732A62 Lab 2

Emil K Svensson & Rasmus Holm

2017-10-02

Assignment 1

a)

```
library(astsa)
library(kernlab)
library(TSA)
library(forecast)

set.seed(12345)
AR3 <- arima.sim(1000, model = list(order = c(3,0,0),
                                    ar = c(0.8, -0.2, 0.1)))

## The theoretical
AR3.pacf <- pacf(AR3, plot=F)
AR3.data <- ts.intersect(xt = AR3, x1 = lag(AR3, 1), x2 =lag(AR3, 2), x3= lag(AR3, 3))

AR.lm <- resid(lm(xt ~ x1 + x2, data = AR3.data))
AR.lm.lag3 <- resid(lm(x3 ~ x1 + x2 , data = AR3.data))

AR3.pacf[3]
```

```
##
## Partial autocorrelations of series 'AR3', by lag
##
##     3
## 0.117
```

```
cat(paste("The theoretical value:",round(cor(AR.lm, AR.lm.lag3), digits = 3)))
```

```
## The theoretical value: 0.115
```

As seen above the theoretical and the output from the pacf-function are very similar.

b)

```
set.seed(12345)
AR2 <- arima.sim(100, model = list(order = c(2,0,0),
                                   ar = c(0.8, 0.1)))

ar2.yw <- ar(AR2, order.max = 2, method = "yw", aic = FALSE)
ar2.ols <- ar(AR2, order.max = 2, method = "ols", aic = FALSE)
ar2.mle <- arima(AR2, order = c(2,0,0), method = "ML")
```

```
ar2.yw$ar
```

```
## [1] 0.8029146 0.1037053
```

```
ar2.ols$ar
```

```
## , , 1
##
##            [,1]
## [1,] 0.8066782
## [2,] 0.1205352
```

```
ar2.mle$coef
```

```
##       ar1       ar2 intercept
## 0.7966919 0.1188537 0.8289656
```

The Yule Walker estimate seems to have the parameters closes to the true parameters given in the assignment.

```
ar2.mle
```

```
##
## Call:
## arima(x = AR2, order = c(2, 0, 0), method = "ML")
##
## Coefficients:
##          ar1     ar2  intercept
##       0.7967  0.1189     0.8290
## s.e.  0.0992  0.1000     1.1385
##
## sigma^2 estimated as 1.126:  log likelihood = -148.71,  aic = 303.41
```

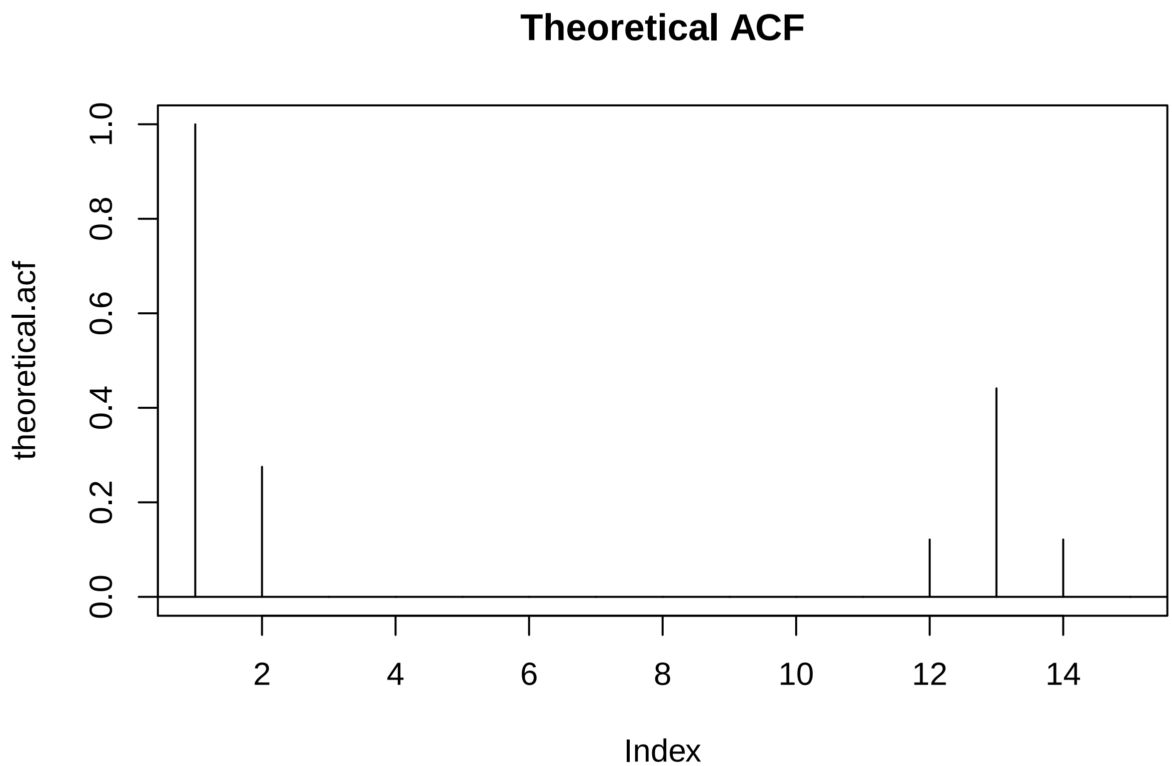
Yes, the theoretical value for

$$\phi_2$$

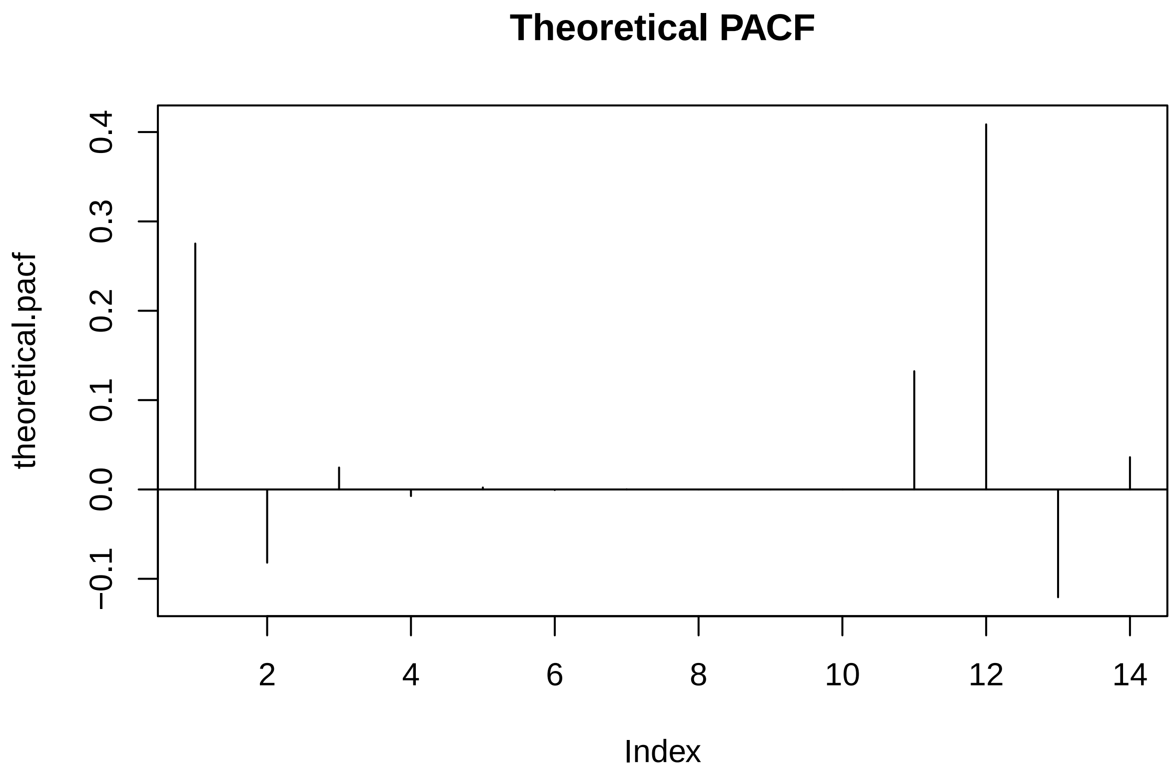
is inside the confidence-intervall for the ML estimate which can seen from the s.e times 1.96 which obviously will cover the true coefficients.

c)

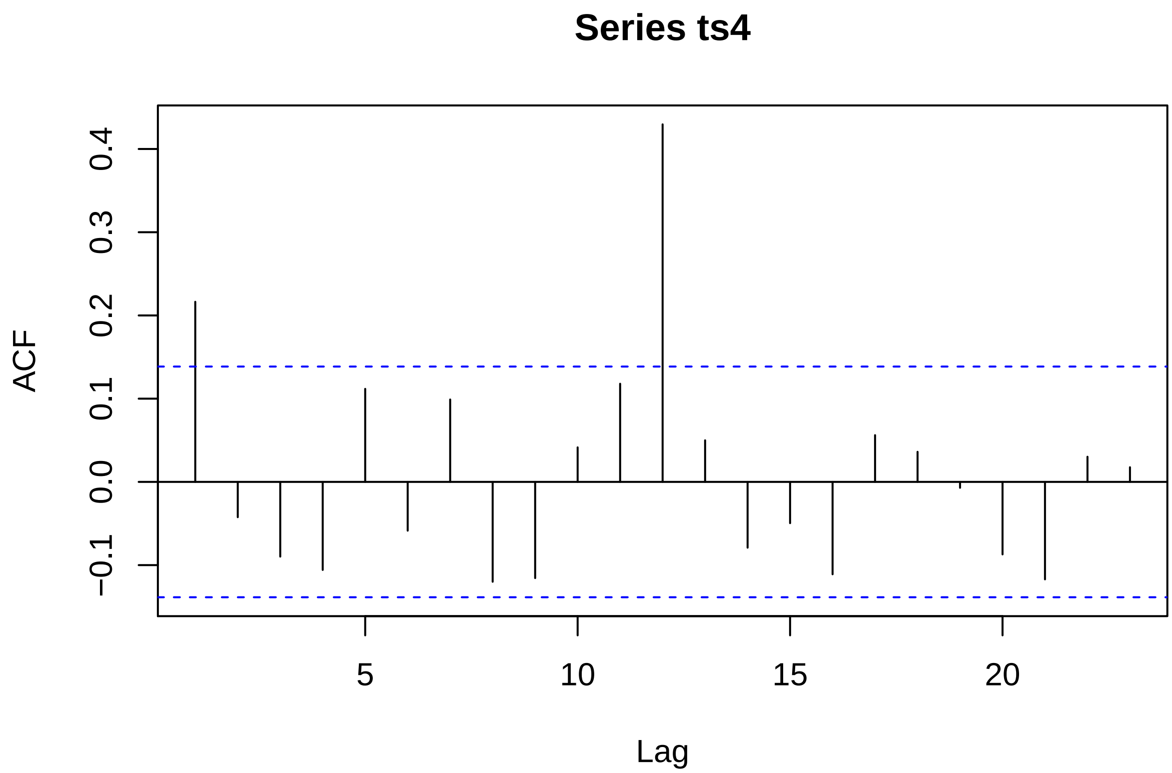
```
set.seed(12345)
ma.coef <- c(0.3, rep(0, 10), 0.6)
ts4 <- arima.sim(n=200, model=list(order=c(0, 0, 12), ma = ma.coef))

theoretical.acf <- ARMAacf(ma=c(ma.coef, 0.3 * 0.6))
theoretical.pacf <- ARMAacf(ma=c(ma.coef, 0.3 * 0.6), pacf=TRUE)

plot(theoretical.acf, type="h", main="Theoretical ACF")
abline(h=0)
```



