Timeseries-lab2

Prudhvi Peddmallu(prupe690),Zhixuan Duan(zhidu838)

23 September 2019

Assignment 1. Computations with simulated data

Assignment 1. Computations with simulated data

a. Generate 1000 observations from AR(3) process with $\phi_1 = 0.8, \phi_2 = -0.2, \phi_3 = 0.1$. Use these data and the definition of PACF to compute ϕ_{33} from the sample, i.e. write your own code that performs linear regressions on necessarily lagged variables and then computes an appropriate correlation. Compare the result with the output of function pacf() and with the theoretical value of ϕ_{33}

b. Simulate an AR(2) series with $\phi_1 = 0.8, \phi_2 = 0.1$ and $n = 100$. Compute the estimated parameters and their standand errors by using three methods: method of moments (Yule-Walker equations), conditional least squares and maximum likelihood (ML) and compare their results to the true values. Which method does seem to give the best result? Does theoretical value for ϕ_2 fall within confidence interval for ML estimate?

c. Generate 200 observations of a seasonal $ARIMA(0,0,1) \times (0,0,1)_{12}$ model with coefficients $\Theta = 0.6$ and $\theta = 0.3$ by using arima.sim(). Plot sample ACF and PACF and also theoretical ACF and PACF. Which patterns can you see at the theoretical ACF and PACF? Are they repeated at the sample ACF and PACF?

d. Generate 200 observations of a seasonal $ARIMA(0,0,1) \times (0,0,1)_{12}$ model with coefficients $\Theta = 0.6$ and $\theta = 0.3$ by using arima.sim(). Fit $ARIMA(0,0,1) \times (0,0,1)_{12}$ model to the data, compute forecasts and a prediction band 30 points ahead and plot the original data and the forecast with the prediction band. Fit the same data with function *gausspr* from package **kernlab** (use default settings). Plot the original data and predicted data from $t = 1$ to $t = 230$. Compare the two plots and make conclusions.

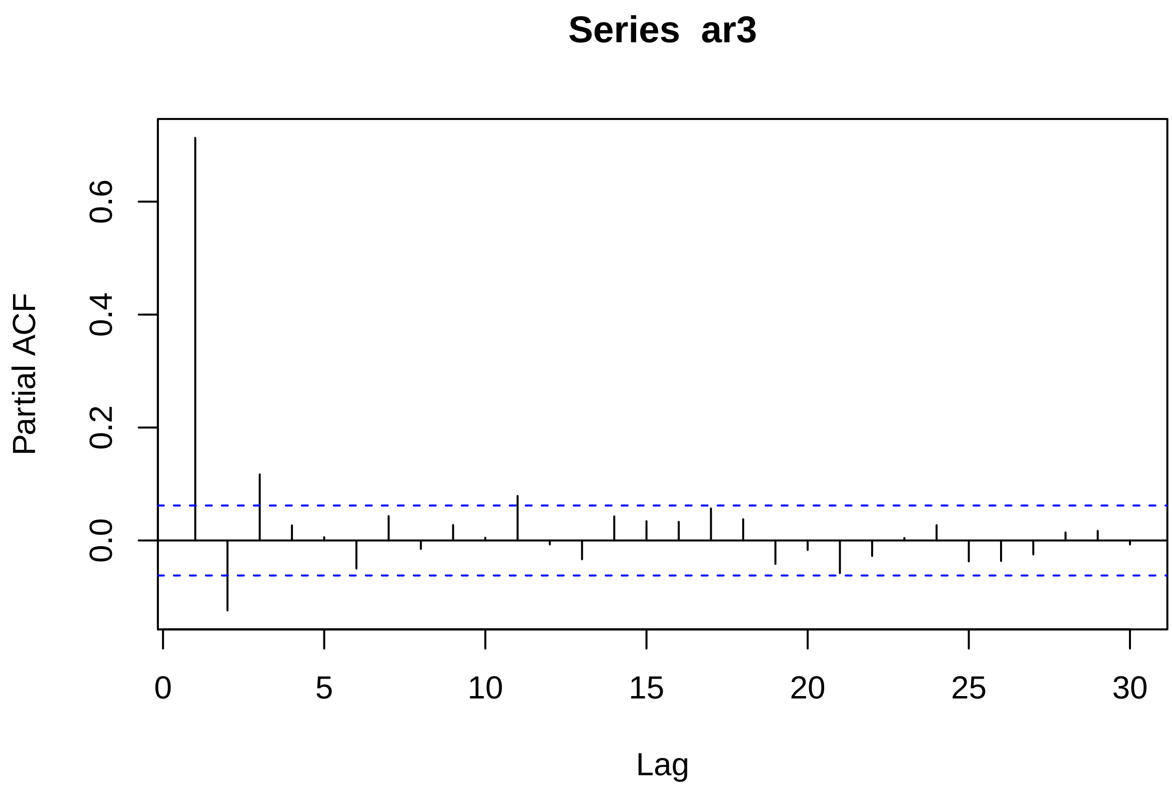
e. Generate 50 observations from ARMA(1,1) process with $\phi = 0.7, \theta = 0.5$. Use first 40 values to fit an ARMA(1,1) model with $\mu = 0$. Plot the data, the 95% prediction band and plot also the true 10 values that you initially dropped. How many of them are outside the prediction band? How can this be interpreted?

1A

a. Generate 1000 observations from AR(3) process with $\phi_1 = 0.8, \phi_2 = -0.2, \phi_3 = 0.1$. Use these data and the definition of PACF to compute ϕ_{33} from the sample, i.e. write your own code that performs linear regressions on necessarily lagged variables and then computes an appropriate correlation. Compare the result with the output of function pacf() and with the theoretical value of ϕ_{33}

```
#1a
#Libraries
library(knitr)#used for kable function
#seed
set.seed(12345)
#Simualte values from AR(3)
ar3 <- arima.sim(model = list(ar = c(0.8,-0.2,0.1)), n = 1000)#data
```

```r
#pacf is the function used for the partial autocorrelations.
ar3partial <- pacf(ar3, plot = FALSE)
#plot PACF
plot(ar3partial)
```




```r
ar3data <- ts.intersect(xt = ar3, x1 = lag(ar3,1), x2 = lag(ar3,2), x3 = lag(ar3, 3))#bind the data
#Correlation, Variance and Covariance
theoretical <- cor(resid(lm(x3 ~ x1 + x2, data = ar3data)),resid(lm(xt ~ x1 + x2,
data = ar3data)))
dfOut <- data.frame("Simulated Value" = ar3partial$acf[3],
"Theoretical Value" = theoretical)
kable(dfOut, caption = "Comparison of correlations using different methods")
```

Table 1: Comparison of correlations using different methods

Simulated.Value	Theoretical.Value
0.1170643	0.1146076

```r
#1A-2method
set.seed(12345)
data=arima.sim(list(ar=c(0.8,-0.2,0.1)), n=1000)
data1=ts.intersect(x=data, x1=lag(data,-1), x2=lag(data,-2), x3=lag(data,-3), dframe = T)
res1=lm(x~x1+x2,data=data1)
res2=lm(x3~x2+x1,data=data1)
r1=residuals(res1)
```

```
r2=residuals(res2)
cor(cbind(r1,r2))
g=pacf(data)
g
```

Analysis:-The theoretical and the simulated seems to be quite similar.

1B

b. Simulate an AR(2) series with $\phi_1 = 0.8, \phi_2 = 0.1$ and $n = 100$. Compute the estimated parameters and their standand errors by using three methods: method of moments (Yule-Walker equations), conditional least squares and maximum likelihood (ML) and compare their results to the true values. Which method does seem to give the best result? Does theoretical value for ϕ_2 fall within confidence interval for ML estimate?

```
#library
library(knitr)
set.seed(12345)
Ar2 <- arima.sim(model = list(ar = c(0.8,0.1)), n = 100)
#Estimating parameters using different methods
estimated_parameters <- function(simDat = Ar2, output = "component", val = c(0.8,0.1)){
# ar is Fit Autoregressive Models to Time Series
YW <- ar(simDat,order = 2, method = "yule-walker", aic = FALSE)#sim data is Ar2 data
CLS <- ar(simDat, order = 2, method = "ols", aic = FALSE)
MLE <- ar(simDat, order = 2, method = "mle", aic = FALSE)
if (output == "component"){

df <- data.frame("Yule-Walker" = c(YW $ar[1],YW $ar[2]),
"Conditional LS" = c(CLS$ar[1],CLS$ar[2]),
"MLE-value" = c(unname(MLE$ar[1]), unname(MLE$ar[2])))
rownames(df) <- c(" 1 component", "2 component")
return(df)
}
if(output == "SE")
  {
mle_model <- arima(simDat, order = c(2,0,0), method = "ML")
CI <- unname(c(mle_model$coef[2]+1.96*(sqrt(mle_model$var.coef[2,2])),
mle_model$coef[2]-1.96*(sqrt(mle_model$var.coef[2,2]))))
ifelse(((0.1-CI[1])*(CI[2]-0.1)>0),
print("theoretical value fall within confidence"),
print("theoretical value  not fall within confidence"))
}
}
kable(estimated_parameters(), caption = "Estimating Parameters using different methods")
```

Table 2: Estimating Parameters using different methods

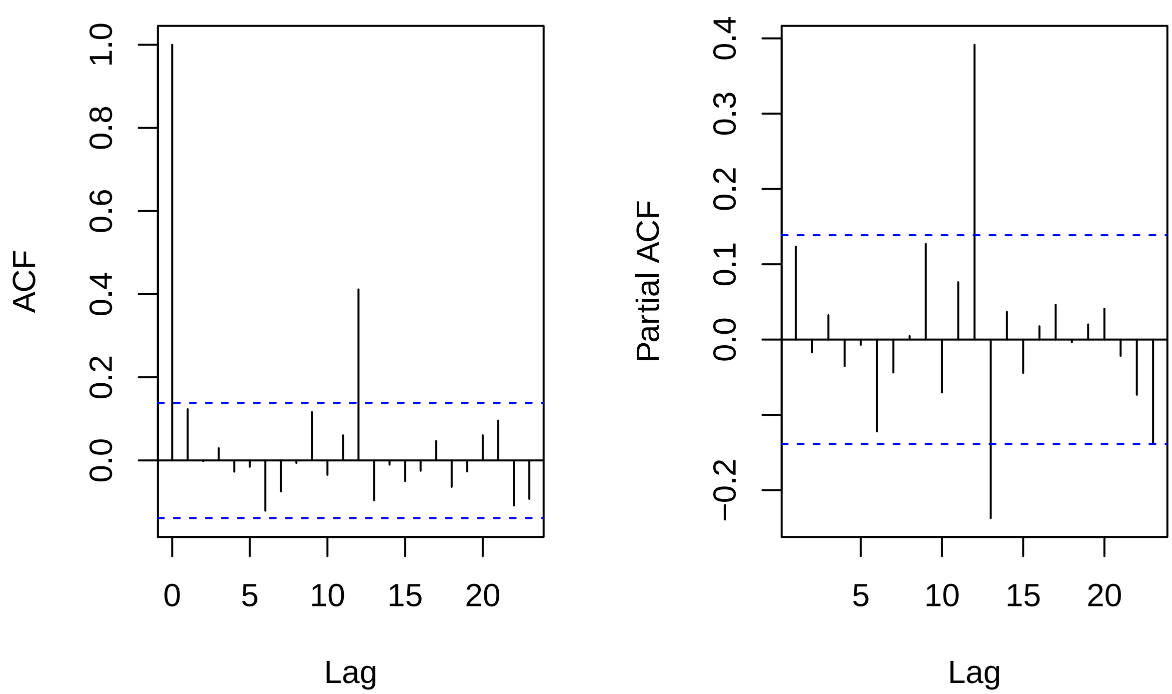
	Yule.Walker	Conditional.LS	MLE.value
1 component	0.8029146	0.8066782	0.7968774
2 component	0.1037053	0.1205352	0.1189369

Analysis:-The theoretical value for falls within the confidence that can be seen in the epar function. Yule Walker's technique seems to be the most precise estimate, but all techniques are close to the real parameters.

