pstat274lab6

2023-11-10

Lab Assignment 6

We will analyze adjusted monthly milk production measured in pounds per from Jan. 1962 to Dec. 1975 from Lab 5. And we can import the dataset from tsdl package as milk in R, and denote the milk time series as Xt. For comparison, we split the dataset into training set train and testing set test. The training set is used for model building, and the testing set is used for prediction verification and comparison.

```
library(tsdl)
library(astsa)
milk <- subset(tsdl, 12, "Agriculture")[[3]]
source("plot.roots.R", local=knitr::knit_global())
```

(a). Split the dataset into training set train and testing set test. The testing set is the last 6 months of the data and training set is the rest of the first 150 months of the data.

```
train_data = milk[c(1: 150)]
test_data = milk[c(151:156)]
```

(b). From previous lab assignment, we determine possible candidate models SARIMA(p, d, q) $\times$ (P, D, Q)s for the series Yt. Fit the training data to the chosen model.(c). For the model you selected, do we need to set any coefficient to zero and why?

From lab 5, we have two candidate model SARIMA(0, 1, 0) $\times$ (2, 1, 1)12 or SARIMA(0, 1, 0) * (4, 1, 1)12

1. For SARIMA(0, 1, 0)(2, 1, 1)12

```
# Fit the SARIMA model to the training data
library(astsa)
fit.i <- sarima(xdata = train_data,p = 0, d = 1, q = 0,
                P = 2, D = 1, Q = 1,S = 12, details = F)
fit.i$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
```

```
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##         sar1    sar2     sma1
##       0.0737  0.0168  -0.7513
## s.e.  0.1609  0.1242   0.1517
##
## sigma^2 estimated as 34.63:  log likelihood = -441.42,  aic = 890.84
```

Checking coefficients: From the above coefficients table,sar1 and sar2 are not significant because the confidence interval of the estimated coefficient contains 0. Therefore, we should set these coefficients to 0.

```
fit.i1 <- sarima(xdata = train_data, p = 0, d = 1, q = 0,
                 P = 2 , D = 1, Q = 1, S = 12,
                 fixed=c(0, 0, NA),
                 details = F)
fit.i1$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##       sar1  sar2     sma1
##          0     0  -0.6989
## s.e.     0     0   0.0804
##
## sigma^2 estimated as 34.79:  log likelihood = -441.54,  aic = 887.07
```

From the above output, we can write the model as: (1-B)(1-B12)Yt = (1-0.6989B12)Zt

2. For SARIMA(0, 1, 0) * (4, 1, 1)12

```
fit.i <- sarima(xdata = train_data,p = 0, d = 1, q = 0,
                 P =4 , D = 1, Q = 1, S = 12, details = F)
fit.i$fit
```

```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##          sar1     sar2     sar3     sar4    sma1
##       -0.6547  -0.4459  -0.3463  -0.4192  0.0068
## s.e.   0.2256   0.1595   0.1205   0.0961  0.2553
##
## sigma^2 estimated as 31.05:  log likelihood = -436.82,  aic = 885.65
```

Checking coefficients: From the above coefficients table,sma1 is not significant because the confidence interval of the estimated coefficient contains 0. Therefore, we should set these coefficients to 0.

```
fit.i1 <- sarima(xdata = train_data, p = 0, d = 1, q = 0,
                 P = 4 , D = 1, Q = 1, S = 12,
                 fixed=c(NA,NA,NA,NA,0),
                 details = F)
fit.i1$fit
```

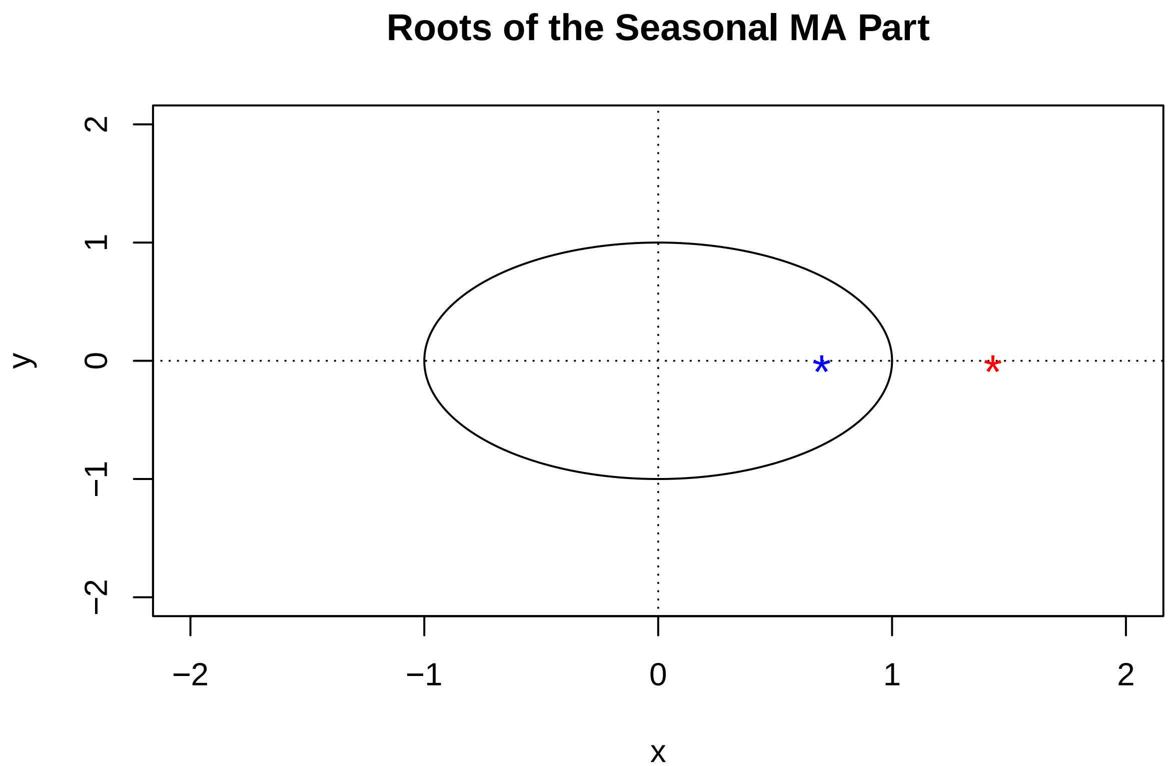
```
##
## Call:
## arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D, Q), period = S),
##     include.mean = !no.constant, transform.pars = trans, fixed = fixed, optim.control = list(trace =
##         REPORT = 1, reltol = tol))
##
## Coefficients:
##          sar1     sar2     sar3     sar4  sma1
##       -0.6491  -0.4425  -0.3445  -0.4186     0
## s.e.   0.0817   0.1001   0.1010   0.0942     0
##
## sigma^2 estimated as 31.05:  log likelihood = -436.82,  aic = 883.65
```

