Empirical Methods HW 2

Group 9: Linqi Huang, Abhesh Kumar, Yu Onohara, Maitrayee Patil, Redmond Xia

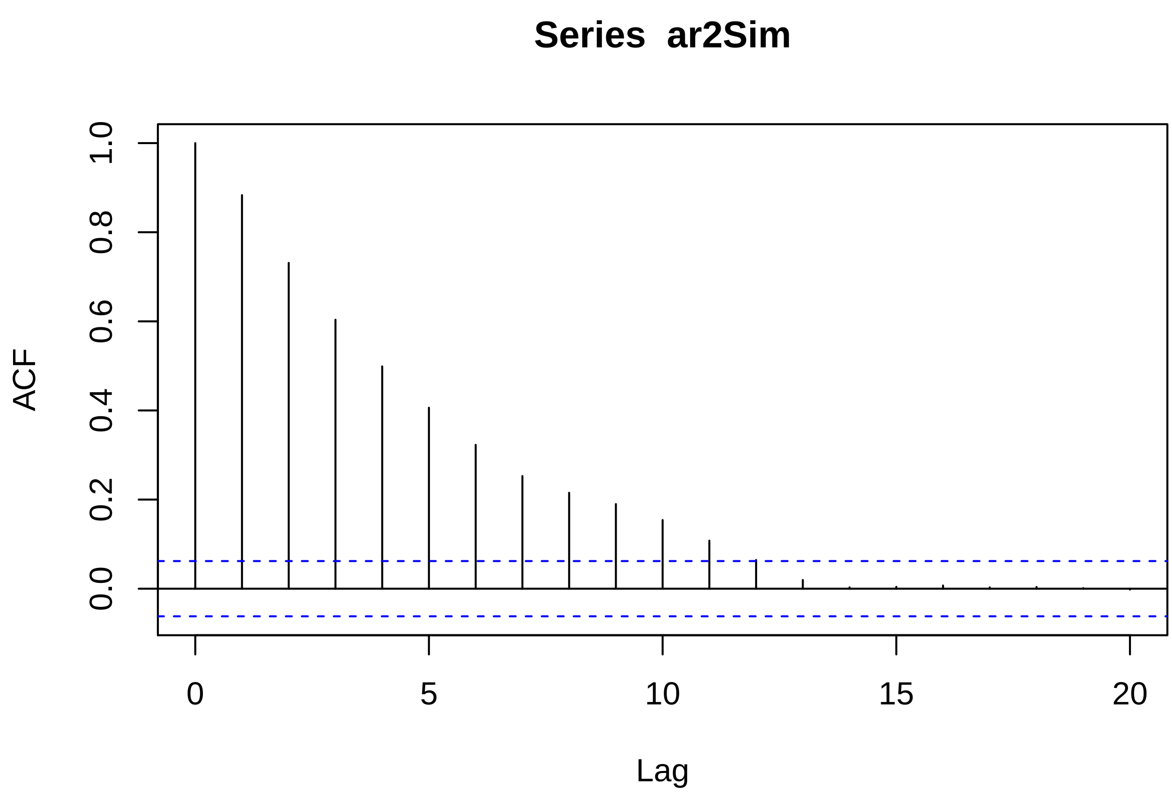
January 19, 2020

Question 1

3a

```r
phi1 <- 1.1
phi2 <- -0.25

ar2Sim <- arima.sim(model = list(ar=c(phi1,phi2)), n = 1000)
acf(ar2Sim,lag.max = 20)
```



#3b

```r
x1 = (phi1 + sqrt(phi1^2 + 4*phi2))/(-2 * phi2)
x2 = (phi1 - sqrt(phi1^2 + 4*phi2))/(-2 * phi2)
w1 <- 1/x1
w2 <- 1/x2
cat("The roots are ",w1," ",w2,'\n')
```

```
## The roots are  0.3208712   0.7791288
```

Thus, the two roots are real and less than one in modulus. So it is a stationary process

3c

The dynamic multiplier for this series is $\frac{\partial X_t}{\partial \varepsilon} = \phi_1^6 + 5\phi_1^4\phi_2 + 6\phi_1^2\phi_2^2 + \phi_2^3$

```
dMultiplier1 <- phi1^6 + 5 * phi1^4 *phi2 + 6 * phi1^2 * phi2^2 + phi2^3
cat("The multiplier is ",dMultiplier1,'\n')
```

```
## The multiplier is  0.379561
```

