Ejercicio Time Series - Clase 2

Daniel Ferreira Zanchetta and Lais Silva Almeida Zanchetta

31/05/2020

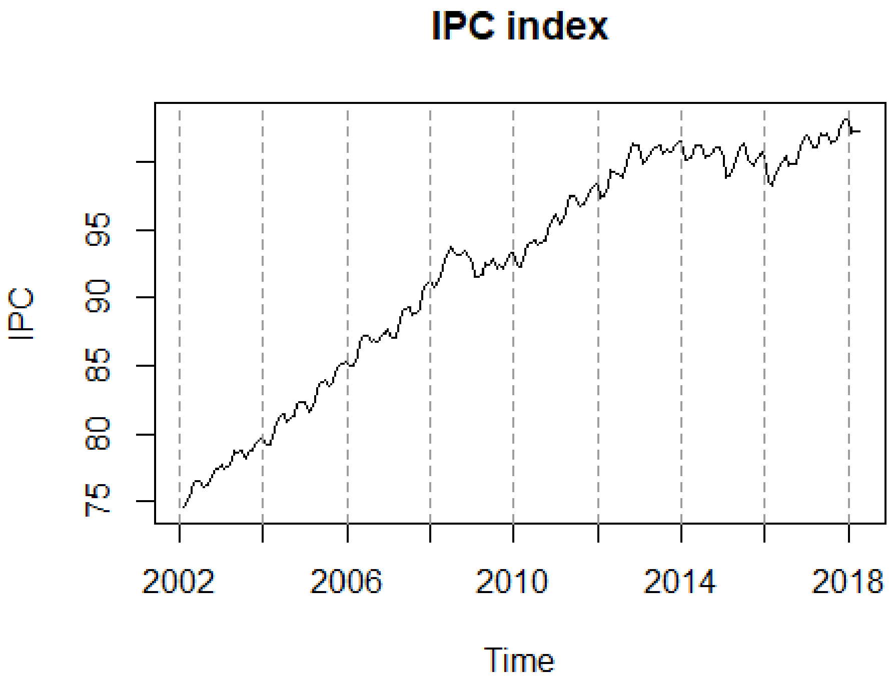
```
#setwd("/Users/laisalmeida/Desktop/Master/Data Analytics/Aula 22 - 08-
05/Time_Series_Datasets/")
setwd("C:/Users/Daniel/Documents/Certificados & Faculdade/UPC Master Big
Data/Data Analytics/Time Series/Sesion 08/05/20/Time_Series_Datasets/")
#Reading and formatting the file
goods <- read.csv2(file = "INE_IPC.csv", sep = ";", header = TRUE, nrows
= 13, skip = 6)
```

*Exercise 1) Read the IPC data as you did in the last homework. Then compute the Inflation as Inflation <- 100*diff(IPC,lag=12)/lag(IPC,k=-12).*

```
cols <- colnames(goods)
end_block_IPCindex <- min(grep(".1",cols,fixed = TRUE))-1
for ( i in 2:end_block_IPCindex){
    cols[i] <-  paste0(cols[i],".0")
}
cols[782] <- sub(cols[782],"X.1","X.4")
```

IPC Index

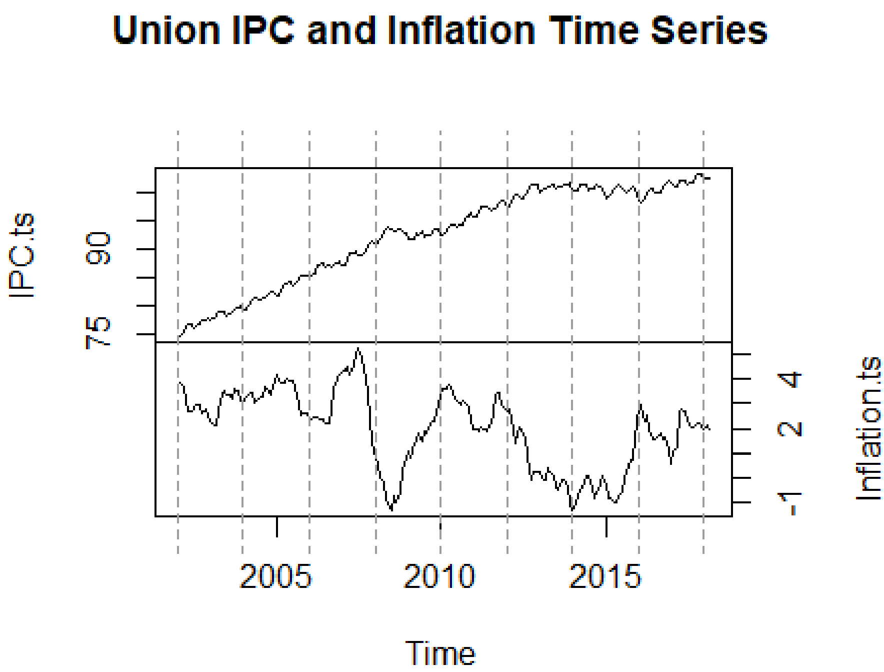
```
end <- max(grep(".0",cols,fixed = TRUE))
cols.0 <- cols[end:2]
# IPC is only taking into consideration the "Indice geral" row
IPC <- as.numeric(
  t(read.csv2(file = "INE_IPC.csv", skip=7,
              dec = "." , header = FALSE,
              encoding = "UTF-8", row.names = 1,
              nrows=1)[1,(end-1):1])
)
year <- as.numeric(substr(cols.0,start=2,stop=5))
month <- as.numeric(substr(cols.0,start=7,stop=8))
tt <- year + month/12
IPC.ts <- ts(IPC, frequency = 12, start=c(year[1],month[1]))
plot(tt,IPC,type="l",main="IPC index", xlab="Time",lab=c(8,5,7))
abline(v=seq(2002,2018,by=2),lty=2,col=8)
```




```
#Computing the inflation
Inflation <- 100*diff(IPC,lag=12)/lag(IPC,k=-12)

## Warning in 100 * diff(IPC, lag = 12)/lag(IPC, k = -12): comprimento do
objeto
## maior não é múltiplo do comprimento do objeto menor
```

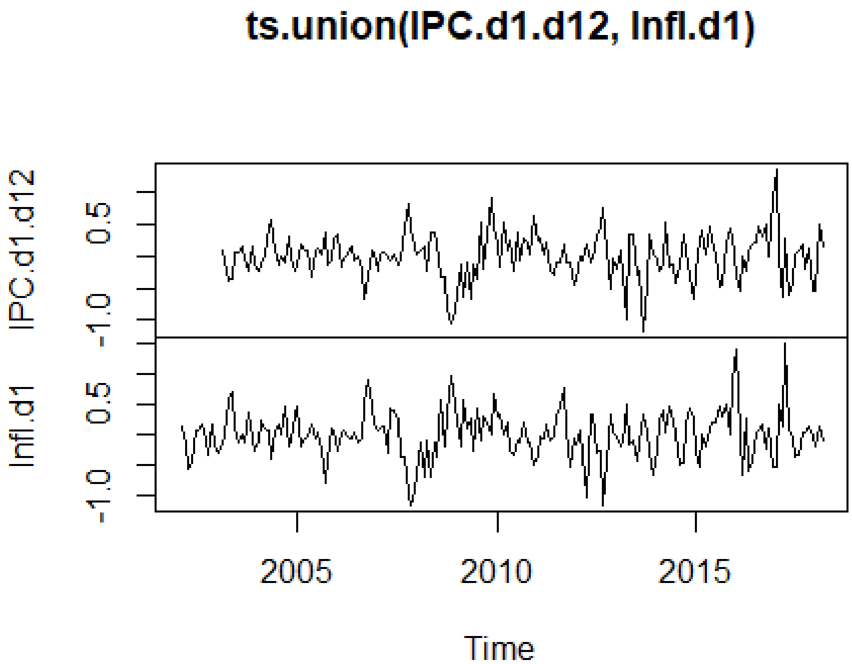
Exercise 2) Join the two previous time series with ts.union and plot the resulting bivariate series.

```
Inflation.ts <- ts(Inflation, start=c(year[1], month[1]), frequency = 12)

tsunion <- ts.union(IPC.ts,Inflation.ts)
plot(tsunion,main="Union IPC and Inflation Time Series",yax.flip=TRUE)
abline(v=seq(2002,2018,by=2),lty=2,col=8)
```



Exercise 3) Define IPC.d1.d12 as the IPC series after having taken difference of orders 1 (regular difference) and 12 (seasonal difference). Define Infl.d1 as the regular difference of Inflation. Check that both series are very similar.

```
IPCindexdiff12.ts <- diff(IPC.ts, lag=12)
IPC.d1.d12 <- diff(IPCindexdiff12.ts, lag=1)
Infl.d1 <- diff(Inflation.ts,lag = 1)

plot(ts.union(IPC.d1.d12, Infl.d1))
```



Exercise 4) Using the function window, cut the time series IPC in two parts, a training part until December 2016, and a test part from January 2017. Call them IPC.tr and IPC.te, respectively.

```
IPC.tr <- window(IPC.ts,start=c(year[1],month[1]),end = c(2016,12))
IPC.te <- window(IPC.ts,start=c(2017,1))
```

Exercise 5) Compute the regular and seasonal difference of IPC.tr and call the resulting series IPC.tr.d1.d12.

