

# 732A62 Lab 2

*Emil K Svensson & Rasmus Holm*

*2017-10-03*

## Assignment 1

a)

```
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

$$\phi_2$$

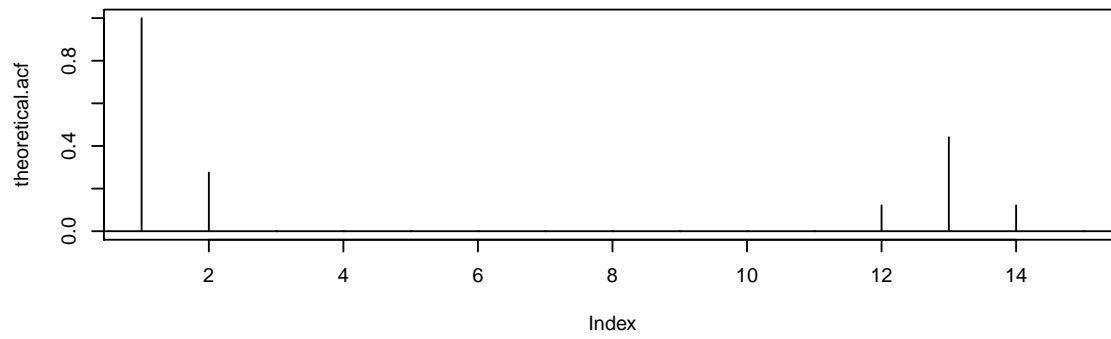
is inside the confidence-intervall for the ML estimate which can seen from the s.e times 1.96 which obviously will cover the true coefficients.

c)

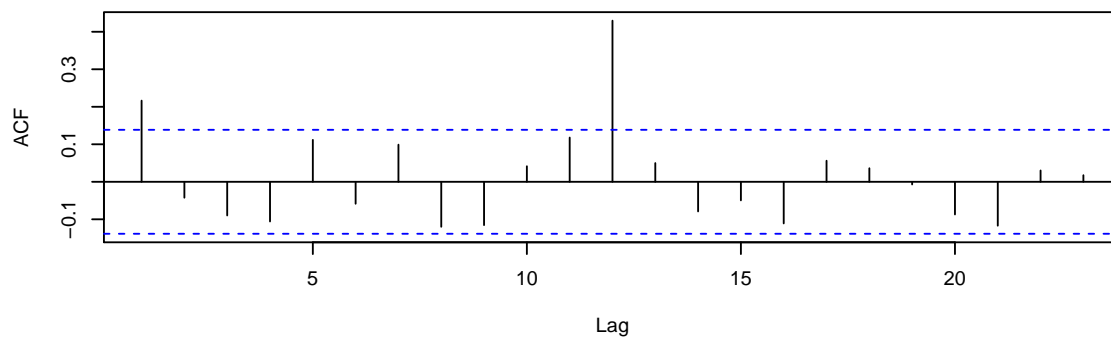
```
set.seed(12345)
ma.coef <- c(0.3, rep(0, 10), 0.6)
ts4 <- arima.sim(n=200, model=list(order=c(0, 0, 12), ma = ma.coef))

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

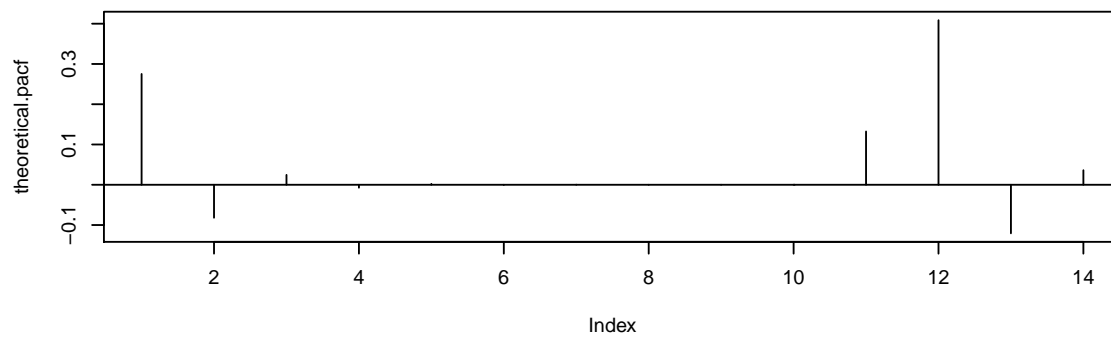
**Theoretical ACF**



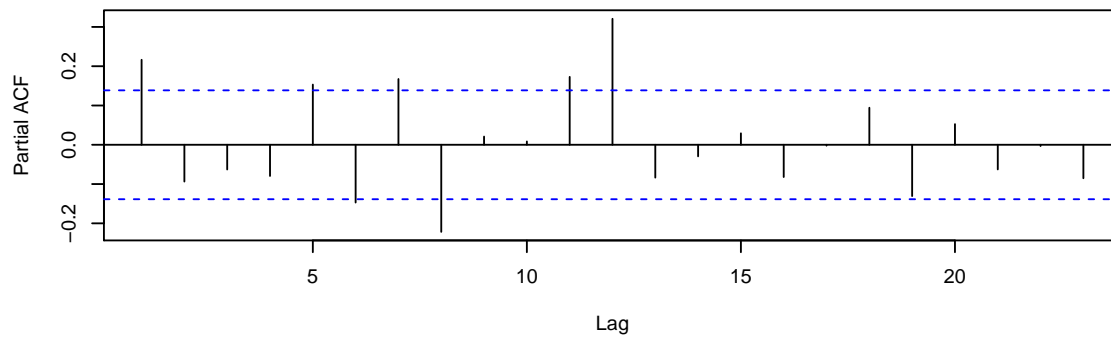
**Empirical ACF**



**Theoretical PACF**



**Empirical PACF**



The patterns seem to be somewhat similar. In the theoretical ACF we can see a large spike at lags 1 and 13 and in the sample ACF we have large spikes at lags 1 and 12 instead. The difference between the two being that the sample ACF has some correlation along the lags although under the confidence interval.

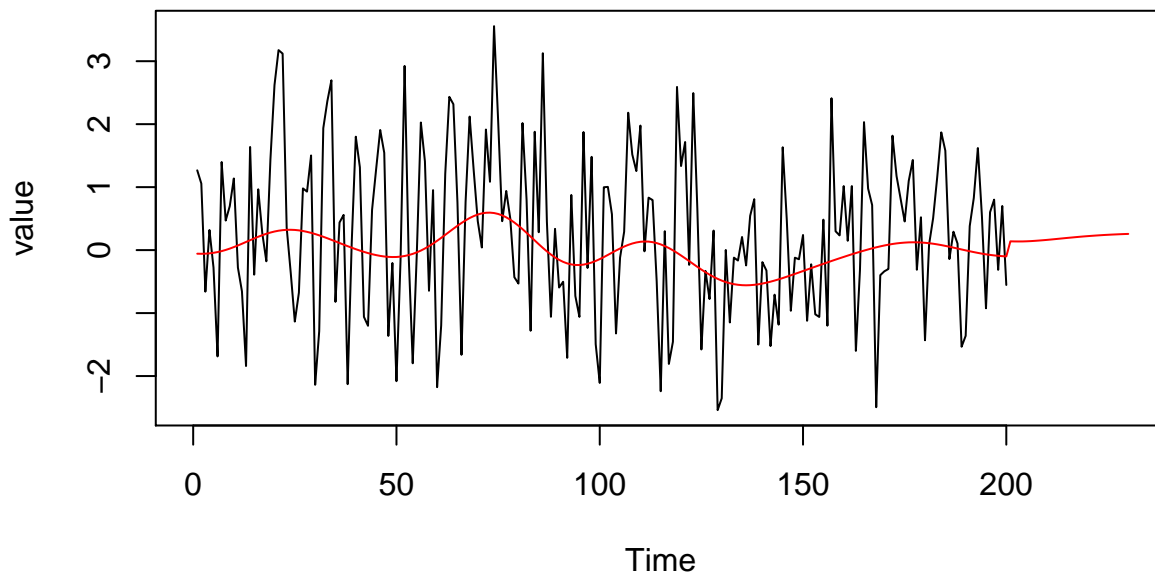
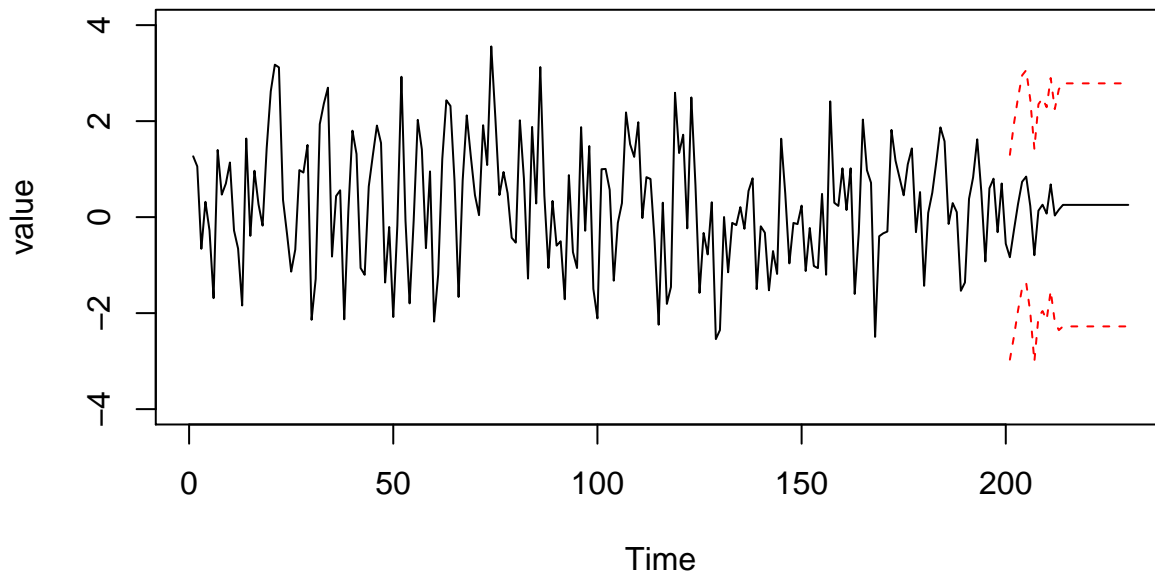
For the PACF we can see a similar observation that there are spikes at lags 1 and 12 for both the theoretical and sample PACF.