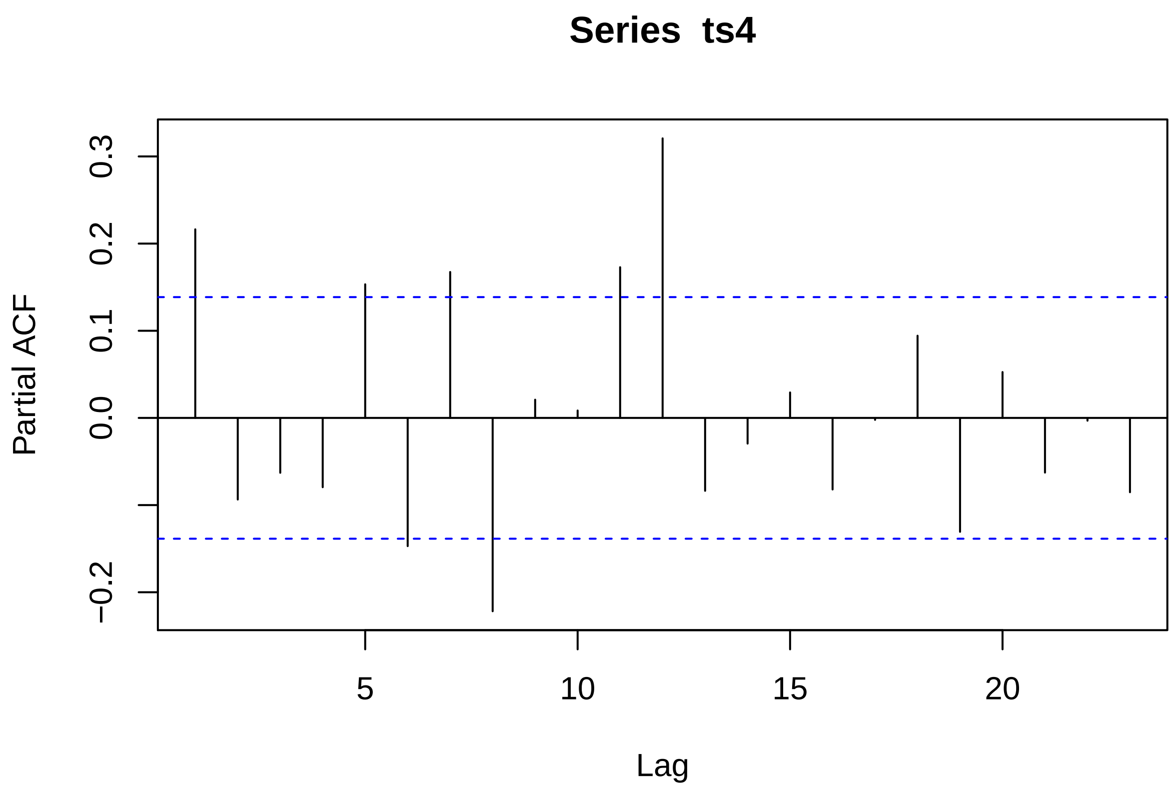

```r
plot(theoretical.pacf, type="h", main="Theoretical PACF")
abline(h=0)
```




```
empirical.acf <- acf(ts4)
```




```
empirical.pacf <- pacf(ts4)
```

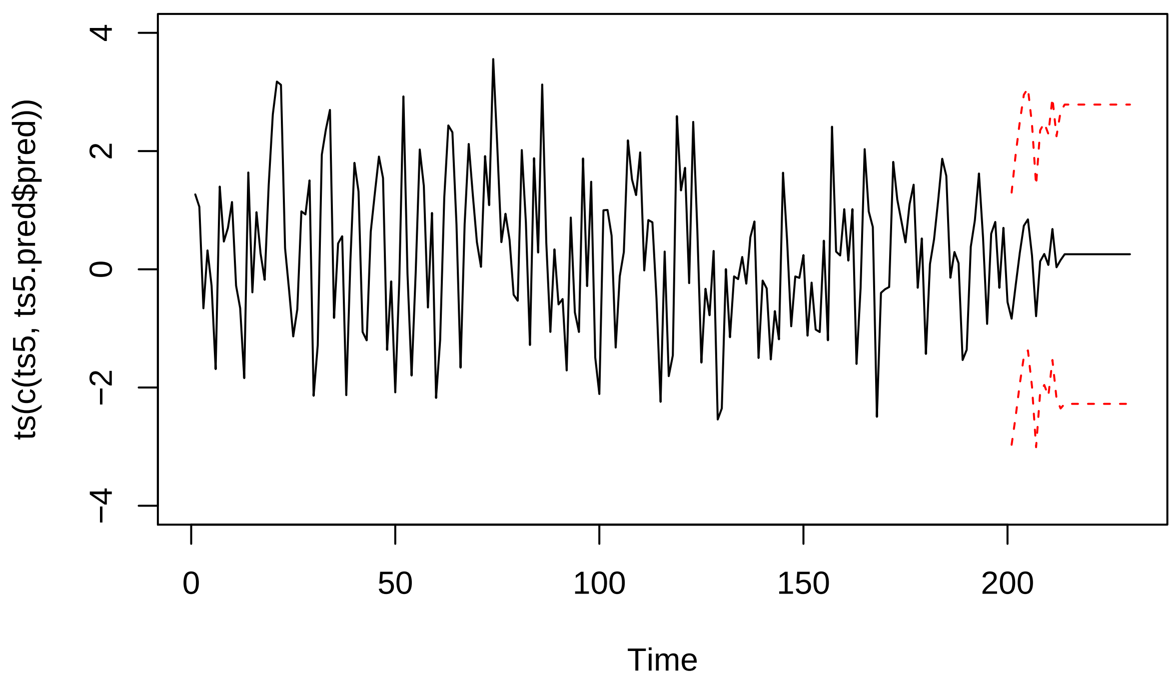


The patterns seem to be somewhat similar. In the theoretical ACF we can see a large spike at lag 14 which also can be seen in the sample ACF. The difference between the two beeing that the sample ACF has some correlation along the lags although under the confidence interval.

For the PACF we can do the same observation as for the relationship between the theoretical and the sample.

d)

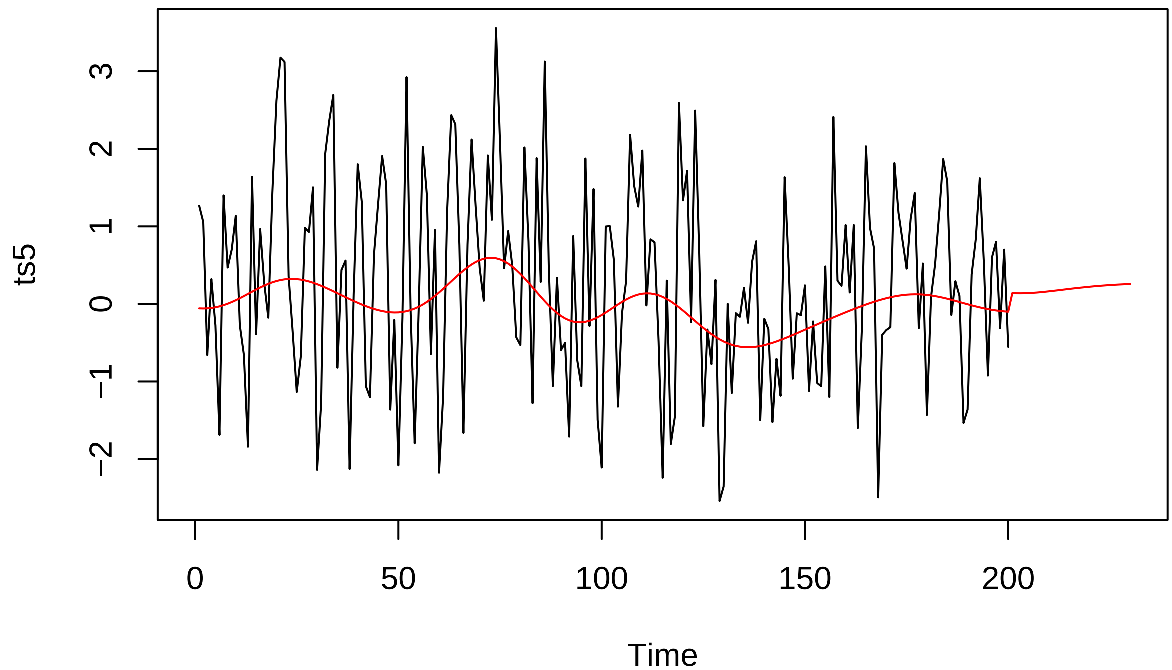
```r
set.seed(12345)
ma.coef <- c(0.3, rep(0, 10), 0.6)
ts5 <- arima.sim(n=200, model=list(order=c(0, 0, 12), ma = ma.coef))

ts5.fit <- arima(ts5, order=c(0, 0, 1), seasonal=list(order=c(0, 0, 1), period=12))
ts5.pred <- predict(ts5.fit, n.ahead=30, se.fit=TRUE)

plot(ts(c(ts5, ts5.pred$pred)), ylim=c(-4, 4))
lines(200 + 1:length(ts5.pred$pred), ts5.pred$pred + 1.96 * ts5.pred$se, lty=2, col="red")
lines(200 + 1:length(ts5.pred$pred), ts5.pred$pred - 1.96 * ts5.pred$se, lty=2, col="red")
```




```
gausspr.data <- data.frame(y=ts5, x=1:200)
gausspr.fit <- gausspr(y ~ x, gausspr.data)
```

```
## Using automatic sigma estimation (sigest) for RBF or laplace kernel
```

```
gausspr.pred <- predict(gausspr.fit, data.frame(x=201:230))

plot(ts5, xlim=c(0, 230))
lines(c(fitted(gausspr.fit), gausspr.pred), , col="red")
```