3d

```
phi1 <- 0.9
phi2 <- 0.8
dMultiplier2 <- phi1^6 + 5 * phi1^4 *phi2 + 6 * phi1^2 * phi2^2 + phi2^3
cat("The multiplier is: ", dMultiplier2,'\n')
```

```
## The multiplier is:  6.778241
```

```
x1 = (phi1 + sqrt(phi1^2 + 4*phi2))/(-2 * phi2)
x2 = (phi1 - sqrt(phi1^2 + 4*phi2))/(-2 * phi2)
w1 <- 1/x1
w2 <- 1/x2
cat("The roots are: ",w1," ",w2,'\n')
```

```
## The roots are:  -0.5512492   1.451249
```

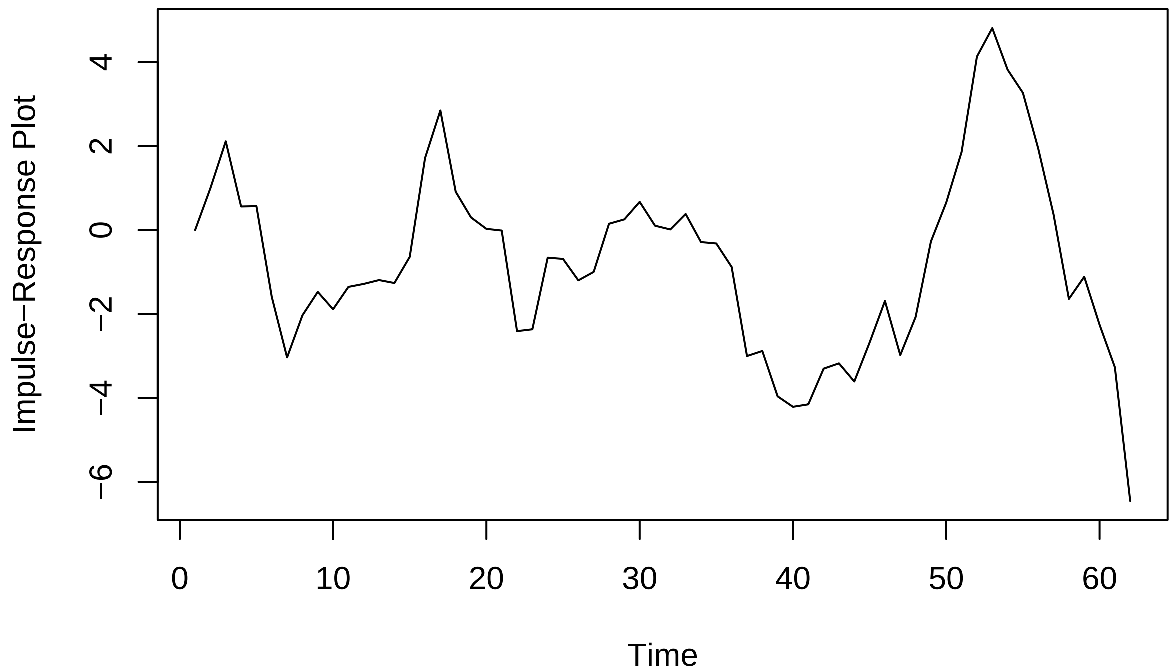
The root is now bigger than one in modulus, thus this process is not stationary. It allows for greater shock and it is not mean reverting.

3e

```
phi1 <- 1.1
phi2 <- -0.25

x <- numeric(63)
x[1] <- 0
x[2] <- 0
x[3] <- x[2] * phi1 + x[1] * phi2 + 1 #epsilon at t is one
for (i in 4:63) {
  x[i] <- x[i-1]*phi1 + x[i-2]*phi2 + rnorm(1,0,1)
}

plot(x[-1],type = "l", xlab = "Time", ylab = "Impulse-Response Plot")
```



