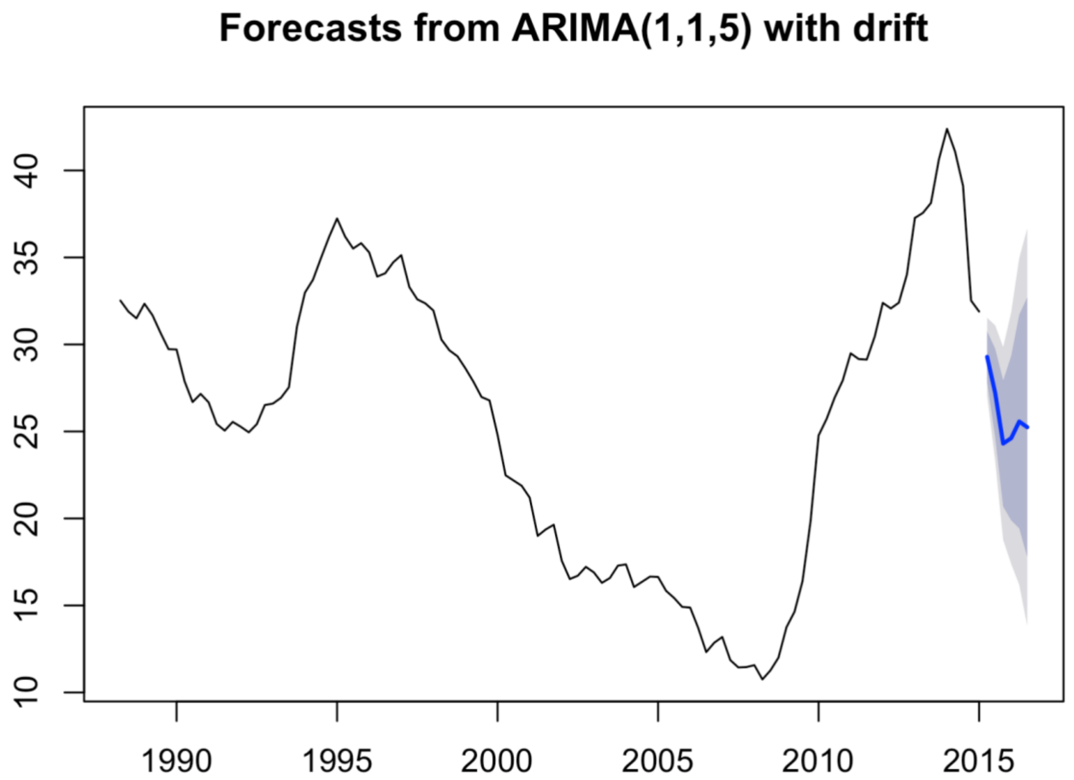
Normality

```
> jarque.bera.test(model.5$residuals)

        Jarque Bera Test

data:  model.5$residuals
X-squared = 272.7, df = 2, p-value < 2.2e-16
```

f)

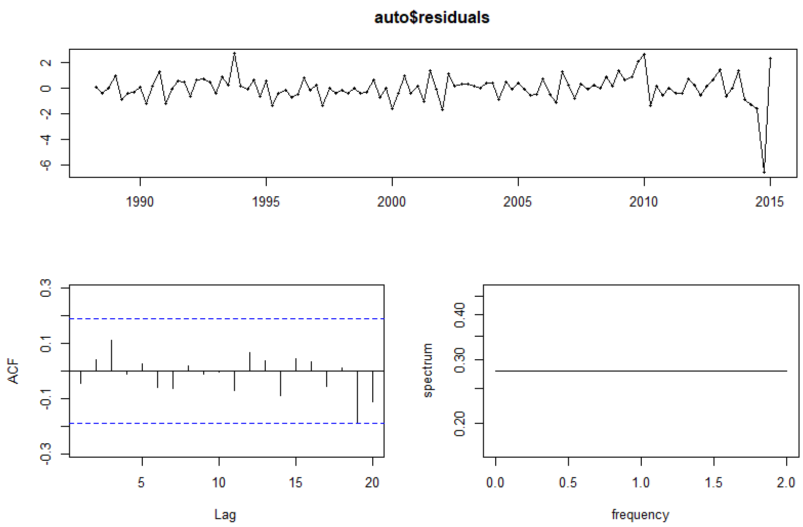


h)

```
> auto=auto.arima(ts,max.p=3,max.q=3,max.P=3,max.Q=3)
> auto
Series: ts
ARIMA(1,1,0)(1,0,1)[4]

Coefficients:
         ar1     sar1     sma1
      0.5770   0.9043  -0.6372
s.e.  0.0832   0.0795   0.1565

sigma^2 estimated as 1.151:  log likelihood=-159.08
AIC=326.16   AICc=326.55   BIC=336.85
```



Here, we tried to remove the seasonal component in order to better observe the cyclic-trend in our dataset