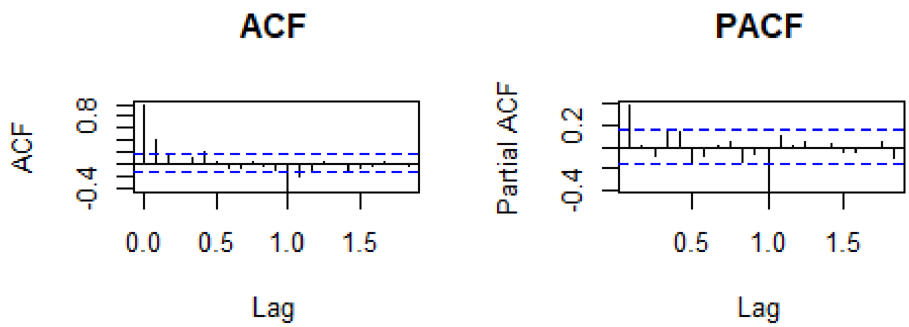
a. Plot this series, as well as its ACF and its PACF.

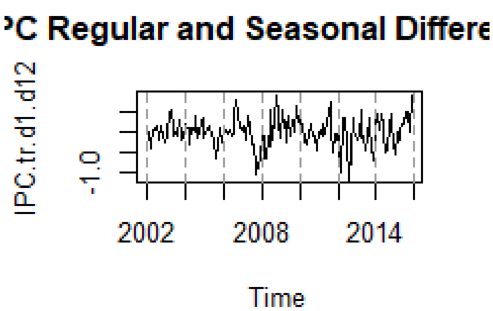
b. Do you think IPC.tr.d1.d12 is white noise?

Reply: At first glance they look like white noise, however we **do not** think it can be seen as white noise, because the ACF and PACF plots show correlations between the data.

```
IPCtrdiff12.ts <- ts(diff(IPC.tr, lag=12), frequency = 12,
start=c(year[1], month[1]))
IPC.tr.d1.d12 <- ts(diff(IPCtrdiff12.ts, lag=1), frequency = 12,
start=c(year[1], month[1]))

op<-par(mfrow=c(2,2))
acf(IPC.tr.d1.d12, main = "ACF")
pacf(IPC.tr.d1.d12, main = "PACF")
plot.ts(IPC.tr.d1.d12,type="l",main="IPC Regular and Seasonal
```

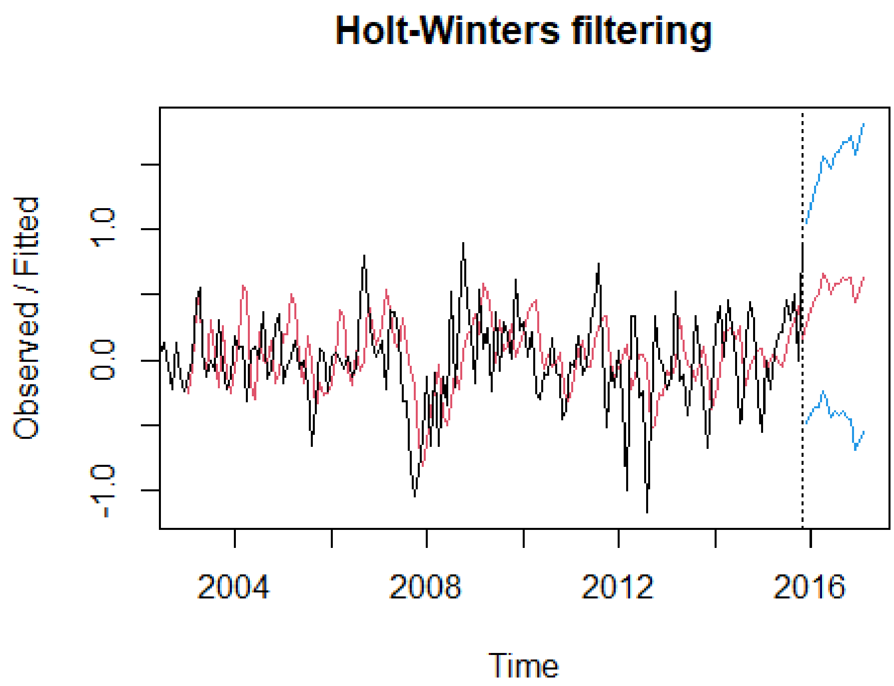
```
Differences", xlab="Time")
abline(v=seq(2002,2018,by=2),lty=2,col=8)
par(op)
```





Exercise 6) Use the functions HoltWinters (with the default parameters) and predict.HoltWinters to predict the next 15 values of IPC.tr.d1.d12 (these are the forecasting of the values corresponding to the period from January 2017 to March 2018). Plot the forecasted object.

```
library(astsa)
IPC.HW <- HoltWinters(IPC.tr.d1.d12)
pred.HW <- predict(IPC.HW, n.ahead=15, prediction.interval = TRUE)

#library(ggplot2)
#library(forecast)
plot(IPC.HW, predicted.values = pred.HW)
```

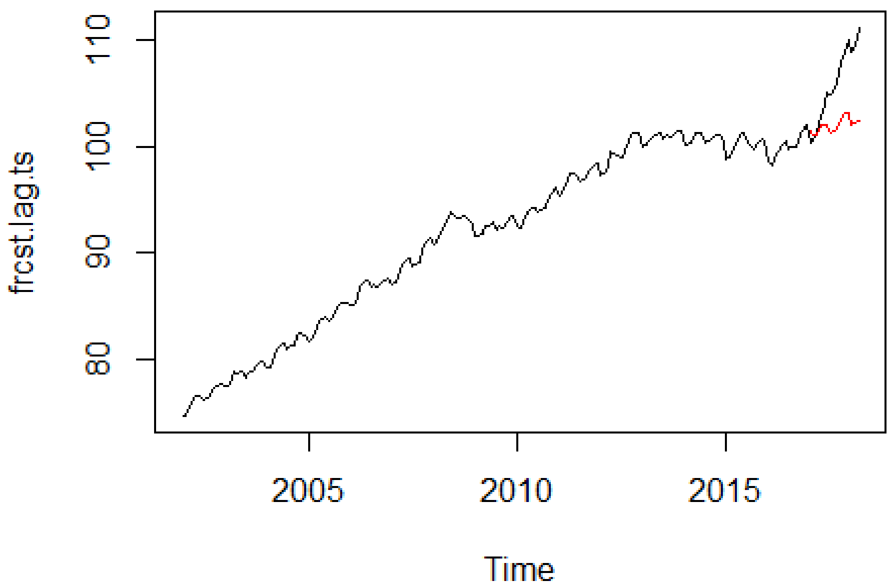


Exercise 7) Use the expression Xt = Xt−1 +Xt−12 −Xt−13 + at to compute the predictions for the IPC j steps ahead, XT+j|T , j = 1, . . . , 15. Compare these predictions with the values of the test values inIPC.te. Take into account the following indications:

- You can use the predicted values for IPC.tr.d1.d12 obtained before as aT+j .

- Once you have computed XT+j|T , you can use this value as an estimation of XT+j when computing XT+h|T for h > j.

