Homework 7

Nikhil Guruji, Bingjie Hu, Jialiang Le

February 24, 2018

Question 1

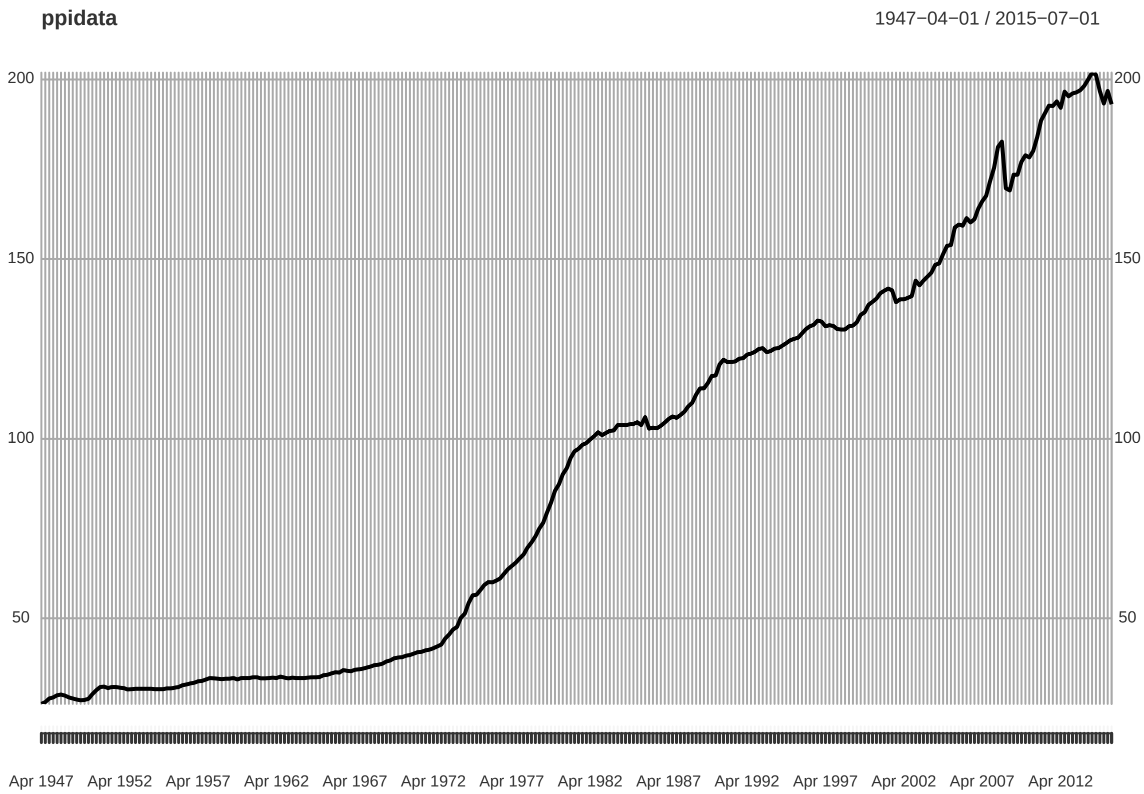
```
library(httr)
```

```
## Warning: package 'httr' was built under R version 3.4.3
```

