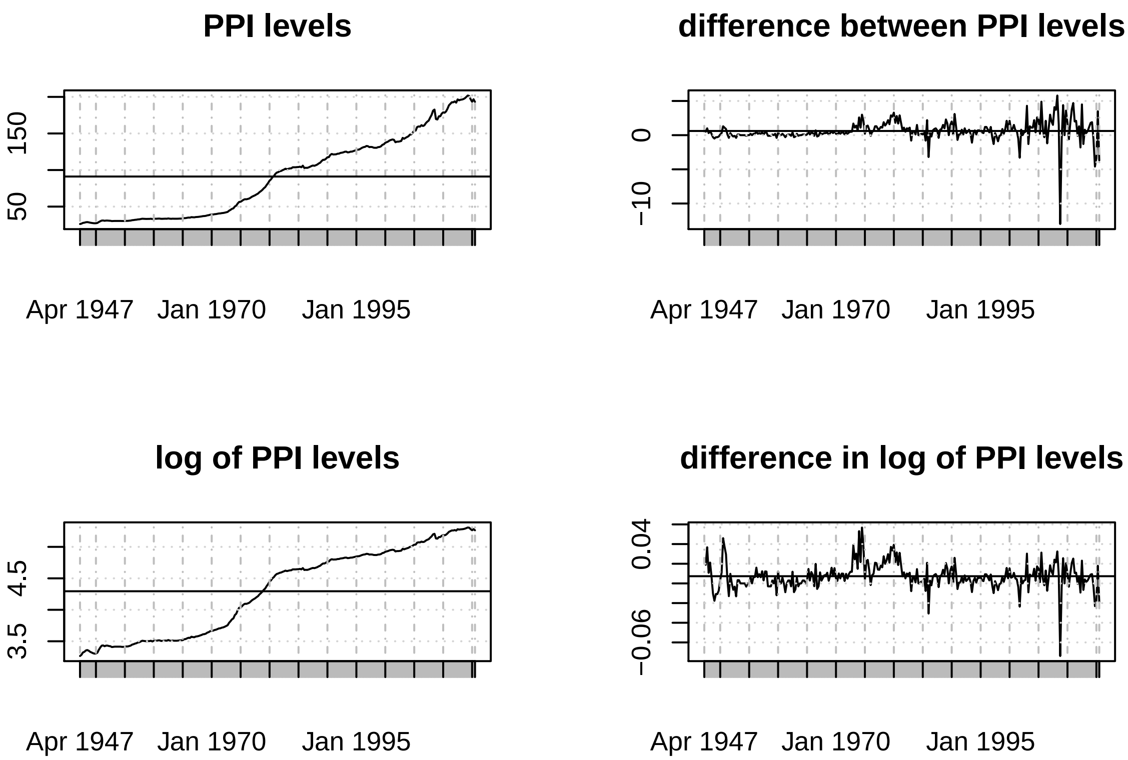
Empirical Assignment 7

Nitish Ramkumar, Ian Laker, PrasanthKumar

1

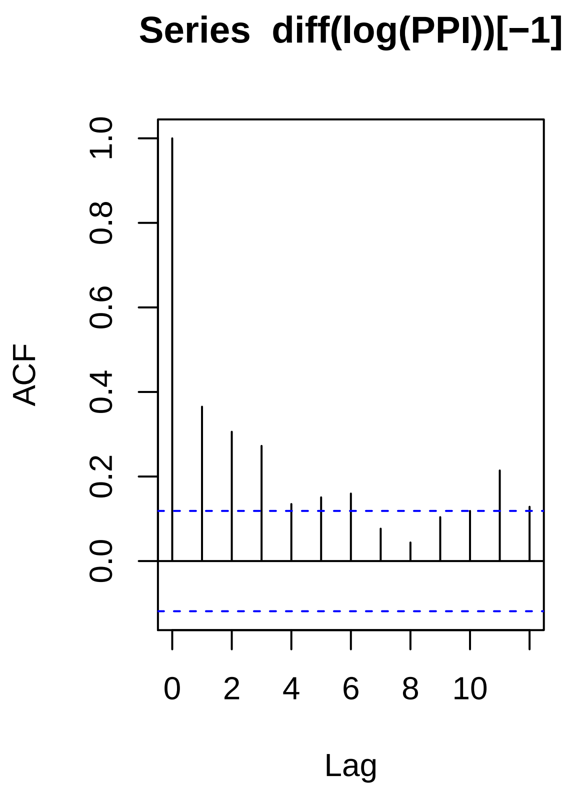
The graphs of the 4 situations are as below