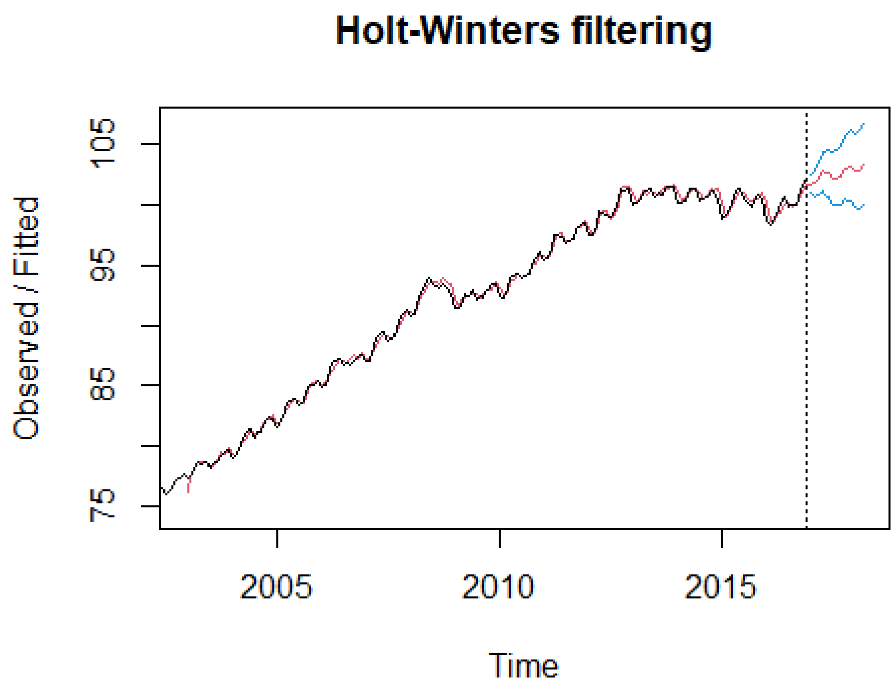
```
frcst.lag = IPC.tr[1:180]

#Para los intervalor de test
for (i in ((length(IPC.tr)+1):(length(IPC.tr)+length(IPC.te)))){
  #Xt = Xt−1 +Xt−12 −Xt−13 + at
  frcst.lag[i] <- frcst.lag[i-1]+frcst.lag[i-12]-frcst.lag[i-
13]+pred.HW[i-180]
  }
frcst.lag.ts <- ts(frcst.lag, frequency = 12, start=c(year[1], month[1]))

plot(frcst.lag.ts)
lines(IPC.te,col="red")
```



Exercise 8) Use the functions HoltWinters (with the default parameters) and predict.HoltWinters to predict the next 15 values of IPC.tr (these are the forecasting of the values corresponding to the period from January 2017 to March 2018).

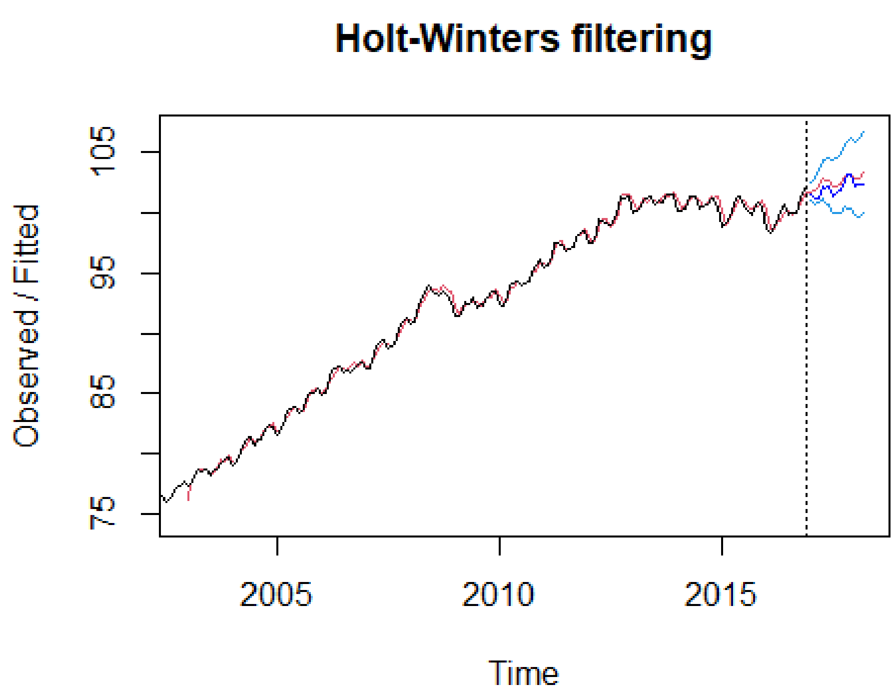
- a. Plot the forecasted object.

```
IPC.tr.HW <- HoltWinters(IPC.tr)
pred.IPC.tr.HW <- predict(IPC.tr.HW,n.ahead=15,prediction.interval=TRUE)

plot(IPC.tr.HW, predicted.values = pred.IPC.tr.HW)
```



- b. Compare these predictions with the values of the test values in IPC.te.

```
plot(IPC.tr.HW, predicted.values = pred.IPC.tr.HW)
lines(IPC.te,col="blue")
```



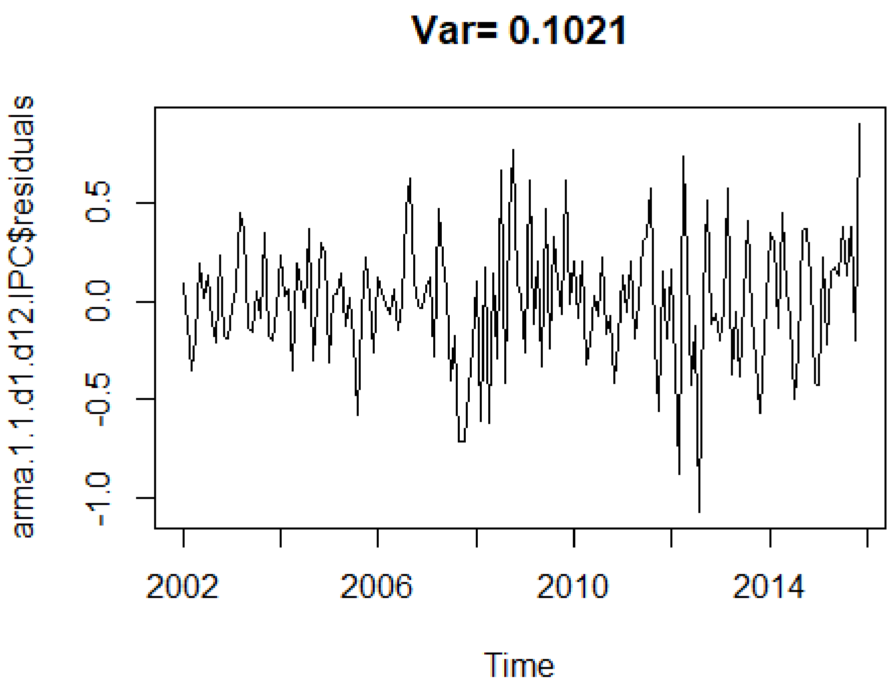
c. (Optional) Which forecasting method is preferable in this case, the one used here or that used in the previous exercise?

Reply: In our experience with both ways to calculate the forecast so far, we prefer the method used here.

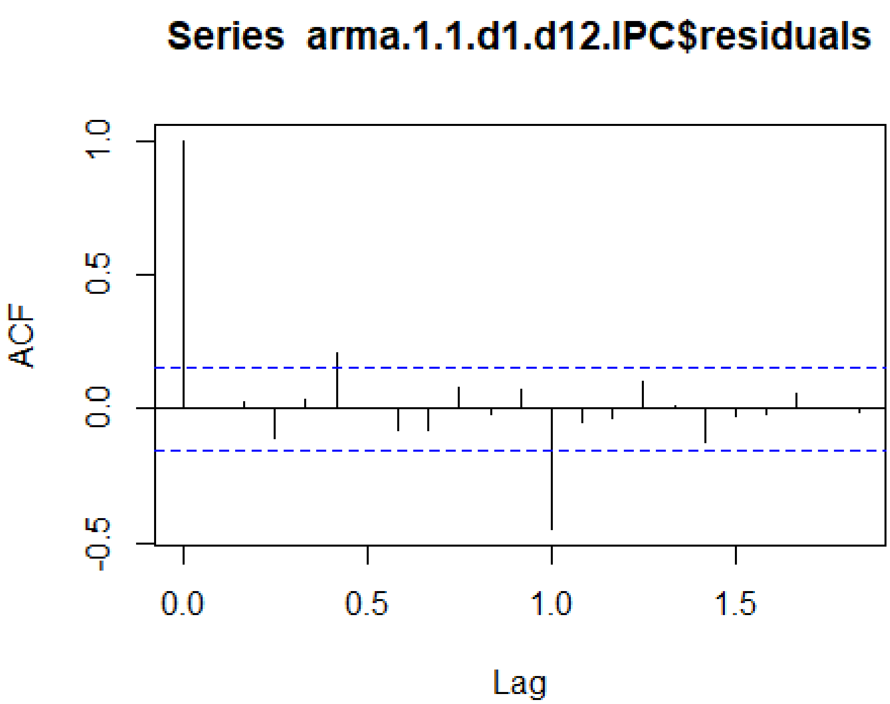
Exercise 9) Based on the ACF and the PACF, propose at least two different ARMA models for IPC.tr.d1.d12.

- a.Estimate the ARMA models you propose. Plot the residuals of them and the residuals ACF and PACF.