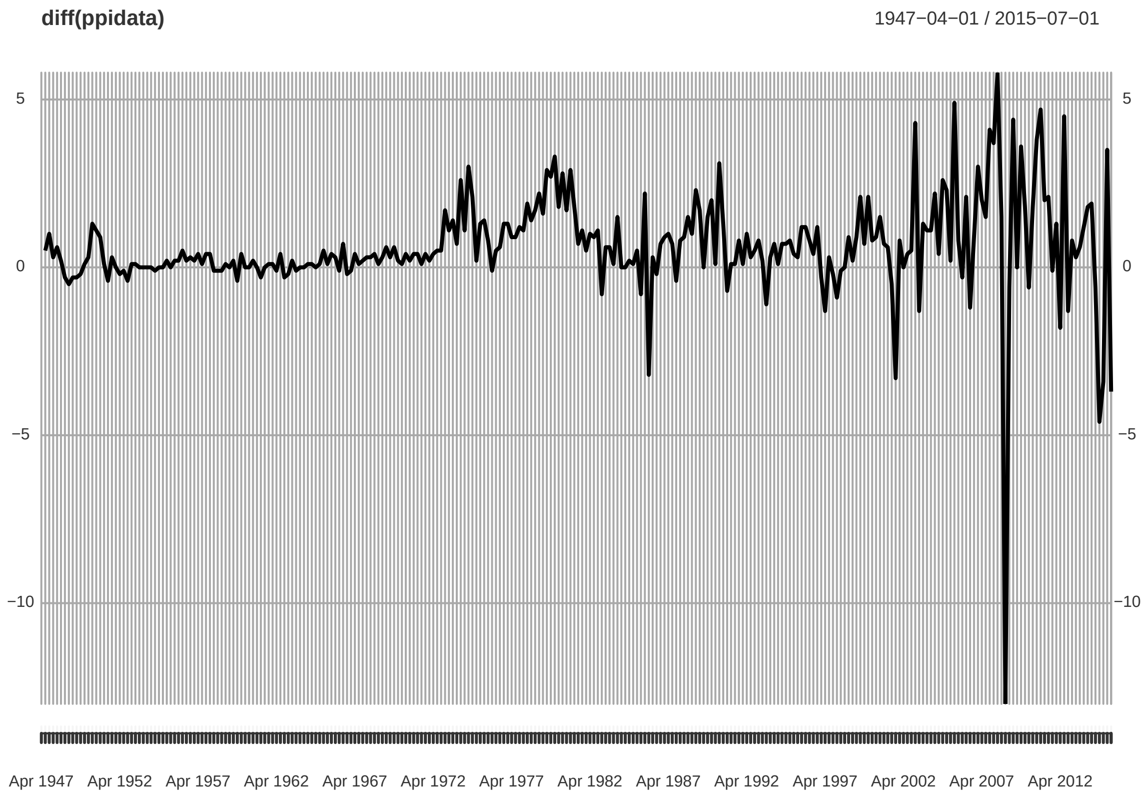
```
library(readxl)
library(dplyr)
library(zoo)
library(xts)
library(lubridate)
library(forecast)
url1="https://github.com/nikhilg12/nikhil/raw/master/Empirical%20HW7/PPIFGS.xls"
GET(url1, write_disk(tf1 <- tempfile(fileext = ".xls")))
```

```
## Response [https://raw.githubusercontent.com/nikhilg12/nikhil/master/Empirical%20HW7/PPIFGS.xls]
##   Date: 2018-02-26 08:49
##   Status: 200
##   Content-Type: application/octet-stream
##   Size: 35.3 kB
## <ON DISK>  C:\Users\bingj\AppData\Local\Temp\RtmpuePJ6UQ\file3b8058117f2d.xls
```

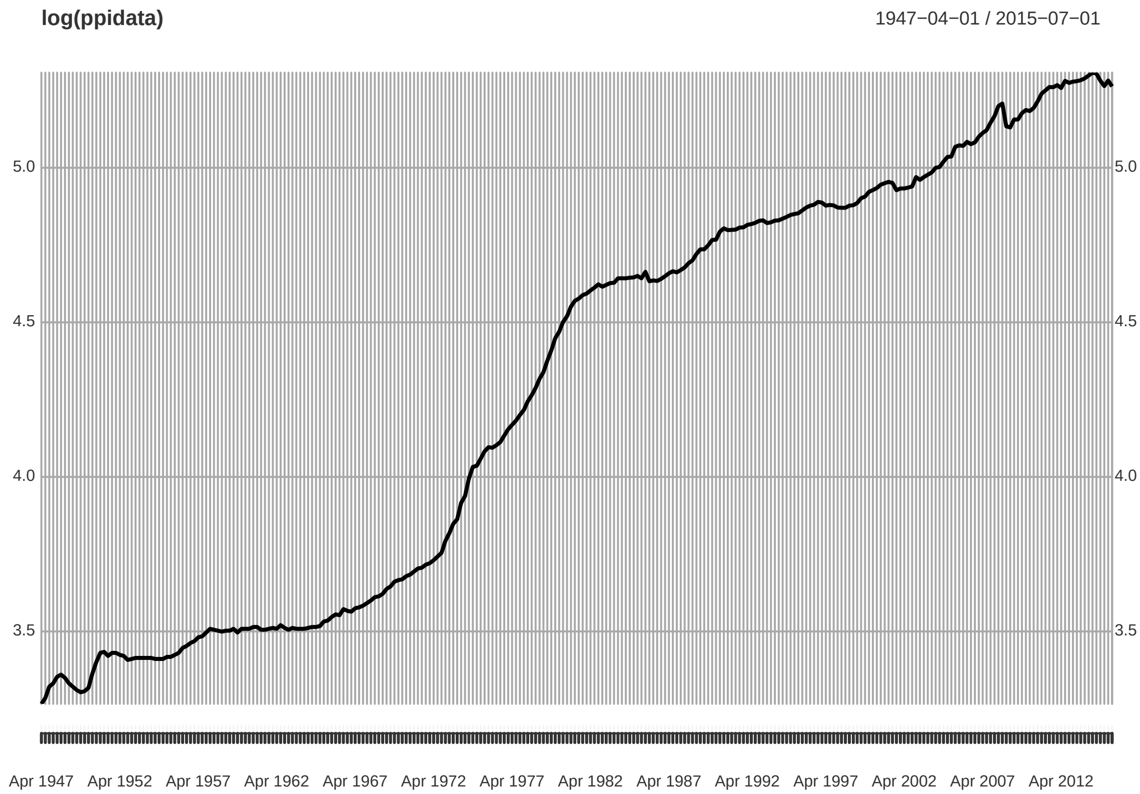
```
ppidata=read_excel(tf1)
ppidata <- xts(ppidata$VALUE,order.by = ppidata$DATE)
plot(ppidata)
```