1C

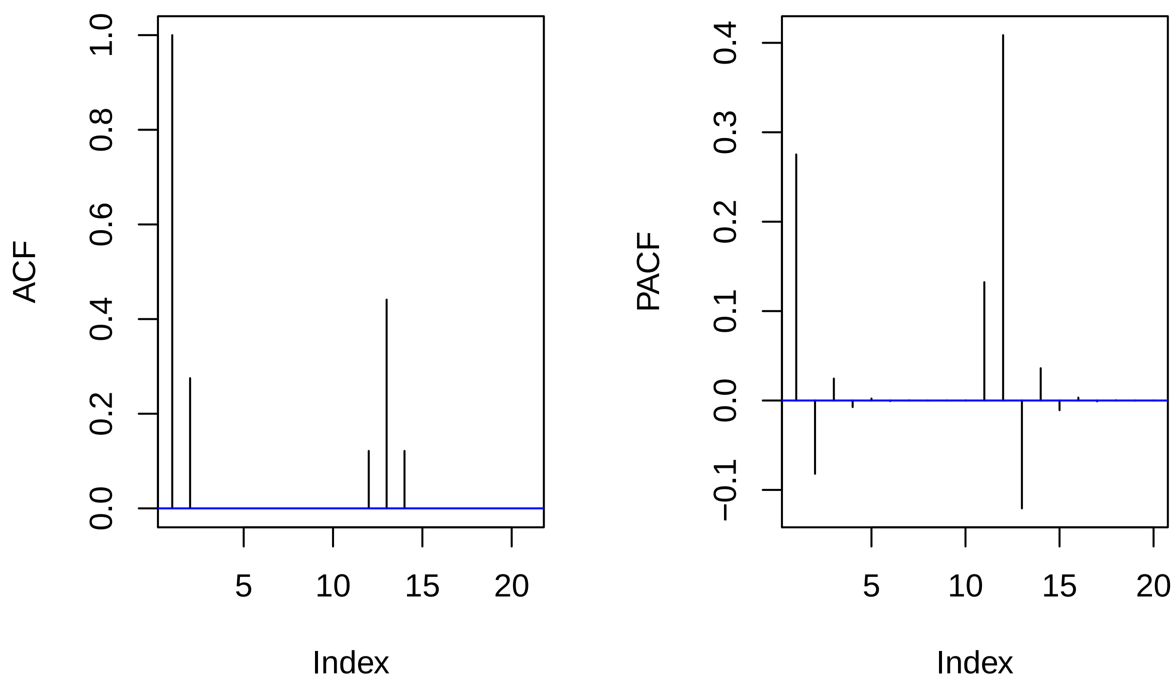
c. Generate 200 observations of a seasonal $ARIMA(0,0,1) \times (0,0,1)_{12}$ model with coefficients $\Theta = 0.6$ and $\theta = 0.3$ by using arima.sim(). Plot sample ACF and PACF and also theoretical ACF and PACF. Which patterns can you see at the theoretical ACF and PACF? Are they repeated at the sample ACF and PACF?

```
Seasonal_ARIMA1<-arima.sim(n = 200, model = list(order = c(0,0,12),  ma = c(0.3,rep(0,10), 0.6)))
#this for 12
#Seasonal_ARIMA1 <- arima.sim(n = 200, model = list(order = c(0,0,13),
#ma = c(0.3,rep(0,10), 0.6, 0.18)))#0.6*0.3=0.18
par(mfrow = c(1,2))
# sample acf plot
acf(Seasonal_ARIMA1, main = NA)
#sample pacf plot
pacf(Seasonal_ARIMA1, main = NA)
```




```
par(mfrow = c(1,1))
#Theoretical ACF and PACF
arima_acf <- ARMAacf(ma = c(0.3,rep(0,10),0.6,0.6*0.3), lag.max = 20)
arima_pacf <- ARMAacf(ma = c(0.3,rep(0,10),0.6,0.6*0.3), pacf = TRUE, lag.max = 20)
par(mfrow = c(1,2))
```

```
#Theoretical plots
plot(arima_acf, type = "h", main = NA, ylab = "ACF")
abline(h = 0, col = "blue")
plot(arima_pacf, type = "h", ylab = "PACF")
abline(h = 0, col = "blue")
```




```
par(mfrow = c(1,1))
```

1D

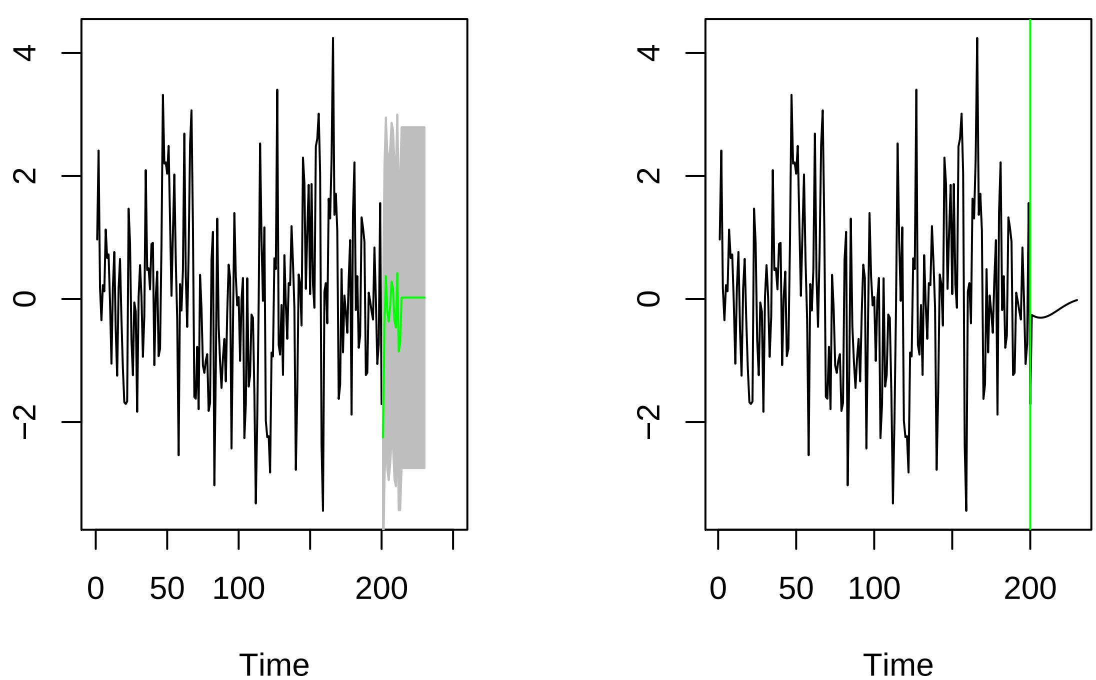
d. Generate 200 observations of a seasonal $ARIMA(0,0,1) \times (0,0,1)_{12}$ model with coefficients $\Theta = 0.6$ and $\theta = 0.3$ by using arima.sim(). Fit $ARIMA(0,0,1) \times (0,0,1)_{12}$ model to the data, compute forecasts and a prediction band 30 points ahead and plot the original data and the forecast with the prediction band. Fit the same data with function *gausspr* from package **kernlab** (use default settings). Plot the original data and predicted data from $t = 1$ to $t = 230$. Compare the two plots and make conclusions.

```
#Library
library(kernlab)
#model
arimaSeason_model <- arima.sim(list(order = c(0,0,12), ma = c(0.7,rep(0,10),0.6)), n = 200)
#fit the seasonal model
arimaSeason_fit<- arima(arimaSeason_model, order = c(0,0,1),
                        seasonal = list(order = c(0,0,1),period = 12))
```

```r
arimaSeason_predict <- predict(arimaSeason_fit, n.ahead = 30)#30 points

#function gausspr from package kernlab
gaussdata <- data.frame(y = as.vector(arimaSeason_model), x = 1:200)
#1:200 because arimaSeasom _model contains [1:200]observations

gausspredict <- gausspr(y ~ x, data = gaussdata)
```

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

```r
gaussprediction <- predict(gausspredict, newdata = data.frame(x = 201:230))#t =1 to 230

par(mfrow = c(1,2))
#compare plots
plot(arimaSeason_model, xlim = c(0,250),
ylab = NA)
#arimaSeason_predict is list of two variables pred and se
upper <- arimaSeason_predict$pred + 1.96*arimaSeason_predict$se
lower <- arimaSeason_predict$pred - 1.96*arimaSeason_predict$se

polygon(c(time(upper),rev(time(upper))),c(lower, rev(upper)),border = 8, col = "grey")
lines(arimaSeason_predict$pred, col = "green")
plot(c(arimaSeason_model,gaussprediction), col = "black", type = "l",
ylab = NA, xlab = "Time")
abline(v = 200, col = "green", lty = 1)
```



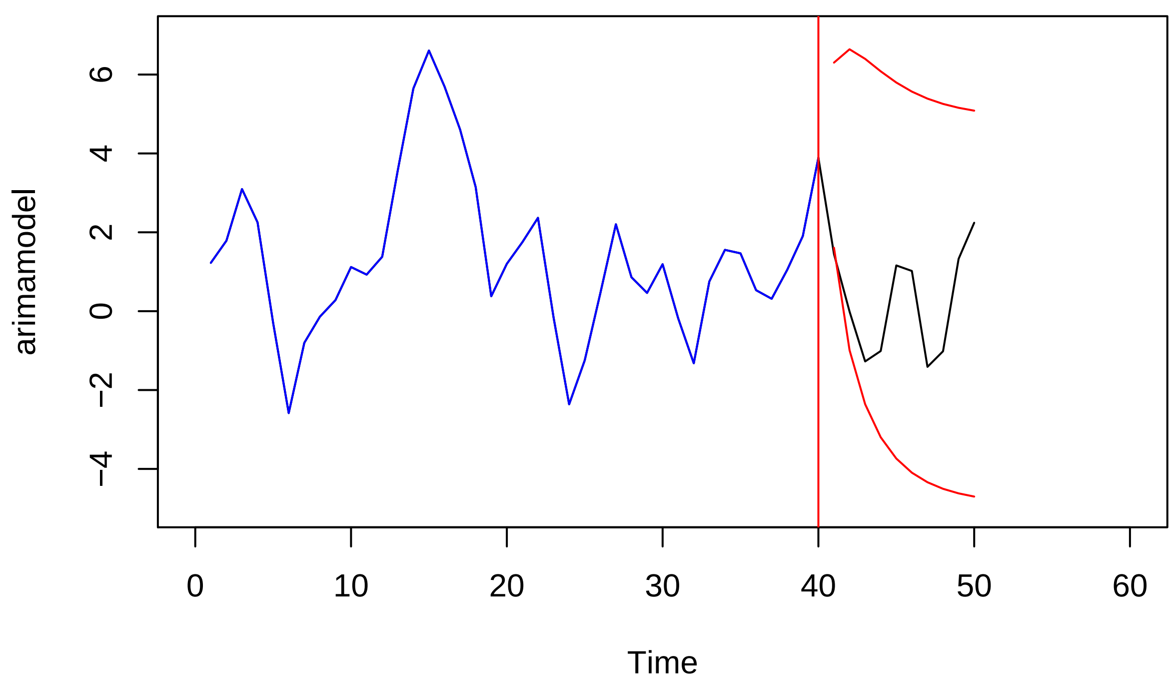
```r
par(mfrow = c(1,1))
```

Analysis:- We can observe that compared to the periodic forecast, the gausspr generates a smoother line of prediction.We can see from the periodic arima forecast that the predictive bands of 95 percent are quite broad. We can see that the periodic arima forecast generates a more choppy prediction line that suits the information charaster better than the Gaussian model's smooth line.