d)

```
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)

gausspr.data <- data.frame(y=ts5, x=1:200)
gausspr.fit <- gausspr(y ~ x, gausspr.data)
gausspr.pred <- predict(gausspr.fit, data.frame(x=201:230))
```



In the first plot we can see the MA models predictions seem reasonable 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)

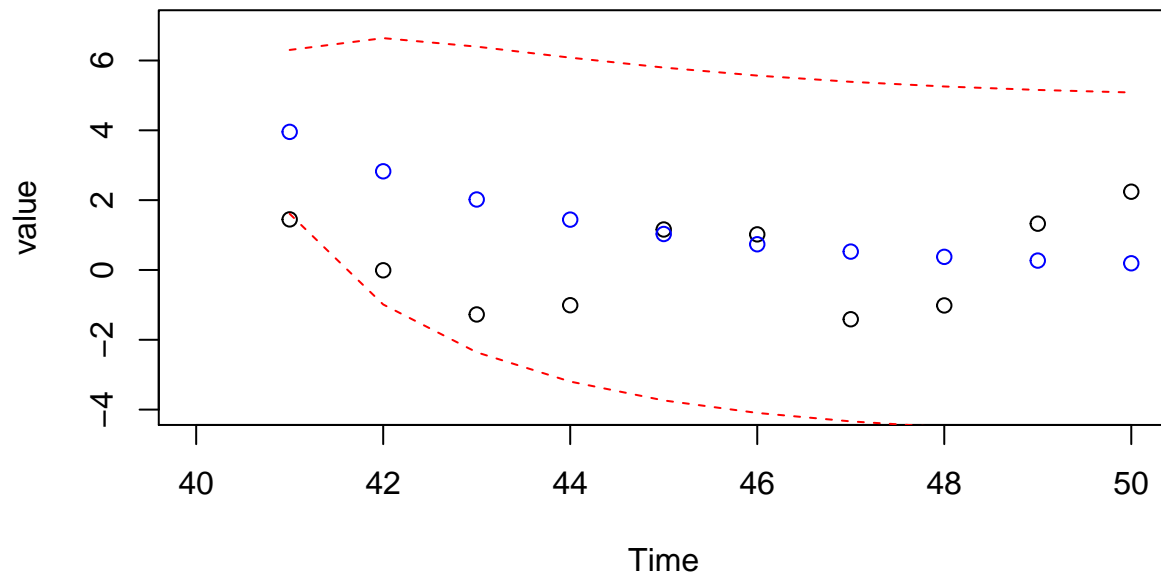
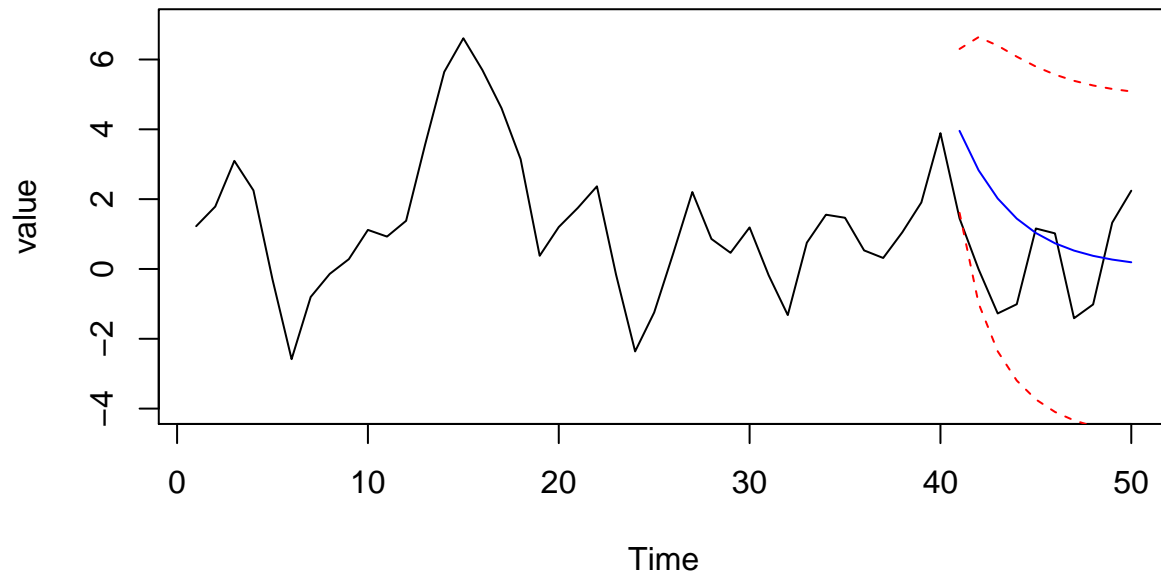
```
set.seed(12345)
ts6 <- arima.sim(model=list(ma=c(0.5), ar=c(0.7)), n=50)
```

```

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)

```



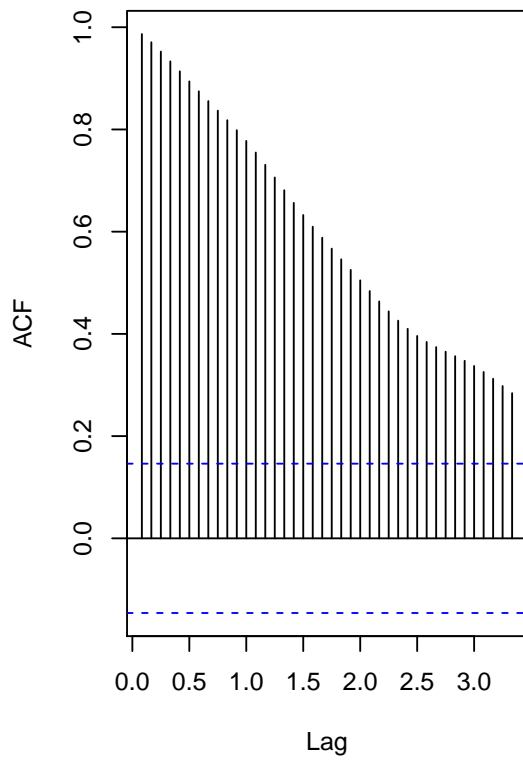
The first of the 10 observations withheld from the fitted model is the only observation that is not in the prediction interval. Since it is a 95 % prediction interval we say that we expect that 5 of 100 observations will be outside the interval so it is not unreasonable that 1 of 10 is outside the interval.