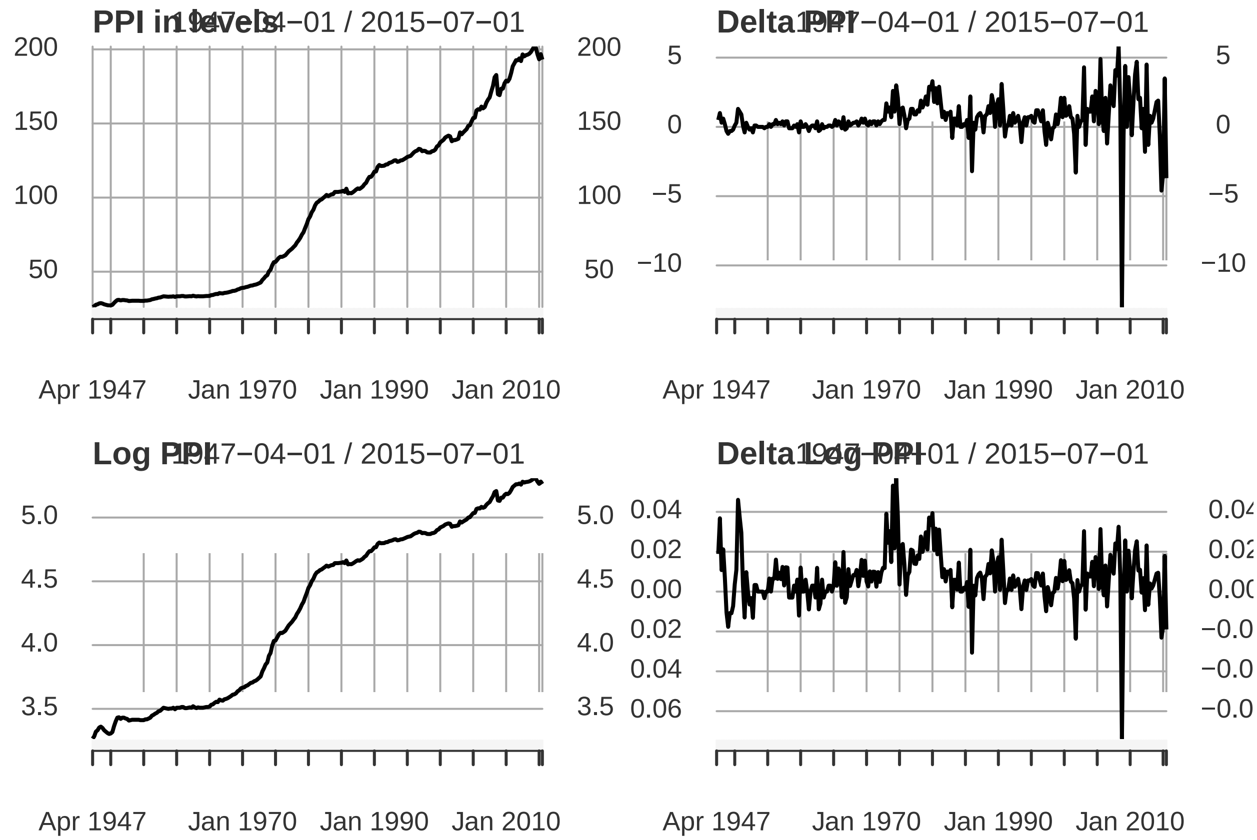
Question 2

Reading in the data

```
suppressMessages(suppressWarnings(library(readxl)))
suppressMessages(suppressWarnings(library(forecast)))
suppressMessages(suppressWarnings(library(xts)))
PPI <- read_xls("PPIFGS.xls")
```

1a,b,c,d

```
PPI_xts <- xts(x = as.double(PPI$VALUE),as.Date(PPI$DATE))
par(mfrow= c(2,2))
plot(PPI_xts, main = "PPI in levels")
plot(diff(PPI_xts), main = "Delta PPI")
plot(log(PPI_xts), main = "Log PPI")
plot(diff(log(PPI_xts)), main = "Delta Log PPI")
```