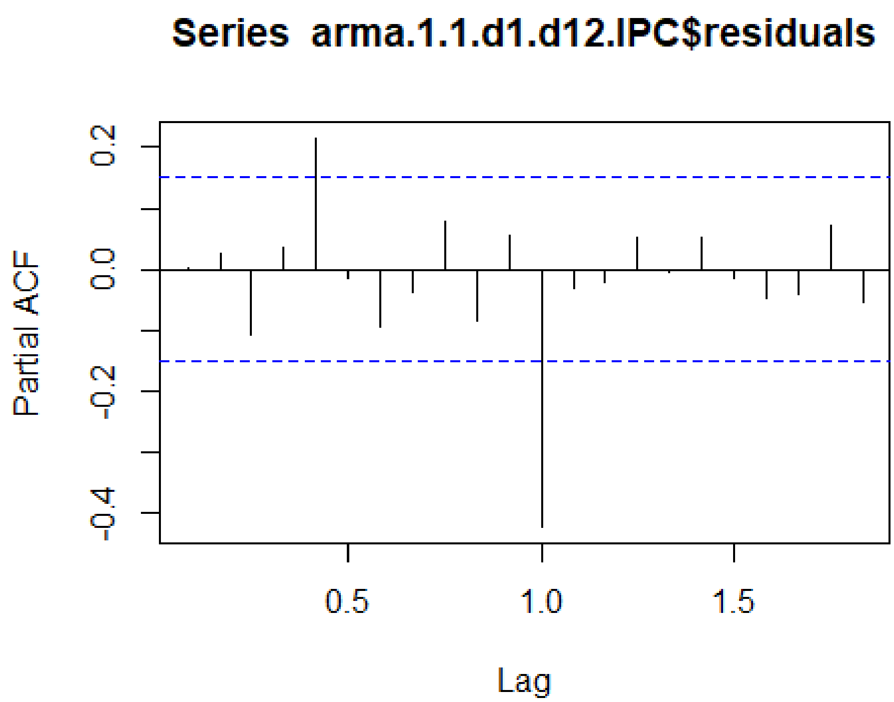
```
#par(mfrow=c(1,2))
#acf(IPC.tr.d1.d12)
#pacf(IPC.tr.d1.d12)

#Model 1) ARMA(1,1)
library(forecast)

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

##
## Attaching package: 'forecast'

## The following object is masked from 'package:astsa':
##
##     gas

arma.1.1.d1.d12.IPC <- Arima(IPC.tr.d1.d12,order=c(1,0,1))

plot(arma.1.1.d1.d12.IPC$residuals,main=paste("Var=",round(var(arma.1.1.d
1.d12.IPC$residuals),4)))
```



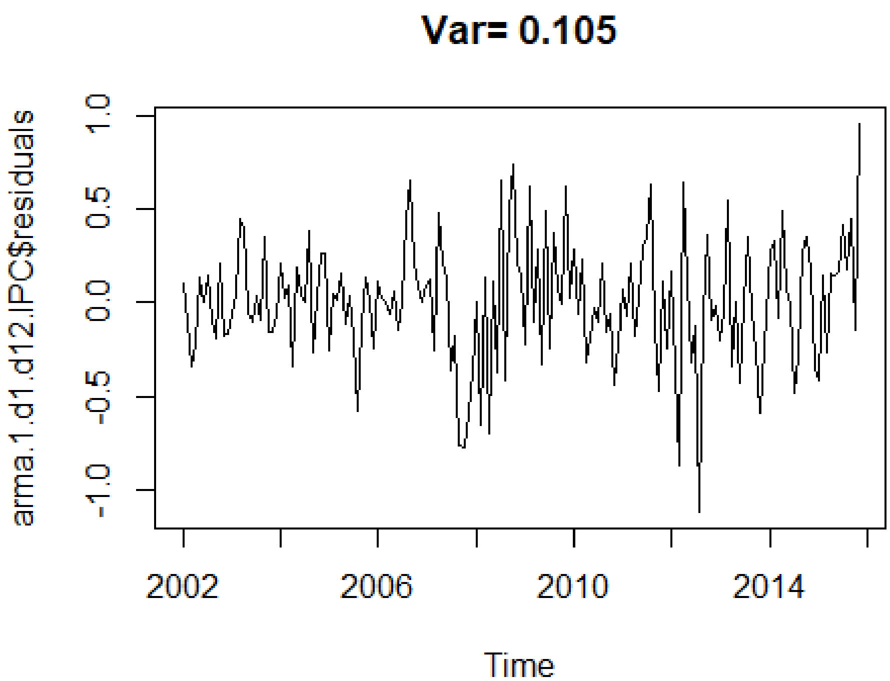

```
acf(arma.1.1.d1.d12.IPC$residuals)
```



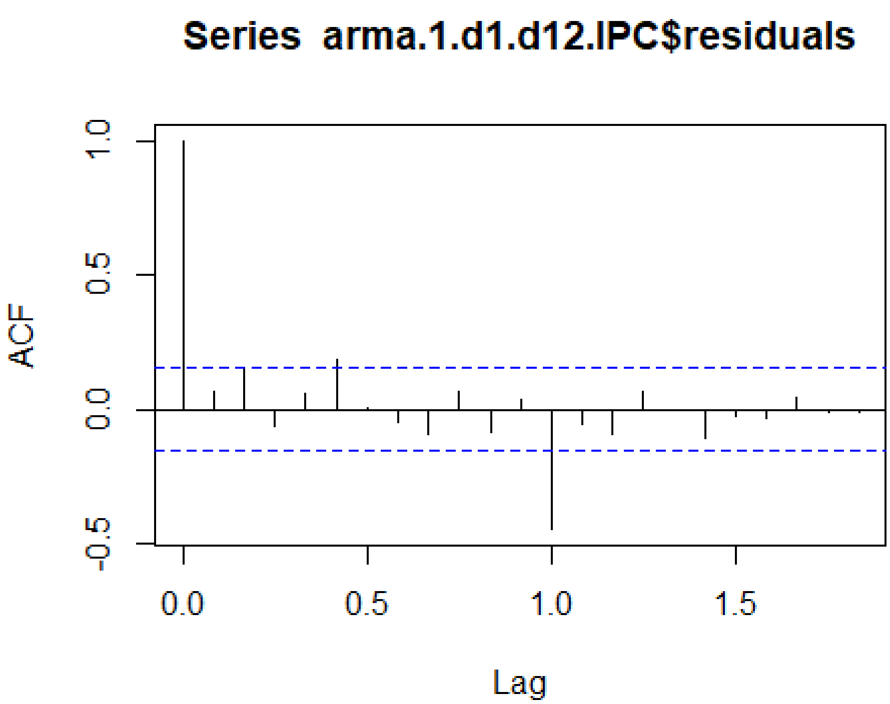

```
pacf(arma.1.1.d1.d12.IPC$residuals)
```



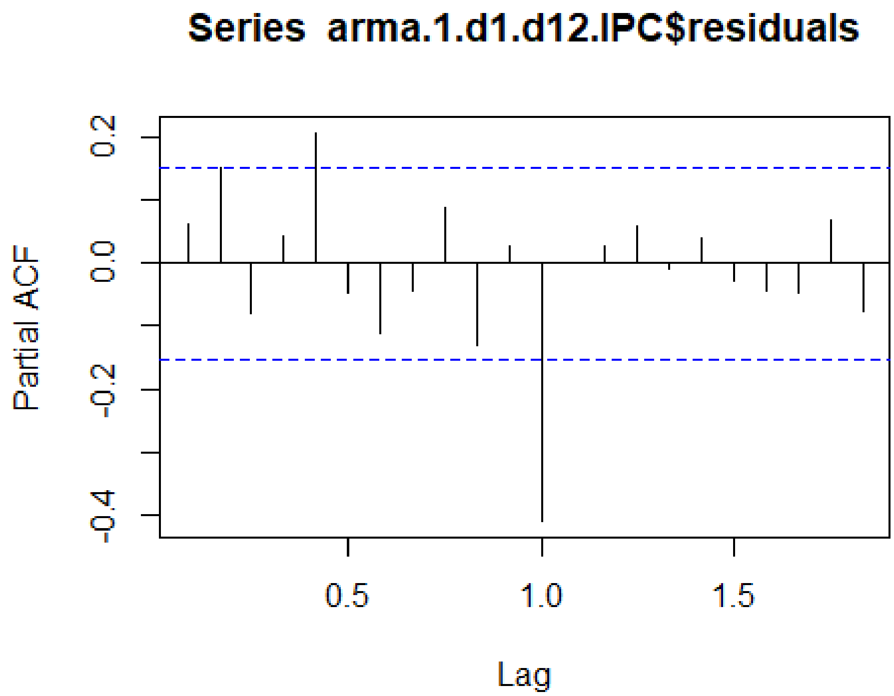

```
#Model 2) ARMA(0,1)
library(forecast)
arma.1.d1.d12.IPC <- Arima(IPC.tr.d1.d12,order=c(0,0,1))

plot(arma.1.d1.d12.IPC$residuals,main=paste("Var=",round(var(arma.1.d1.d1
2.IPC$residuals),4)))
```




```r
acf(arma.1.d1.d12.IPC$residuals)
```