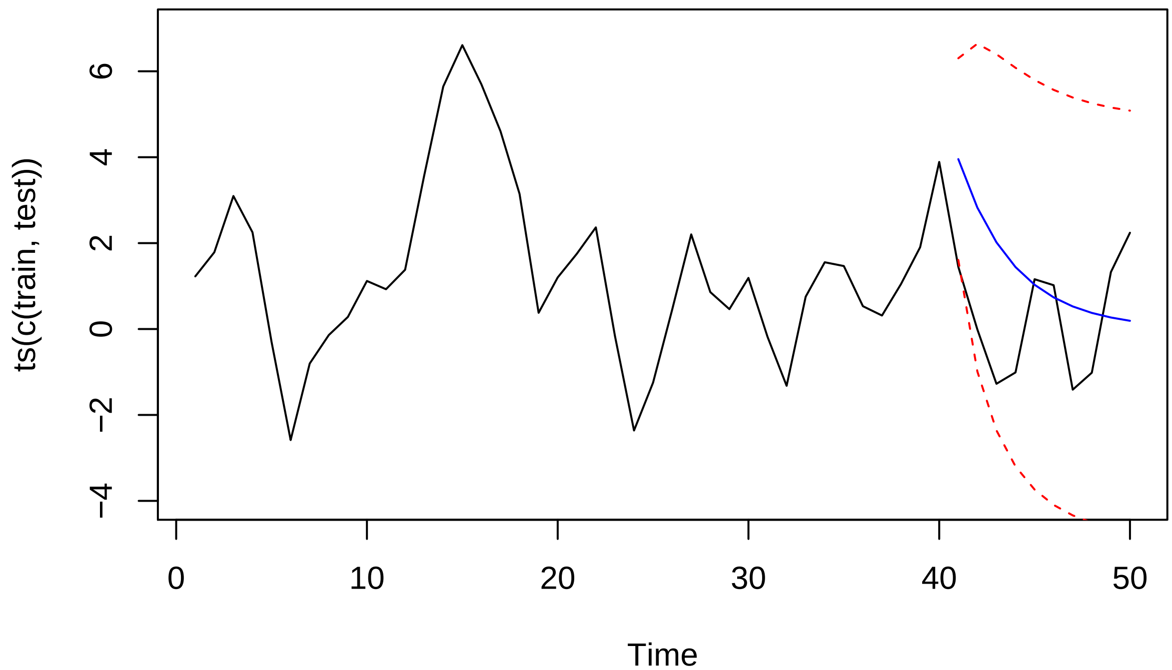


In the first plot we can see the MA models predicions seem resonable but after a while the predictions just tails of in to a mean function at y = 0.

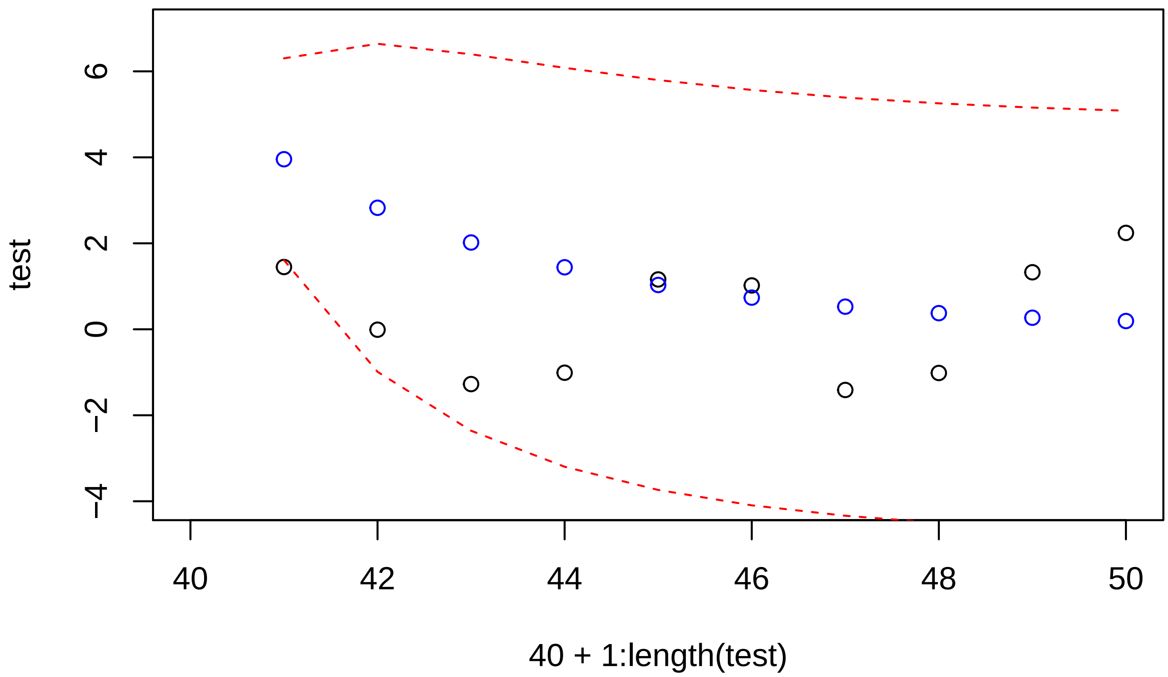
For the gaussian process there is an initial jump in the predictions and afterwards continues with a smoothed pattern similar to the fitted line in the observed data.

e)

```
set.seed(12345)
ts6 <- arima.sim(model=list(ma=c(0.5), ar=c(0.7)), n=50)

train <- ts(ts6[1:40])
test <- ts(ts6[41:50])

ts6.fit <- arima(train, order=c(1, 0, 1), include.mean = F)
ts6.pred <- predict(ts6.fit, n.ahead=10)

plot(ts(c(train, test)), ylim=c(-4, 7), type="l")
lines(40 + 1:length(test), ts6.pred$pred, col="blue")
lines(40 + 1:length(test), ts6.pred$pred + 1.96 * ts6.pred$se, lty=2, col="red")
lines(40 + 1:length(test), ts6.pred$pred - 1.96 * ts6.pred$se, lty=2, col="red")
```




```r
plot(40 + 1:length(test), test, ylim=c(-4, 7), xlim=c(40, 50), type="p")
points(40 + 1:length(test), ts6.pred$pred, col="blue")
lines(40 + 1:length(test), ts6.pred$pred + 1.96 * ts6.pred$se, lty=2, col="red")
lines(40 + 1:length(test), ts6.pred$pred - 1.96 * ts6.pred$se, lty=2, col="red")
```

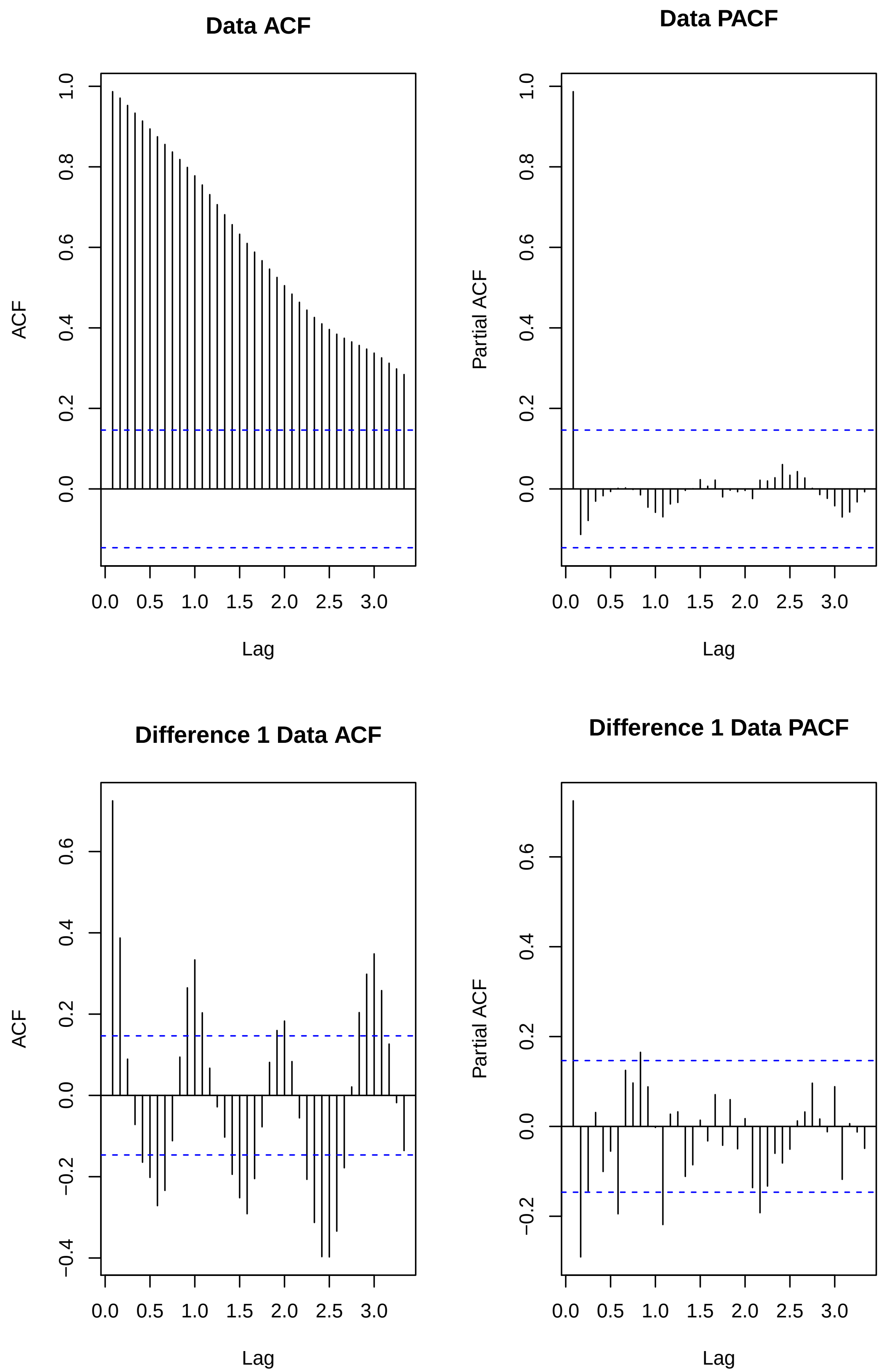


The first of the 10 observations withheld from the fitted model is the only observation that is not in the prediction interval. Since it is a 95 % prediction interval we say that we expect that 5 of 100 observations will be outside the interval it is not unresonable that 1 of 10 is outside the interval.

Assignment 2

```
assignment2 <- function(data){
    old <- par(mfrow = c(2, 2))
    acf(data, lag.max = 40, main="Data ACF")
    pacf(data, lag.max = 40, main="Data PACF")
    acf(diff(data, lag = 1), lag.max = 40, main="Difference 1 Data ACF")
    pacf(diff(data, lag = 1), lag.max = 40, main="Difference 1 Data PACF")
    par(old)
}
```

Chicken


Data ACF
ACF
Lag
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Data PACF
Partial ACF
Lag
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Difference 1 Data ACF
ACF
Lag
−0.4 −0.2 0.0 0.2 0.4 0.6
0.0 0.5 1.0 1.5 2.0 2.5 3.0
Difference 1 Data PACF
Partial ACF
Lag
−0.2 0.0 0.2 0.4 0.6
0.0 0.5 1.0 1.5 2.0 2.5 3.0

Data ACF

The ACF on the original data suggests an AR or ARMA model since the ACF tails off.

Data PACF

The PACF on the original data cuts off after lag 1 suggesting an AR(1) model.