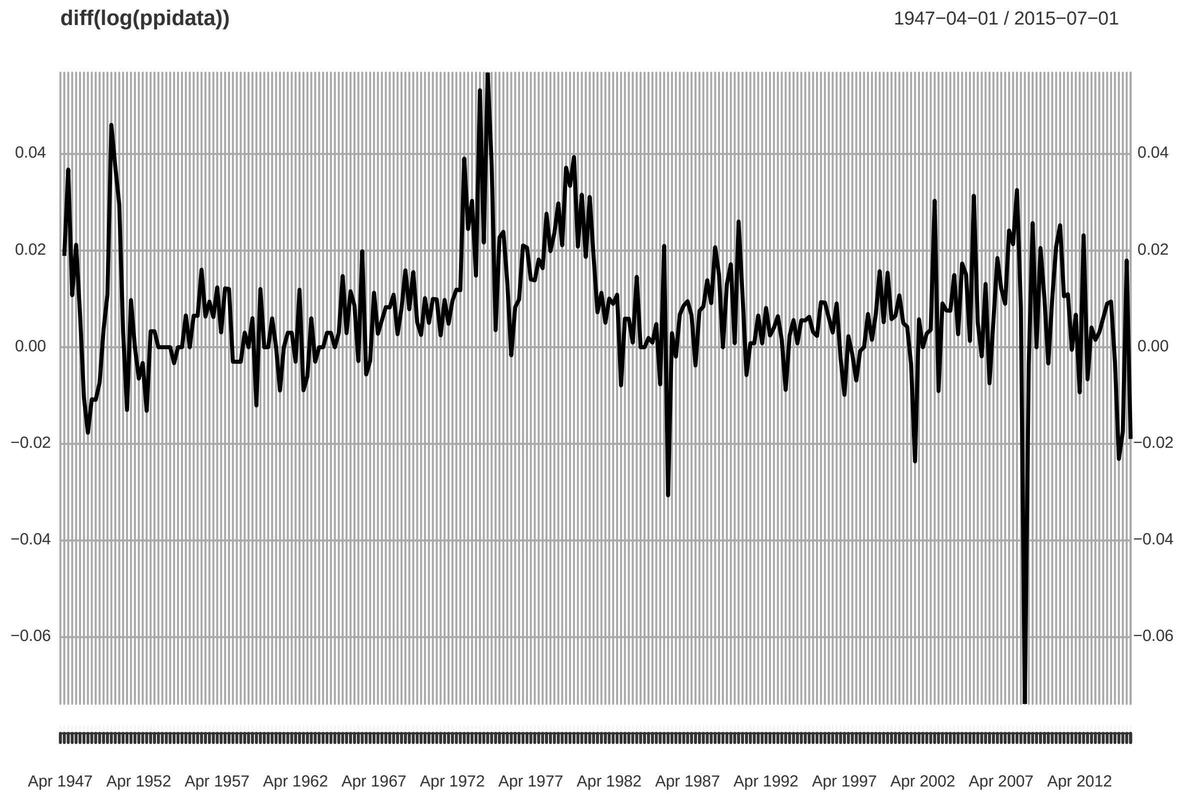

```
plot(diff(ppidata))
```




```
plot(log(ppidata))
```




```r
plot(diff(log(ppidata)))
```