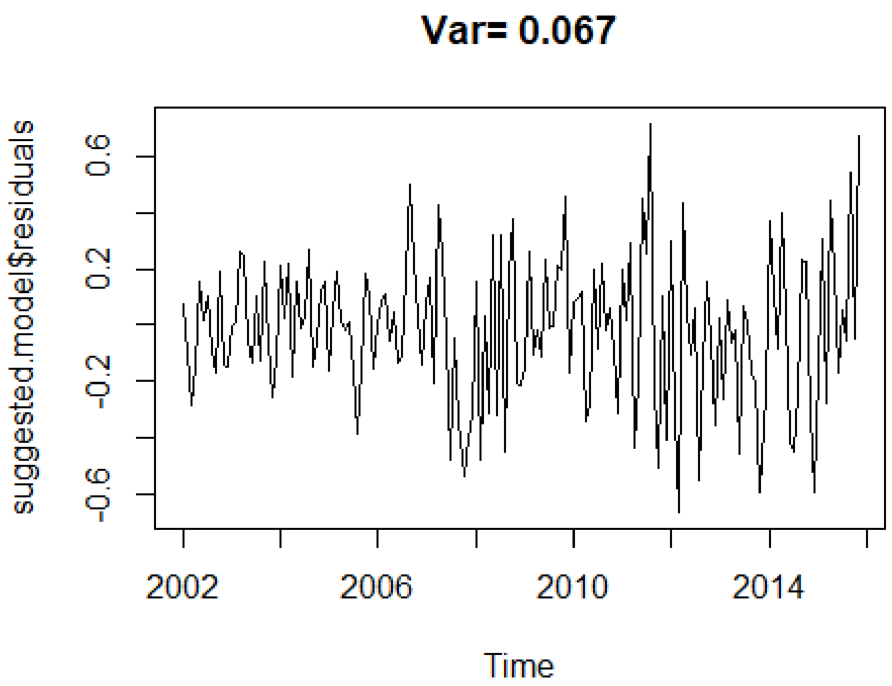



```r
pacf(arma.1.d1.d12.IPC$residuals)
```

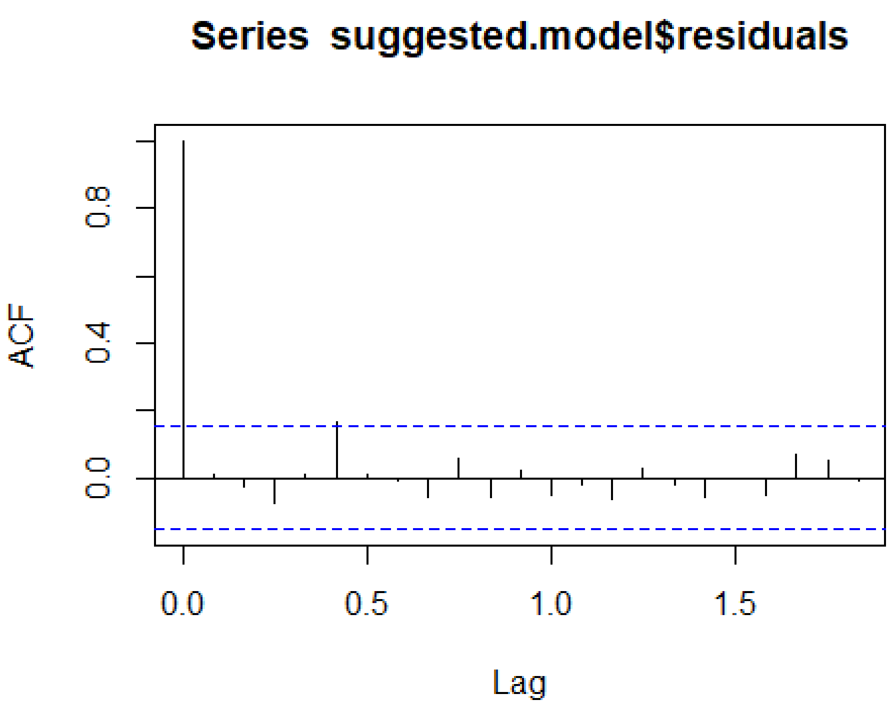


b. Which is the model suggested by auto.arima? Plot the residuals of them and the residuals ACF and PACF.
Reply: ARIMA(1,0,0)(0,0,1)[12] with zero mean

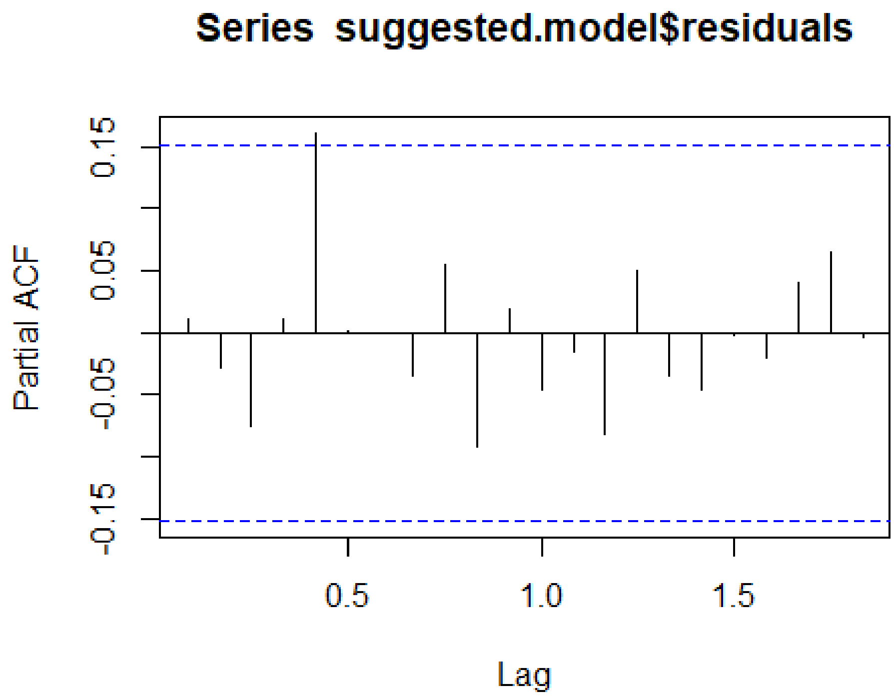
```
suggested.model <- auto.arima(IPC.tr.d1.d12)
print(suggested.model)

## Series: IPC.tr.d1.d12
## ARIMA(1,0,0)(0,0,1)[12] with zero mean
##
## Coefficients:
##          ar1     sma1
##       0.4134  -0.7176
## s.e.  0.0719   0.0665
##
## sigma^2 estimated as 0.06771:  log likelihood=-15.57
## AIC=37.13   AICc=37.28   BIC=46.49

plot(suggested.model$residuals,main=paste("Var=",round(var(suggested.mode
l$residuals),4)))
```




```
acf(suggested.model$residuals)
```




```
pacf(suggested.model$residuals)
```



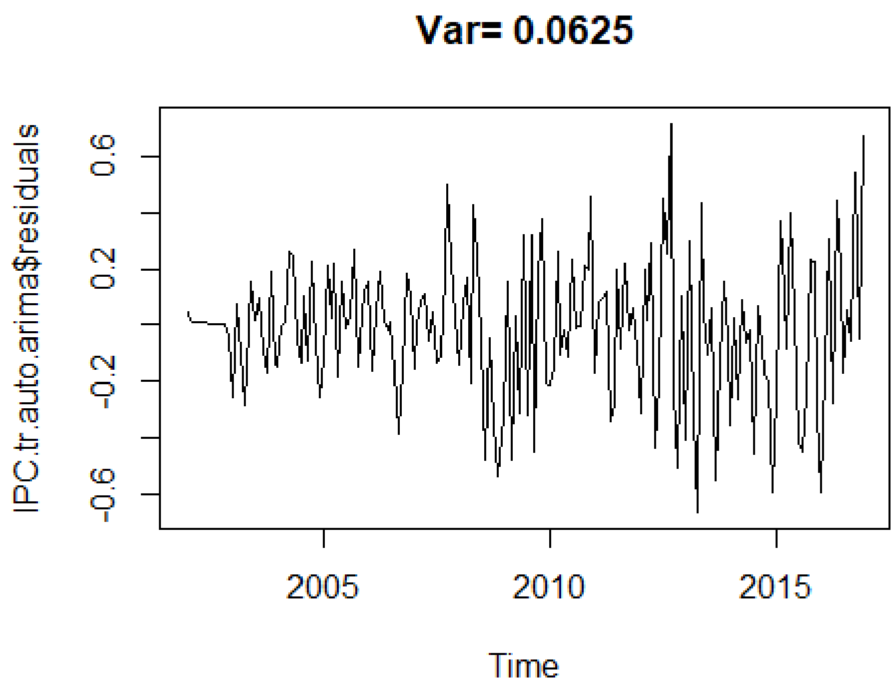
c. Which ARMA model do you chose finally for IPC.tr.d1.d12?