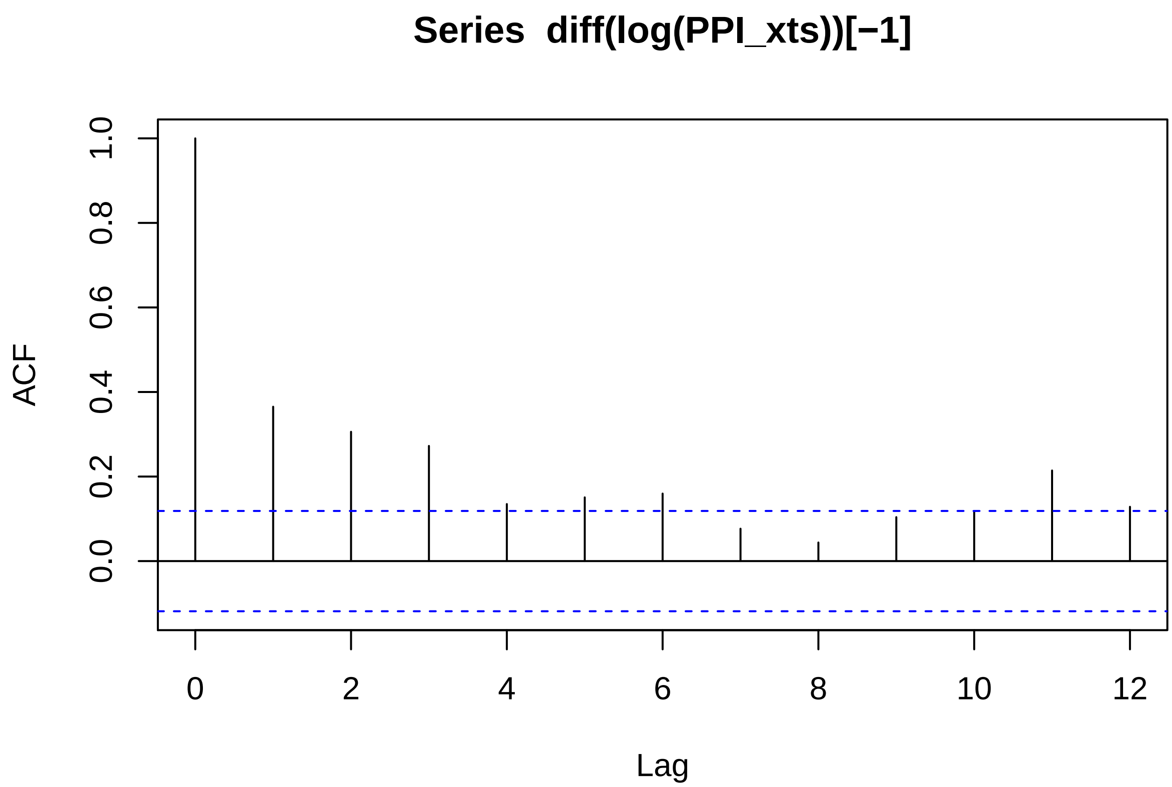


2

The series 1a,c are most likely not covariance stationary because the mean changes over time. Between 1b and 1d, d is more likely to be covariance stationary because its volatility is more stable or constant choose $y_t = f(PPI_t)$