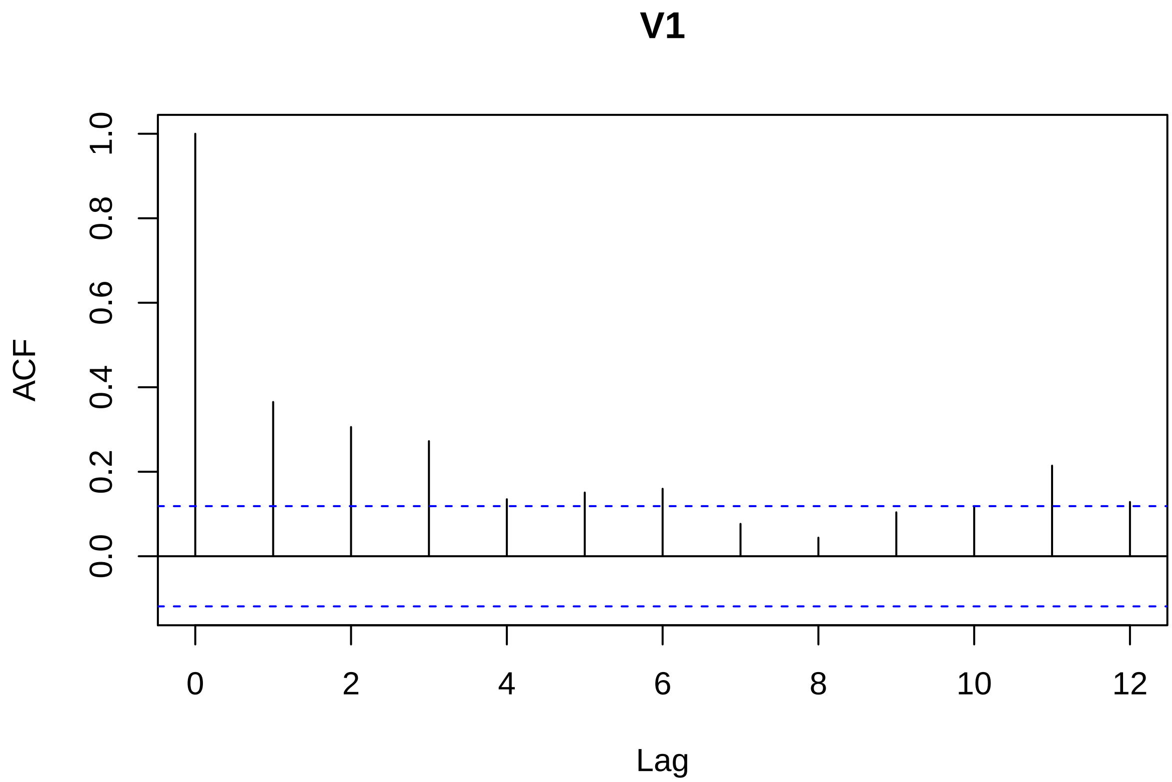


Question 2

The $\triangle logPPI$ looks covariance-stationary because the (a) and (c) don't seem to have the same mean and the (b) have different covariance at the end of the period comparing to the previous periods.

Question 3

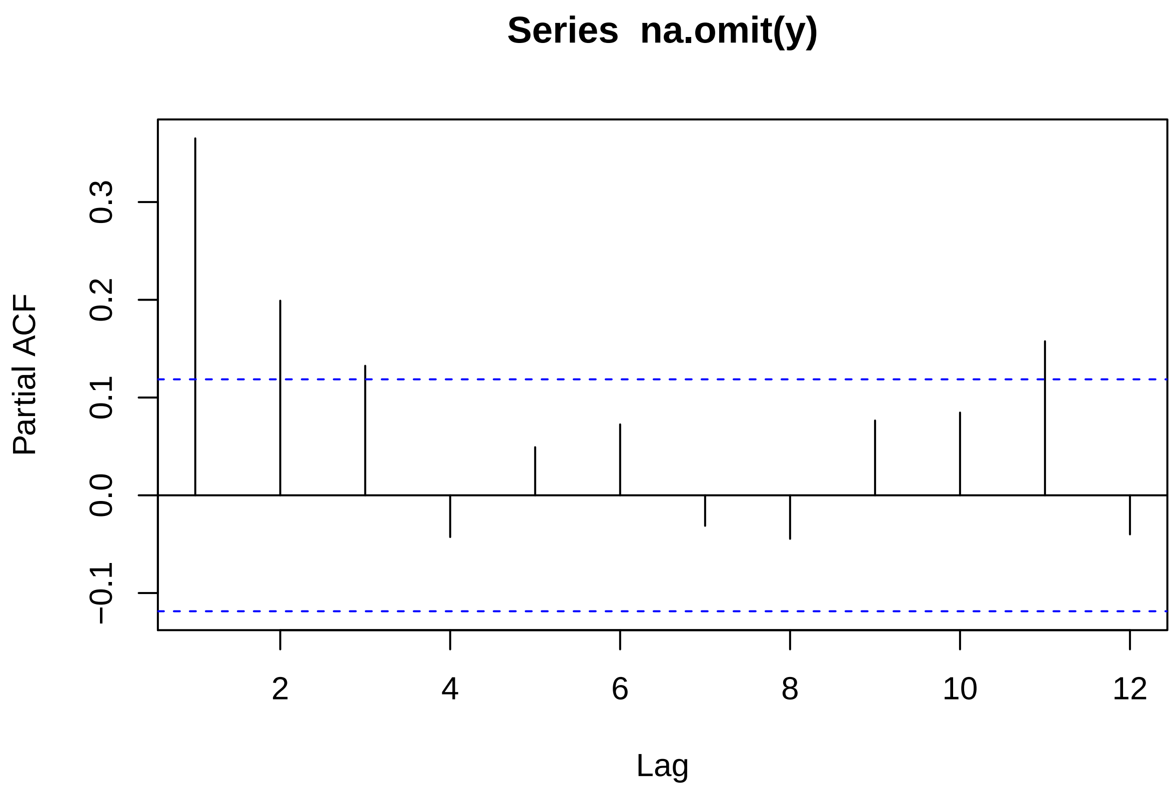
```
y <- as.data.frame(diff(log(ppidata)))
acf(na.omit(y),lag.max = 12)
```



From the ACF, we can conclude the autocorrelation converges very quickly, and it dies out after three periods, which means that MA(3) is appropriate. If the ACF converges very slowly, y_t might not be covariance stationary.