## Assignment 2

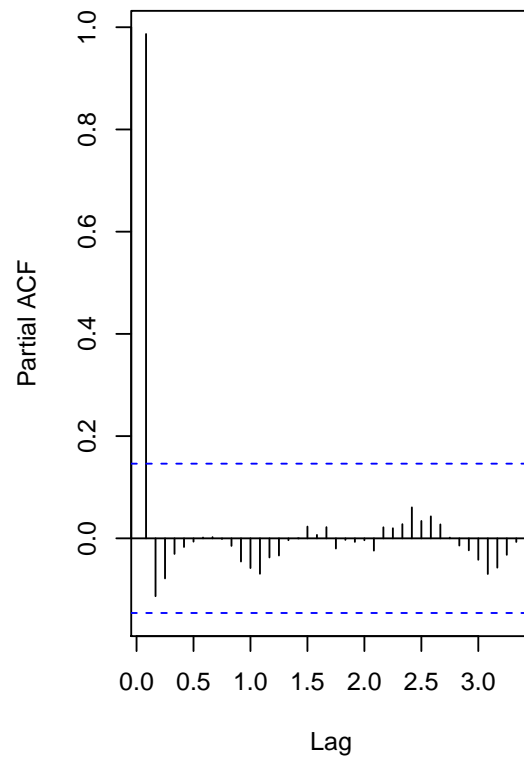
```
assignment2 <- function(data){  
  old <- par(mfrow = c(2, 2))  
  acf(data, lag.max = 40, main="Data ACF")  
  pacf(data, lag.max = 40, main="Data PACF")  
  acf(diff(data, lag = 1), lag.max = 40, main="Difference 1 Data ACF")  
  pacf(diff(data, lag = 1), lag.max = 40, main="Difference 1 Data PACF")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

## Chicken

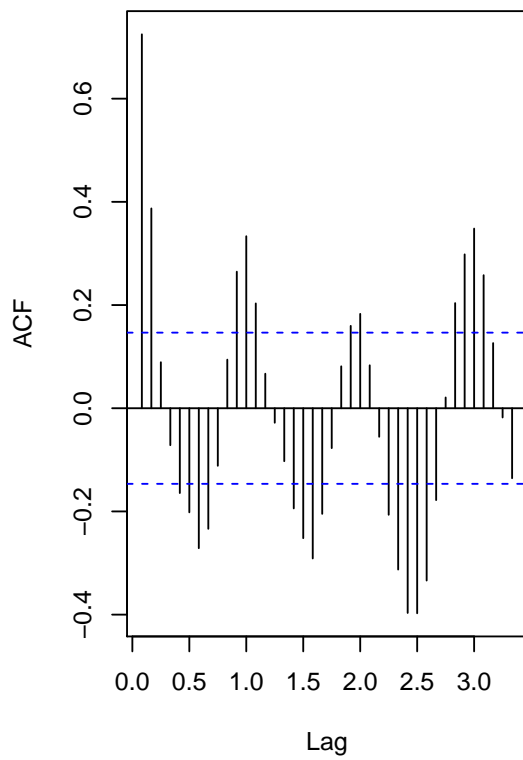
**Data ACF**



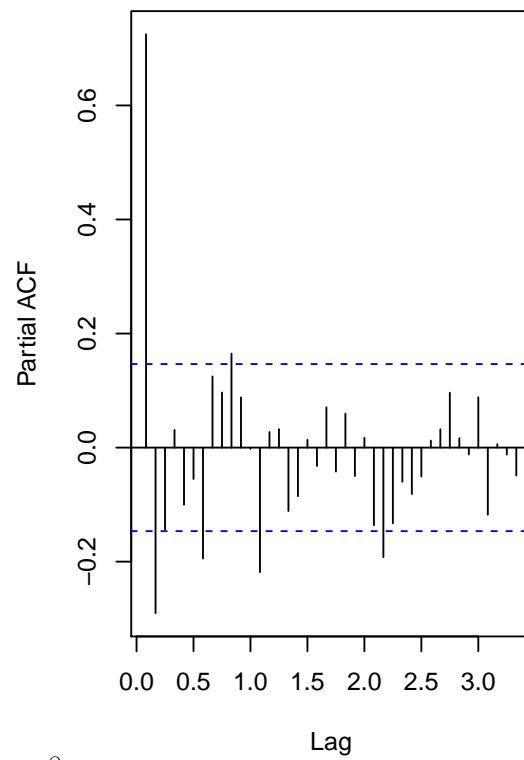
**Data PACF**



**Difference 1 Data ACF**



**Difference 1 Data PACF**





### **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 that tails off for a while then goes up again. This may indicate an AR model.

### **Difference 1 Data PACF**

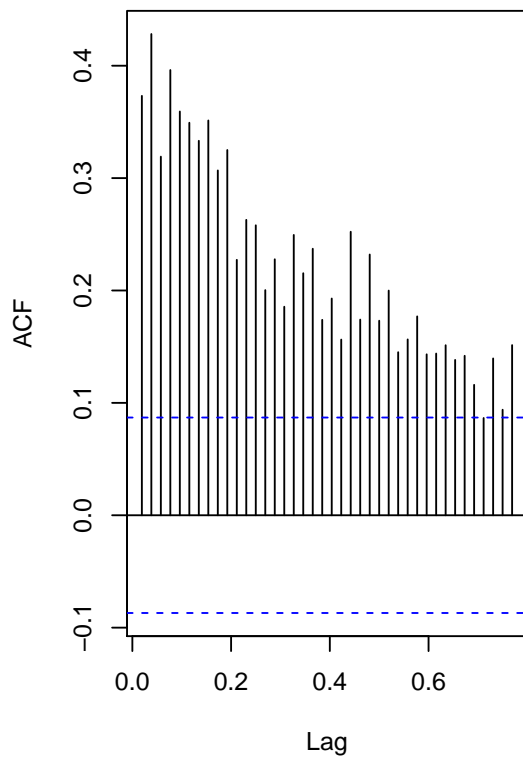
The PACF shows that the seasonality cuts off after lag 12 which is  $1 * 12$  indicating an AR(1) model.

### **Final Verdict**

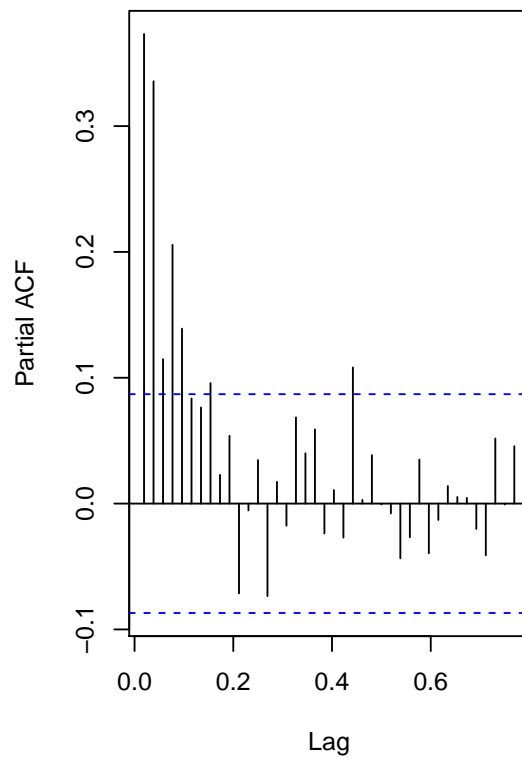
Needs further investigation  $ARIMA(1, 0, 0) \times (1, 1, 0)_{12}$

so2

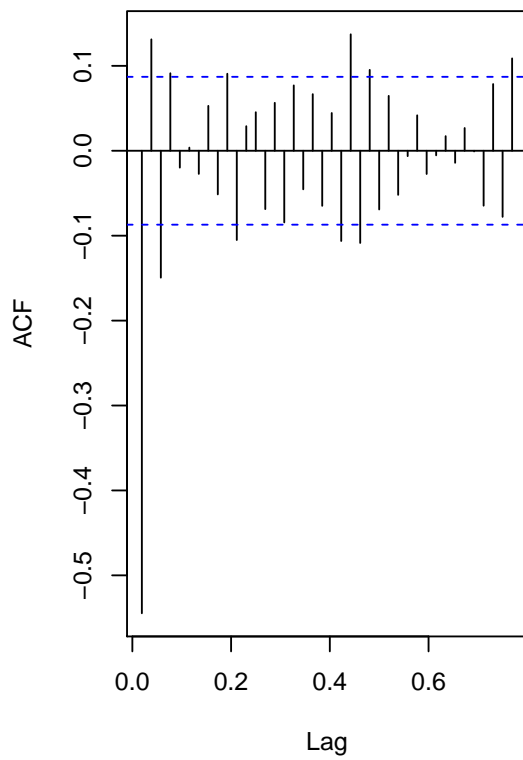
**Data ACF**



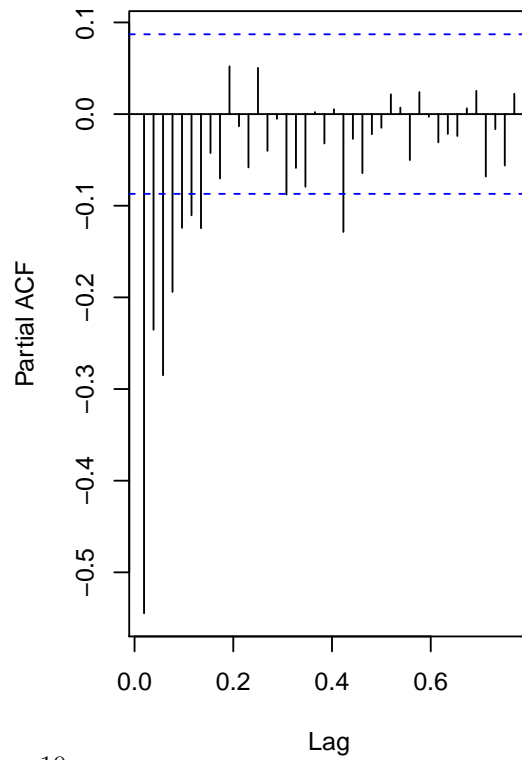
**Data PACF**



**Difference 1 Data ACF**



**Difference 1 Data PACF**



**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**

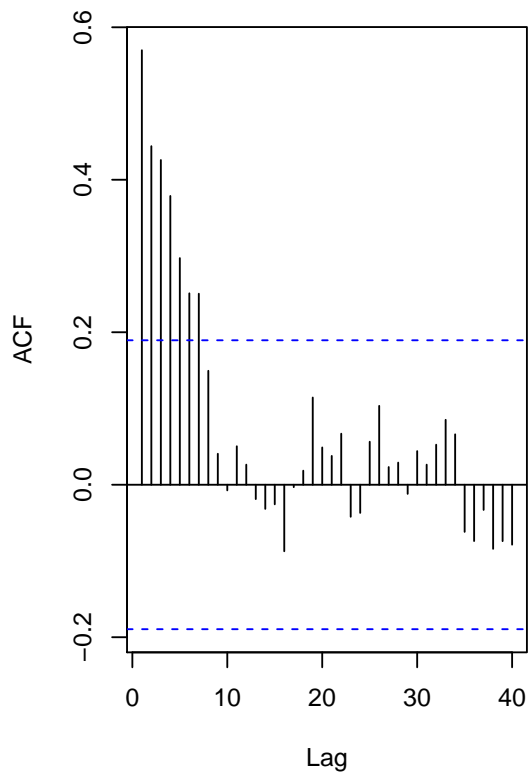
