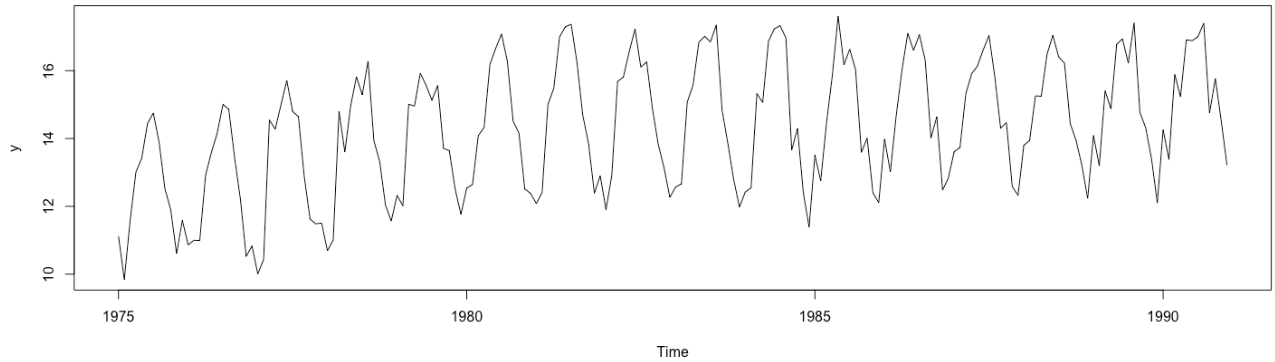
Project work for the teaching Time Series Analysis – Luca Pennella

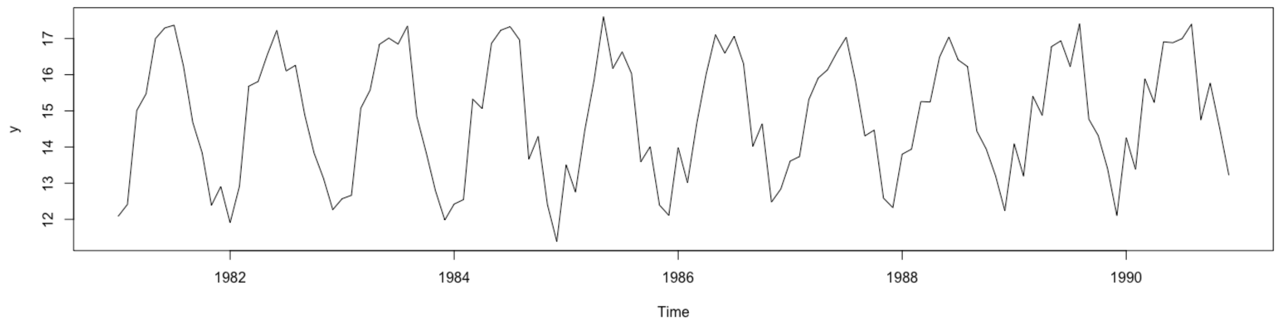
Second Time Series Analysis

Exploratory Data Analysis

The analysis concerns the study of time series relating to monthly beer sales in millions of barrels, from 01/1975 to 12/1990. The dataset shows the information with a monthly frequency (from 1 to 192), the data range starts from a minimum value of 9.841 up to a maximum of 17.604 with an average value of 14.279.
The source of the data is *Frees, EW, Data Analysis Using Regression Models, Prentice Hall, 1996*.