Difference 1 Data ACF

After having performed difference of order 1 we can clearly see that there is a seasonal trend in the data. The ACF suggests a seasonality of 10 but it does not seem to tail off.

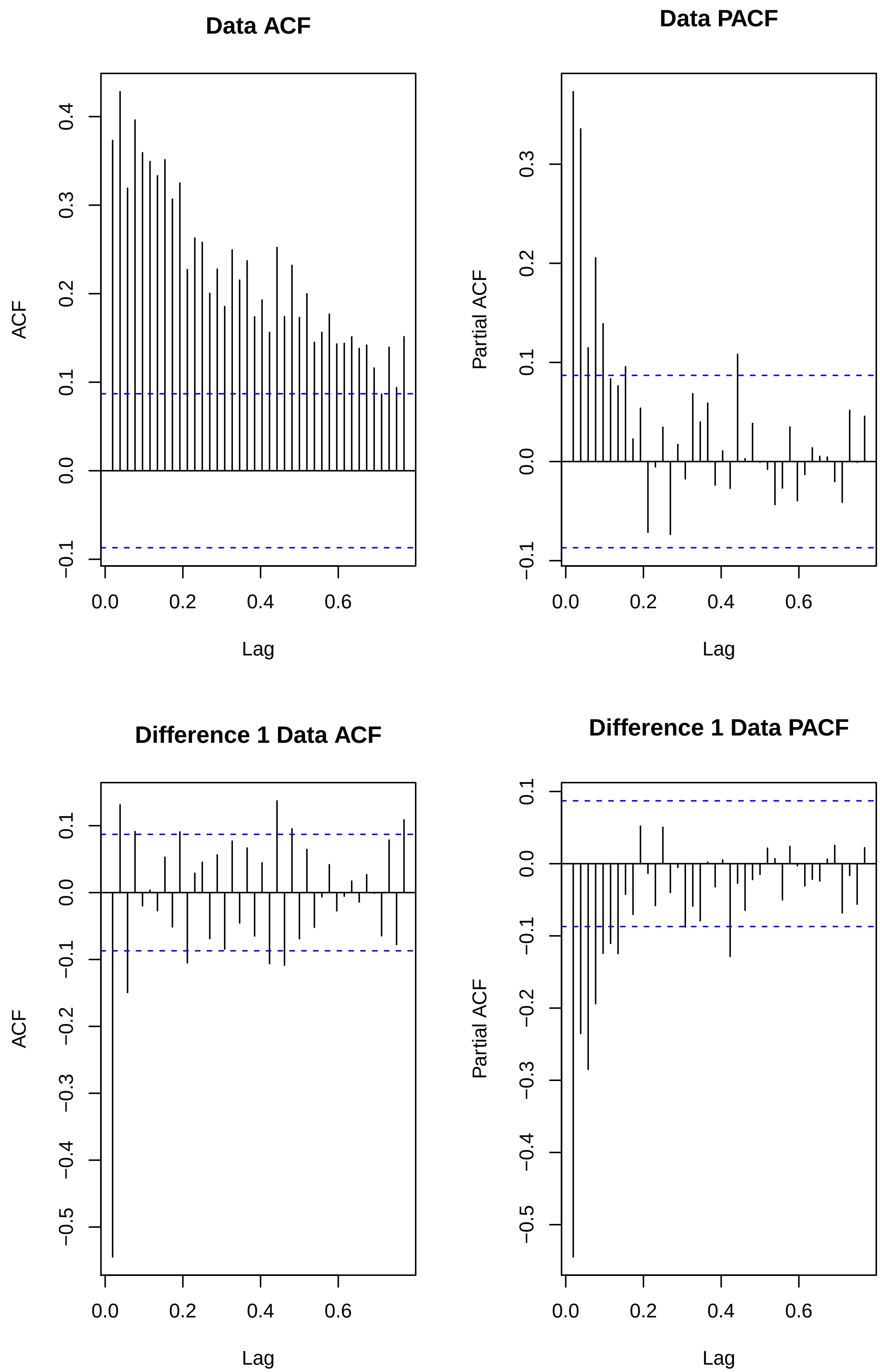
Difference 1 Data PACF

The PACF indicates that ...

Final Verdict

Needs further investigation ARIMA(1, 0, 0) x $(\_, 1, \_)_{1}0$

so2



Data ACF
ACF
Lag
Data PACF
Partial ACF
Lag
Difference 1 Data ACF
ACF
Lag
Difference 1 Data PACF
Partial ACF
Lag

Data ACF

The ACF tails off suggesting either an AR or ARMA model.

Data PACF

The PACF tails off as well suggesting an ARMA model.

Difference 1 Data ACF

The ACF after difference cuts off after lag 1 suggesting a MA(1) model.

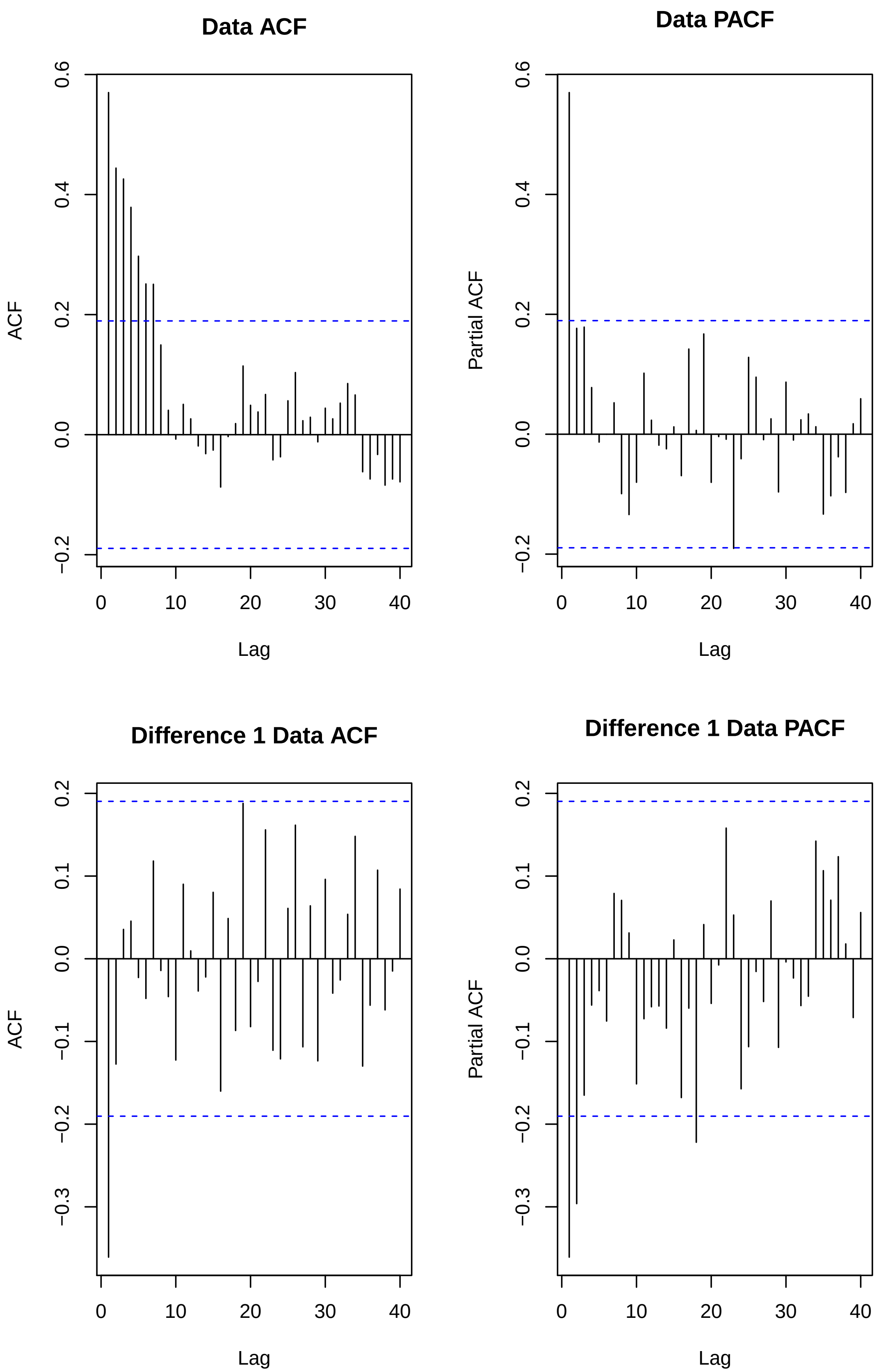
Difference 1 Data PACF

The PACF after difference tails off further suggesting a MA(1) model.

Final Verdict

ARIMA(0, 1, 1)

EQcount


Data ACF
ACF
Lag
Data PACF
Partial ACF
Lag
Difference 1 Data ACF
ACF
Lag
Difference 1 Data PACF
Partial ACF
Lag

Data ACF

The ACF tails off suggesting an AR or ARMA model.

Data PACF

The PACF cuts off after lag 1 suggesting AR(1) model.

Difference 1 Data ACF

The ACF after difference cuts off after lag 1 suggesting a MA(1) model.