Reply: As per the results seen we would choose the model provided by auto.arima function. The values seem more adjusted, and even when comparing ACF and PACF, the lags are within the boundaries.

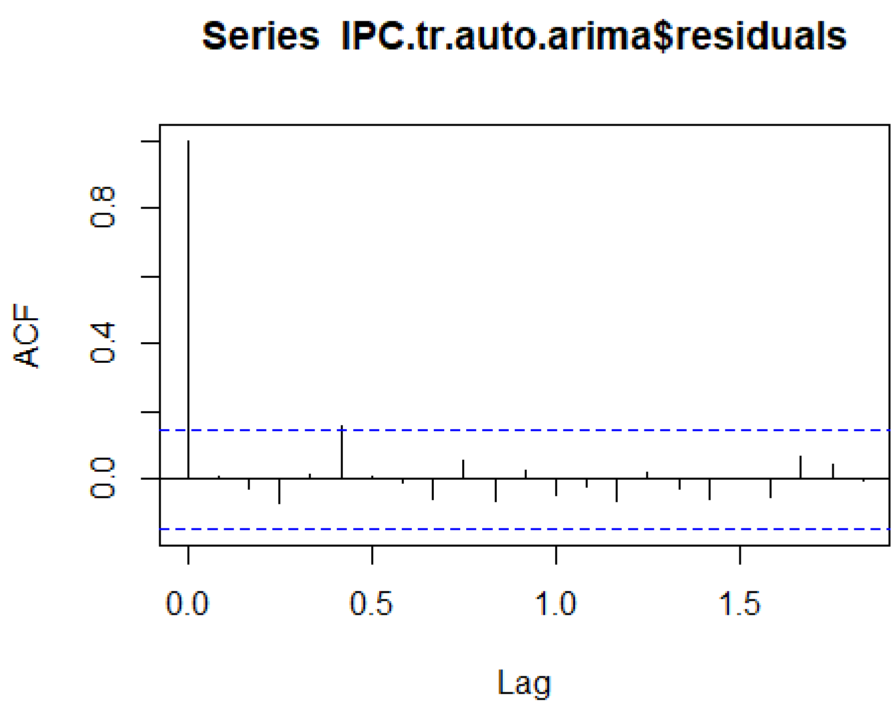
Exercise 10) Which is the model suggested by auto.arima for the time series IPC.tr? Plot the residuals of this model and the residuals ACF and PACF.

Reply: ARIMA(1,1,0)(0,1,1)[12]

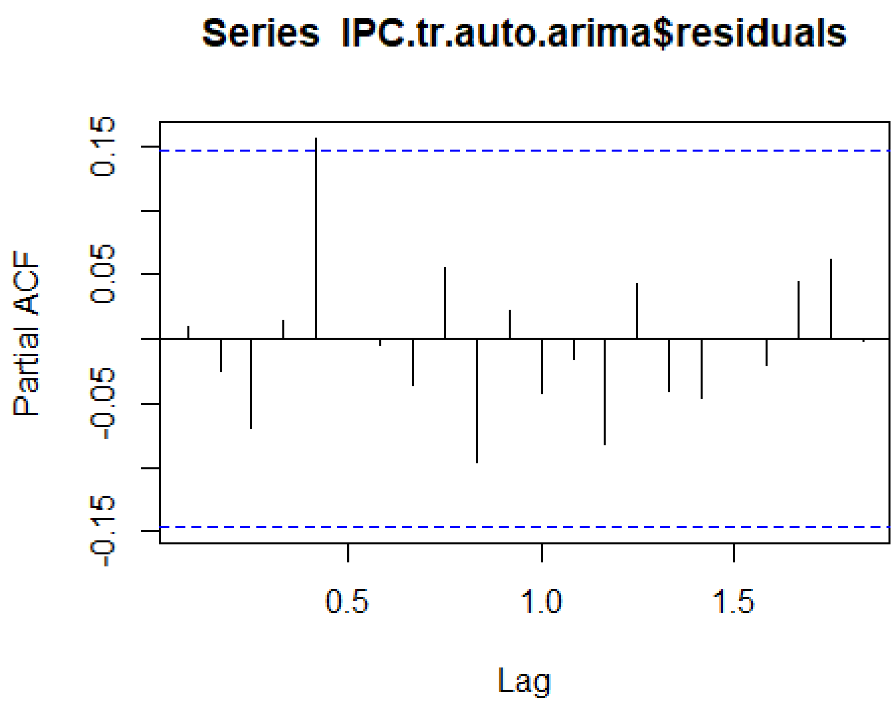
```
IPC.tr.auto.arima <- auto.arima(IPC.tr)
print(IPC.tr.auto.arima)

## Series: IPC.tr
## ARIMA(1,1,0)(0,1,1)[12]
##
## Coefficients:
##          ar1     sma1
##       0.4134  -0.7175
## s.e.  0.0719   0.0665
##
## sigma^2 estimated as 0.06814:  log likelihood=-15.57
## AIC=37.14   AICc=37.29   BIC=46.5

plot(IPC.tr.auto.arima$residuals,main=paste("Var=",round(var(IPC.tr.auto.
arima$residuals),4)))
```




```r
acf(IPC.tr.auto.arima$residuals)
```

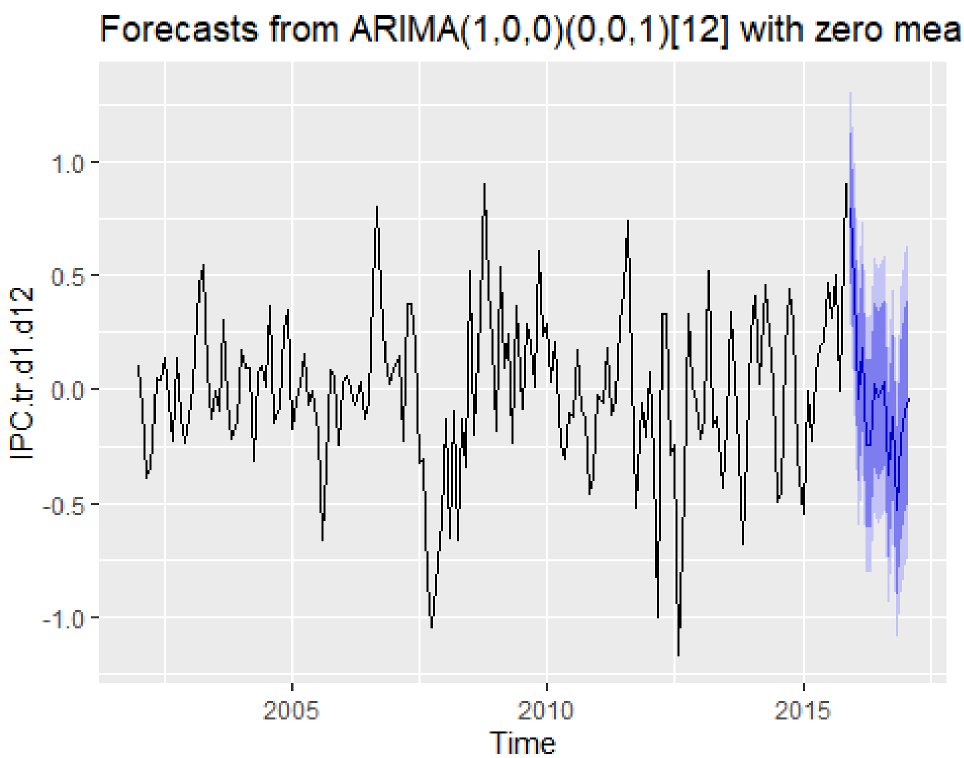



```r
pacf(IPC.tr.auto.arima$residuals)
```

