fish\_dataset\_time\_series\_forecast.R

tokol

2022-02-28

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.2

## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5 v purrr 0.3.4  
## v tibble 3.1.6 v dplyr 1.0.8  
## v tidyr 1.2.0 v stringr 1.4.0  
## v readr 2.1.2 v forcats 0.5.1

## Warning: package 'ggplot2' was built under R version 4.1.2

## Warning: package 'tibble' was built under R version 4.1.2

## Warning: package 'tidyr' was built under R version 4.1.2

## Warning: package 'readr' was built under R version 4.1.2

## Warning: package 'purrr' was built under R version 4.1.2

## Warning: package 'dplyr' was built under R version 4.1.2

## Warning: package 'stringr' was built under R version 4.1.2

## Warning: package 'forcats' was built under R version 4.1.2

## -- Conflicts ------------------------------------------ tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

library(fpp)

## Warning: package 'fpp' was built under R version 4.1.2

## Loading required package: forecast

## Warning: package 'forecast' was built under R version 4.1.2

## Registered S3 method overwritten by 'quantmod':  
## method from  
## as.zoo.data.frame zoo

## Loading required package: fma

## Warning: package 'fma' was built under R version 4.1.2

## Loading required package: expsmooth

## Warning: package 'expsmooth' was built under R version 4.1.2

## Loading required package: lmtest

## Warning: package 'lmtest' was built under R version 4.1.2

## Loading required package: zoo

## Warning: package 'zoo' was built under R version 4.1.2

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

## Loading required package: tseries

## Warning: package 'tseries' was built under R version 4.1.2

library(forecast)  
library(backtest)

## Warning: package 'backtest' was built under R version 4.1.2

## Loading required package: grid

## Loading required package: lattice

library(quantmod)

## Warning: package 'quantmod' was built under R version 4.1.2

## Loading required package: xts

## Warning: package 'xts' was built under R version 4.1.2

##   
## Attaching package: 'xts'

## The following objects are masked from 'package:dplyr':  
##   
## first, last

## Loading required package: TTR

## Warning: package 'TTR' was built under R version 4.1.2

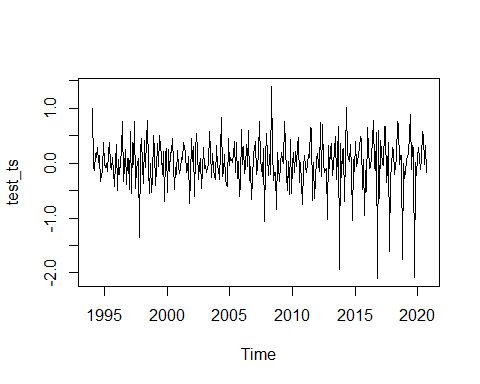
library(lubridate)

## Warning: package 'lubridate' was built under R version 4.1.2

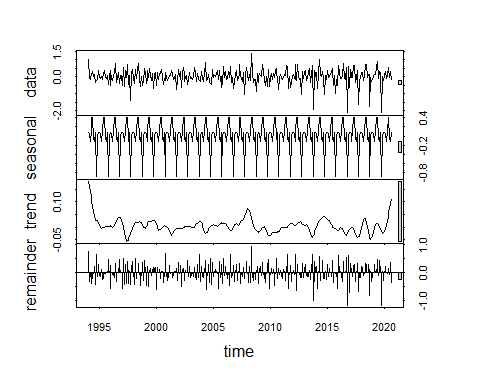
##   
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':  
##   
## date, intersect, setdiff, union

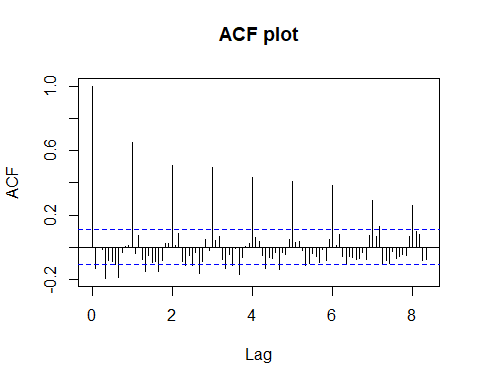
library(dplyr)  
  