1E

e. Generate 50 observations from ARMA(1,1) process with $\phi = 0.7, \theta = 0.5$. Use first 40 values to fit an ARMA(1,1) model with $\mu = 0$. Plot the data, the 95% prediction band and plot also the true 10 values that you initially dropped. How many of them are outside the prediction band? How can this be interpreted?

```r
set.seed(12345)
arimamodel <- arima.sim(list(order = c(1,0,1), ar = 0.7,ma = 0.5), n = 50)
arimasample40 <- arimamodel[1:40]#40 values
arimamodel40 <- arima(arimasample40, order = c(1,0,1), include.mean = 0)#mu =0
prediction10 <- predict(arimamodel40, n.ahead = 10)#predict 10 values
plot(arimamodel, type = "l", xlim = c(0,60), ylim = c(-5,7))
lines(prediction10$pred + 1.96*prediction10$se, col = "red")
lines(prediction10$pred - 1.96*prediction10$se, col = "red")
lines(arimasample40, col = "blue")
abline(v = 40, col = "red")
```



Analysis:- That there should be at least 95% of the true observations within the bands of trust, but in our case all the true observations are within trust.

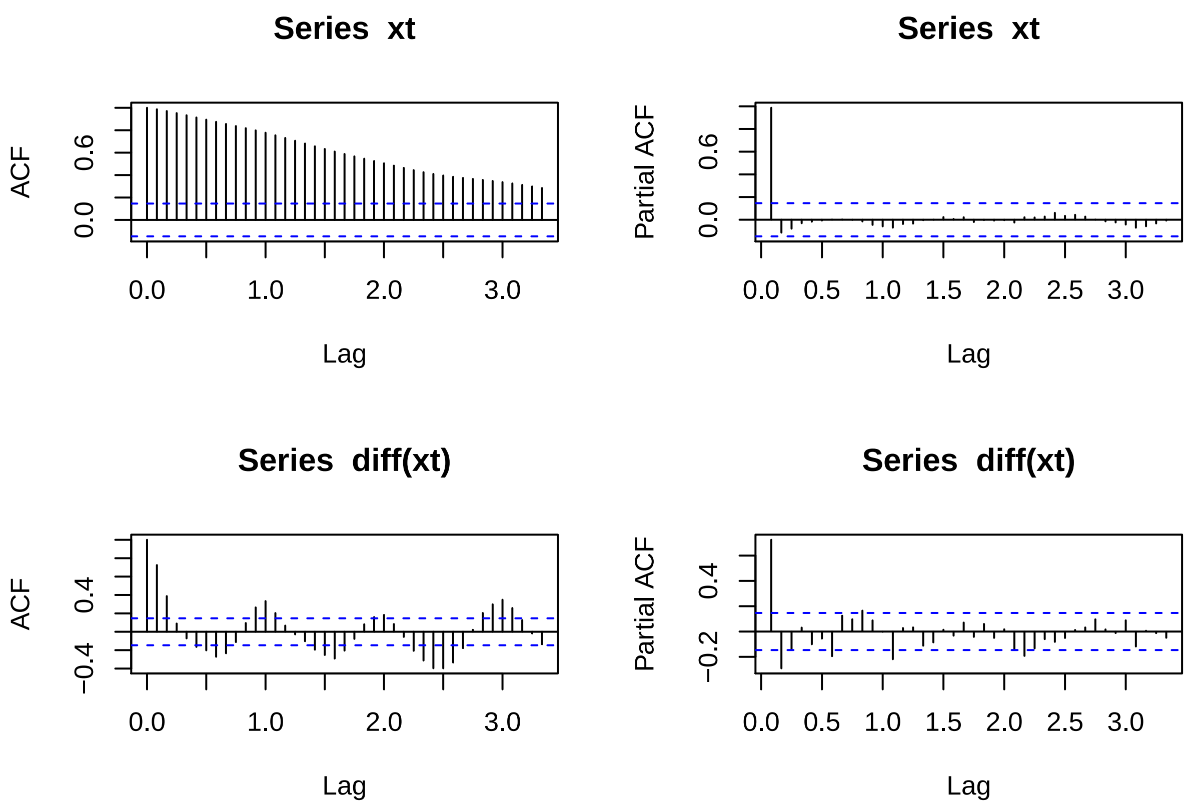
Assignment 2. ACF and PACF diagnostics

2A & B

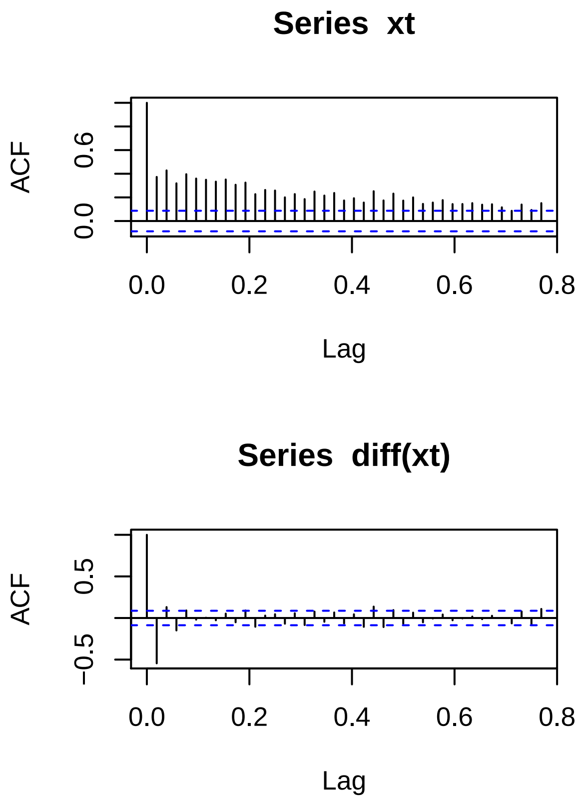
Assignment 2. ACF and PACF diagnostics.

a. For data series *chicken* in package **astsa** (denote it by x_t), plot 4 following graphs up to 40 lags: ACF(x_t), PACF(x_t), ACF(∇x_t), PACF(∇x_t) (group them in one graph). Which $ARIMA(p, d, q)$ or $ARIMA(p, d, q) \times (P, D, Q)_s$ models can be suggested based on this information only? Motivate your choice.
b. Repeat step 1 for the following datasets: *so2, EQcount, HCT* in package **astsa.**

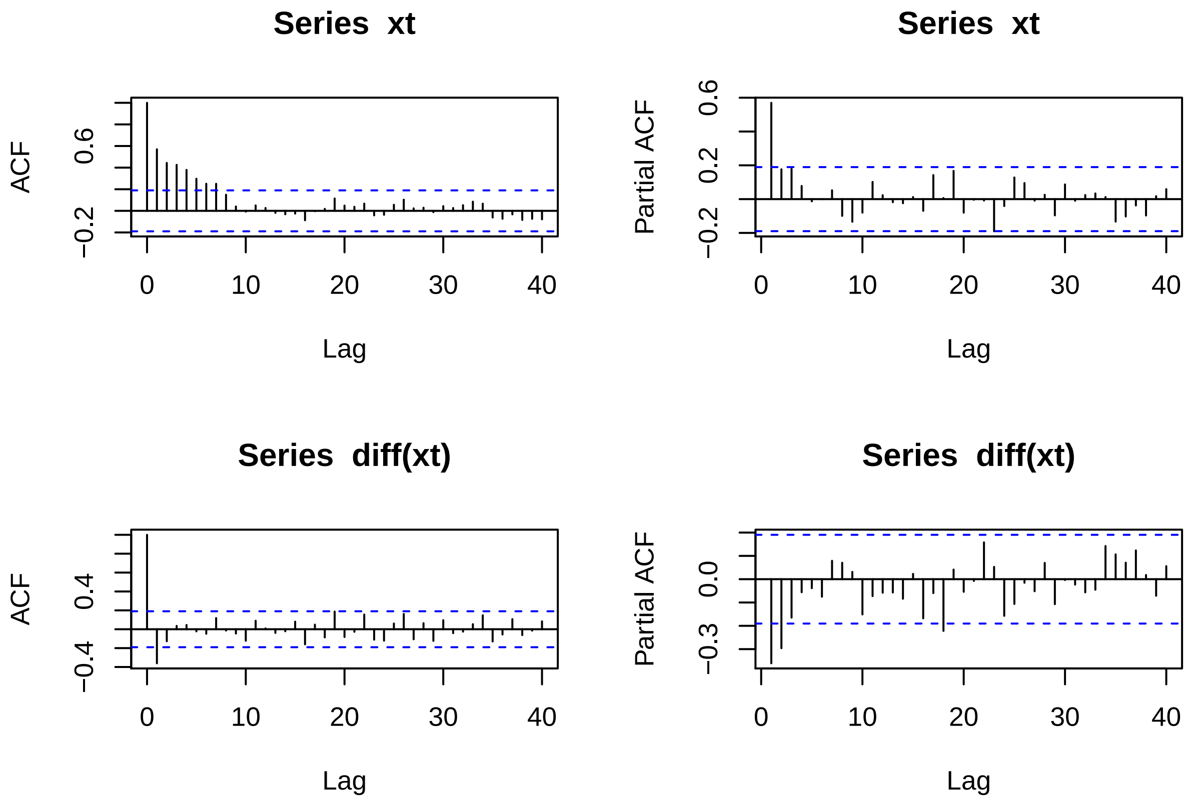
```r
#Library
library(astsa)
#Plots of ACF PACF of xt and ACF PACF diff xt
plots <- function(xt, maxlag = 40){ #max lag 40
par(mfrow = c(2, 2))
acf(xt, lag.max = maxlag)
pacf(xt, lag.max = maxlag)
acf(diff(xt), lag.max = maxlag) ## difference?
pacf(diff(xt), lag.max = maxlag)
}
plots(chicken)
```