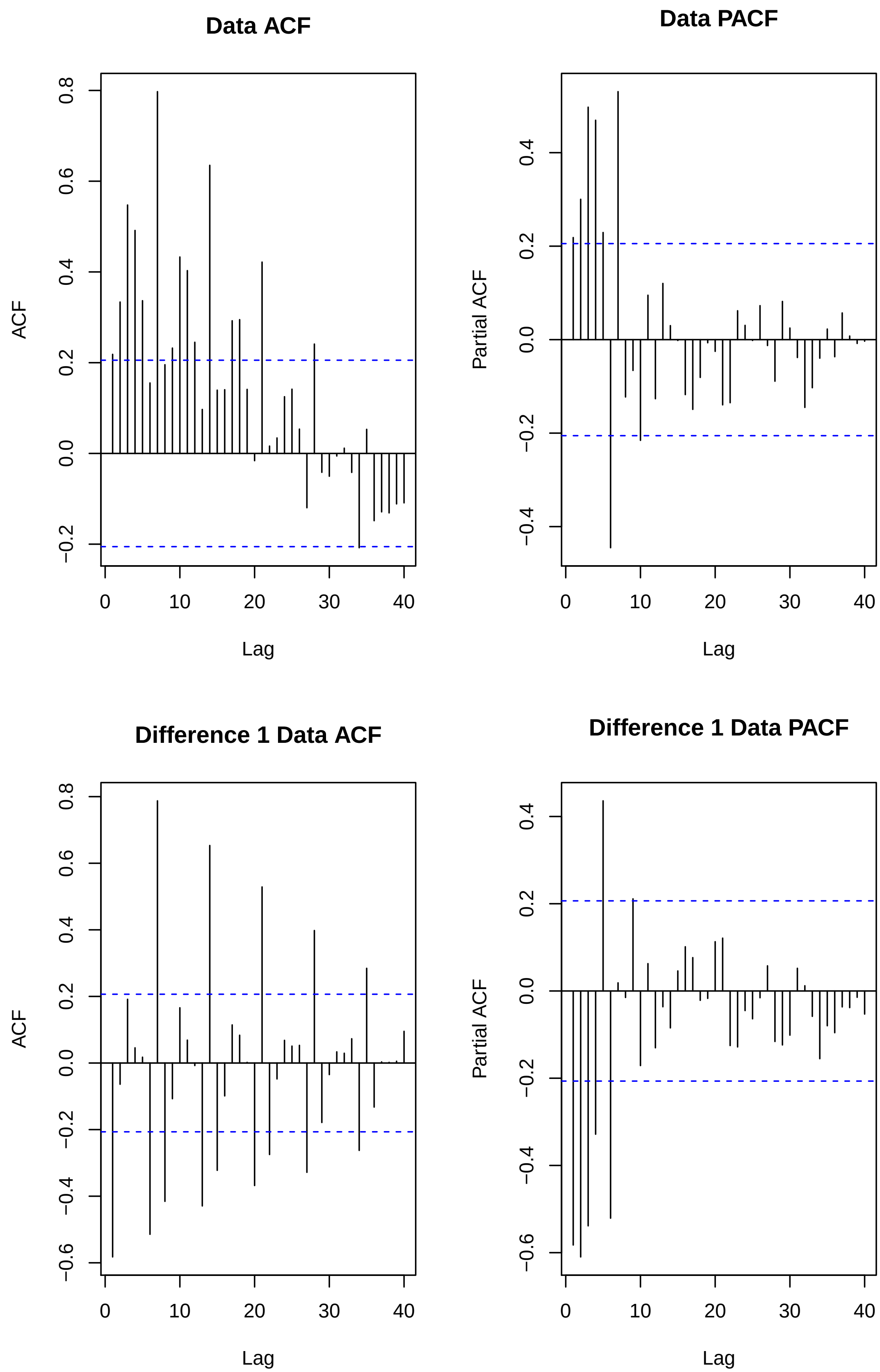
Difference 1 Data PACF

The PACF after difference tails off further suggesting a MA model.

Final Verdict

Either a ARMA(1, 0, 0) or ARMA(0, 1, 1)

HCT



Data ACF
ACF
Lag
Data PACF
Partial ACF
Lag
Difference 1 Data ACF
ACF
Lag
Difference 1 Data PACF
Partial ACF
Lag

Data ACF

The ACF tails off suggesting either an AR or ARMA model.

Data PACF

The PACF cuts off after lag 7 suggesting an AR(7) model.

Difference 1 Data ACF

The ACF suggests seasonality that tails off after lag 7 suggesting an seasonality of 7.

Difference 1 Data PACF

The PACF cuts off after 6 lags suggesting an AR(6) seasonality model.

Final Verdict

ARIMA(7, 0, 0) x $(1, 1, 0)_7$

Assignment 3

```
plot_helper <- function(data, title) {
    old <- par(mfrow=c(4, 1))
    plot(data, main=title)
    acf(data, lag.max=40, main="")
    pacf(data, lag.max=40, main="")
    qqnorm(data, main="", las=1)
    qqline(data)
    par(old)
}

test_helper <- function(data) {
    print(Box.test(data, lag = 1, type = "Ljung-Box"))
    print(suppressWarnings(adf.test(data)))
    e <- eacf(data)
}

fit_plot <- function(model) {
    pred <- predict(model, n.ahead=20, se.fit=TRUE)
    upper_band <- pred$pred + 1.96 * pred$se
    lower_band <- pred$pred - 1.96 * pred$se

    plot(c(model$x, pred$pred), type="l",
         xlim=c(500, length(oil) + 20),
         ylim=c(min(lower_band), max(upper_band)))
    lines(length(oil) + 1:20, upper_band, lty=2, col="red")
    lines(length(oil) + 1:20, lower_band, lty=2, col="red")
}
```

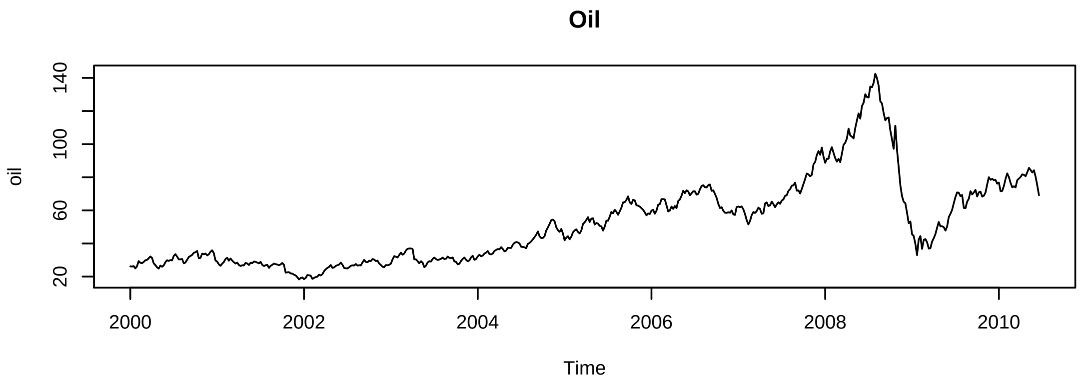
a)

```
loil <- log(oil)
doil <- diff(oil)
ddoil <- diff(oil, 2)
dloil <- diff(loil)
ddloil <- diff(loil, 2)
```

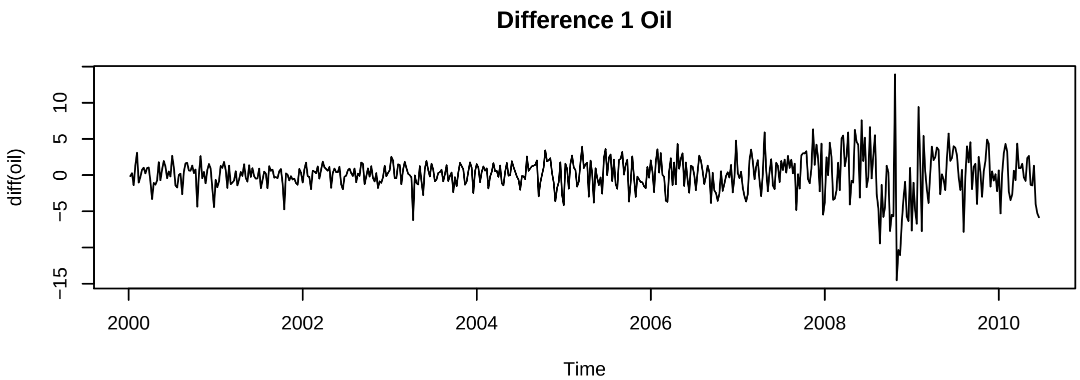
```
test_helper(doil)
test_helper(ddoil)
test_helper(dloil)
test_helper(ddloil)
```

```
fit1 <- Arima(loil, order=c(0, 2, 1))
fit1

fit_plot(fit1)
```

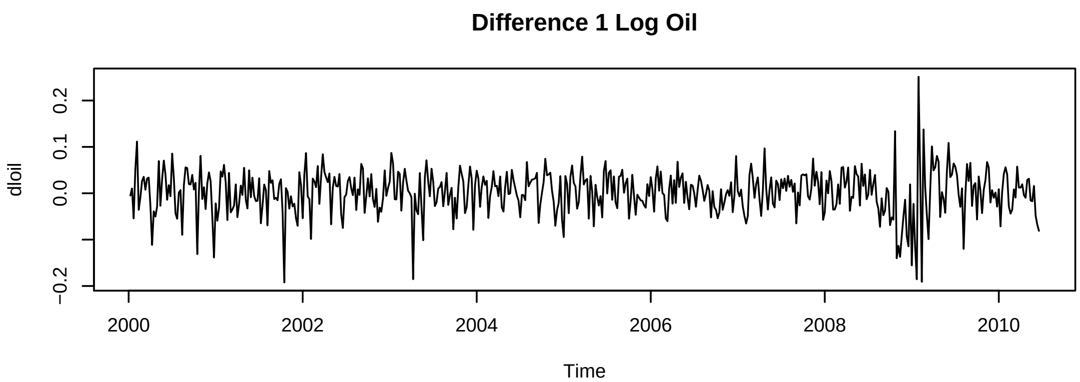
```
fit2 <- Arima(loil, order=c(0, 1, 3))
fit2

fit_plot(fit2)
```



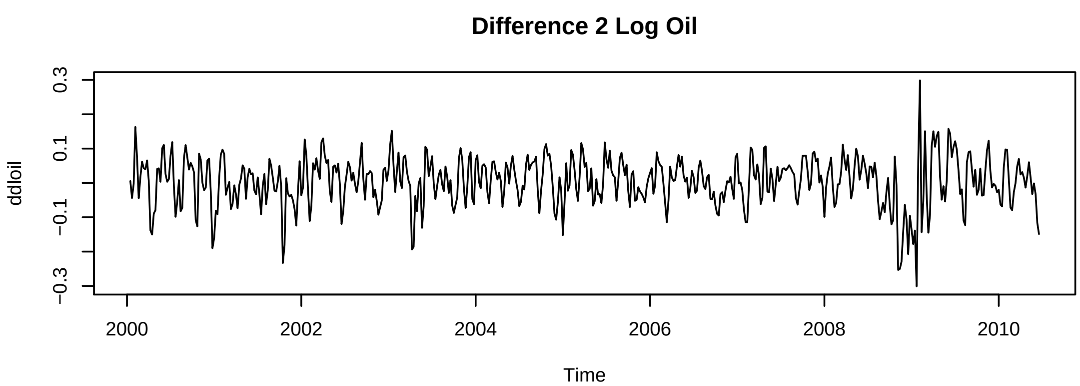
Oil
oil
20
60
100
140
2000
2002
2004
2006
2008
2010
Time



Difference 1 Oil
diff(oil)
−15
−5
0
5
10
2000
2002
2004
2006
2008
2010
Time



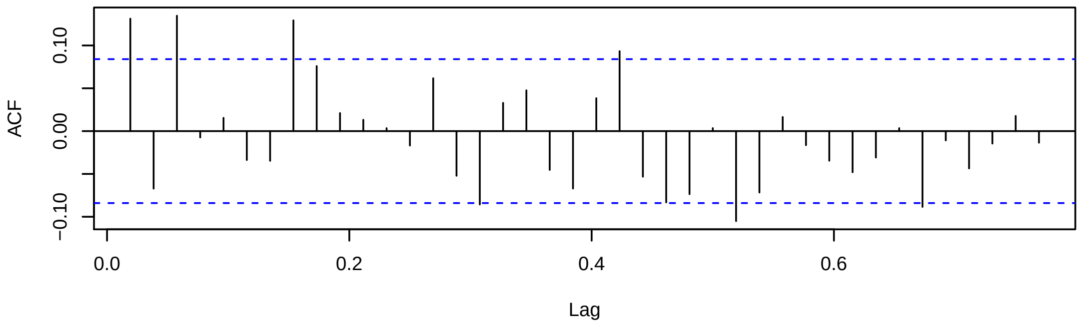
Difference 1 Log Oil
dloil
−0.2
0.0
0.1
0.2
2000
2002
2004
2006
2008
2010
Time



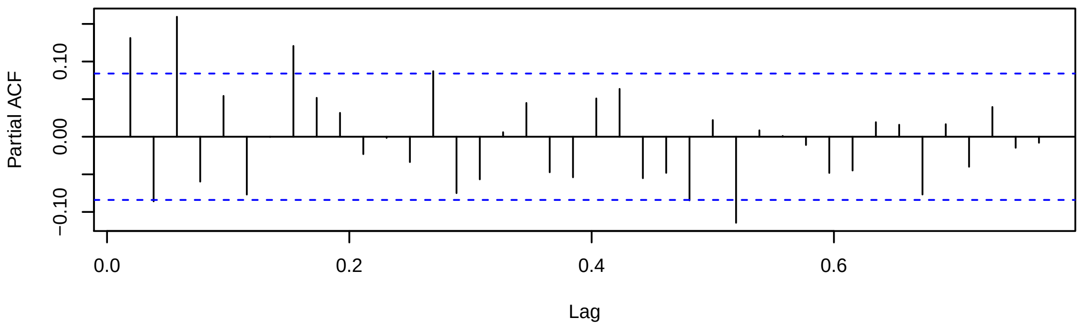
Difference 2 Log Oil
ddloil
−0.3
−0.1
0.1
0.3
2000
2002
2004
2006
2008
2010
Time

Clearly difference log is the data we should work with. bla, bla, ...