3

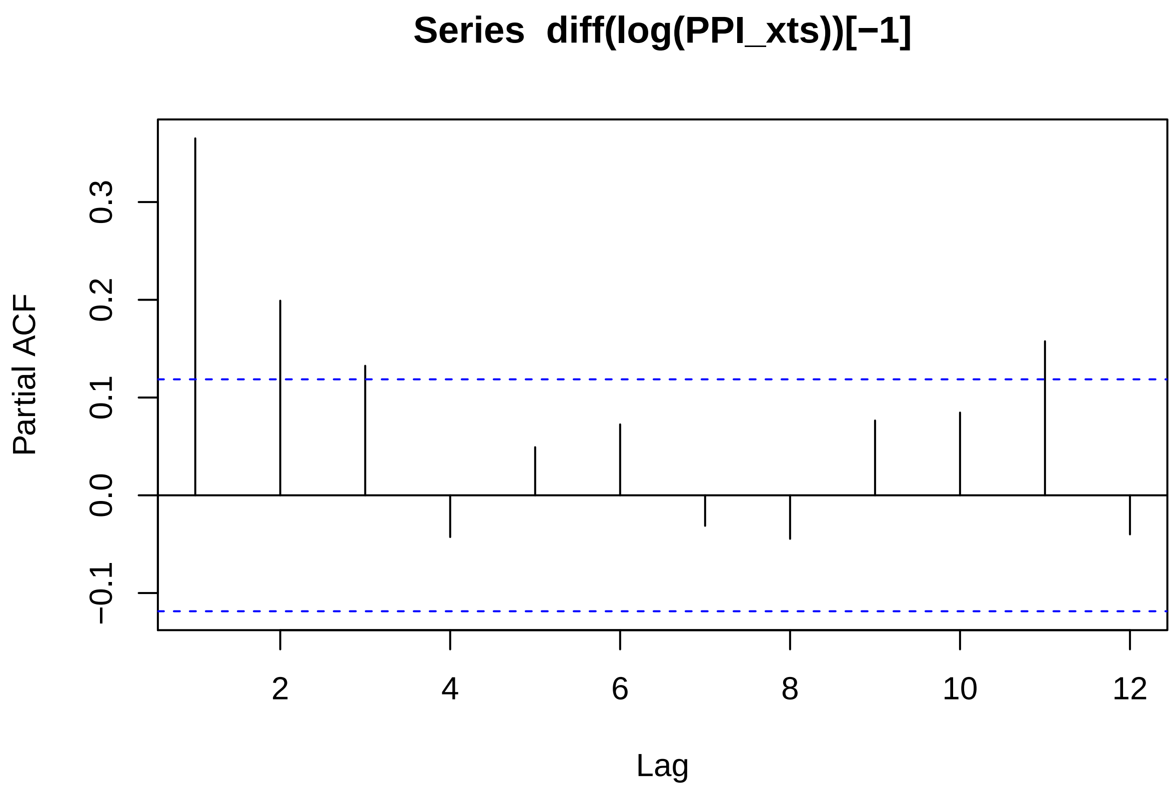
```
par(mfrow = c(1,1))
acf(diff(log(PPI_xts))[-1], lag.max = 12)
```



We notice that this plot starts to converge slowly after a year. This means there might be some seasonality or short term memory. We conclude a short term ar process may be useful.

4

```r
pacf(diff(log(PPI_xts))[-1], lag.max = 12)
```



We conclude that there is still significant lags for the first two lags and for the 11th. A short term lag may be useful and includes lag 11, which is the outlier.

5a

We will fit the ar(3) and ar(1,2,3,11) model based on the acf graph. Lastly, we check what ar is selected by the AIC criterion for up to 20 lags.

```
ar_3 <- arima(diff(log(PPI_xts)),order = c(3,0,0))
ar_3
```

```
##
## Call:
## arima(x = diff(log(PPI_xts)), order = c(3, 0, 0))
##
## Coefficients:
##          ar1     ar2     ar3  intercept
##       0.2687  0.1607  0.1402     0.0073
## s.e.  0.0602  0.0628  0.0614     0.0016
##
## sigma^2 estimated as 0.0001389:  log likelihood = 824.87,  aic = -1639.75
```

```
cat("s.e. for the model", sqrt(ar_3$sigma2),'\n')
```

```
## s.e. for the model 0.0117843
```

```
ar_12311 <- arima(diff(log(PPI_xts)),order = c(11,0,0),
                                        fixed = c(NA,NA,NA,0,0,0,0,0,0,0,NA,NA),
```

```r
                                          transform.pars = FALSE)
ar_12311
```

```
##
## Call:
## arima(x = diff(log(PPI_xts)), order = c(11, 0, 0), transform.pars = FALSE, fixed = c(NA,
##     NA, NA, 0, 0, 0, 0, 0, 0, 0, NA, NA))
##
## Coefficients:
##          ar1     ar2     ar3  ar4  ar5  ar6  ar7  ar8  ar9  ar10    ar11
##       0.2523  0.1469  0.1396    0    0    0    0    0    0     0  0.1746
## s.e.  0.0594  0.0618  0.0602    0    0    0    0    0    0     0  0.0563
##       intercept
##          0.0070
## s.e.     0.0024
##
## sigma^2 estimated as 0.000134:  log likelihood = 829.57,  aic = -1647.15
```

```r
cat("s.e. for the model", sqrt(ar_12311$sigma2),'\n')
```

```
## s.e. for the model 0.01157443
```

```r
ar_20 <- ar(diff(log(PPI_xts))[-1], order.max = 20, aic = TRUE)
ar_20
```

```
##
## Call:
## ar(x = diff(log(PPI_xts))[-1], aic = TRUE, order.max = 20)
##
## Coefficients:
##      1       2       3
## 0.2661  0.1603  0.1325
##
## Order selected 3  sigma^2 estimated as  0.0001417
```

```r
# ar(3) is used

ar_3Roots <- polyroot(c(1,-ar_3$coef[1:3]))
ar_12311Roots <- polyroot(c(1,-ar_12311$coef[1:11]))
cat("Modulus for ar3: ", Mod(1/ar_3Roots),'\n')
```

```
## Modulus for ar3:  0.740977 0.4349771 0.4349771
```