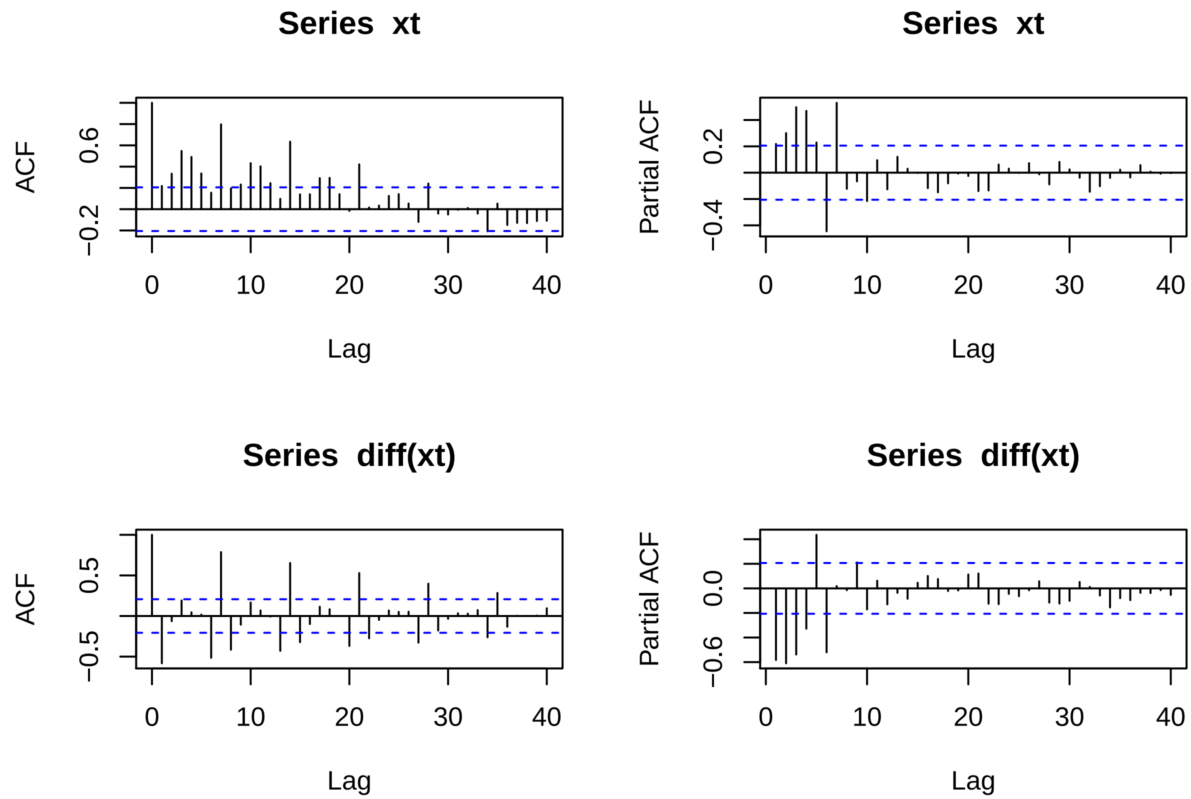

```
#2b-Repeat step 1 for the following datasets: so2, EQcount, HCT in package astsa
plots(so2)
```




```
plots(EQcount)
```


Series xt
ACF
0.6
-0.2
0
10
20
30
40
Lag
Series xt
Partial ACF
0.6
0.2
-0.2
0
10
20
30
40
Lag
Series diff(xt)
ACF
0.4
-0.4
0
10
20
30
40
Lag
Series diff(xt)
Partial ACF
0.0
-0.3
0
10
20
30
40
Lag

```
plots(HCT)
```



Analysis:-

Chicken We can see from the ACF that the arrangement appears to be non-stationary and so we ought to take the first difference for the arima show. From the differenced ACF ready to se that it cuts of after two slacks which mean that we ought to have an AR(2) component in our arima demonstrate. We will moreover see from the differenced ACF that we might have a regular effect with a period of 12. We will consider an regular autoregressive ARIMA(2, 1, 0) with a regular componend with a period of 12. so2 ACF and PACF (from the differenced arrangement) appears that this might be an AR. From the PACF able to see that is kind of tailing off so it moreover recommend an MA. From the distinction of arrange 1 we are able see a cut off on ACF after slack 1 whereas the PACF tailes of which propose an ARIMA(0,1,1) with no regular component. EQcount From the ACF ready to see that it tails off truly quick and from the PACF ready to see that there's a significant correlation at slack one taken after by not noteworthy relationships, this recommend a ARIMA(7, 0, 0)(0, 0, 1)7

Assignment 3. ARIMA modeling cycle

Assignment 3. ARIMA modeling cycle.

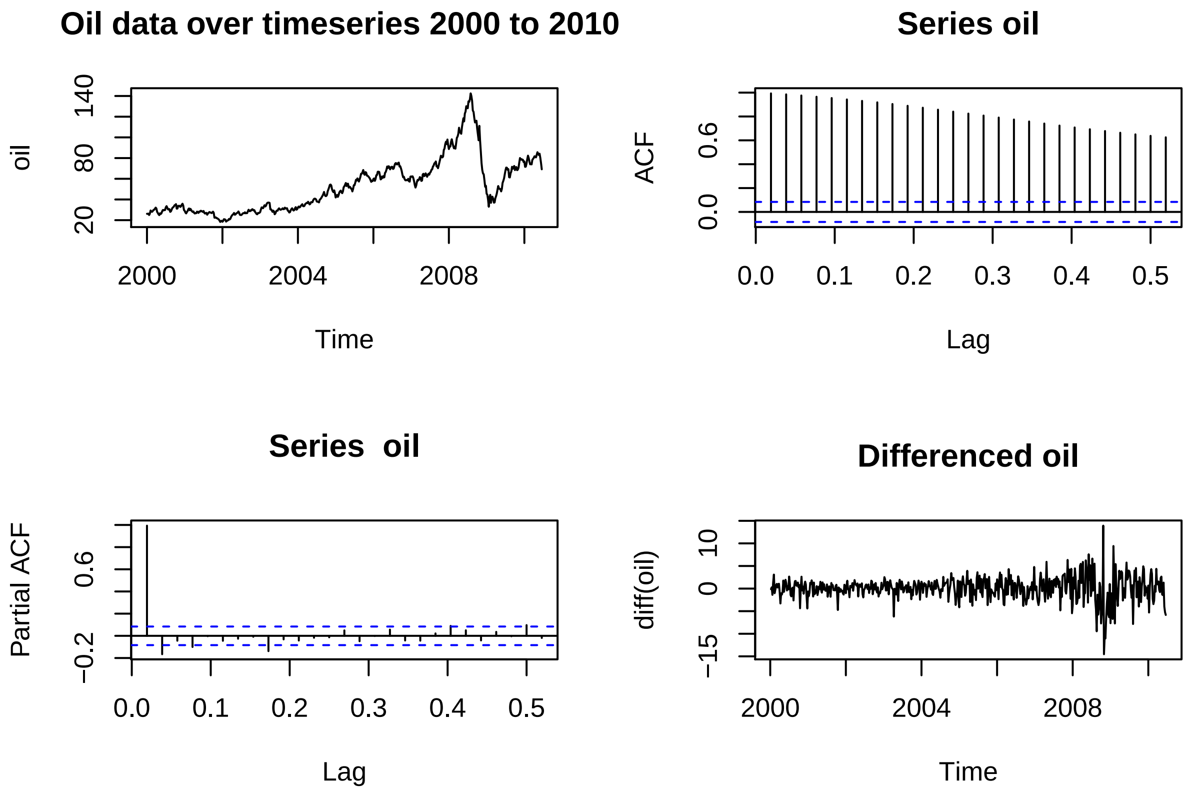
In this assignment, you are assumed to apply a complete ARIMA modeling cycle starting from visualization and detrending and ending up with a forecasting.

a. Find a suitable $ARIMA(p, d, q)$ model for the data set *oil* present in the library **astsa.** Your modeling should include the following steps in an appropriate order: visualization, unit root test, detrending by differencing (if necessary), transformations (if necessary), ACF and PACF plots when needed, EACF analysis, Q-Q plots, Box-Ljung test, ARIMA fit analysis, control of the parameter redundancy in the fitted model. When performing these steps, always have 2 tentative models at hand and select one of them in the end. Validate your choice by AIC and BIC and write down the equation of the selected model. Finally, perform forecasting of the model 20 observations ahead and provide a suitable plot showing the forecast and its uncertainty.
b. Find a suitable $ARIMA(p, d, q) \times (P, D, Q)_s$ model for the data set *unemp* present in the library **astsa.** Your modeling should include the following steps in an appropriate order: visualization, detrending by differencing (if necessary), transformations (if necessary), ACF and PACF plots when needed, EACF analysis, Q-Q plots, Box-Ljung test, ARIMA fit analysis, control of the parameter redundancy in the fitted model. When performing these steps, always have 2 tentative models at hand and select one of them in the end. Validate your choice by AIC and BIC and write down the equation of the selected model (write in the backshift operator notation without expanding the brackets). Finally, perform forecasting of the model 20 observations ahead and provide a suitable plot showing the forecast and its uncertainty.