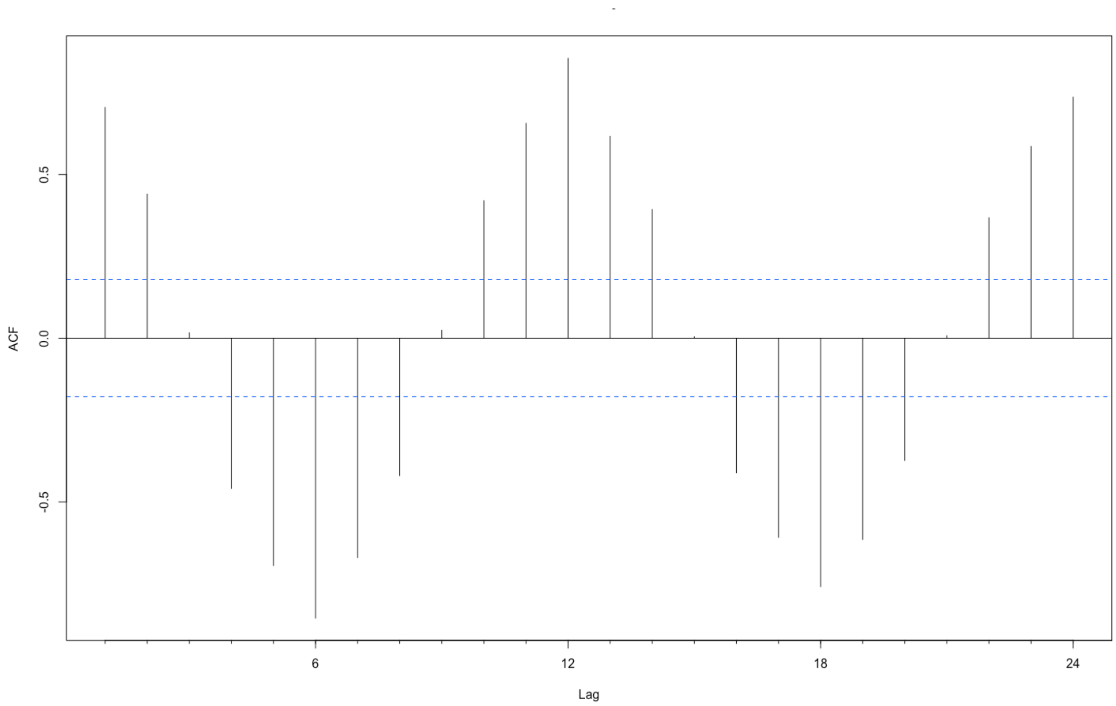


The series seems to show a slightly growing trend, plus a seasonal pattern at the year level is evident. This behavior is realistic as we can expect more beer consumption in the summer months.
To make the analysis easier, I decided to consider the time series from 1980 as the average seems to stabilize.