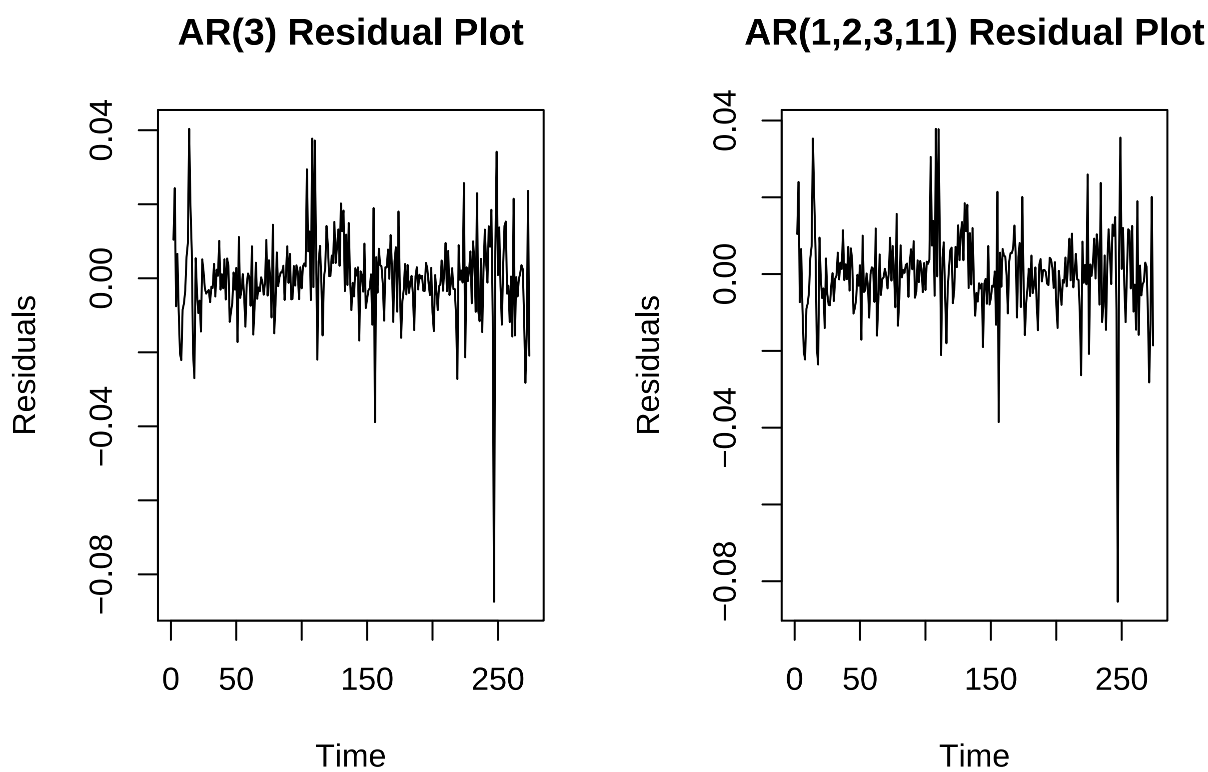
```r
cat("Modulus for ar(1,2,3,11): \n", Mod(1/ar_12311Roots),'\n')
```

```
## Modulus for ar(1,2,3,11):
##  0.8336421 0.8502911 0.8502911 0.8688112 0.8428416 0.8428416 0.9304615 0.8688112 0.8346636 0.8336421
```

Since all of the characteristic roots are smaller than 1, we conclude these models are stationary

5b

```r
par(mfrow = c(1,2))
plot(ar_3$residuals,ylab = "Residuals", main = "AR(3) Residual Plot")
plot(ar_12311$residuals, ylab = "Residuals", main = "AR(1,2,3,11) Residual Plot")
```