Sale\_of\_fish <- read.csv("Fish dataset.csv",header=T)  
  
#remove columns that are not interested  
adj <- select(Sale\_of\_fish, -c("year","month","end.of.period","value..E.","unsold..Kg."))  
#remove the space from the number and convert to numeric format  
adj$sales <- as.numeric(gsub(' ','',adj$sales))  
#convert to day, month and year format  
adj$start <- dmy(adj$start)  
#log(sales from the last period) - log(sales of the current period)  
adj$logr <- log(lag(adj$sales)) - log(adj$sales)  
  
#remove NA  
narm <- function (x)  
{  
 x[is.na(x)] <- 1  
 return(x)  
}  
  
adj$logr <- narm(adj$logr)  
  
max\_date <- max(adj$start)  
min\_date <- min(adj$start)  
  
test\_ts <- ts(adj$logr,end = c(year(max\_date),month(max\_date)),start = c(year(min\_date),month(min\_date)), frequency = 12)  
  
plot(test\_ts)



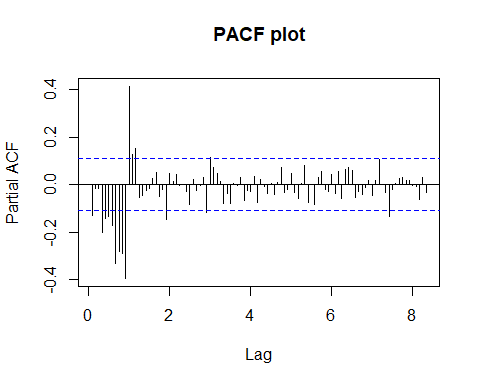
#raw data, seasonality, trend, residuals  
plot(stl(test\_ts,s.window ="periodic"))



acf = acf(test\_ts, main = 'ACF plot', lag.max = 100)



pacf.logr = pacf(test\_ts,main = 'PACF plot',lag.max = 100)



#Augmented Dickey-Fuller(ADF) Test  
print(adf.test(test\_ts))

## Warning in adf.test(test\_ts): p-value smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: test\_ts  
## Dickey-Fuller = -10.997, Lag order = 6, p-value = 0.01  
## alternative hypothesis: stationary

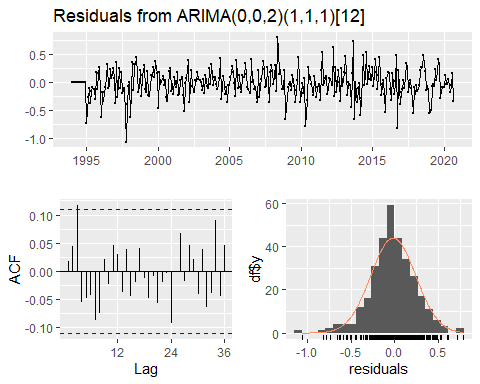
#The obtained p-value < 0.05, therefore it is safe to assume that the dataset is stationary  
  
m1 <- auto.arima(test\_ts, seasonal = TRUE)  
summary(m1)

## Series: test\_ts   
## ARIMA(0,0,2)(1,1,1)[12]   
##   
## Coefficients:  
## ma1 ma2 sar1 sma1  
## -0.5810 -0.1680 0.1845 -0.7589  
## s.e. 0.0589 0.0673 0.0818 0.0548  
##   
## sigma^2 = 0.06985: log likelihood = -29.19  
## AIC=68.38 AICc=68.58 BIC=87.05  
##   
## Training set error measures:  
## ME RMSE MAE MPE MAPE MASE  
## Training set -0.01977306 0.257616 0.1944489 125.3062 246.7707 0.7168303  
## ACF1  
## Training set 0.01697301

accuracy(forecast(m1))

## ME RMSE MAE MPE MAPE MASE  
## Training set -0.01977306 0.257616 0.1944489 125.3062 246.7707 0.7168303  
## ACF1  
## Training set 0.01697301

checkresiduals(m1)



##   
## Ljung-Box test  
##   
## data: Residuals from ARIMA(0,0,2)(1,1,1)[12]  
## Q\* = 20.655, df = 20, p-value = 0.4177  
##   
## Model df: 4. Total lags used: 24