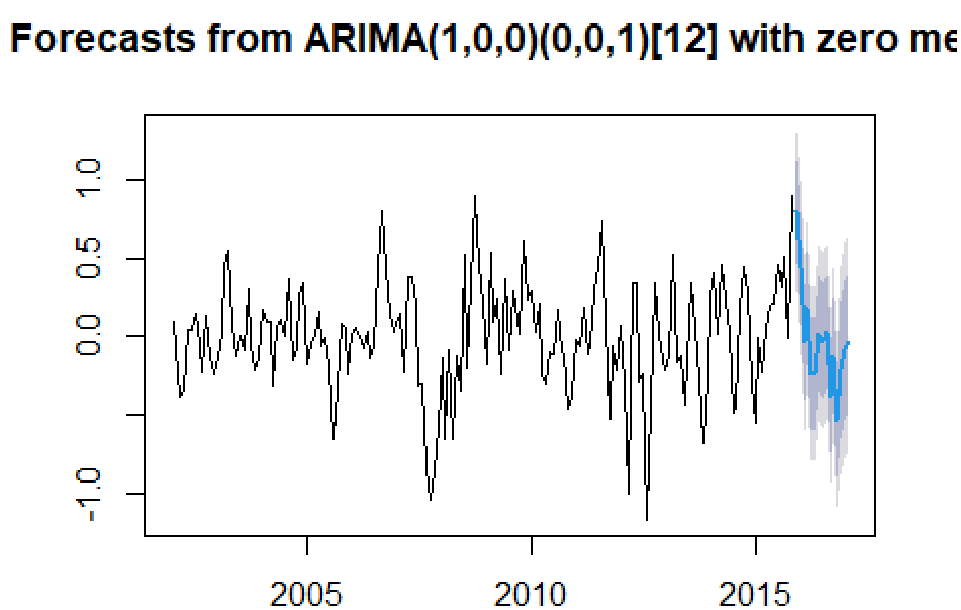


Exercise 11) Consider the ARMA model suggested and estimated by auto.arima for the time series IPC.tr.d1.d12.Use the function forecast from library forecast to predict the next 15 values of IPC.tr.d1.d12 (these are the forecasting of the values corresponding to the period from January 2017 to March 2018).Plot the forecasted object using plot and autoplot.

```
library(forecast)
library(ggplot2)
frcst.suggested.model <- forecast(suggested.model,h=15)
autoplot(frcst.suggested.model)
```



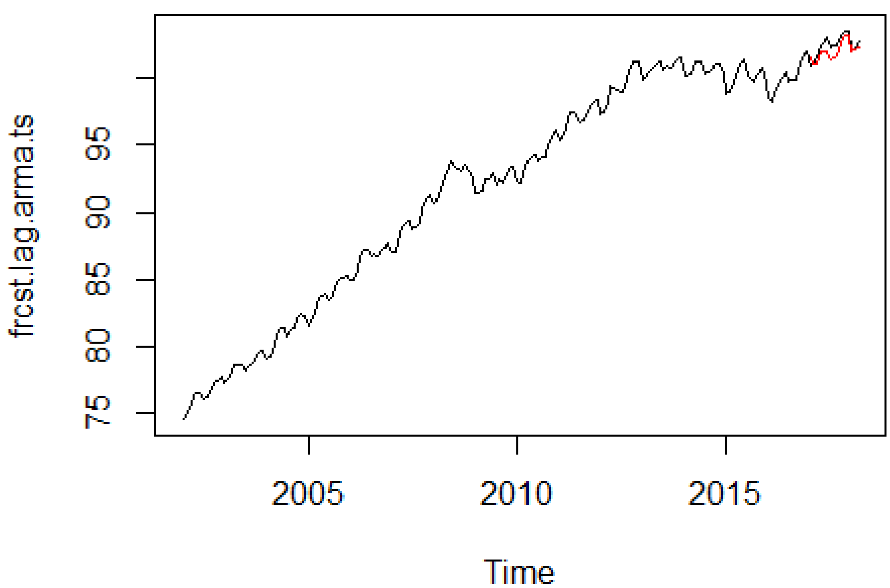

```
plot(frcst.suggested.model)
```



Exercise 12) Let Xt be the time series IPC.tr. Use the expression Xt = Xt−1 + Xt−12 − Xt−13 + at to compute the predictions for the IPC j steps ahead, XT+j|T , j = 1, . . . , 15. Compare these predictions with the values of of the test values in IPC.te. Take into account the following indications:

- You can use the predicted values for IPC.tr.d1.d12 obtained from the previous ARMA model.

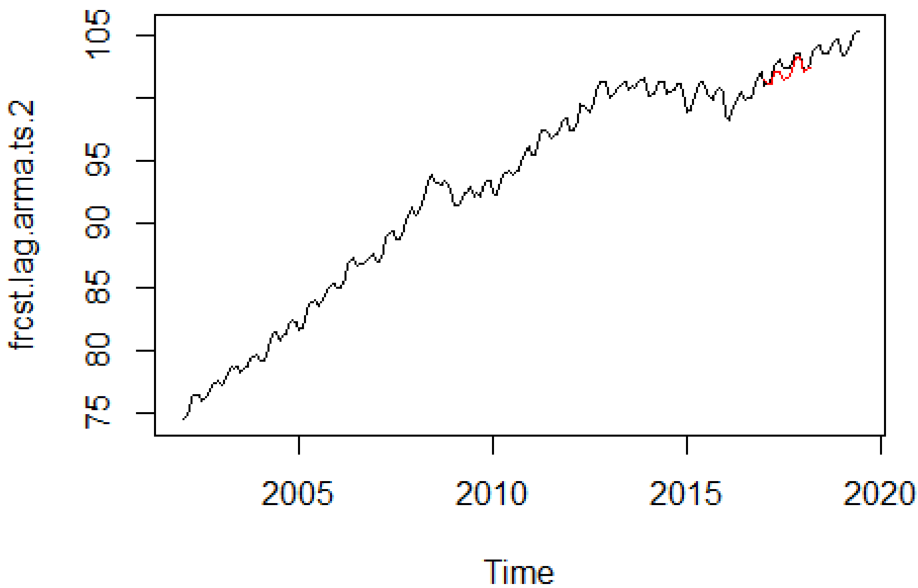
```
frcst.IPC.tr.arma = IPC.tr[1:180]

for (i in ((length(IPC.tr)+1):(length(IPC.tr)+length(IPC.te)))){
  frcst.IPC.tr.arma[i] <- frcst.IPC.tr.arma[i-1] + frcst.IPC.tr.arma[i-
12] - frcst.IPC.tr.arma[i-13] + frcst.suggested.model$mean[i-180]
}
frcst.lag.arma.ts <- ts(frcst.IPC.tr.arma, frequency = 12,
start=c(year[1], month[1]))

plot(frcst.lag.arma.ts)
lines(IPC.te,col="red")
```



- Once you have computed XT+j|T , you can use this value as an estimation of XT+j when computing XT+h|T for h > j.

```
###
h = length(frcst.IPC.tr.arma)+15
frcst.IPC.tr.arma.2 = frcst.lag.arma.ts[1:195]
```

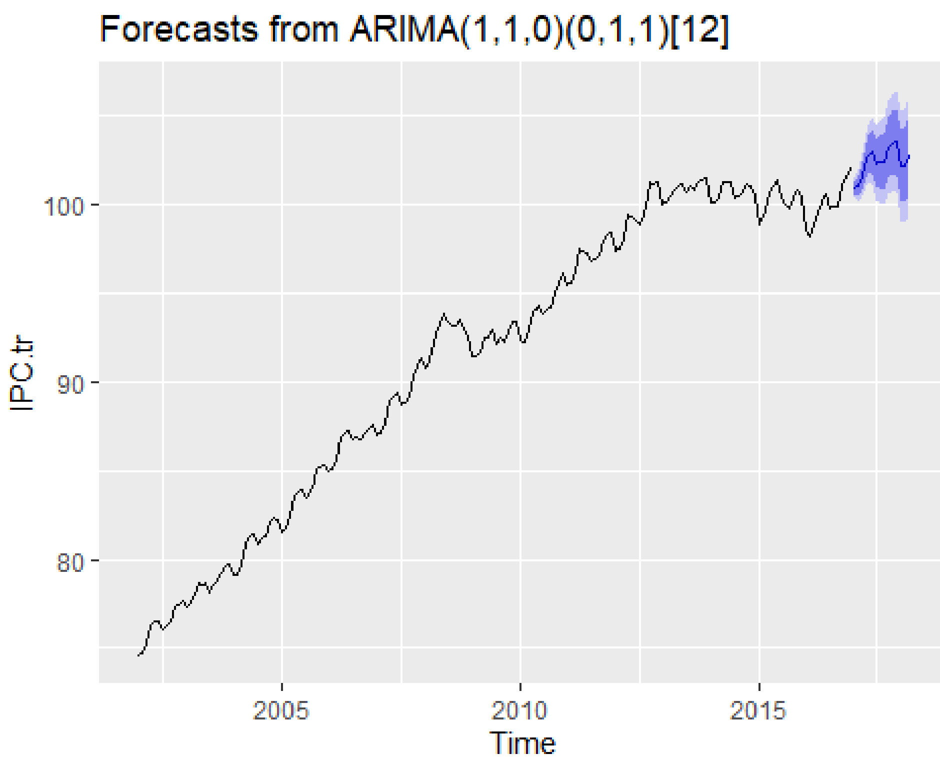
```
for (i in ((length(frcst.IPC.tr.arma.2)+1):h)){
  frcst.IPC.tr.arma.2[i] <- frcst.IPC.tr.arma.2[i-1] +
frcst.IPC.tr.arma.2[i-12] - frcst.IPC.tr.arma.2[i-13]
}
frcst.lag.arma.ts.2 <- ts(frcst.IPC.tr.arma.2, frequency = 12,
start=c(year[1], month[1]))

plot(frcst.lag.arma.ts.2)
lines(IPC.te,col="red")
```