The PACF after difference tails off further suggesting a MA(1) model.

**Final Verdict**

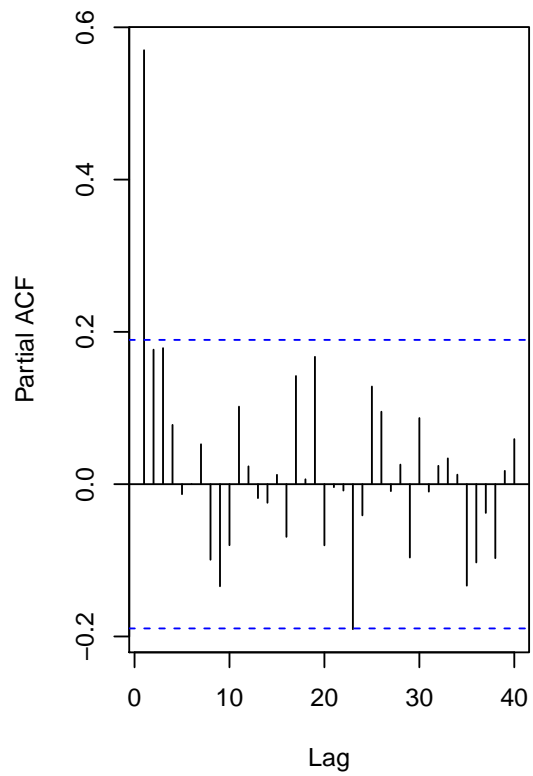
ARIMA(0, 1, 1)

EQcount

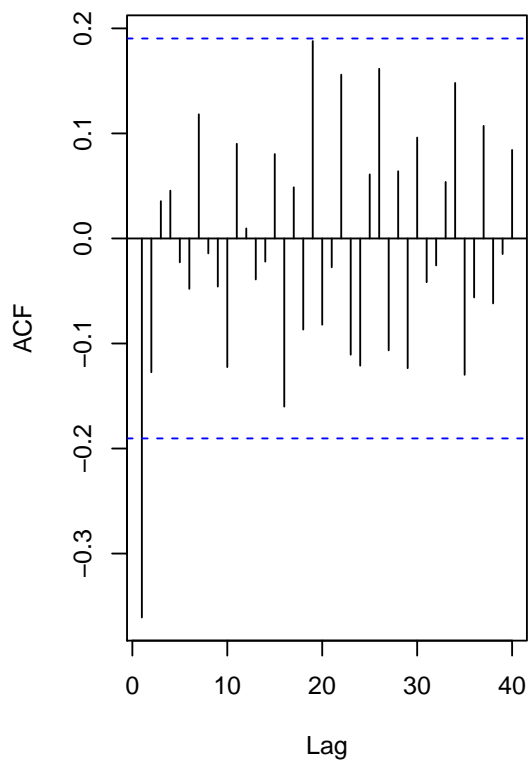
**Data ACF**



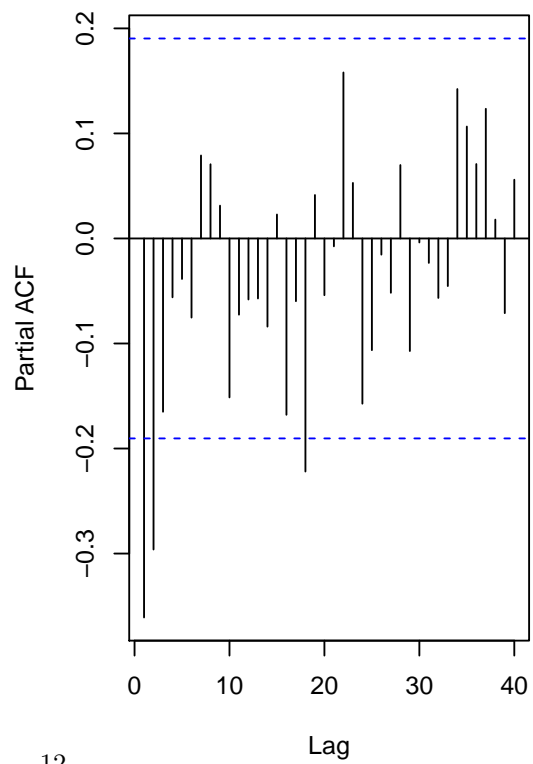
**Data PACF**



**Difference 1 Data ACF**



**Difference 1 Data PACF**



**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**

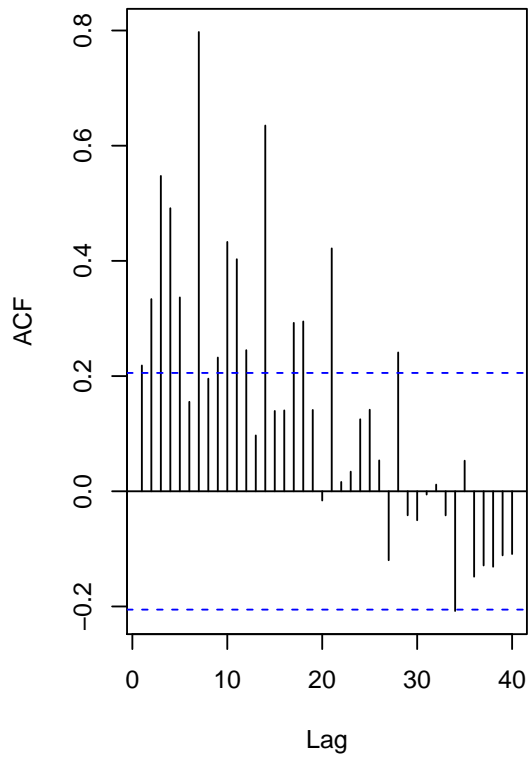
The PACF after difference tails off further suggesting a MA model.

**Final Verdict**

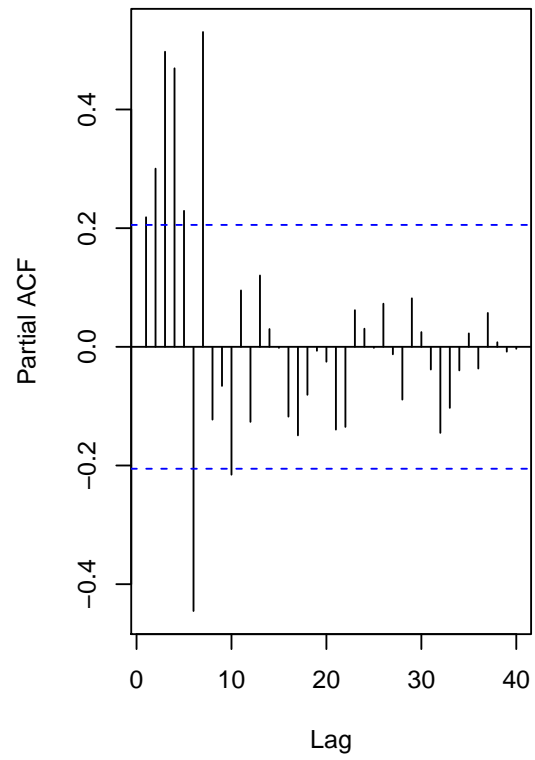
Either a ARMA(1, 0, 0) or ARMA(0, 1, 1)

HCT

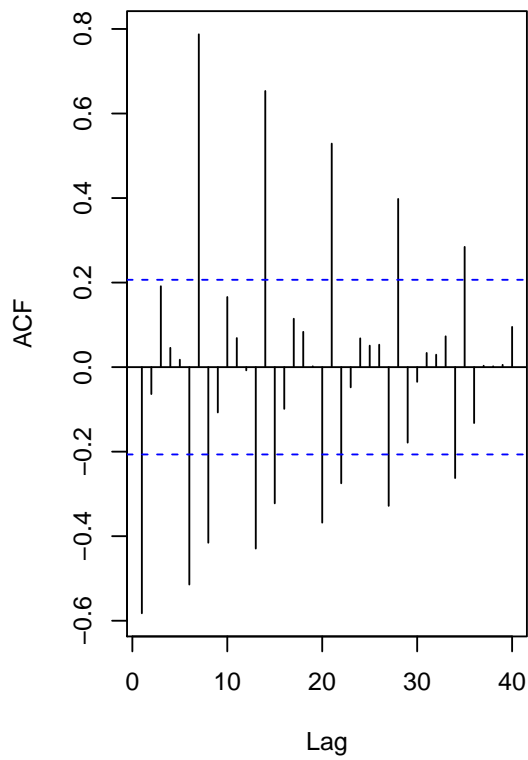
**Data ACF**



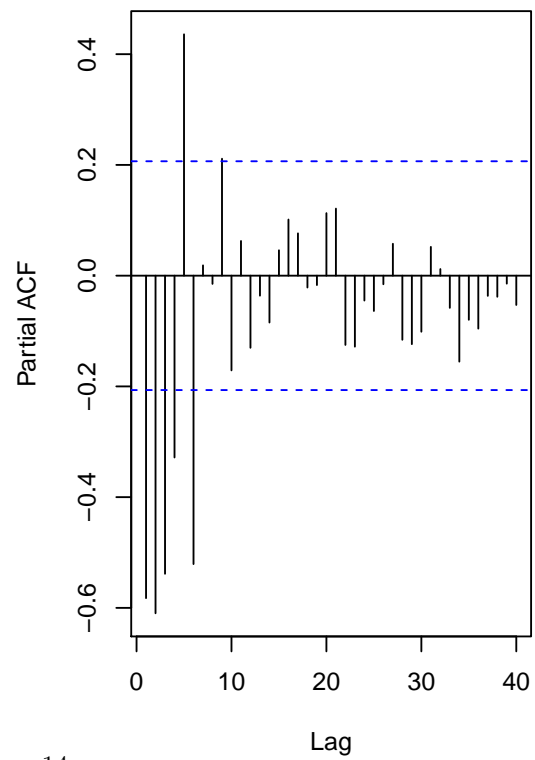
**Data PACF**



**Difference 1 Data ACF**



**Difference 1 Data PACF**



**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)  $\times$  (1, 1, 0)<sub>7</sub>

## 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

  n <- length(model$x)

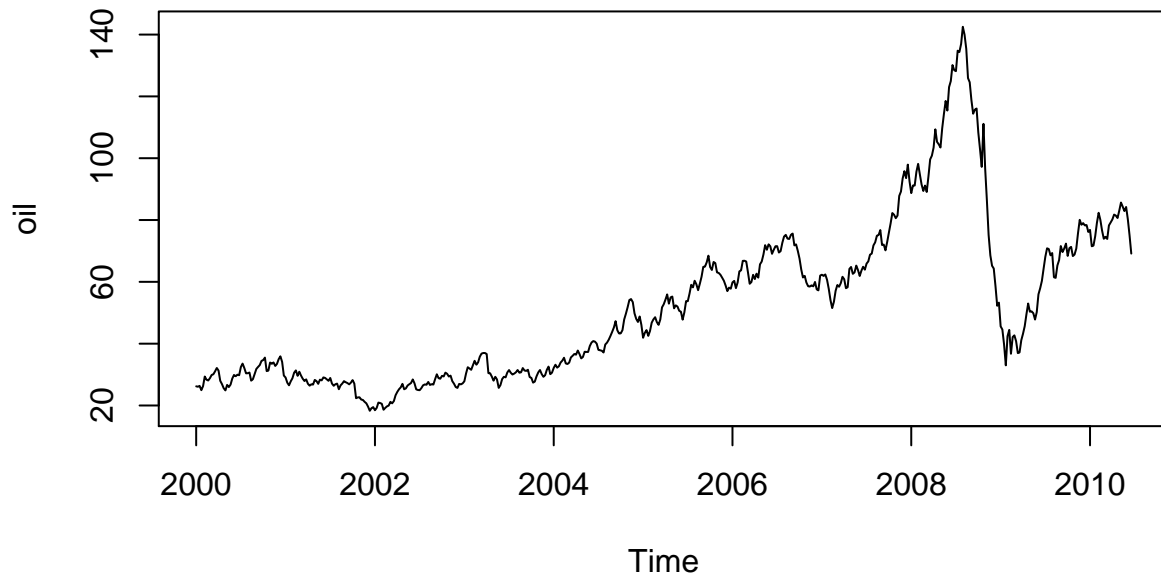
  plot(c(model$x, pred$pred), type="l",
       xlim=c(n - 20, n + 20),
       ylim=c(min(lower_band), max(upper_band)), ylab="Value", xlab="Time")
  lines(n + 1:20, upper_band, lty=2, col="red")
  lines(n + 1:20, lower_band, lty=2, col="red")
}
```