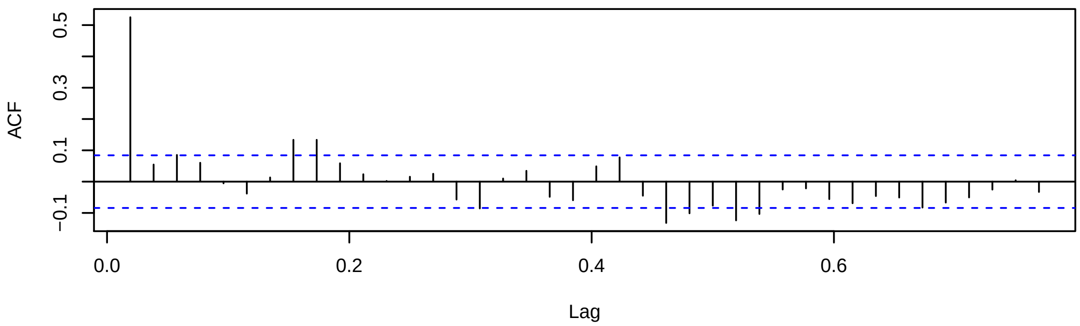
Difference 1 Log Oil ACF



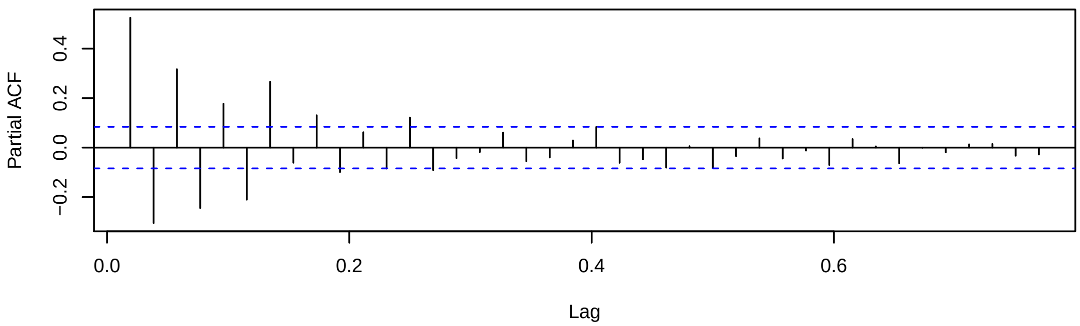
Difference 1 Log Oil PACF



Difference 2 Log Oil ACF



Difference 2 Log Oil PACF

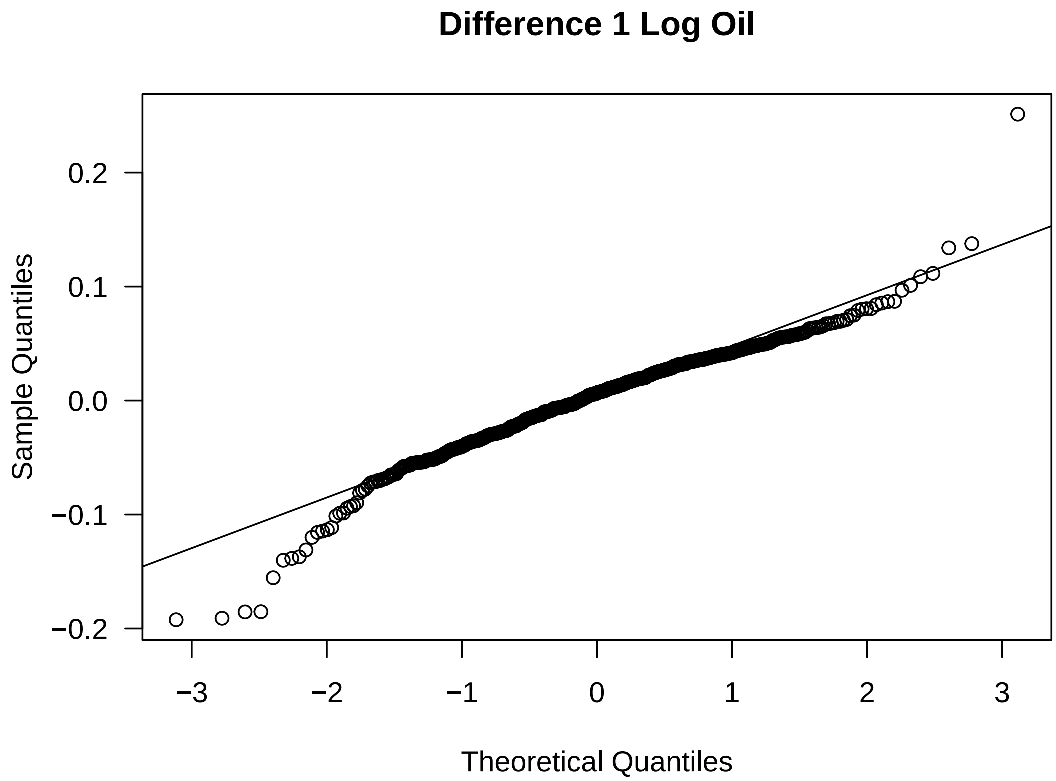


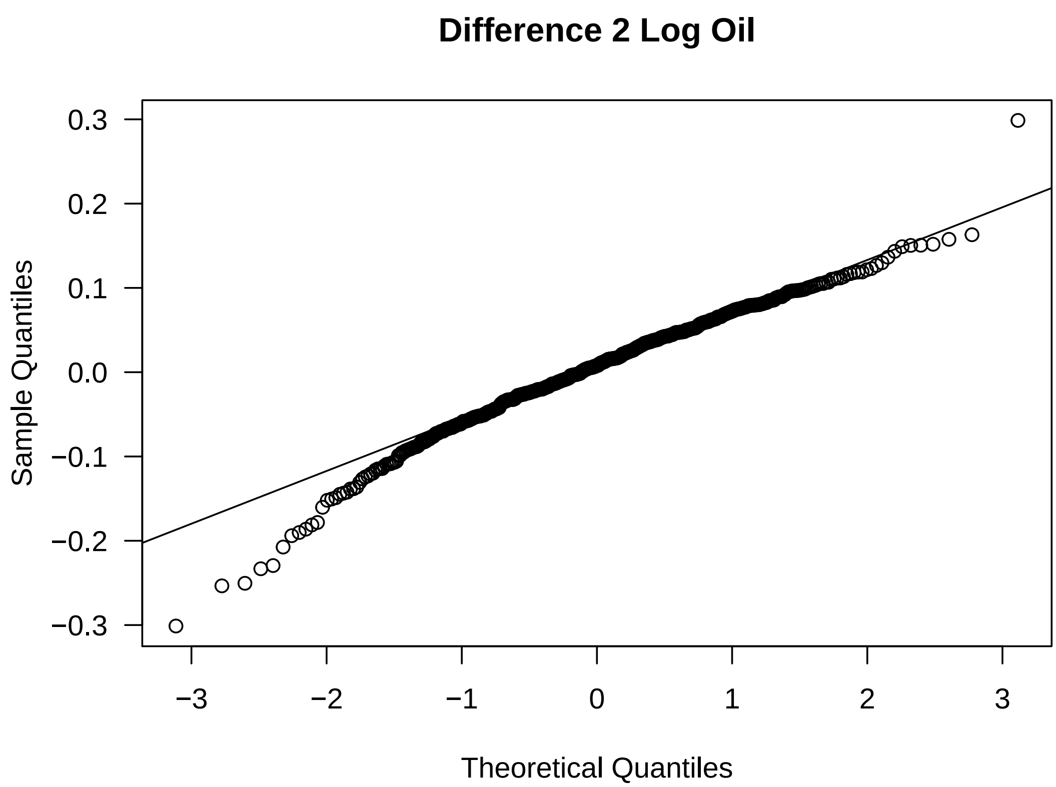
```r
eacf(dloil)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o x o o o o x o o o  o  o  o
## 1 x o x o o o o x o o o  o  o  o
## 2 x x x o o o o x o o o  o  o  o
## 3 x x x o o o o x o o o  o  o  o
## 4 x o x o o o o x o o o  o  o  o
## 5 x x x o x o o x o o o  o  o  o
## 6 o x x o x x o x o o o  o  o  x
## 7 o x x x x x x x o x o  o  o  o
```

```r
eacf(ddloil)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o o o o o x x o o  o  o  o
## 1 x x o o o o o x x o o  o  o  o
## 2 x x x o o o o x x o o  o  o  o
## 3 x x x x o o o x x o o  o  o  o
## 4 x x x x o o o x x o o  o  o  o
## 5 x o x x x o o x x o o  o  o  o
## 6 x o x x x x x x o x o  o  o  o
## 7 x x x x x x x x o x o  o  o  o
```


Difference 1 Log Oil
Sample Quantiles
Theoretical Quantiles
0.2
0.1
0.0
−0.1
−0.2
−3
−2
−1
0
1
2
3