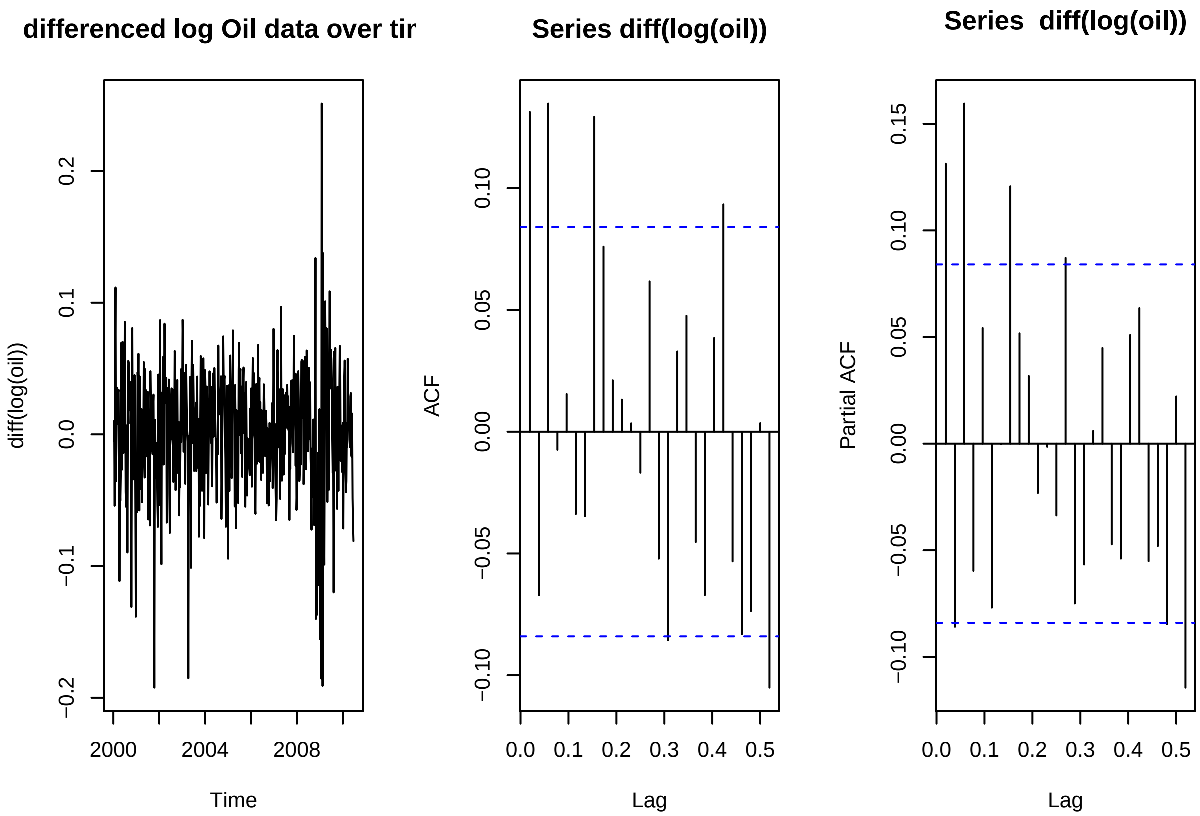
3A

a. Find a suitable $ARIMA(p, d, q)$ model for the data set *oil* present in the library **astsa.** Your modeling should include the following steps in an appropriate order: visualization, unit root test, detrending by differencing (if necessary), transformations (if necessary), ACF and PACF plots when needed, EACF analysis, Q-Q plots, Box-Ljung test, ARIMA fit analysis, control of the parameter redundancy in the fitted model. When performing these steps, always have 2 tentative models at hand and select one of them in the end. Validate your choice by AIC and BIC and write down the equation of the selected model. Finally, perform forecasting of the model 20 observations ahead and provide a suitable plot showing the forecast and its uncertainty.

```
#Library
library(tseries)
library(knitr)
library(TSA)
library(astsa)
#oil data is in astsa library
oil<-oil#[1:545] observations in oil data
par(mfrow = c(2,2))
plot(oil, main = "Oil data over timeseries 2000 to 2010 ")
#acf and pacf
acf(oil)
pacf(oil)

plot(diff(oil), main = "Differenced oil")
```




```r
par(mfrow = c(1,1))
par(mfrow = c(1,3))
plot(diff(log(oil)), main = "differenced log Oil data over time")
#diff plot of acf and pacf
acf(diff(log(oil)))
pacf(diff(log(oil)))
```



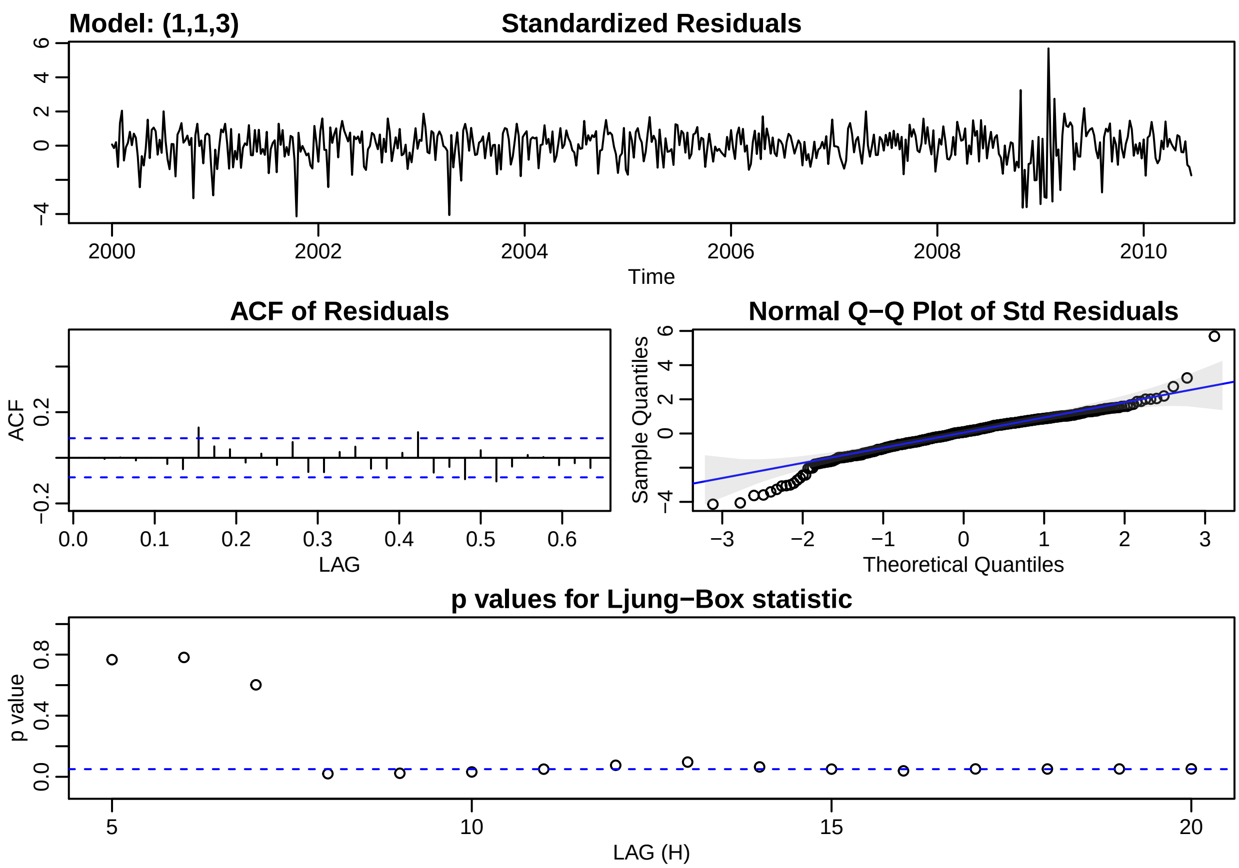

```r
par(mfrow = c(1,1))
#extended acf
eacf(diff(log(oil)))
```

```
## 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
#Suggested Models
#ar is the row ma is the column
model3A <- sarima(log(oil), 1,1,3)#by taking eacf matrix we need to take the values p,d,q etc
```

```
## initial  value -3.057594
## iter   2 value -3.081639
## iter   3 value -3.086469
## iter   4 value -3.086671
## iter   5 value -3.086741
## iter   6 value -3.086743
## iter   7 value -3.086743
## iter   8 value -3.086746
```

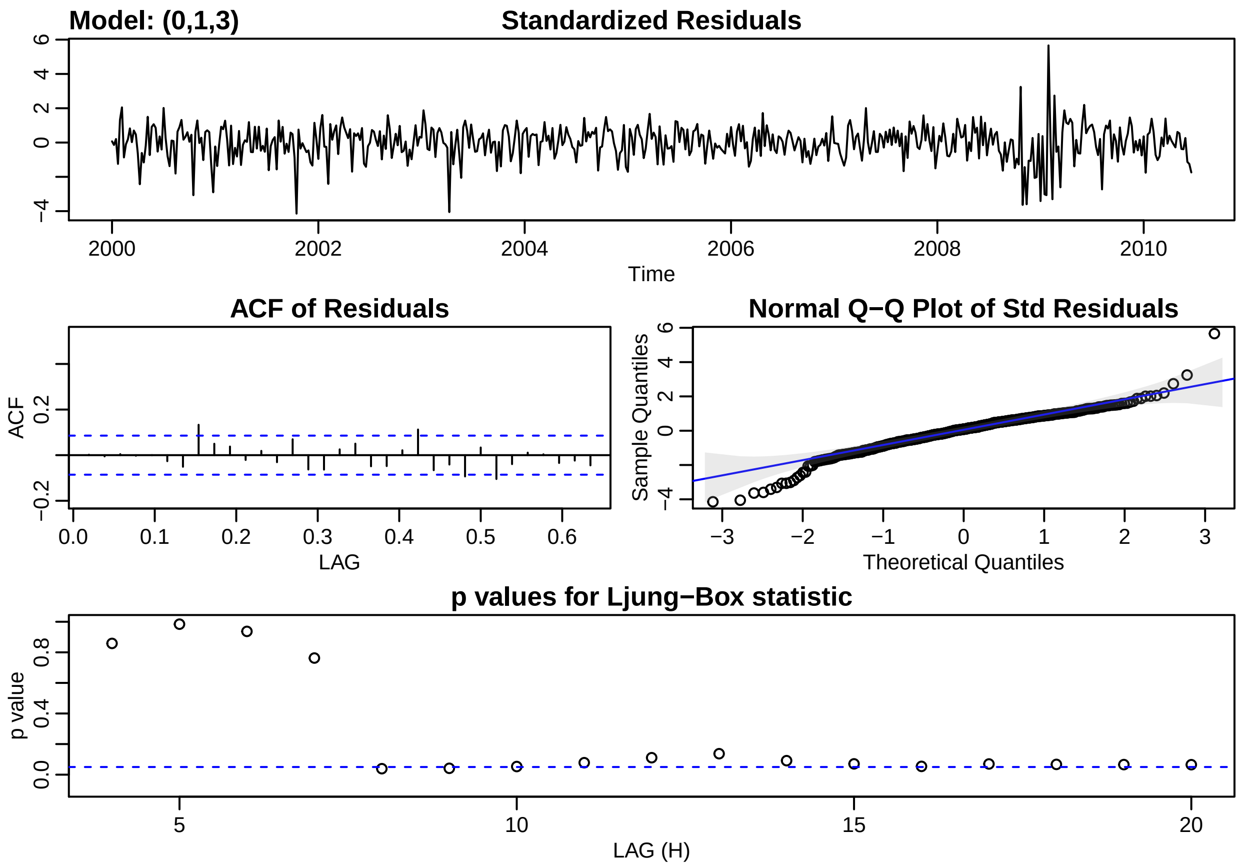
```
## iter   9 value -3.086748
## iter  10 value -3.086749
## iter  11 value -3.086749
## iter  12 value -3.086750
## iter  13 value -3.086750
## iter  14 value -3.086750
## iter  15 value -3.086750
## iter  15 value -3.086750
## iter  15 value -3.086750
## final  value -3.086750
## converged
## initial  value -3.087502
## iter   2 value -3.087503
## iter   3 value -3.087503
## iter   4 value -3.087503
## iter   5 value -3.087503
## iter   6 value -3.087503
## iter   6 value -3.087503
## iter   6 value -3.087503
## final  value -3.087503
## converged
```




```r
model3B <- sarima(log(oil), 0,1,3)
```

```
## initial  value -3.058495
## iter   2 value -3.086110
## iter   3 value -3.086980
```

```
## iter   4 value -3.087501
## iter   5 value -3.087521
## iter   6 value -3.087521
## iter   7 value -3.087522
## iter   8 value -3.087522
## iter   9 value -3.087522
## iter   9 value -3.087522
## iter   9 value -3.087522
## final  value -3.087522
## converged
## initial  value -3.087448
## iter   2 value -3.087448
## iter   3 value -3.087449
## iter   3 value -3.087449
## iter   3 value -3.087449
## final  value -3.087449
## converged
```




```r
#ADF test
adf.test(model3A$fit$residuals)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  model3A$fit$residuals
## Dickey-Fuller = -6.7508, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

```r
adf.test(model3B$fit$residuals)
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  model3B$fit$residuals
## Dickey-Fuller = -6.7187, Lag order = 8, p-value = 0.01
## alternative hypothesis: stationary
```