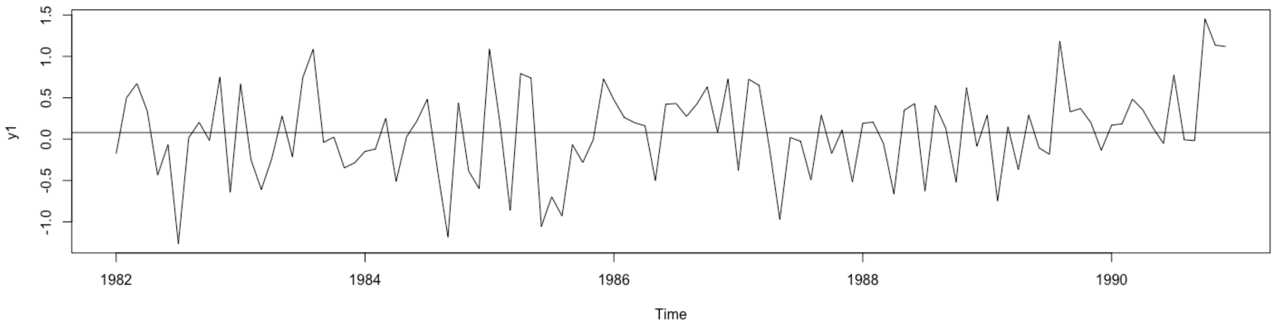


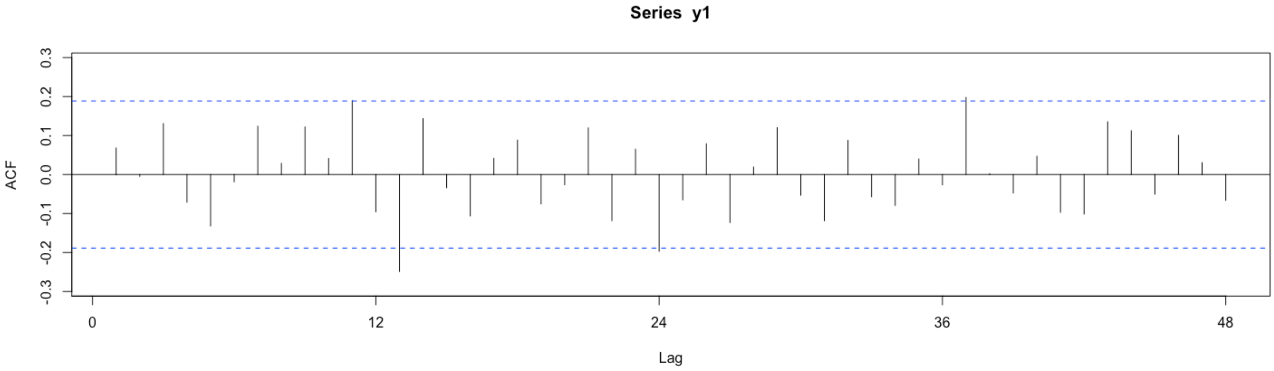
Find the Models

The first phase is the control of the stationarity of the series through the study of the ACF and the application of some tests.



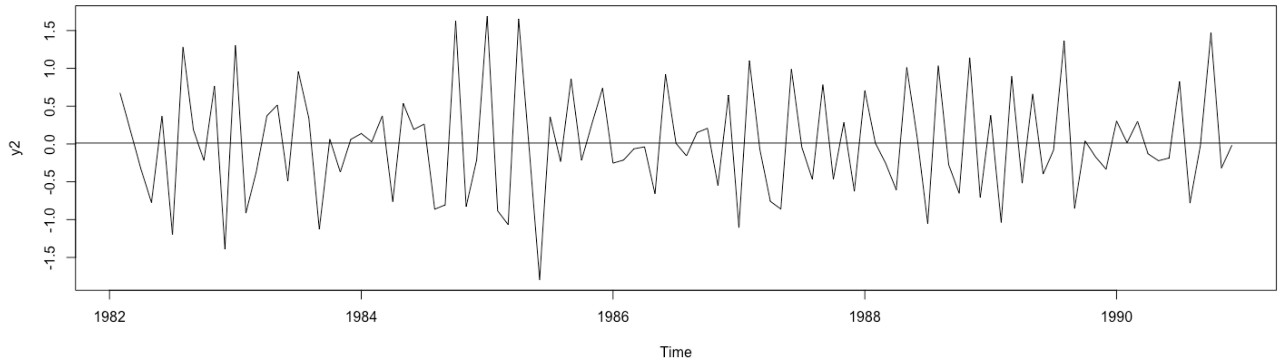
By observing the graph it is possible to see how the ACF presents a seasonal trend every 6 lag, with a peak every 6 lag and with the 4 significant lag around it.
Given the obvious seasonality, I proceed by making the seasonal differences for m = 12.