Difference 2 Log Oil
Sample Quantiles
Theoretical Quantiles
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−3
−2
−1
0
1
2
3

```r
fit1 <- Arima(loil, order=c(1, 1, 1))
fit1
```

```
## Series: loil
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1     ma1
##       -0.5253  0.7142
## s.e.   0.0872  0.0683
##
## sigma^2 estimated as 0.002112:  log likelihood=904.58
## AIC=-1803.15   AICc=-1803.11   BIC=-1790.25
```

```r
fit2 <- Arima(loil, order=c(0, 1, 3))
fit2
```

```
## Series: loil
## ARIMA(0,1,3)
##
## Coefficients:
##          ma1      ma2     ma3
##       0.1696  -0.0886  0.1458
## s.e.  0.0424   0.0424  0.0429
##
## sigma^2 estimated as 0.002094:  log likelihood=907.41
## AIC=-1806.83   AICc=-1806.75   BIC=-1789.63
```

```r
fit3 <- Arima(loil, order=c(0, 2, 1))
fit3
```

```
## Series: loil
## ARIMA(0,2,1)
##
## Coefficients:
##           ma1
##       -1.0000
## s.e.   0.0061
##
## sigma^2 estimated as 0.002213:  log likelihood=886.63
## AIC=-1769.26   AICc=-1769.24   BIC=-1760.67
```

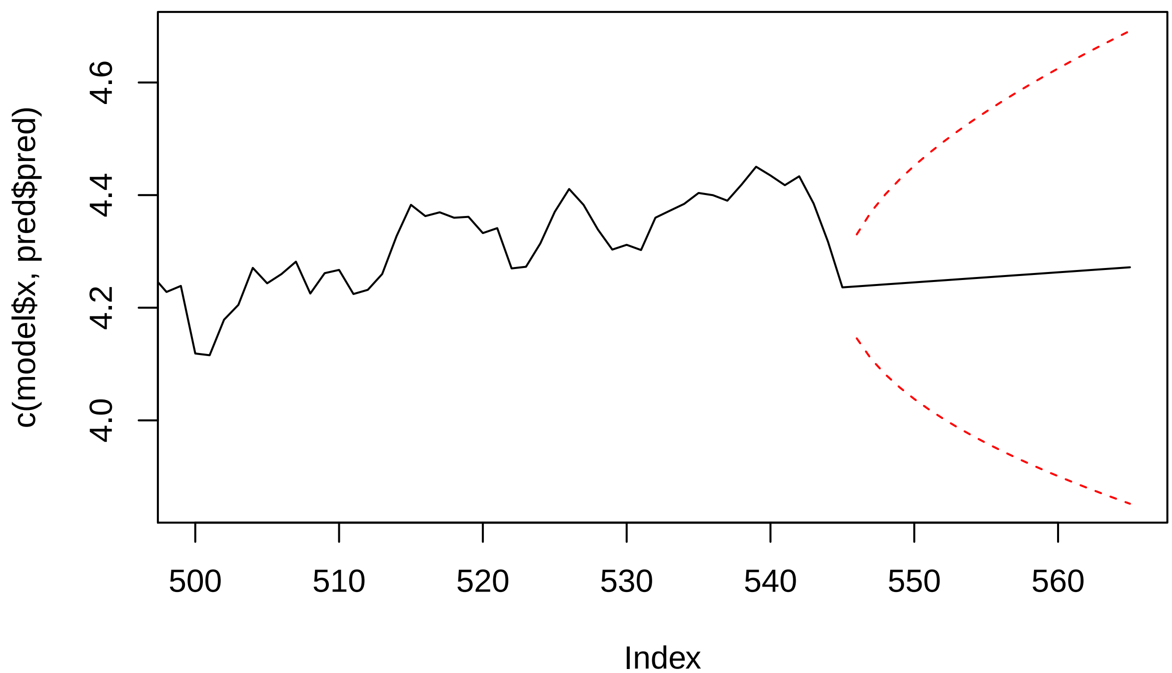
```
complex_dist <- function(x) {
    sqrt(Re(x)^2 + Im(x)^2)
}

sapply(polyroot(c(1, -2, 1)), complex_dist)
```

```
## [1] 1 1
```

```
sapply(polyroot(c(1, -1)), complex_dist)
```

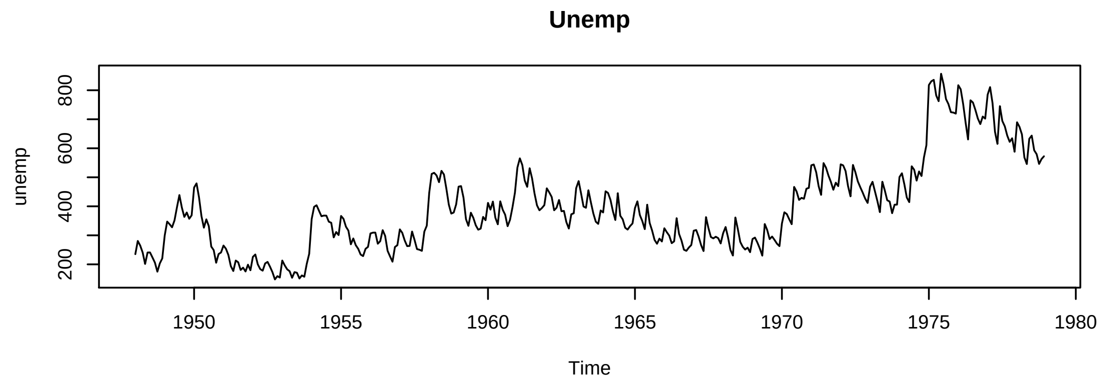
```
## [1] 1
```

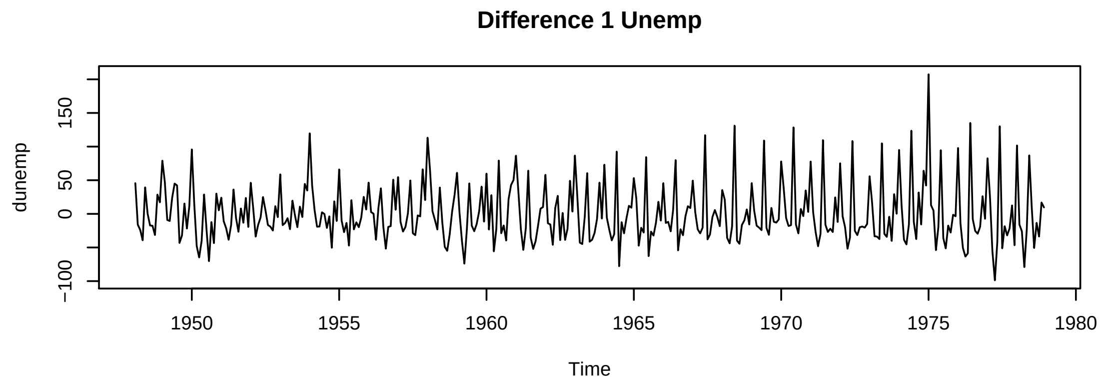
```
fit_plot(fit3)
```

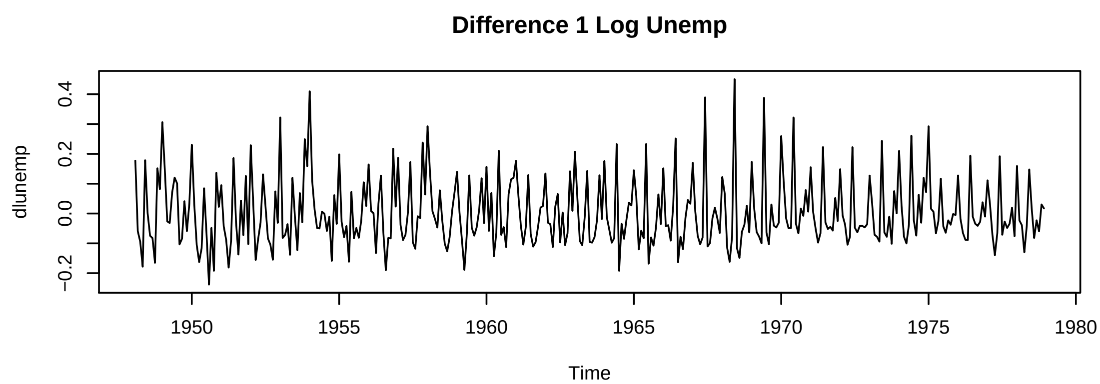


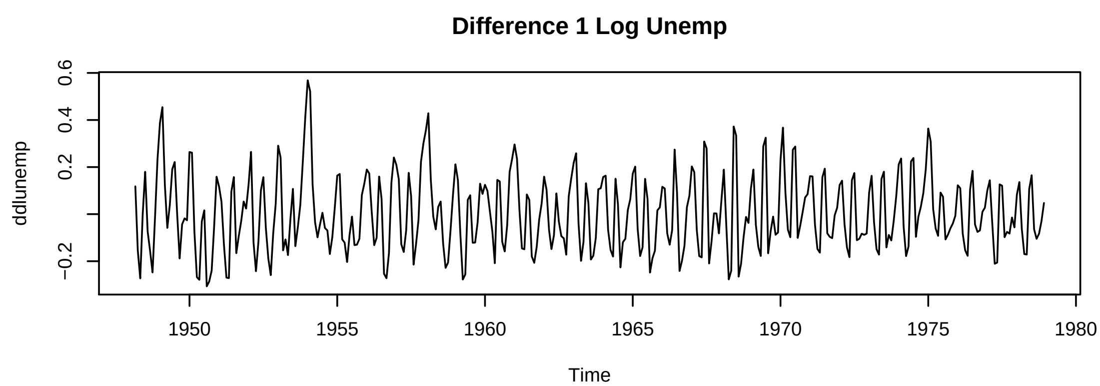
b)

```r
lunemp <- log(unemp)
dunemp <- diff(unemp)
dlunemp <- diff(lunemp, 1)
ddlunemp <- diff(lunemp, 2)
```