a)

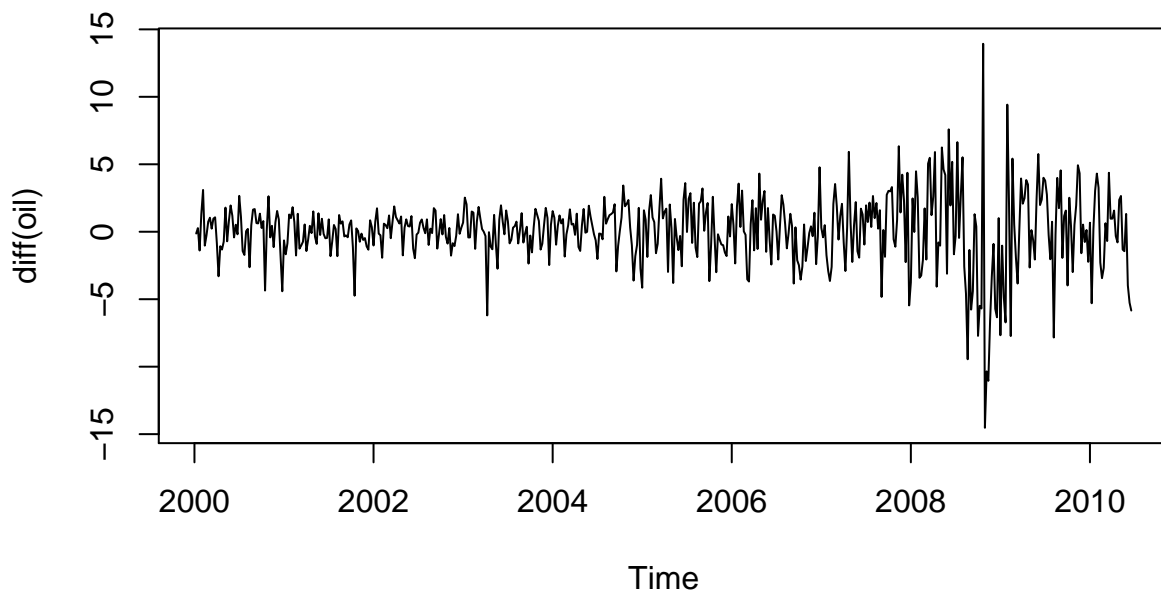
```
loil <- log(oil)
doil <- diff(oil)
ddoil <- diff(oil, 2)
dloil <- diff(loil)
ddloil <- diff(loil, 2)
```



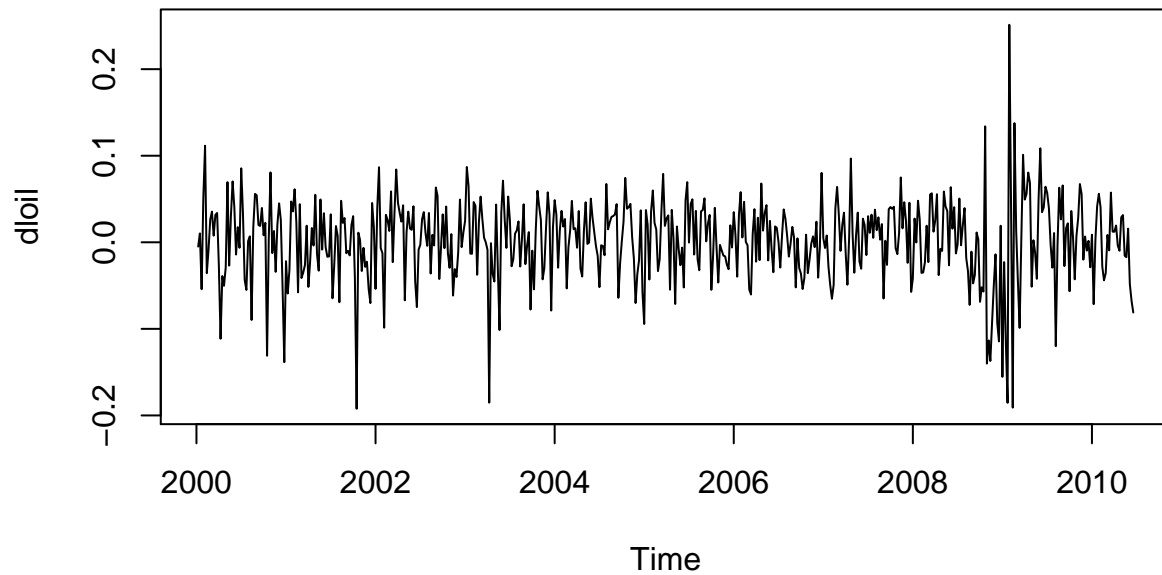
## Oil



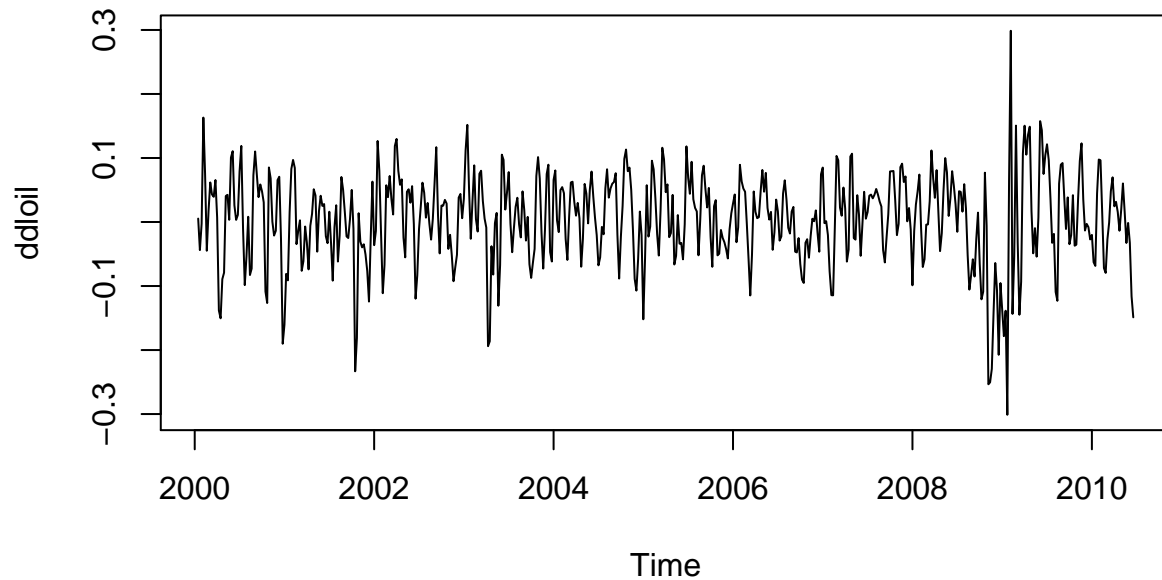
## Difference 1 Oil



### Difference 1 Log Oil



### Difference 2 Log Oil



Clearly difference log is the data we should work with, a second difference does not do any visual change to the data that can already be considered stationary at diff 1 with a log transformation but we keep both and keep both as tentative models.

```
adf.test(oil)
```

```
##  
## Augmented Dickey-Fuller Test  
##
```

```
## data: oil
## Dickey-Fuller = -3.4217, Lag order = 8, p-value = 0.04983
## alternative hypothesis: stationary
```

```
adf.test(loil)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: loil
## Dickey-Fuller = -2.775, Lag order = 8, p-value = 0.2503
## alternative hypothesis: stationary
```

```
adf.test(dloil)
```

```
## Warning in adf.test(dloil): p-value smaller than printed p-value
```

```
##
## Augmented Dickey-Fuller Test
##
## data: dloil
## Dickey-Fuller = -6.3708, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

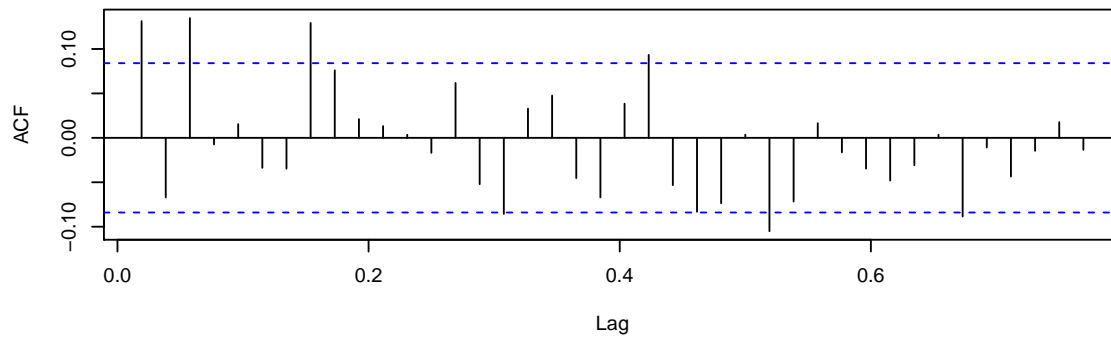
```
adf.test(ddloil)
```