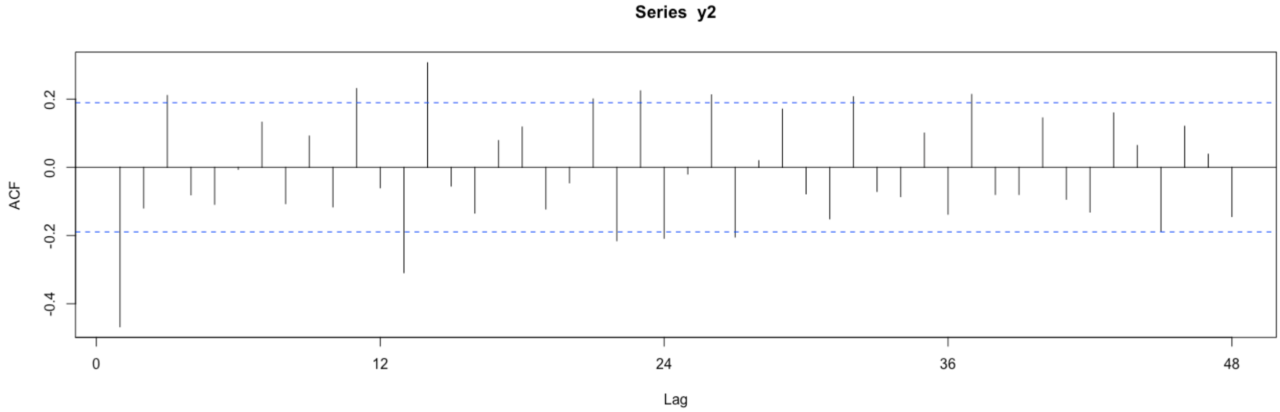




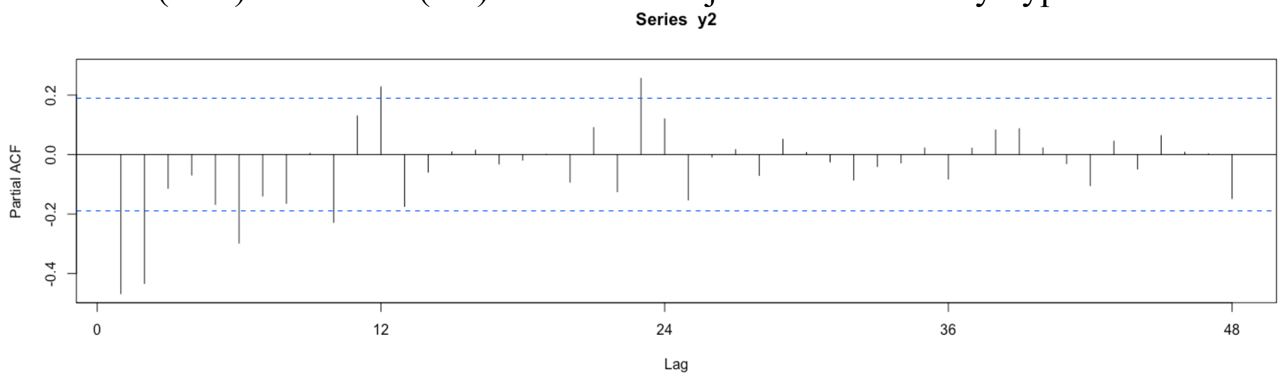
The series appears to be stationary on average and in variance.
Looking at the ACF graph, we see that the seasonal peaks are almost all within the confidence intervals.
The ADF test (0.01) rejects the non-stationarity hypothesis while the KPSS has a borderline value (0.08) which does not allow to safely reject the hypothesis of non-stationarity.

Given the test results we proceed by applying the differences on a non-seasonal level.





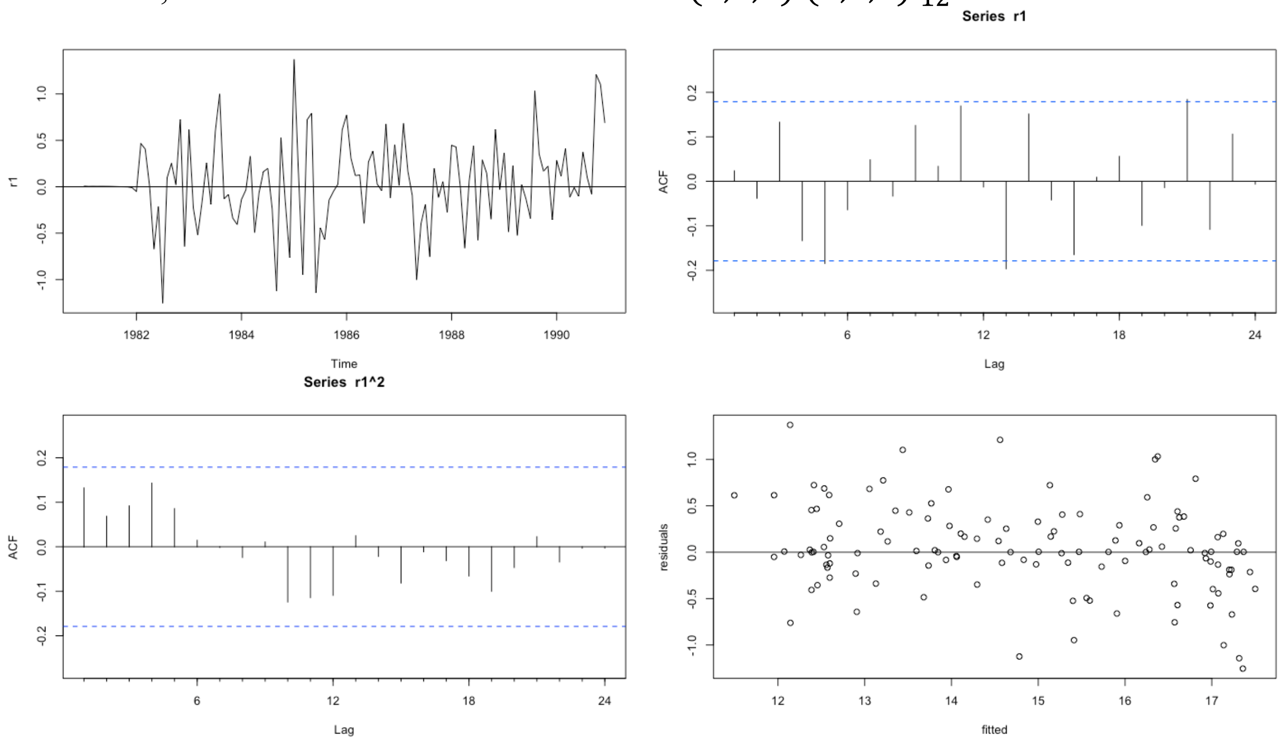
The time chart shows no evidence, while the ACF seems to indicate an order of seasonal MA (SMA) equal to 2 and MA equal to 1.
The ADF (0.01) and KPSS (0.1) tests do not reject the stationarity hypothesis.