Question 4

```
pacf(na.omit(y),12)
```



From the PACF, we can conclude that the partial acf converges after three periods, which means that AR(3) is appropriate.

Question 5

The four testing models are ARMA(0,1), ARMA(1,0), ARMA(1,1), ARMA(3,3) ####(a)

```
library(tseries)
y <- na.omit(y)
m1 <- arima(y,c(0,0,1))
m2 <- arima(y,c(1,0,0))
m3 <- arima(y,c(1,0,1))
m4 <- arima(y,c(3,0,3))
m1
```

```
##
## Call:
## arima(x = y, order = c(0, 0, 1))
##
## Coefficients:
##          ma1  intercept
##       0.2821     0.0073
## s.e.  0.0524     0.0010
##
## sigma^2 estimated as 0.000154:  log likelihood = 810.87,  aic = -1615.74
```

```r
m2
```

```
##
## Call:
## arima(x = y, order = c(1, 0, 0))
##
## Coefficients:
##          ar1  intercept
##       0.3704     0.0073
## s.e.  0.0566     0.0012
##
## sigma^2 estimated as 0.0001476:  log likelihood = 816.66,  aic = -1627.32
```

```r
m3
```

```
##
## Call:
## arima(x = y, order = c(1, 0, 1))
##
## Coefficients:
##          ar1      ma1  intercept
##       0.8190  -0.5435     0.0073
## s.e.  0.0729   0.1068     0.0018
##
## sigma^2 estimated as 0.0001395:  log likelihood = 824.22,  aic = -1640.44
```

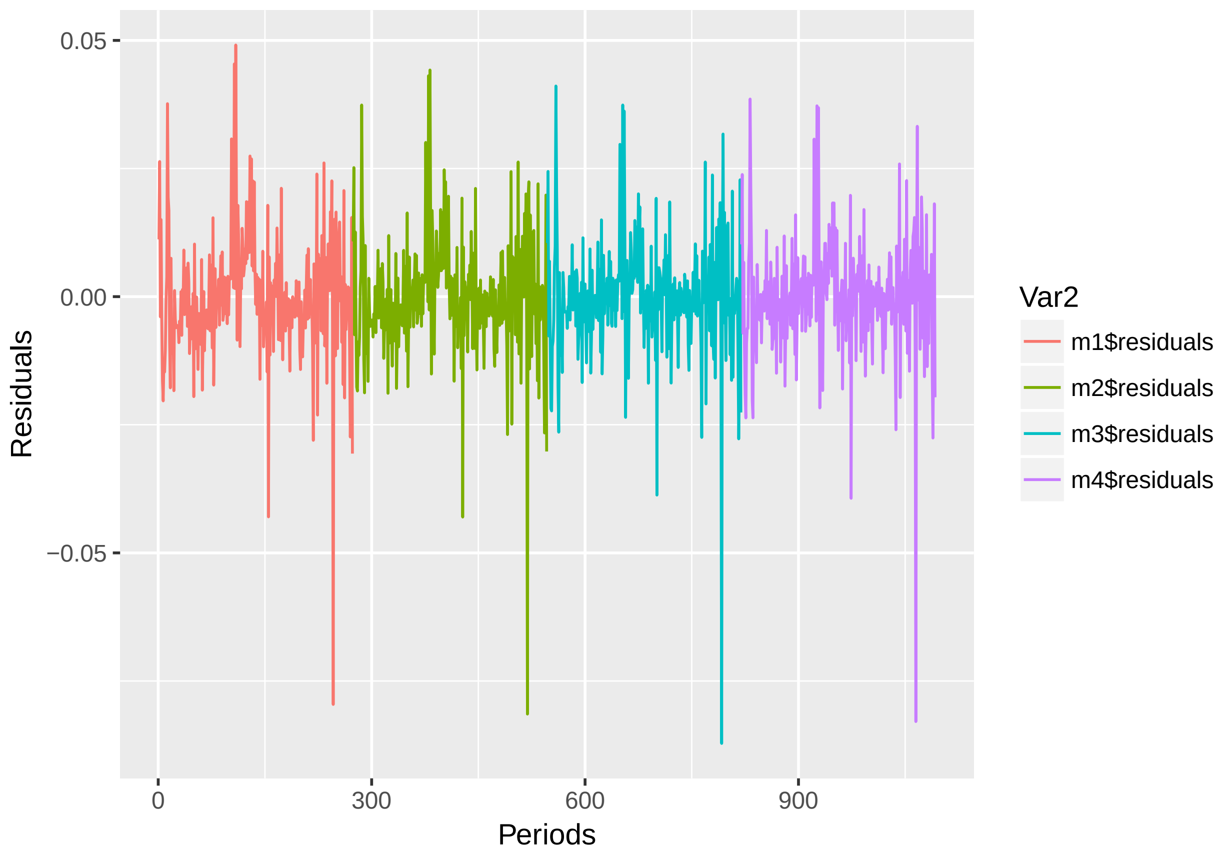
```r
m4
```

```
##
## Call:
## arima(x = y, order = c(3, 0, 3))
##
## Coefficients:
##           ar1     ar2     ar3     ma1     ma2      ma3  intercept
##       -0.5054  0.0550  0.7905  0.8037  0.3197  -0.4680     0.0073
## s.e.   0.0767  0.1019  0.0752  0.1098  0.1502   0.1093     0.0017
##
## sigma^2 estimated as 0.0001341:  log likelihood = 829.14,  aic = -1642.29
```