```
## Warning in adf.test(ddloil): p-value smaller than printed p-value
```

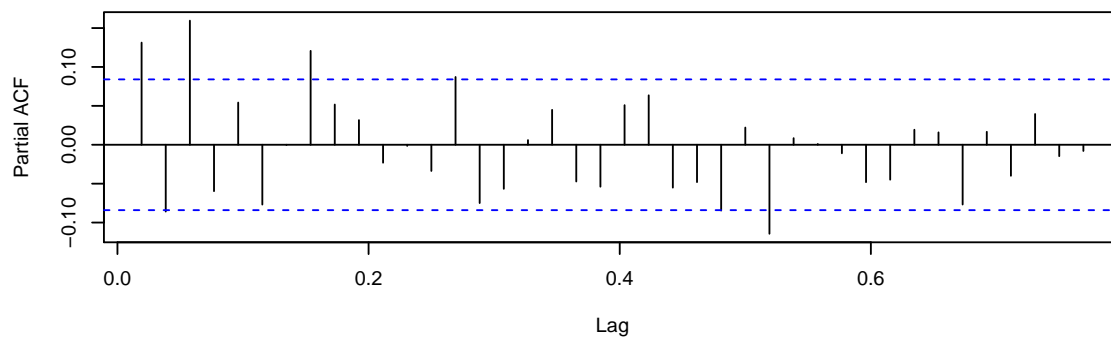
```
##
## Augmented Dickey-Fuller Test
##
## data: ddloil
## Dickey-Fuller = -5.6251, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

The Augmented Dickey-Fuller test also indicates that we should use differencing to make the data stationary.

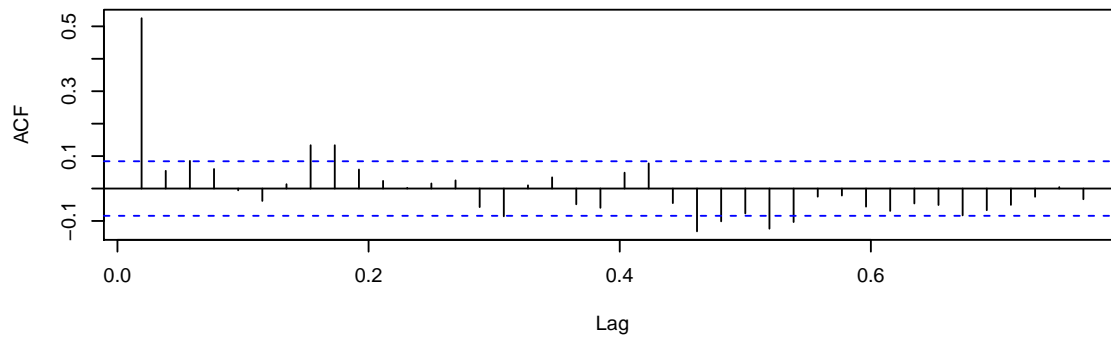
**Difference 1 Log Oil ACF**



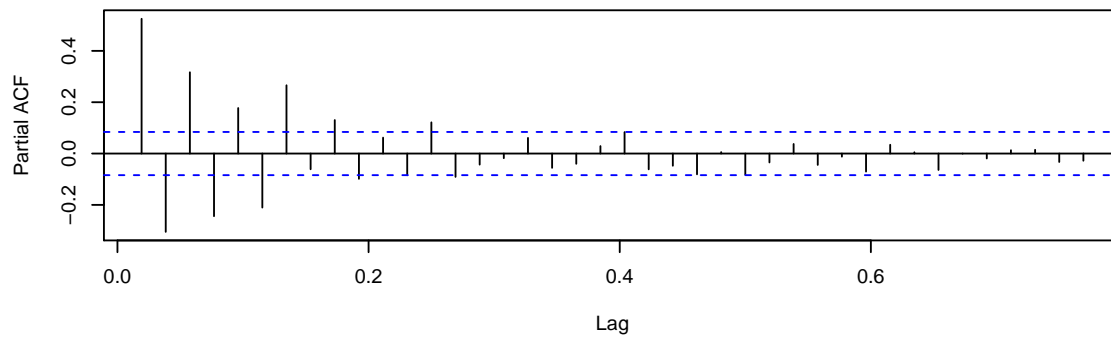
**Difference 1 Log Oil PACF**



**Difference 2 Log Oil ACF**



**Difference 2 Log Oil PACF**



Here the ACF and PACF for the first order differencing doesn't seem to have any clear defined patterns compared to the second order differencing where we can see that the ACF cuts off at lag 1 and trails off in the PACF suggesting a  $ARIMA(0,2,1)$ .

```
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
```

```
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
```

The EACF doesn't seem to help alleviate the ambiguous ACF/PACF for the first order diff series and suggest a wide range of models making the general pattern more like a box than a triangular shape.

For the second order diff we can see the triangular shape with a point at a  $AR(1)$  which would confirm previous statements in the ACF.

```
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
```

```
fit2 <- Arima(loil, order=c(1, 2, 2))
fit2
```

```
## Series: loil
## ARIMA(1,2,2)
##
## Coefficients:
##          ar1      ma1      ma2
##      -0.5247 -0.2861 -0.7139
## s.e.   0.0873  0.0686  0.0685
##
## sigma^2 estimated as 0.002118:  log likelihood=899.7
## AIC=-1791.4  AICc=-1791.32  BIC=-1774.21
```

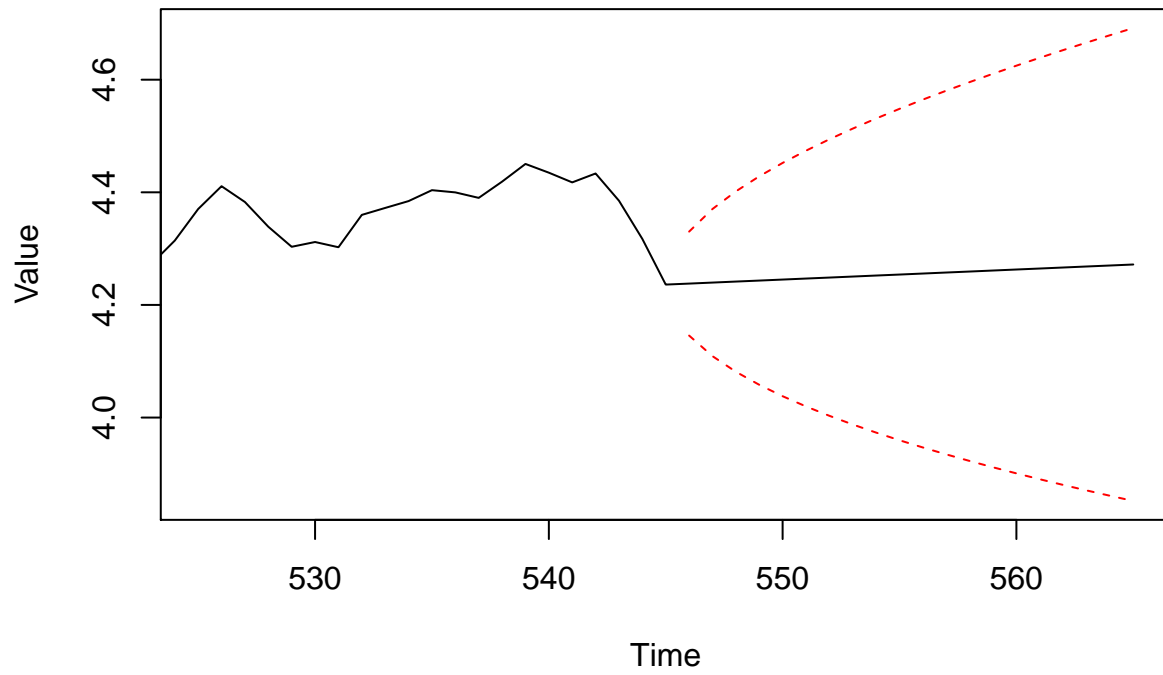
```
fit3 <- Arima(loil, order=c(0, 2, 1))
fit3
```

```
## 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
```