```r
summary(model3A$fit)
```

```
##           Length Class  Mode
## coef        5    -none- numeric
## sigma2      1    -none- numeric
## var.coef   25    -none- numeric
## mask        5    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 545    ts     numeric
## call        8    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

```r
summary(model3B$fit)
```

```
##           Length Class  Mode
## coef        4    -none- numeric
## sigma2      1    -none- numeric
## var.coef   16    -none- numeric
## mask        4    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 545    ts     numeric
## call        8    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

```r
#BIC & AIC
BIC(model3A$fit)
```

```
## [1] -1777.605
```

```r
BIC(model3B$fit)
```

```
## [1] -1783.844
```

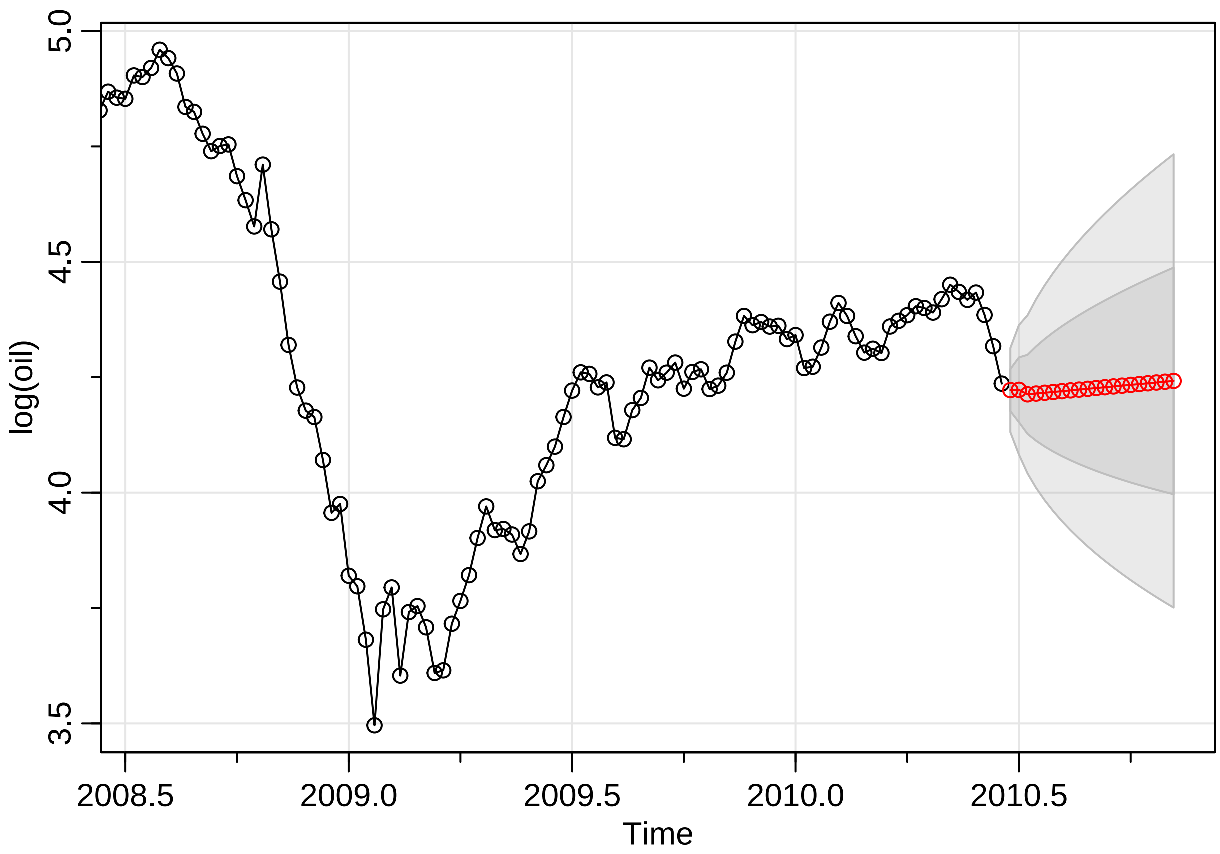
```r
AIC(model3A$fit)
```

```
## [1] -1803.398
```

```r
AIC(model3B$fit)
```

```
## [1] -1805.339
```

```r
#Forecasting
sarima.for(log(oil), 0,1,3, n.ahead = 20)
```




```
## $pred
## Time Series:
## Start = c(2010, 26)
## End = c(2010, 45)
## Frequency = 52
##  [1] 4.222141 4.222731 4.212938 4.214647 4.216356 4.218066 4.219775
##  [8] 4.221485 4.223194 4.224904 4.226613 4.228323 4.230032 4.231741
## [15] 4.233451 4.235160 4.236870 4.238579 4.240289 4.241998
##
## $se
## Time Series:
## Start = c(2010, 26)
## End = c(2010, 45)
## Frequency = 52
##  [1] 0.04561249 0.07016150 0.08569792 0.10226755 0.11650396 0.12918085
##  [7] 0.14072033 0.15138273 0.16134203 0.17072132 0.17961149 0.18808192
## [13] 0.19618697 0.20397021 0.21146718 0.21870731 0.22571532 0.23251220
## [19] 0.23911597 0.24554218
```

Analysis:-

We can see that the orignal acf indicates that the information is not stationary, so we inspect the differentiated sequence that can be seen in the plot at the bottom right. With growing variance as a function of growing time, the variance of the differentiated series appears to change over time. We can also note that the variance seems to decrease at the end of 2008

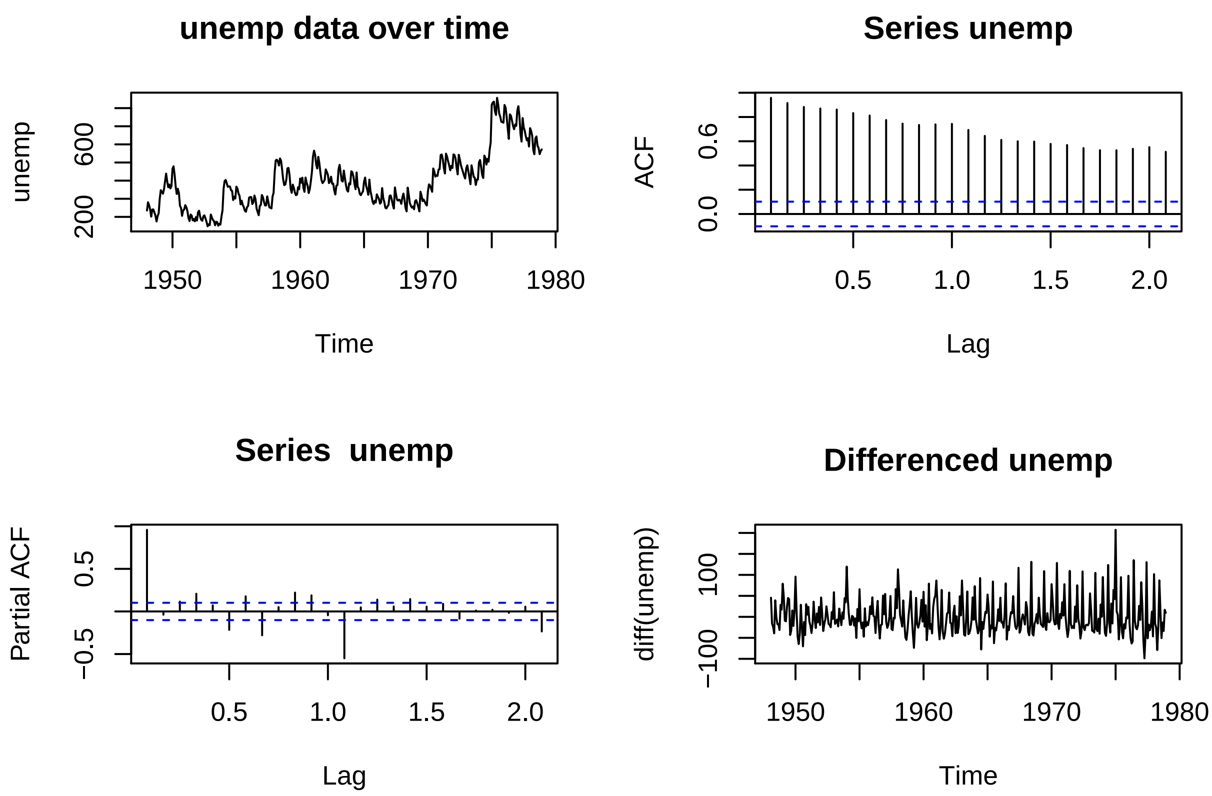
The differentiated log transformations with respective time series, acf and pacf are presented here. From the left graph representing the different time series of logs, we can see that it is not completely stationary, but it seems good enough.

We propose an ARIMA(0,1,3) and ARIMA(1,1,3) based on the eacf's triangles. We can see that the ACF appears to be good and stationary for the first model ARIMA(1, 1, 3) that can be verified from the adf test. The QQ plots seems to be relatively normal and the unit root test shows significance which means stationarity according to the test. The result in model ARIMA(0, 1, 3) shows really similar,diagnosis of the model in both instances. We can see from the AIC and BIC that the result is almost the same, so we will choose the simpler model as our final model (0, 1, 3). For both ljung box tests we can see that residuals are independent for the first lags so we these models can be considered as a good fit.

3B

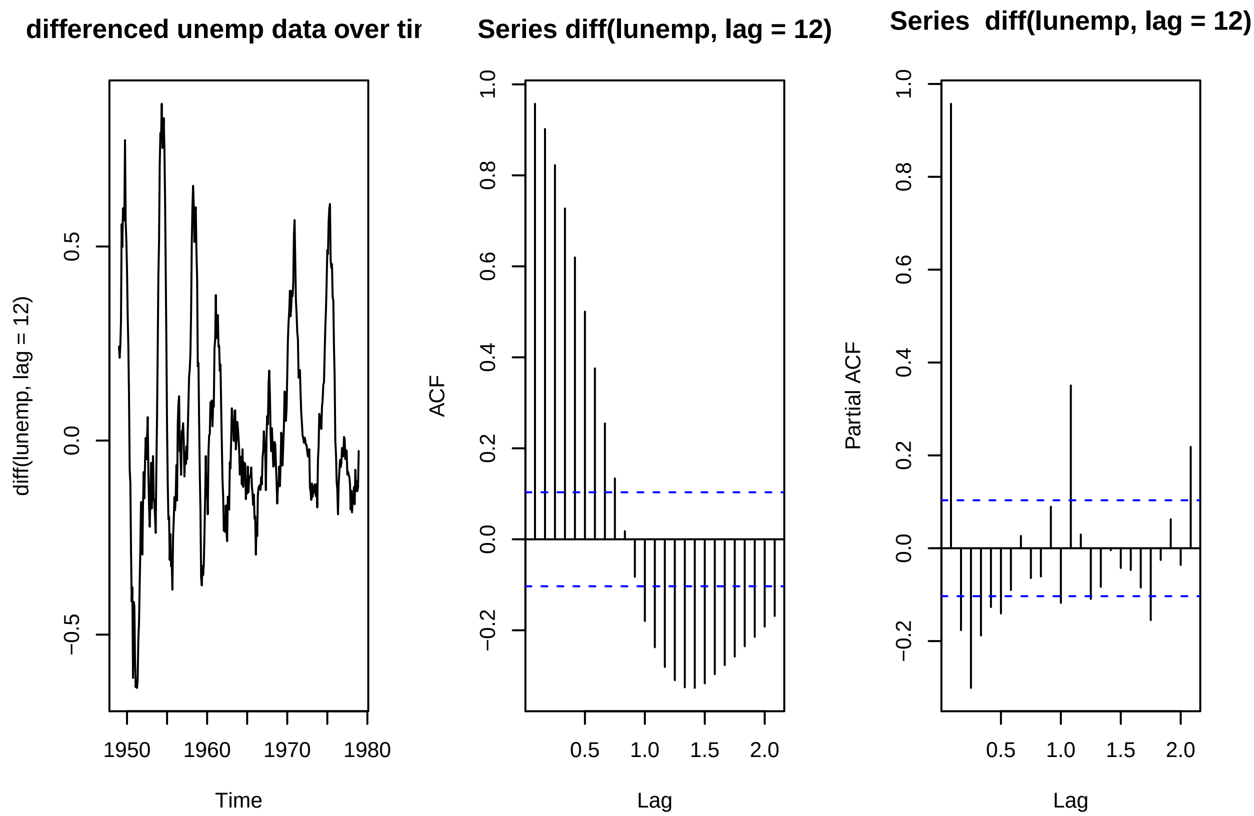
b. Find a suitable $ARIMA(p, d, q) \times (P, D, Q)_s$ model for the data set *unemp* present in the library **astsa.** Your modeling should include the following steps in an appropriate order: visualization, detrending by differencing (if necessary), transformations (if necessary), ACF and PACF plots when needed, EACF analysis, Q-Q plots, Box-Ljung test, ARIMA fit analysis, control of the parameter redundancy in the fitted model. When performing these steps, always have 2 tentative models at hand and select one of them in the end. Validate your choice by AIC and BIC and write down the equation of the selected model (write in the backshift operator notation without expanding the brackets). Finally, perform forecasting of the model 20 observations ahead and provide a suitable plot showing the forecast and its uncertainty.