As the four tables show, the parameter estimates are all stationary.

(b)

```r
library(ggplot2)
library(reshape2)
rsd <- cbind(m1$residuals,m2$residuals,m3$residuals,m4$residuals)
periods <- nrow(rsd)
rsd <- melt(rsd)
p <- ggplot(data = rsd,mapping = aes(x = 1:nrow(rsd),y = value,color = Var2))+geom_line()+xlab("Periods
p
```



(c)

```
library(DataAnalytics)
cmpQValue <- function(residuals,object){
  rsd1 <- back(residuals,8)
  rsd2 <- back(residuals,12)
  return(data.frame(Q8 = Box.test(rsd1,type = "Ljung")$statistic,p8 = Box.test(rsd1,type = "Ljung")$p.va
}
rsd <- cbind(m1$residuals,m2$residuals,m3$residuals,m4$residuals)
cmpQValue(as.vector(rsd[,1]),m1)
```

```
##                  Q8        p8      Q12       p12       AIC       BIC
## X-squared 0.8778827 0.3487822 1.269563 0.2598487 -1615.743 -1604.915
```

```
cmpQValue(as.vector(rsd[,2]),m2)
```

```
##                 Q8        p8       Q12      p12       AIC       BIC
## X-squared 1.353206 0.2447184 0.9302075 0.334809 -1627.316 -1616.488
```

```
cmpQValue(as.vector(rsd[,3]),m3)
```

```
##                    Q8       p8        Q12       p12       AIC       BIC
## X-squared 0.001725613 0.966865 0.01011536 0.9198877 -1640.442 -1626.004
```

```
cmpQValue(as.vector(rsd[,4]),m4)
```

```
##                    Q8        p8         Q12       p12       AIC       BIC
## X-squared 0.004743327 0.9450917 0.005862673 0.9389671 -1642.289 -1613.413
```

According to the tables, the last model, which is arma(3,3) is the best one because both the AIC and BIC are the smallest among the four models, and the p values are all larger than 0.05, which means that the residuals are not correlated.

Question 6

```
ppidata2 <- ppidata["1947-07-01::2005-10-01"]
y2 <- diff(log(ppidata2))
y2 <- as.ts(na.omit(y2))
m12 <- arima(y2,c(0,0,1))
m22 <- arima(y2,c(1,0,0))
m32 <- arima(y2,c(1,0,1))
m42 <- arima(y2,c(3,0,3))
nahead <- nrow(y)-length(y2)
forecastArima <- function(object,nahead){
  pm <- predict(object,nahead)
  mspe <- sum((pm$pred-y[(length(y2)+1):nrow(y),1])^2)/nahead
  return(mspe)
}
mspe1 <- forecastArima(m12,nahead)
mspe2 <- forecastArima(m22,nahead)
mspe3 <- forecastArima(m32,nahead)
mspe4 <- forecastArima(m42,nahead)
c(mspe1,mspe2,mspe3,mspe4)
```

```
## [1] 0.0003228116 0.0003230052 0.0003275092 0.0003287220
```

```
pm_random_walk <- rep(y2[length(y2)],nahead)+cumsum(rnorm(nahead))
mspe_random_walk <- sum(pm_random_walk-y[(length(y2)+1):nrow(y),1])^2/nahead
mspe_random_walk
```