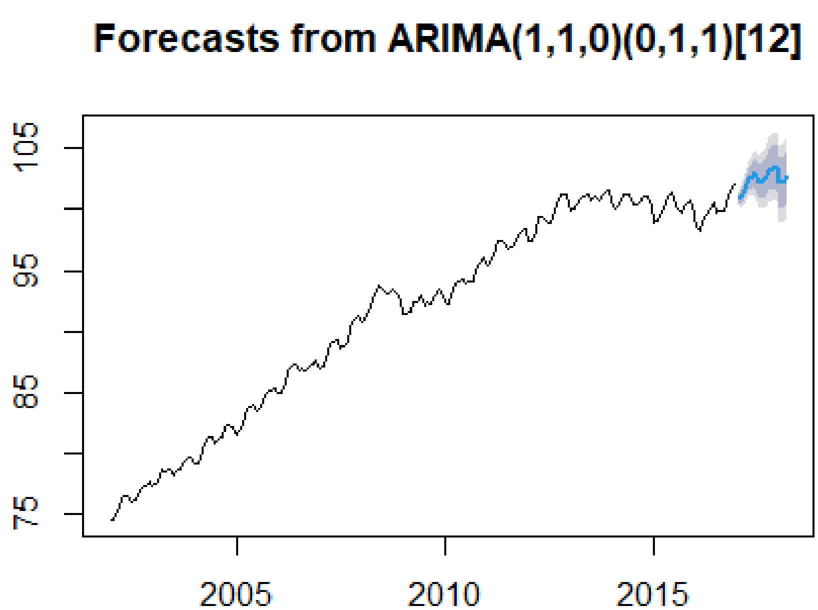
Exercise 13) Consider the ARIMA model suggested and estimated by auto.arima for the time series IPC.tr. Use the function forecast from library forecast to predict the next 15 values of IPC.tr (these are the forecasting of the values corresponding to the period from January 2017 to March 2018).

- a. Plot the forecasted object using plot and autoplot.

```
frcst.IPC.tr.auto.arima <- forecast(IPC.tr.auto.arima,h=15)
autoplot(frcst.IPC.tr.auto.arima)
```

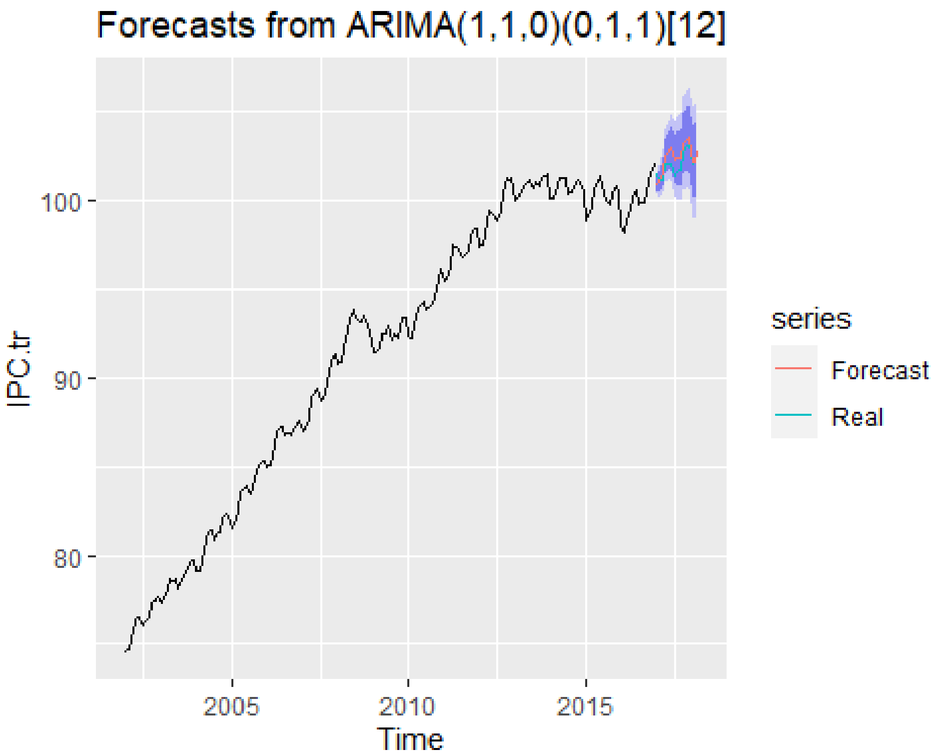



```
plot(frcst.IPC.tr.auto.arima)
```



b. Compare these predictions with the values of the test values in IPC.te.

```r
autoplot(frcst.IPC.tr.auto.arima) +
  autolayer(IPC.te, series="Real") +
  autolayer(frcst.IPC.tr.auto.arima$mean, series="Forecast")
```

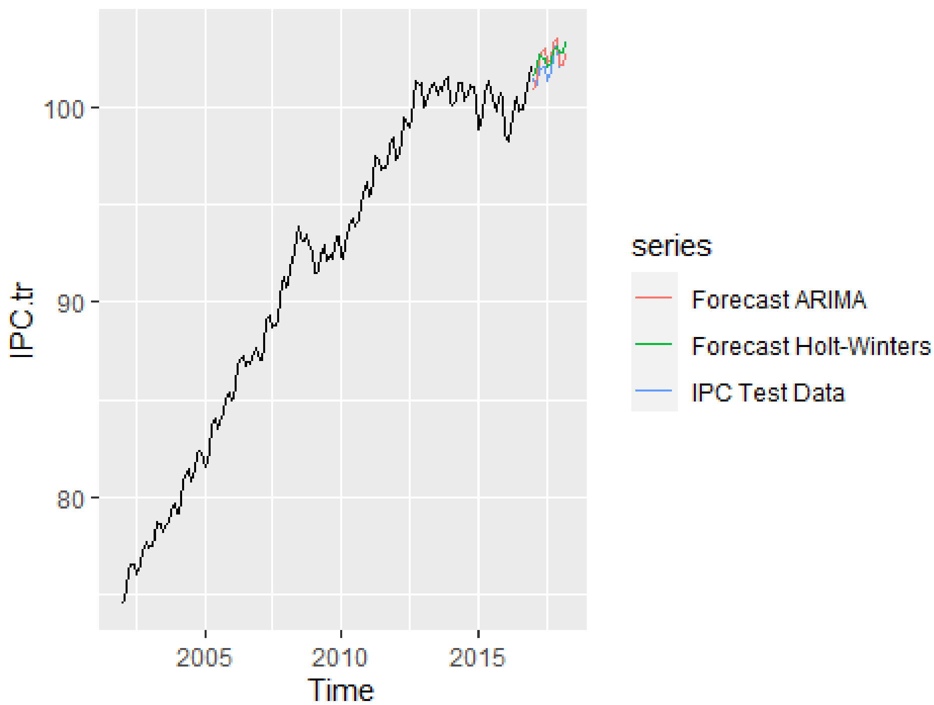


c. (Optional) Do they coincide the forecasting done here and that done in the previous exercise?

Reply: They coincide a lot. We've observed a minimum difference in the predicted values, which are totally irrelevant tough.

Exercise 14) Compare the predictions obtained by Holt-Winters and by the ARIMA model with the true values of IPC (IPC.te).

```r
autoplot(IPC.tr) +
  autolayer(IPC.te, series="IPC Test Data") +
  autolayer(frcst.IPC.tr.auto.arima$mean, series="Forecast ARIMA") +
  autolayer(pred.IPC.tr.HW[,1], series = "Forecast Holt-Winters")
```



Reply: We thought it would be a good way to compare the two forecats with the real test data by plotting it. At first glance, we may see that apparently the prediction by the Holt-Winters model gets closer to the IPC Test data.