```r
#Library
library(tseries)
library(knitr)
library(TSA)
library(astsa)
#oil data is in astsa library
umemp<-unemp#[1:372] observations in oil data
par(mfrow = c(2,2))
plot(unemp, main = "unemp data over time")
acf(unemp)
pacf(unemp)
plot(diff(unemp), main = "Differenced unemp")
```

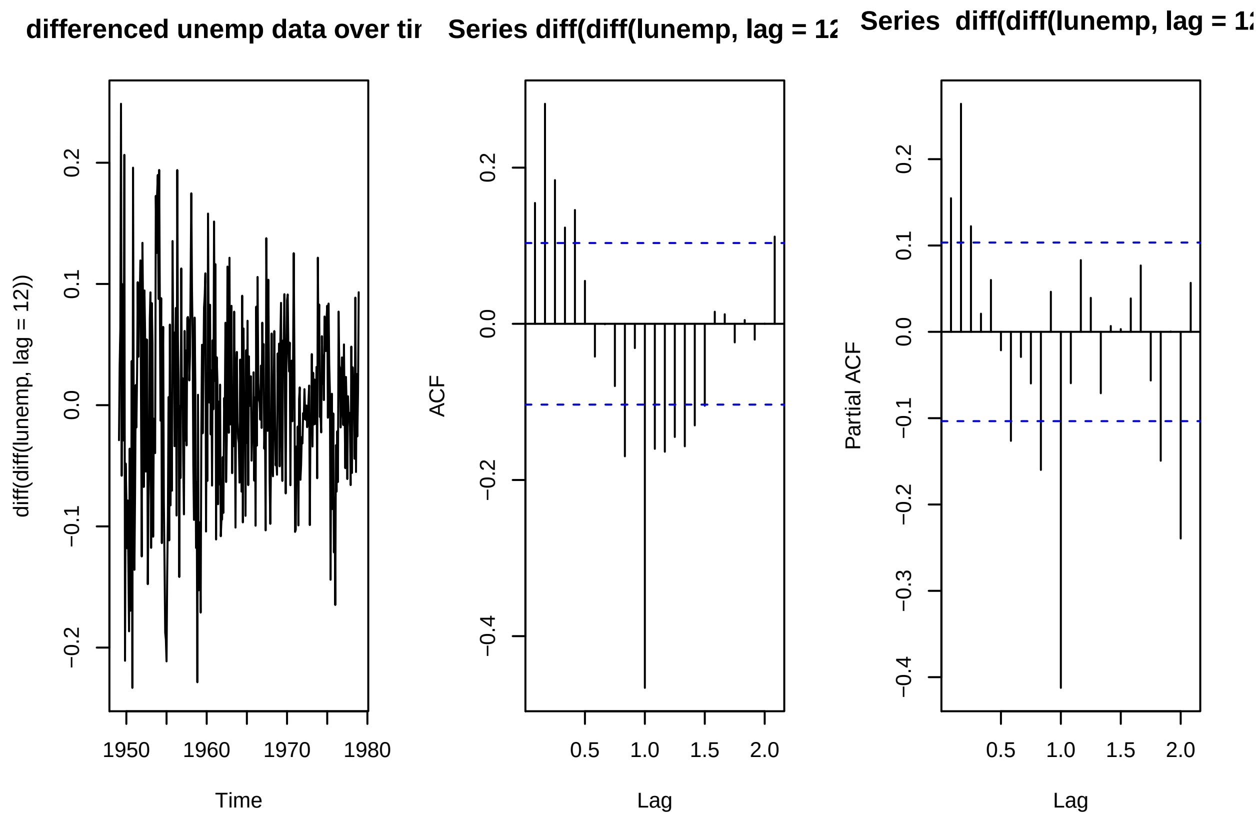



```r
par(mfrow = c(1,1))
```

```r
lunemp <- log(unemp)
par(mfrow = c(1,3))
plot(diff(lunemp, lag = 12), main = "differenced unemp data over time")
acf(diff(lunemp, lag = 12))
pacf(diff(lunemp, lag = 12))
```




```
par(mfrow = c(1,3))
plot(diff(diff(lunemp, lag = 12)), main = "differenced unemp data over time")
acf(diff(diff(lunemp, lag = 12)))
pacf(diff(diff(lunemp, lag = 12)))
```



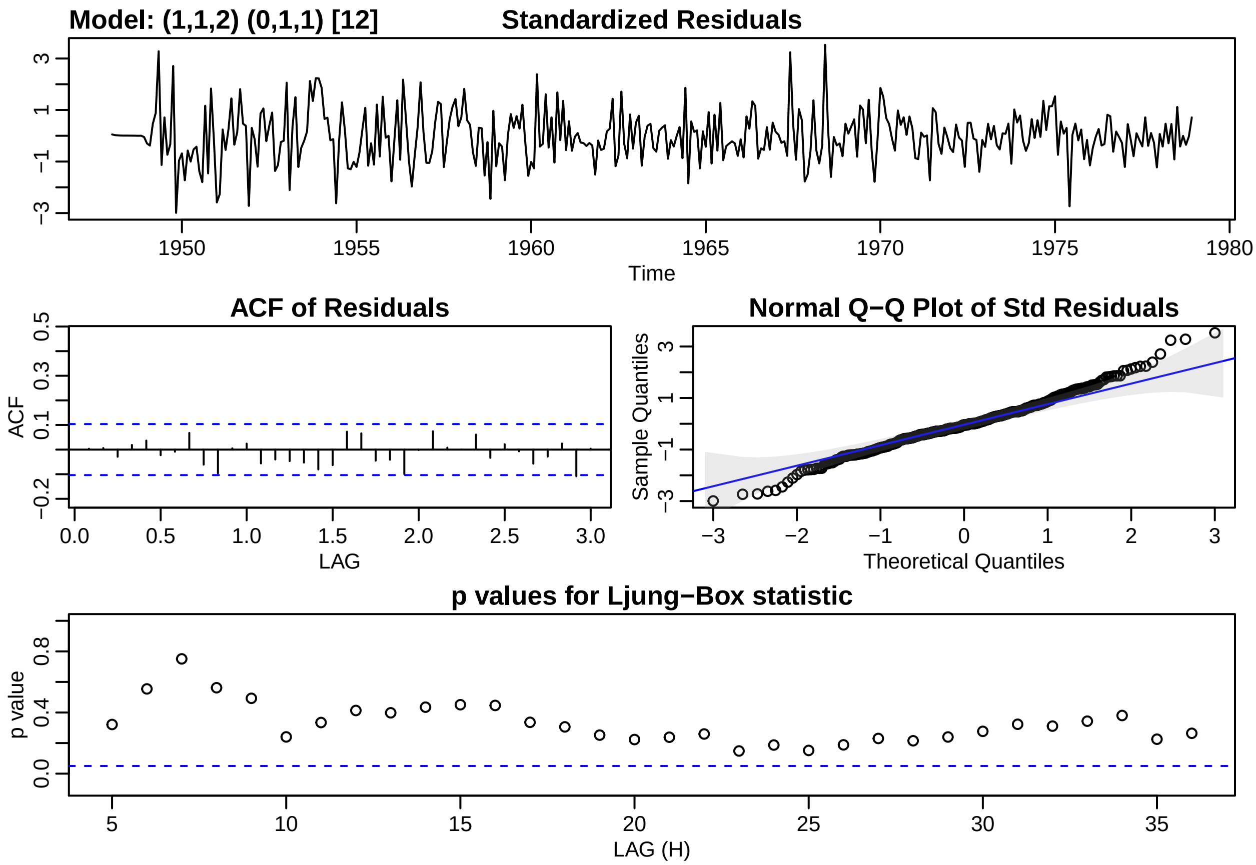

```r
par(mfrow = c(1,1))
d2 <- diff(diff(lunemp, lag = 12))
eacf(d2)
```

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