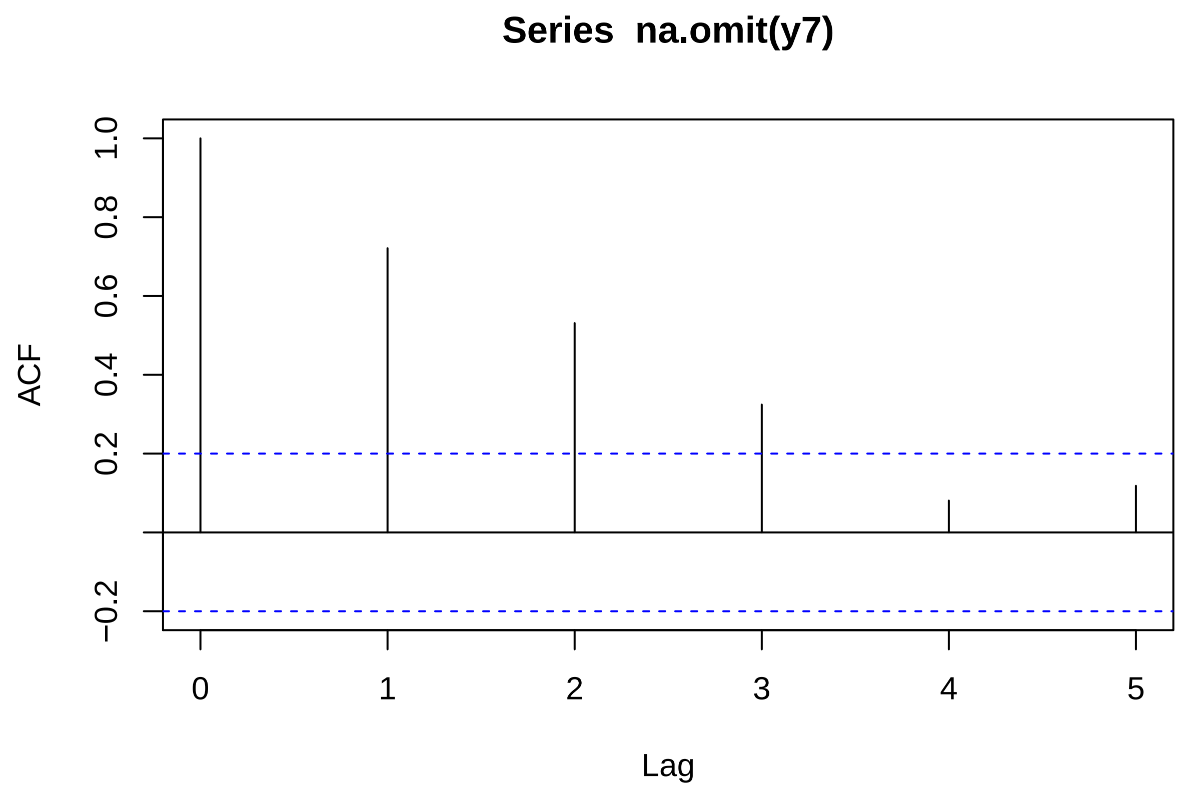
```
## [1] 192.8194
```

The MSPE of random walk model is much larger than the other four models, which means random walk is not a good model for the time series data.

Question 7

(1)

```
epsilon7 <- rnorm(100)
x7 <- epsilon7
e <- cumsum(x7)
y7 <- e-back(e,4)
autocovariance <- acf(na.omit(y7),lag = 5)$acf*var(na.omit(y7))
```