Observing the PACF it seems there may be an AR order of 1 or 2 but to be sure we proceed with the selection through the information criteria.

```
Series: y
ARIMA(0,1,1)(0,1,2)[12]

Coefficients:
          ma1     sma1     sma2
      -0.9125  -0.2028  -0.3231
s.e.   0.0547   0.1133   0.1160

sigma^2 = 0.2568:  log likelihood = -80.46
AIC=168.92   AICc=169.31   BIC=179.61
> m1$coef/sqrt(diag(m1$var.coef))  ## z statistics
       ma1       sma1       sma2
-16.691995  -1.789624  -2.785311
```

The hypotheses made regarding MA orders have been confirmed while there are no AR orders, the model identified is an ARIMA (0,1,1) (0,1,2) $_{12}$.



Residue diagnostics do not exhibit abnormal behavior.

As an alternative to the SARIMA model, we try to model the series with an ARMA with a purely deterministic trend.
The first step to build an ARMA with a deterministic trend is the realization of ARIMA (0,0,0) which has time as a regressor, once this is done it is necessary to check the significance of the model:

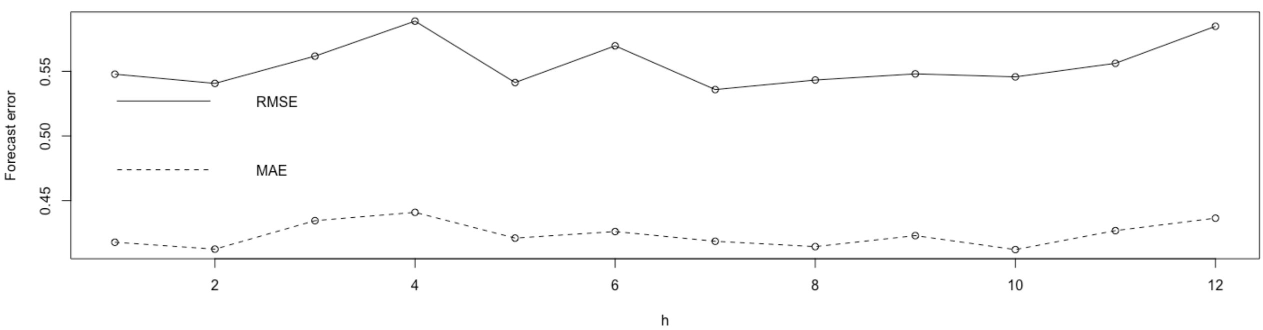
```
Series: y
Regression with ARIMA(0,0,0) errors

Coefficients:
      intercept    xreg
        14.6067  0.0039
s.e.     0.3134  0.0045

sigma^2 = 2.959:  log likelihood = -234.36
AIC=474.71   AICc=474.92   BIC=483.07
> m0$coef/sqrt(diag(m0$var.coef))
 intercept       xreg
46.6087141  0.8583074
```

The linear trend parameter is not significant, even observing the time series there does not seem to be a linear trend, a further attempt can be the use of a cubic trend.