The best model if we evaluate the BIC criterion we choose the ARIMA(0,2,1) which looks like following

$$\nabla^2 x_t = (1 + \theta B)w_t$$

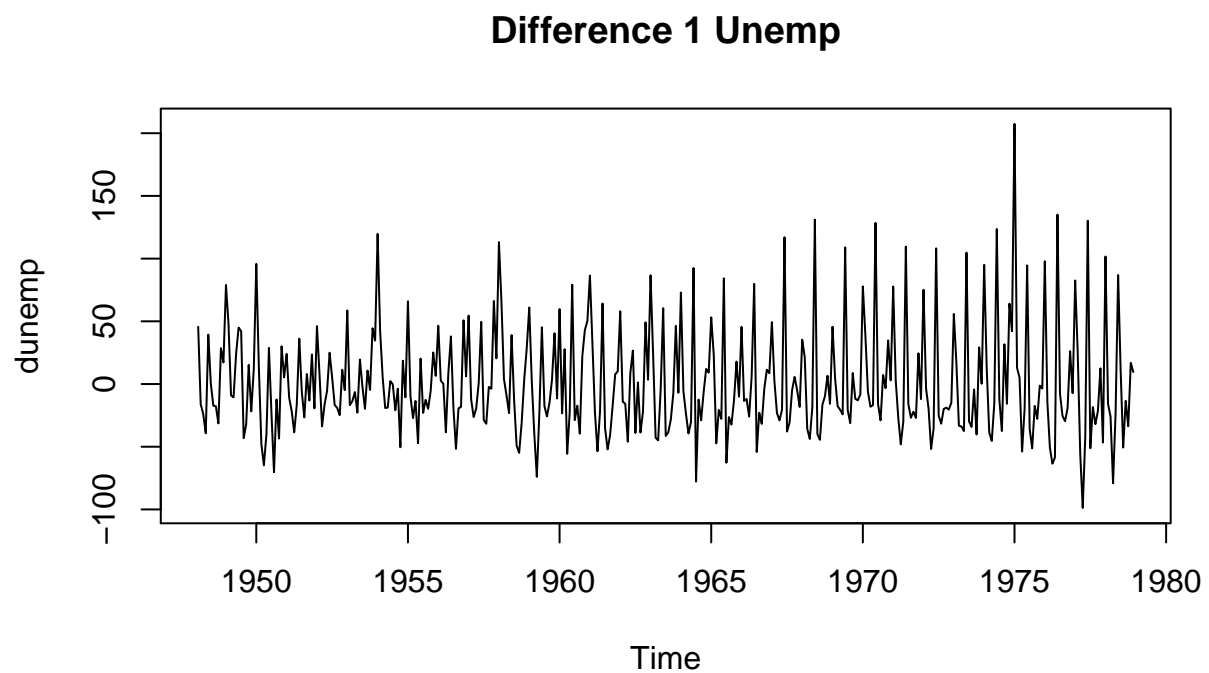
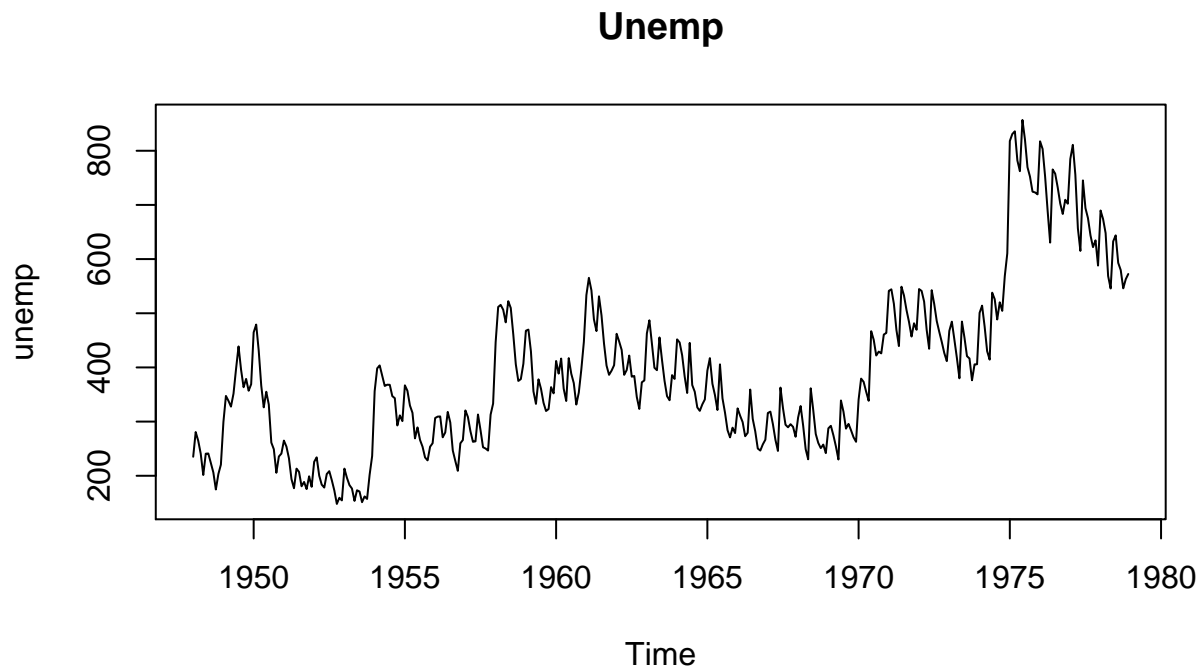
```
fit_plot(fit3)
```



The figure above displays 20 steps ahead prediction. The intervals looks resonable and the prediction as well.

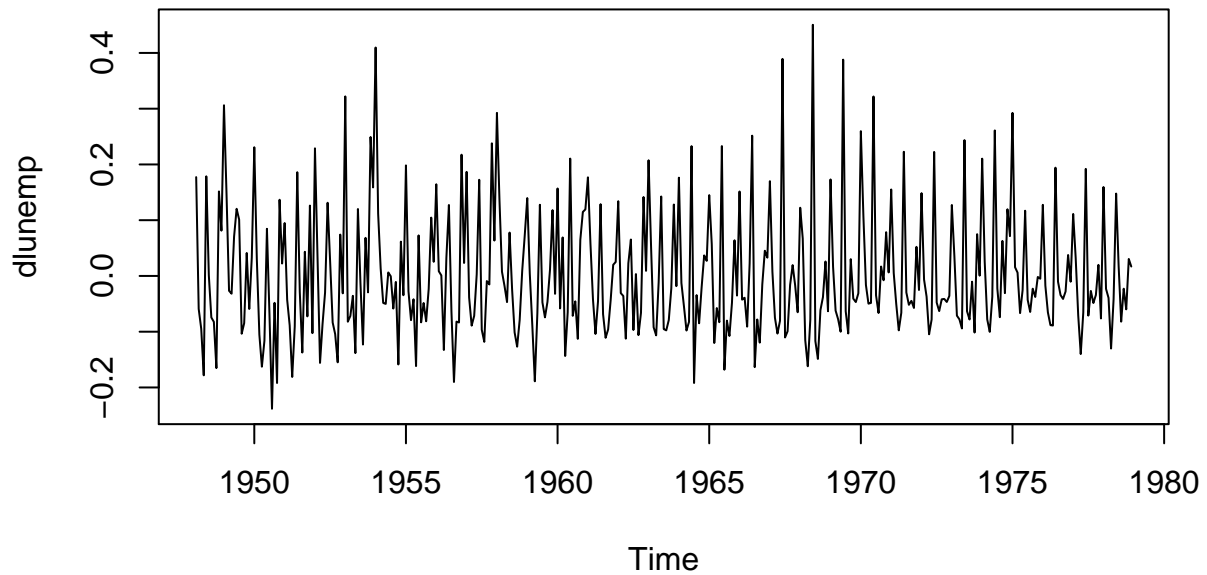
b)