```r
#Suggested Models
modelA3 <- sarima(lunemp, 1,1,2,0,1,1,12)
```

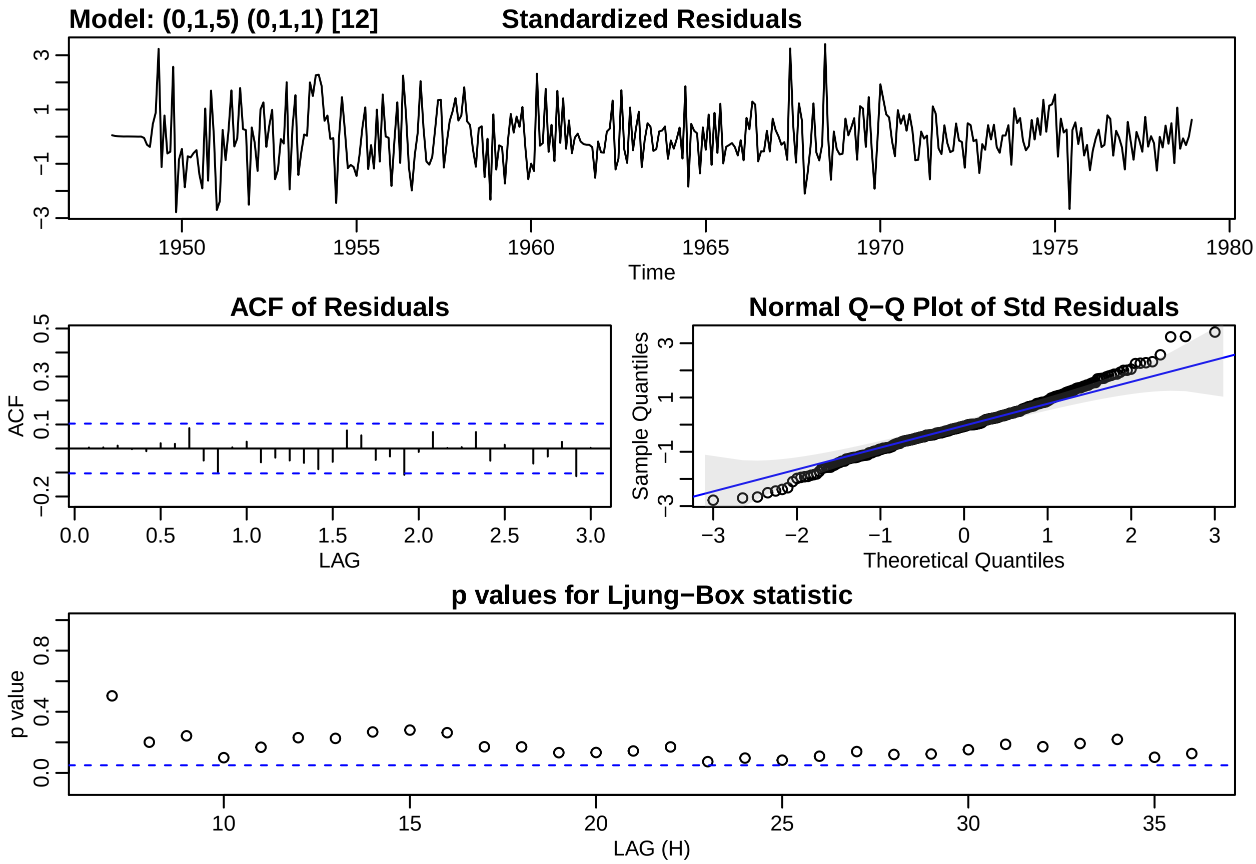
```
## initial  value -2.551638
## iter   2 value -2.718836
## iter   3 value -2.720851
## iter   4 value -2.762012
## iter   5 value -2.770952
## iter   6 value -2.775221
## iter   7 value -2.775460
## iter   8 value -2.775572
## iter   9 value -2.776560
```

```
## iter  10 value -2.778622
## iter  11 value -2.780027
## iter  12 value -2.780851
## iter  13 value -2.780957
## iter  14 value -2.781116
## iter  15 value -2.782147
## iter  16 value -2.782462
## iter  17 value -2.782616
## iter  18 value -2.782623
## iter  19 value -2.782624
## iter  19 value -2.782624
## final  value -2.782624
## converged
## initial  value -2.795103
## iter   2 value -2.796333
## iter   3 value -2.796574
## iter   4 value -2.796804
## iter   5 value -2.796858
## iter   6 value -2.796904
## iter   7 value -2.797054
## iter   8 value -2.797126
## iter   9 value -2.797164
## iter  10 value -2.797165
## iter  10 value -2.797165
## final  value -2.797165
## converged
```



```r
modelB3 <- sarima(lunemp, 0,1,5,0,1,1,12)
```

```
## initial  value -2.552840
## iter   2 value -2.754457
## iter   3 value -2.771161
## iter   4 value -2.781707
## iter   5 value -2.784181
## iter   6 value -2.784503
## iter   7 value -2.784510
## iter   8 value -2.784511
## iter   8 value -2.784511
## iter   8 value -2.784511
## final  value -2.784511
## converged
## initial  value -2.797231
## iter   2 value -2.799250
## iter   3 value -2.799446
## iter   4 value -2.799453
## iter   5 value -2.799454
## iter   5 value -2.799454
## iter   5 value -2.799454
## final  value -2.799454
## converged
```




```r
#ADF test
adf.test(modelA3$fit$residuals)
```

```
## Warning in adf.test(modelA3$fit$residuals): p-value smaller than printed p-
## value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  modelA3$fit$residuals
## Dickey-Fuller = -6.2129, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```r
adf.test(modelB3$fit$residuals)
```

```
## Warning in adf.test(modelB3$fit$residuals): p-value smaller than printed p-
## value
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  modelB3$fit$residuals
## Dickey-Fuller = -5.9288, Lag order = 7, p-value = 0.01
## alternative hypothesis: stationary
```

```r
#Redundancy check
summary(modelA3$fit)
```

```
##           Length Class  Mode
## coef        4    -none- numeric
## sigma2      1    -none- numeric
## var.coef   16    -none- numeric
## mask        4    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 372    ts     numeric
## call        8    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
## model      10    -none- list
```

```r
summary(modelB3$fit)
```

```
##           Length Class  Mode
## coef        6    -none- numeric
## sigma2      1    -none- numeric
## var.coef   36    -none- numeric
## mask        6    -none- logical
## loglik      1    -none- numeric
## aic         1    -none- numeric
## arma        7    -none- numeric
## residuals 372    ts     numeric
## call        8    -none- call
## series      1    -none- character
## code        1    -none- numeric
## n.cond      1    -none- numeric
## nobs        1    -none- numeric
```

```
## model      10    -none- list
```

```r
#BIC & AIC
BIC(modelA3$fit)
```

```
## [1] -960.15
```

```r
BIC(modelB3$fit)
```

```
## [1] -950.0266
```

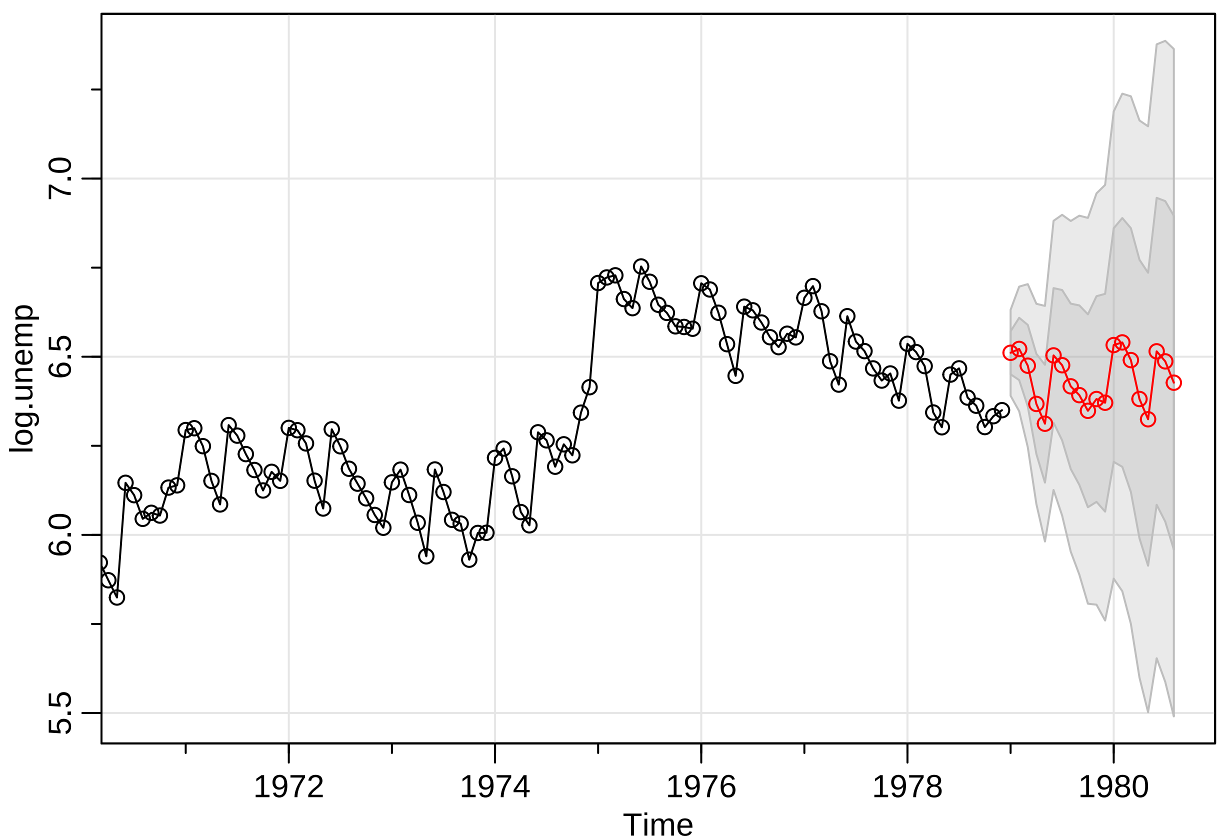
```r
AIC(modelA3$fit)
```

```
## [1] -979.5667
```

```r
AIC(modelB3$fit)
```

```
## [1] -977.2098
```

```r
#Forecasting
log.unemp <- lunemp
sarima.for(log.unemp, 1,1,2,0,1,1,12, n.ahead = 20)
```




```
## $pred
##           Jan      Feb      Mar      Apr      May      Jun      Jul
## 1979 6.511037 6.521885 6.474852 6.367511 6.312002 6.503737 6.476216
## 1980 6.532803 6.540151 6.490560 6.381349 6.324472 6.515208 6.486955
##           Aug      Sep      Oct      Nov      Dec
## 1979 6.416888 6.392187 6.348212 6.381297 6.370862
## 1980 6.427093
```

```
##
## $se
##             Jan        Feb        Mar        Apr        May        Jun
## 1979 0.06026437 0.08744082 0.11445234 0.14051447 0.16533249 0.18882900
## 1980 0.32785034 0.34907398 0.37016138 0.39087701 0.41109021 0.43073737
##             Jul        Aug        Sep        Oct        Nov        Dec
## 1979 0.21102999 0.23200981 0.25186274 0.27068806 0.28858231 0.30563552
## 1980 0.44979658 0.46827113
```

Analysis:-

We begin by plotting the time series together with ACf and PACF in the same manner as in 3a and we can see that the time series is not stationary and there seems to be a trend in the time series

Several time series of records with the corresponding acf and pacf can be seen here. The distinction is made with a 12-year period as we know that we have monthly US unemployment, so we use a seasonal difference to take care of this potential effect that can occur with the unemployment rate. The ACF is still non-stationary and we make another distinction in the next step this time with lag=1.

Based on the triangles discovered in the eacf, ARIMA(1, 1, 2)(0, 1, 1)12 and ARIMA(0, 1, 5)(0, 1, 1)12 are suggested.We can see that for the first model, the ACF appears to be good and stationary, which the adf test can verify. The Q plots appear to be relatively normal, and the root unit test shows that the test is stationary.Model results (1, 1, 2)(0,1, 1)12 show very similar diagnostic plots. From the AIC and BIC, we can see that the result is almost the same, so we will choose the first model as our final model, which is (1, 1, 2)(0, 1, 1)12. We can see that residuals are autonomous for a few lags for both ljung box exams,it seems to be more independent in the (1, 1, 2)(0, 1, 1)12.