5c

```r
Box.test(ar_12311$residuals, lag = 8, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  ar_12311$residuals
## X-squared = 4.6288, df = 8, p-value = 0.7964
```

```r
Box.test(ar_3$residuals, lag = 8, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  ar_3$residuals
## X-squared = 5.346, df = 8, p-value = 0.72
```

```r
Box.test(ar_12311$residuals, lag = 12, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  ar_12311$residuals
## X-squared = 4.7277, df = 12, p-value = 0.9665
```

```
Box.test(ar_3$residuals, lag = 12, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  ar_3$residuals
## X-squared = 13.829, df = 12, p-value = 0.3118
```

```
cat("The AIC for AR(3): ", ar_3$aic, ". The BIC for AR(3): ", BIC(ar_3),
    '\n',"The AIC for AR(1,2,3,11): ", ar_12311$aic, ". The BIC for AR(1,2,3,11): ", BIC(ar_12311),'\n')
```

```
## The AIC for AR(3):  -1639.746 . The BIC for AR(3):  -1621.699
##  The AIC for AR(1,2,3,11):  -1647.148 . The BIC for AR(1,2,3,11):  -1625.491
```

AR(1,2,3,11) scored both lower on AIC and BIC criterion. At the same time the p-values for AR(1,2,3,11) is greater for each of the respective lags. Choose AR(1,2,3,11)

6

```
isDateBefore2005 <- which(PPI$DATE <= "2005-12-31")
isDateAfter2005 <- which(PPI$DATE > "2005-12-31")
indexB05 <- length(isDateBefore2005)
qtr_ahead <- length(isDateAfter2005)
ar3_pred <- numeric(qtr_ahead + 1)
ar12311_pred <- numeric(qtr_ahead + 1)

ar3_pred[1] = as.numeric(PPI[indexB05,"VALUE"])
ar12311_pred[1] = as.numeric(PPI[indexB05,"VALUE"])
for(i in 1:qtr_ahead) {
  ar3 <- arima(diff(log(as.numeric(xts(PPI[1:(indexB05+i),],
                                      PPI[1:(indexB05+i),]$DATE)[,-1]$VALUE))), order = c(3,0,0))
  ar3_pred[i+1] <- exp(diffinv(forecast(ar3, h = 1)$mean))[2] *
                   as.numeric(PPI[indexB05+i,'VALUE']) # This transforms the value back


  ar12311 <- arima(diff(log(as.numeric(xts(PPI[1:(indexB05+i),],
                                          PPI[1:(indexB05+i),]$DATE)[,-1]$VALUE))), order=c(11,0,0),
                                          fixed=c(NA,NA,NA,0,0,0,0,0,0,0,NA,NA), transform.pars = FALSE)
  ar12311_pred[i+1] <- exp(diffinv(forecast(ar12311, h = 1)$mean))[2] *
                       as.numeric(PPI[indexB05+i,'VALUE']) #This transform the values back
}

# transforming the values back
mspe_ar3 <- sum((PPI[isDateAfter2005,]$VALUE - ar3_pred[-1])^2) / qtr_ahead
mspe_ar12311 <- sum((PPI[isDateAfter2005,]$VALUE - ar12311_pred[-1])^2) / qtr_ahead
cat("MSPE for AR(3): ", mspe_ar3, '\n',"MSPE for AR(1,2,3,11): ", mspe_ar12311)
```

```
## MSPE for AR(3):  3.051553
##  MSPE for AR(1,2,3,11):  3.044605
```

We notice the MSPE for AR(1,2,3,11) is lower than the AR(3), which is consistent with 5c Next, we simulate a random walk model.

```
randomWalk_forecast <- numeric(qtr_ahead)
randomWalk_forecast[1] <- as.numeric(PPI[indexB05+1,"VALUE"]) + rnorm(1,1,1)
for(i in 2:qtr_ahead){
```

```r
  randomWalk_forecast[i] <- randomWalk_forecast[i - 1] + rnorm(1,1,1)
}
mspe_RW <- sum((PPI[isDateAfter2005,]$VALUE - randomWalk_forecast)^2) / qtr_ahead
cat("random walk mspe: ", mspe_RW)
```

```
## random walk mspe:  23.09169
```

Comparing the MSPE, we see that the random walk performs way worse as expected. We conclude, that PPI does not follow a random walk model, and could be reasonable predicted using the AR model. If we take the sqrt of the mspe then the we have an average error of 1.745 which is roughly 1% off on average.