Here is the model (1-0.6491B12-0.4425B24-0.3445B36-0.4186B12)(1-B)(1-B12)Xt=Zt

(d). For the model you selected, check the model stationarity and invertibility.

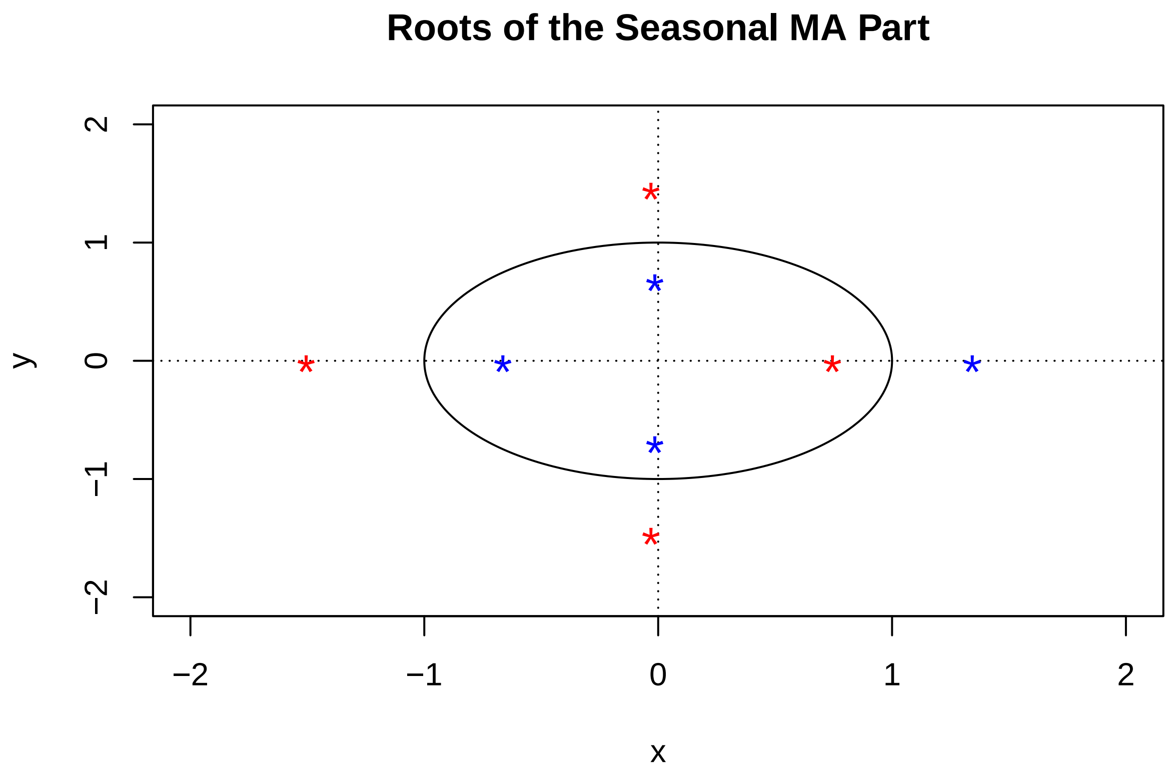
For model (1-B)(1-B12)Yt = (1-0.6989B12)Zt, it is is stationary because it includes differencing terms and has no autoregressive components. By checking the roots of the moving average part, we find all roots lie outside the unit circle, confirming that the model is invertible.

```
# check (1-B)(1-B12)Yt = (1-0.6989B12)Zt' s root
plot.roots(NULL,polyroot(c(1,-0.6989)), main="Roots of the Seasonal MA Part")
```



For model (1-0.6491B12-0.4425B24-0.3445B36-0.4186B12)(1-B)(1-B12)Xt=Zt, it has no moving average terms, so it's automatically invertible. The stationarity of the model is in question because not all autoregressive roots lie outside the unit circle, indicating the model is not stationary.

```
# check stantionarity
plot.roots(NULL,polyroot(c(1,-0.6491,-0.4425,-0.3445,-0.4186)), main="Roots of the Seasonal MA Part")
```



In conclusion, The final fitting model would be (1-B)(1-B12)Yt = (1-0.6989B12)Zt.