2

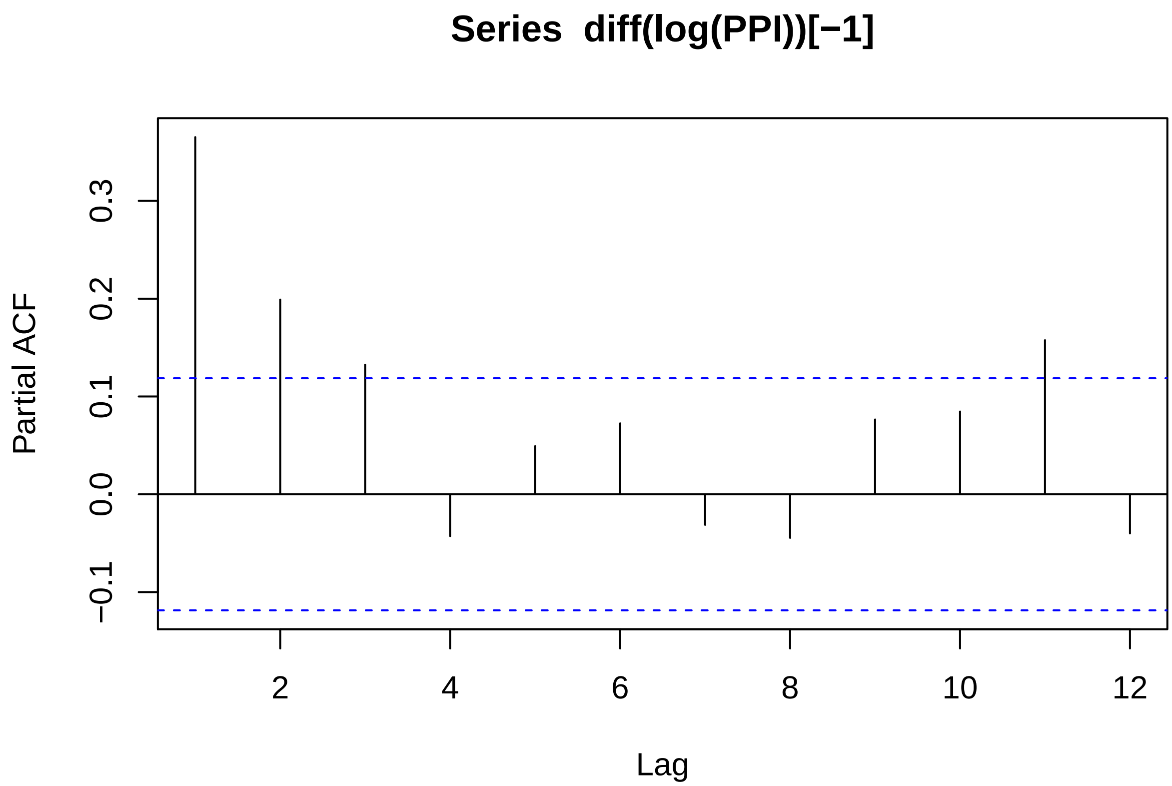
From the graph, we can see that the mean reversion properly happens for differences in PPI levels and difference in log of PPI levels. Amongst the 2, lets use diff(log(PPI)) as it looks more covariance stationary So y_t = diff(log(PPI))

3



From the graph, 0 and 1 are the significant lags. These value will be the lag for MA model.

4



The PACF graph confirms that 1,2 lags would be ideal values for the AR model.

5

The various models which can be used are as below:
Model 1: p=1, q=0
Model 2: p=2, q=0
Model 3: p=1, q=1
Model 4: p=2, q=1

Model 1 - p=1, q=0

Coefficients and stationary check

```
##         ar1   intercept
## 0.370401835 0.007287272
```