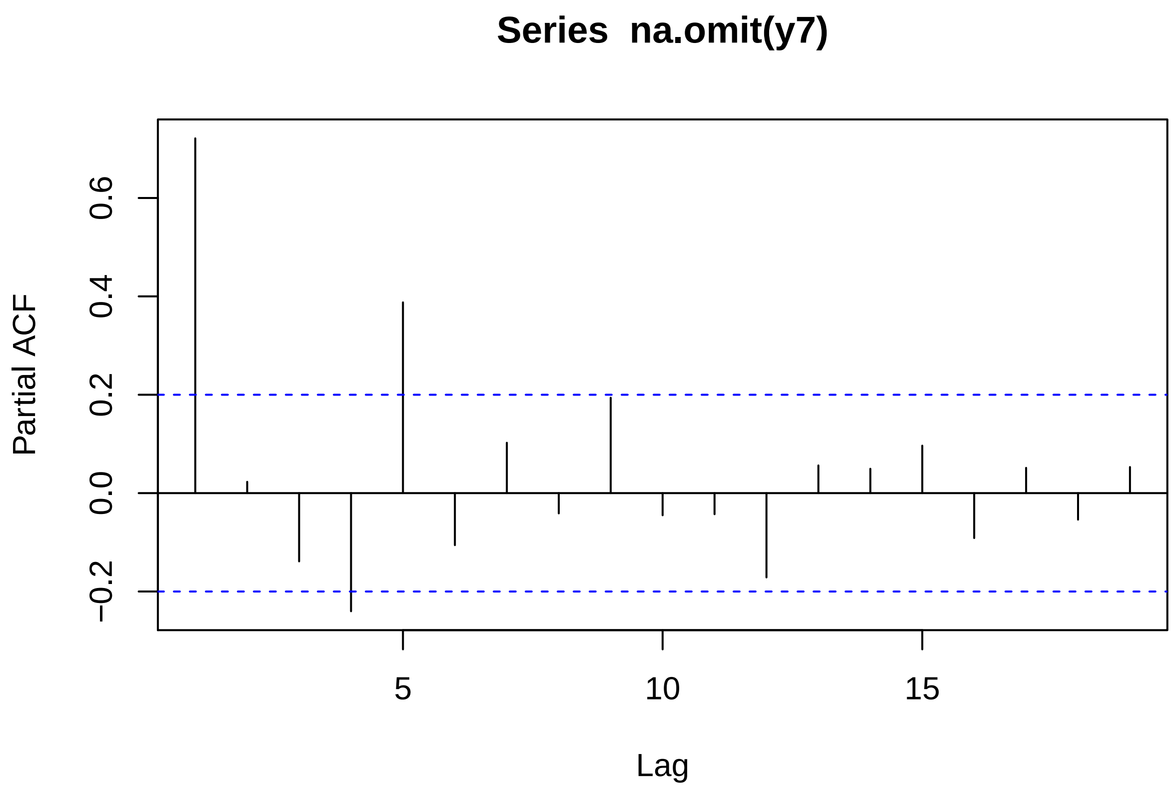



```
autocovariance
```

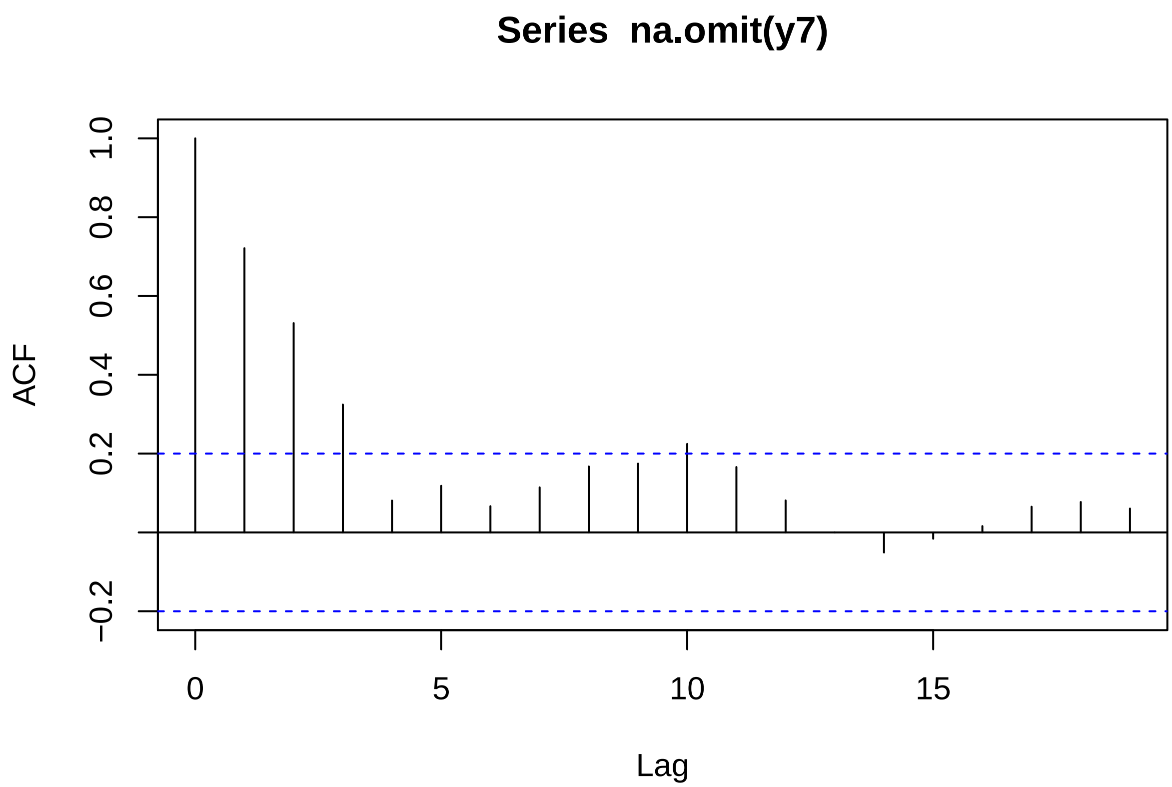
```
## , , 1
##
##            [,1]
## [1,] 4.2511380
## [2,] 3.0663083
## [3,] 2.2584857
## [4,] 1.3791156
## [5,] 0.3433682
## [6,] 0.5028775
```

(2)

```
pacf(na.omit(y7))
```




```r
acf(na.omit(y7))
```




```r
m <- arima(y7,c(3,0,2))
m
```

```
##
## Call:
## arima(x = y7, order = c(3, 0, 2))
##
## Coefficients:
##          ar1      ar2     ar3      ma1     ma2  intercept
##       0.6636  -0.4125  0.2594  -0.0227  0.9855     0.1547
## s.e.  0.1169   0.1375  0.1150   0.0522  0.1365     0.4834
##
## sigma^2 estimated as 1.45:  log likelihood = -156.98,  aic = 327.96
```

According to the ACF and PACF graphs, the AR lags should be 3, and the MA lags should be 2.