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

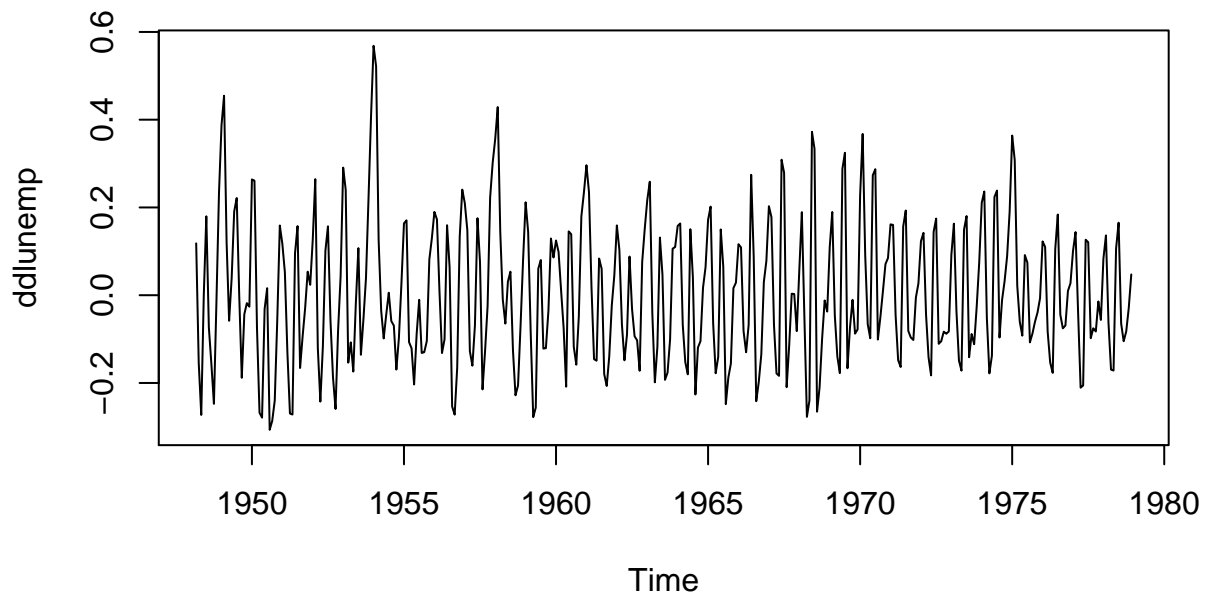




### Difference 1 Log Unemp

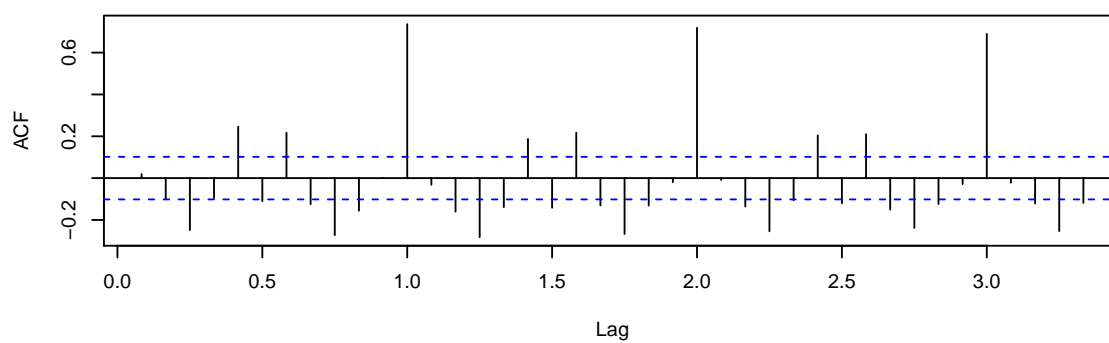


### Difference 1 Log Unemp

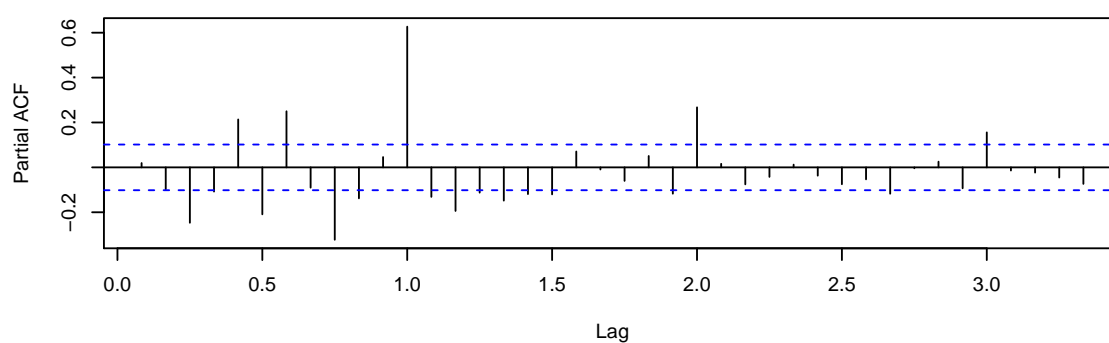


The data is clearly not stationary and to try to make it stationary we do differencing of order 1. The variance seem to increase by time which we try to negative by transforming the data on log-scale. We also used differencing of order 2 which gives a smoother result than differencing of order 1.

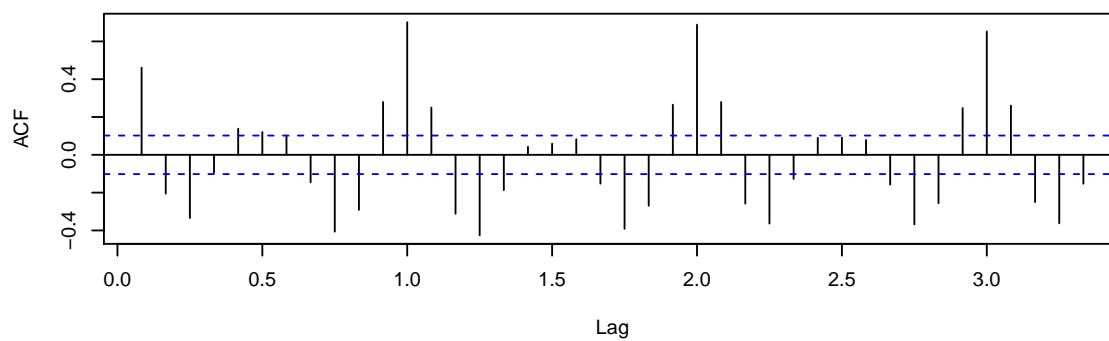
**Difference 1 Log Unemp ACF**



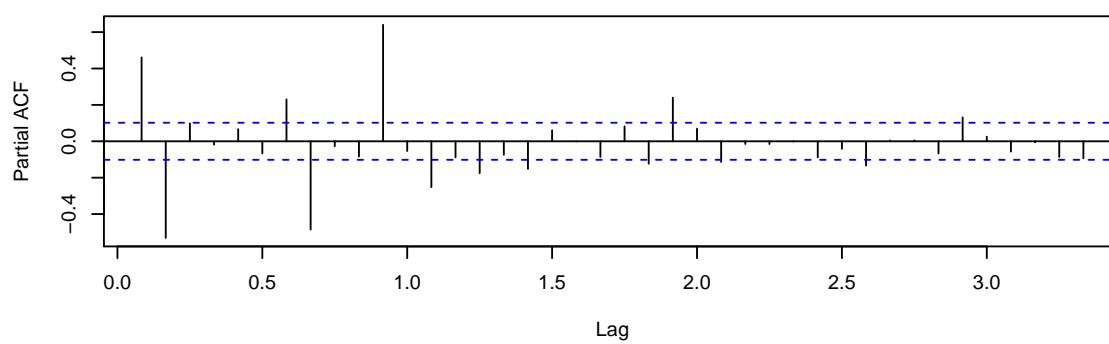
**Difference 1 Log Unemp PACF**



**Difference 2 Log Unemp ACF**



**Difference 2 Log Unemp PACF**



## 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).

**Non-seasonality behavior:** There is no distinct non-seasonal pattern to be seen.

Our model for this data is  $SARMA(3, 1, 0)_{12}$ .

## Difference 2

**Seasonality behavior:** We can see that there are large spikes at lags 9, 12, 15 and then it gets insignificant until the pattern repeats at lags 21, 24, 27. So there is definitely a seasonality pattern in the data. The PACF shows that the pattern tails off over time with spikes at two 2 multiples which indicates AR(2).

**Non-seasonality behavior:** Apart from the seasonal behavior the ACF tails off and the PACF cuts off after lag 2 indicating an AR(2) model.

```
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
```

```
eacf(ddlunemp)
```

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

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

```
fit1 <- Arima(lunemp, order=c(0, 0, 0), seasonal=c(3, 1, 0))
fit1
```

```
## Series: lunemp
## ARIMA(0,0,0)(3,1,0)[12]
##
## Coefficients:
##          sar1      sar2      sar3
##       -0.2589  -0.2801  -0.2872
## s.e.    0.0522   0.0526   0.0547
##
```

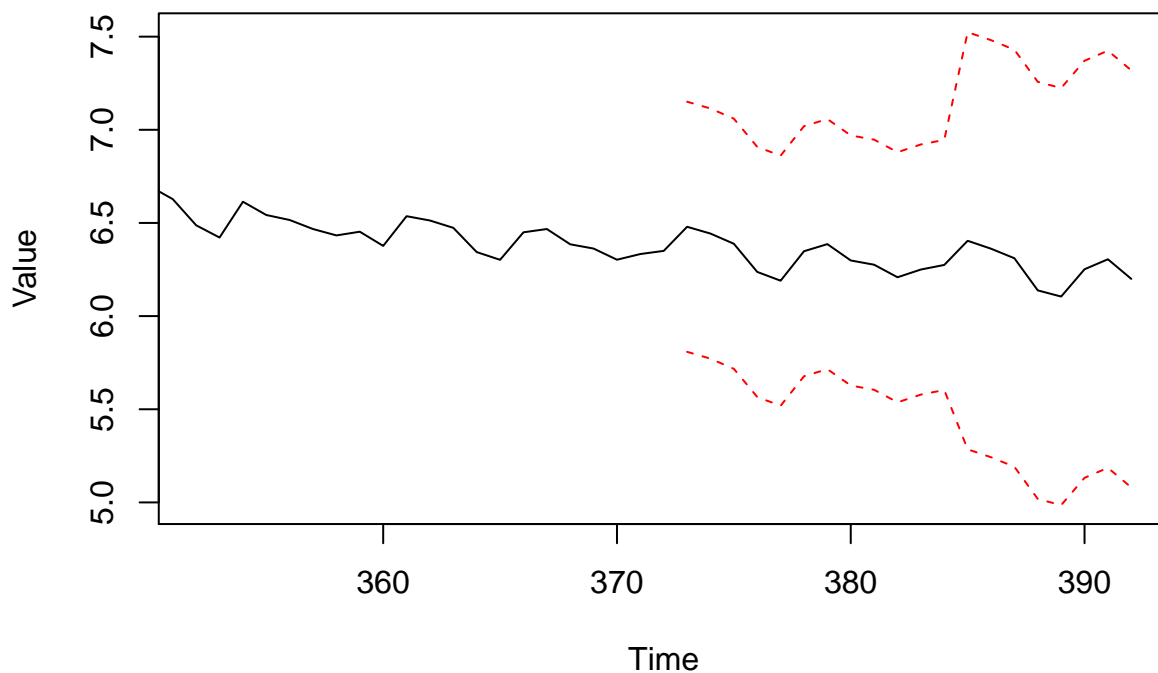
```
## sigma^2 estimated as 0.06287: log likelihood=-13.62
## AIC=35.24 AICc=35.36 BIC=50.79
fit2 <- Arima(lunemp, order=c(0, 0, 0), seasonal=c(2, 2, 0))
fit2

## Series: lunemp
## ARIMA(0,0,0)(2,2,0)[12]
##
## Coefficients:
##          sar1      sar2
##       -0.6655  -0.3370
## s.e.   0.0548   0.0553
##
## sigma^2 estimated as 0.1173: log likelihood=-123.05
## AIC=252.1 AICc=252.17 BIC=263.66
```

The second fit has better BIC/AIC which is then our final model. It can be written formally as

$$(1 + 0.6655B + 0.337B^2) \nabla^2 x_t = w_t.$$

```
fit_plot(fit2)
```



The predictions look reasonable good.