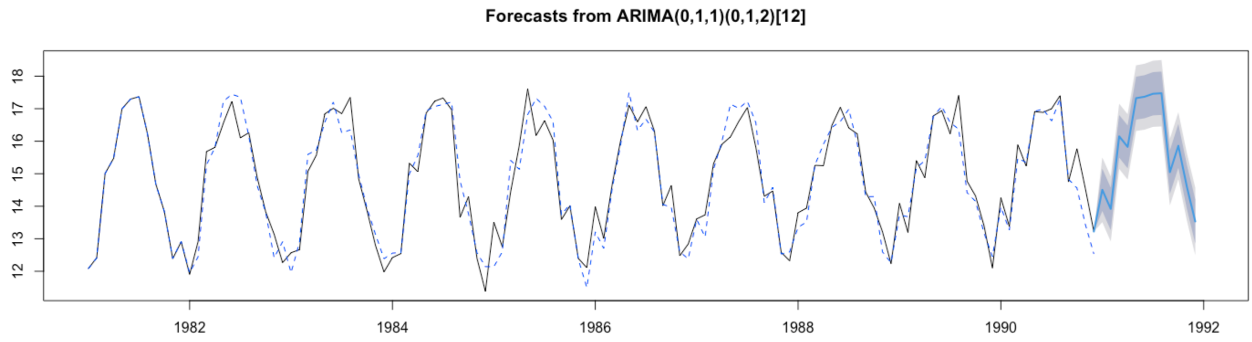
```
Series: y
Regression with ARIMA(0,0,0) errors

Coefficients:
      intercept    time   time2  time3
        14.6823  0.0068  -2e-04  0e+00
s.e.     0.6416  0.0457   9e-04  1e-04

sigma^2 = 2.998:  log likelihood = -234.12
AIC=478.25   AICc=478.78   BIC=492.19
> m0$coef/sqrt(diag(m0$var.coef))
  intercept        time       time2       time3
22.88345995  0.14832598 -0.21537787  0.01667999
```

The cubic trend also has no significant parameters.
Given the poor relationship of a linear or cubic trend to the series, we proceed using the SARIMA identified previously.

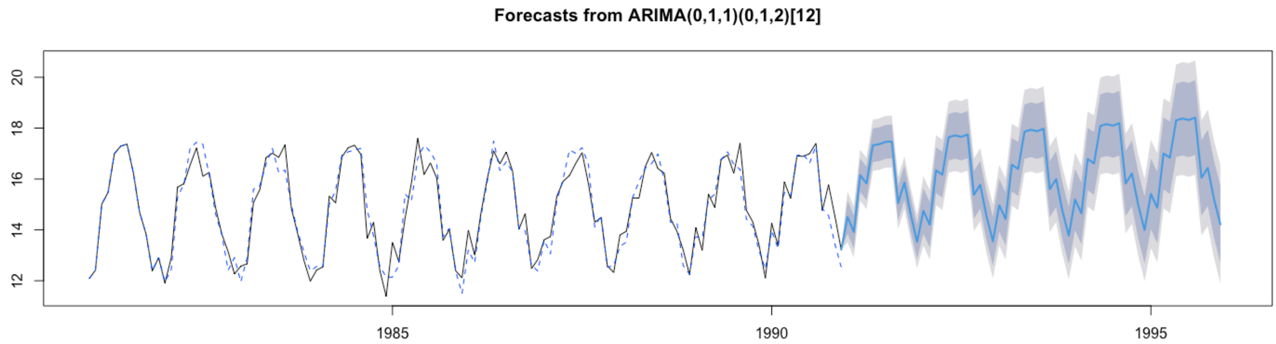
Forecast Section



We observe very low MAE and RMSE values as the forecast tends to repeat the seasonality of the series, this is more easily observable by looking at the graph with the forecast.



The forecast of such a strongly seasonal series tends to reproduce the seasonal trend seen in previous years. The variance appears to be stable except for the peak period where there is greater uncertainty.



If we extend the forecast from one year to five years, we notice how the seasonal trend continues to be repeated but the uncertainty increases but in any case we notice a convergence of the expected value while the variance tends to diverge.