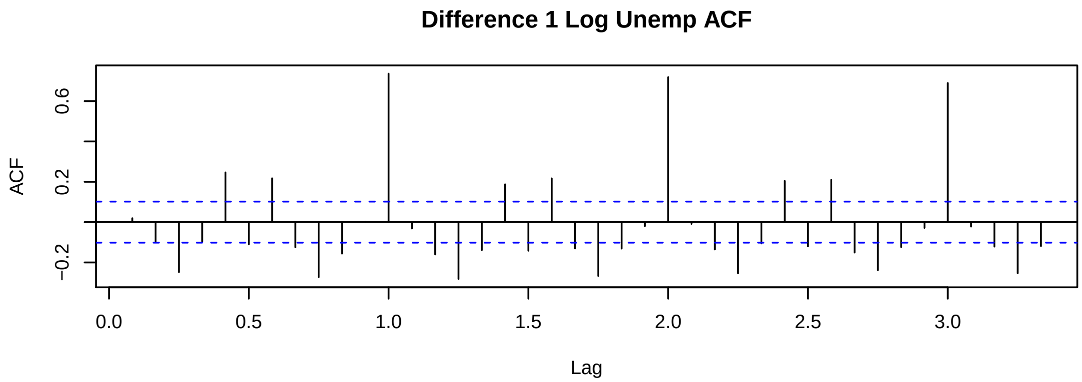

Unemp
unemp
200
400
600
800
1950
1955
1960
1965
1970
1975
1980
Time

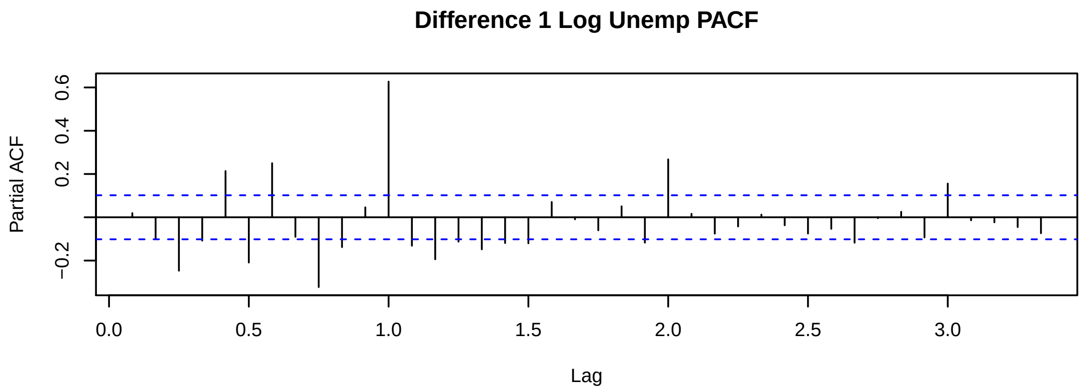

Difference 1 Unemp
dunemp
−100
0
50
150
1950
1955
1960
1965
1970
1975
1980
Time

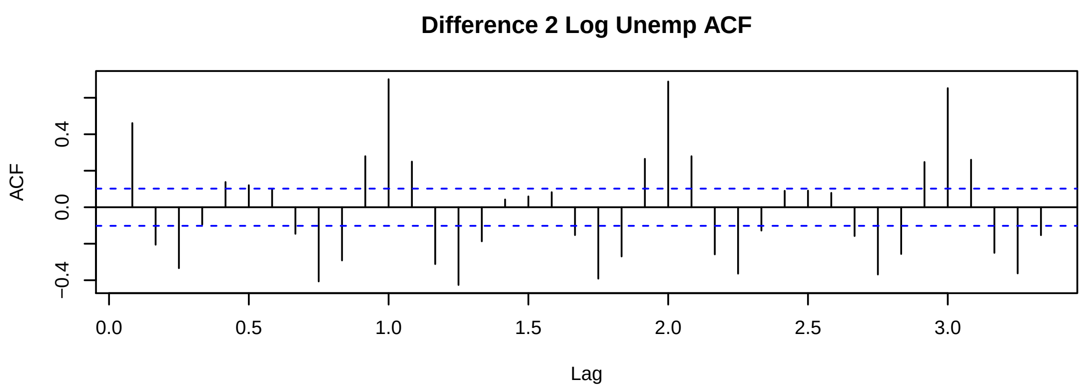

Difference 1 Log Unemp
dlunemp
−0.2
0.0
0.2
0.4
1950
1955
1960
1965
1970
1975
1980
Time

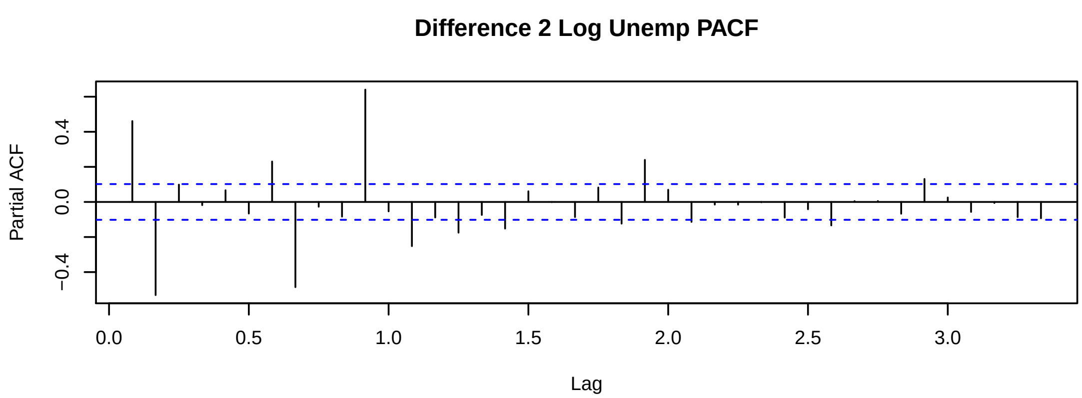

Difference 1 Log Unemp
ddlunemp
−0.2
0.2
0.4
0.6
1950
1955
1960
1965
1970
1975
1980
Time

The data is clearly not stationary and to try to make it stationary we do differencing of order 1. The variance seem to increase by time which we try to negative by transforming the data on log-scale. We also used differencing of order 2 which gives a smoother result than differencing of order 1.


Difference 1 Log Unemp ACF
ACF
0.6
0.2
−0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


Difference 1 Log Unemp PACF
Partial ACF
0.6
0.4
0.2
−0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


Difference 2 Log Unemp ACF
ACF
0.4
0.0
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag


Difference 2 Log Unemp PACF
Partial ACF
0.4
0.0
−0.4
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Lag

Difference 1

Seasonality behavior: The ACF plot suggests a seasonality trend of 12 lags that tails off both in the ACF and the PACF. This suggests an $ARMA_12$ seasonality component. The PACF spikes at 3 multiples which is indicative of AR(3).

There is a seasonality trend that occurs at lags 3, 5, 7, 9, 12, 15, 17, 19, 21, 24.

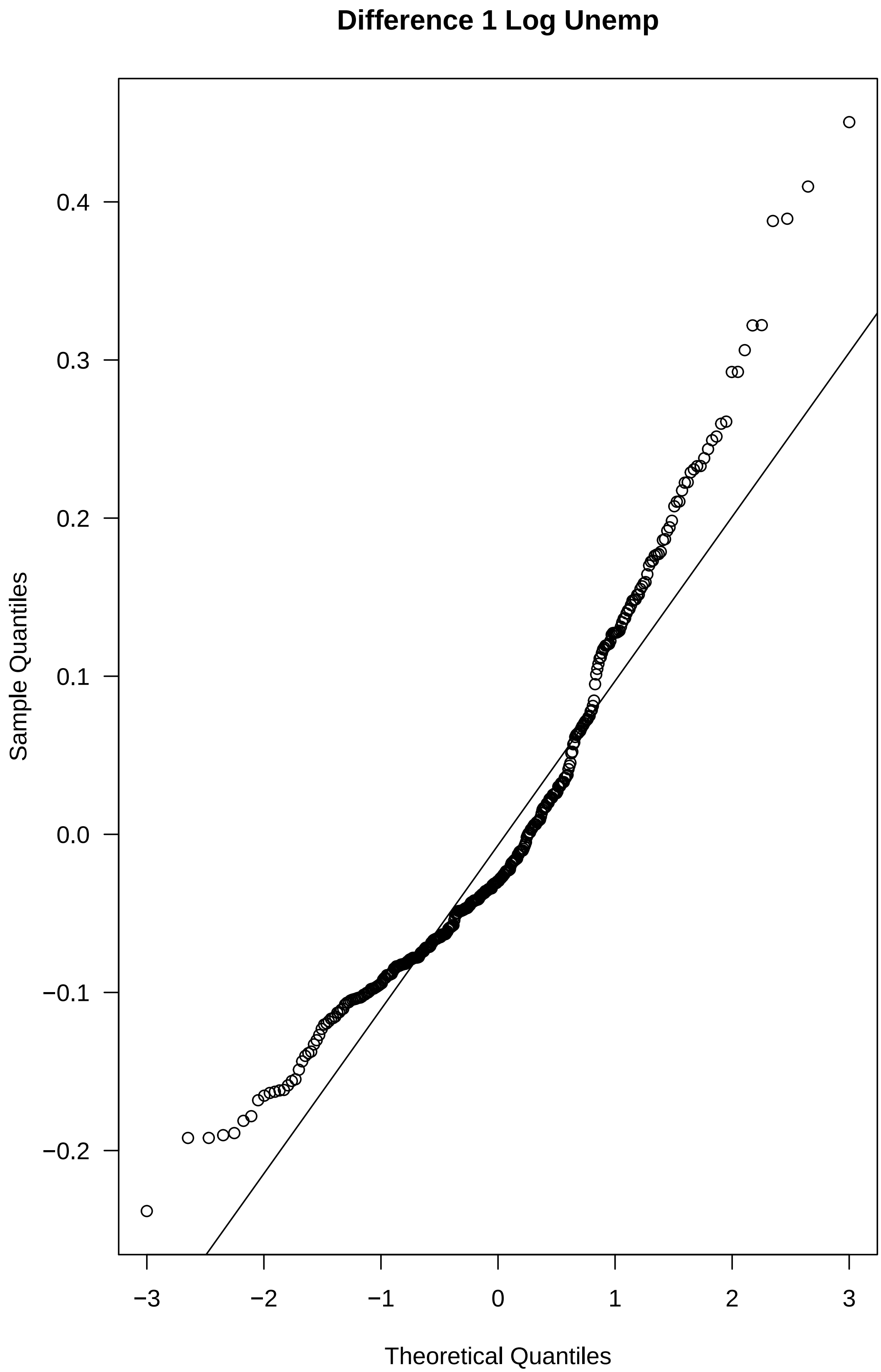
Non-seasonality behavor:

Difference 2

```r
eacf(dlunemp)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 o o x o x x x x x x o  x  o  x
## 1 x o x o x o x x x x o  x  x  o
## 2 x x o x x o o x o o o  x  x  x
## 3 x x x x x o o x o o o  x  x  x
## 4 x x o x o x o o o o o  x  o  x
## 5 x x o x o x o o o o o  x  o  x
## 6 x x x o o x o o o o o  x  o  x
## 7 x x o x x o x o o o o  x  o  o
```

The EACF further suggest that there is no normal ARMA model that is well suited for the data.


Difference 1 Log Unemp
Sample Quantiles
Theoretical Quantiles
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−3
−2
−1
0
1
2
3

```r
fit1 <- Arima(lunemp, order=c(1, 1, 1))
fit1
```

```
## Series: lunemp
## ARIMA(1,1,1)
##
## Coefficients:
##           ar1     ma1
##       -0.7592  0.8157
## s.e.   0.0952  0.0796
##
## sigma^2 estimated as 0.01289:  log likelihood=281.71
## AIC=-557.42   AICc=-557.35   BIC=-545.67
```

```r
fit2 <- Arima(lunemp, order=c(0, 1, 3))
fit2
```

```
## Series: lunemp
## ARIMA(0,1,3)
##
## Coefficients:
##           ma1     ma2      ma3
##       -0.0079  0.0277  -0.3629
## s.e.   0.0470  0.0506   0.0481
##
## sigma^2 estimated as 0.01177:  log likelihood=298.85
## AIC=-589.7   AICc=-589.59   BIC=-574.03
```

```r
fit3 <- Arima(lunemp, order=c(0, 2, 1))
fit3
```

```
## Series: lunemp
## ARIMA(0,2,1)
##
## Coefficients:
##          ma1
##       -1.000
## s.e.   0.007
##
## sigma^2 estimated as 0.01298:  log likelihood=276.19
## AIC=-548.39   AICc=-548.36   BIC=-540.56
```