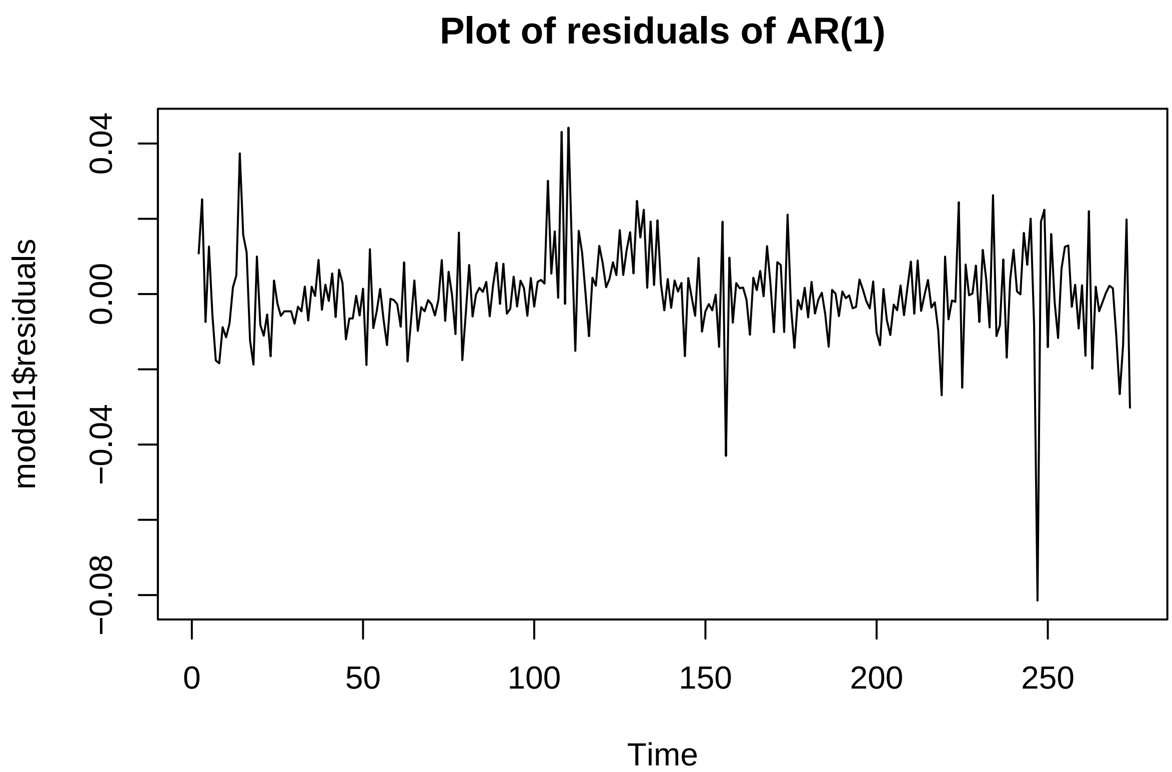
As ϕ_1 is <1, it is stationary

Standard errors

```
##         ar1   intercept
```

```
## 0.056611112 0.001166816
```

Residual Plot



Model 2 - p=2, q=0

Coefficients

```
##         ar1         ar2   intercept
## 0.294200805 0.204868017 0.007321179
```

Stationary check

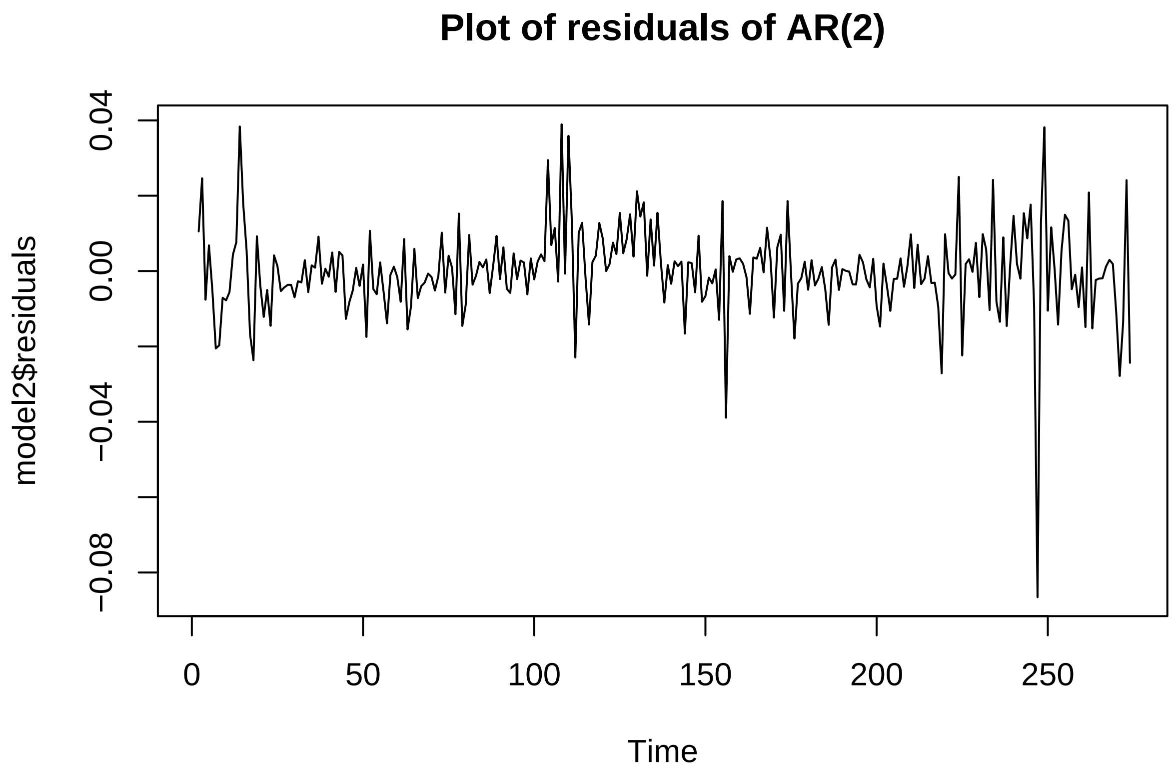
```
## [1] 0.6230274 0.3288266
```

As all the characteristic roots are <1, this is stationary

Standard errors

```
##         ar1         ar2   intercept
## 0.059765358 0.060355150 0.001431467
```

Residual plot



Model 3 - p=1, q=1

Coefficients and stationary proof

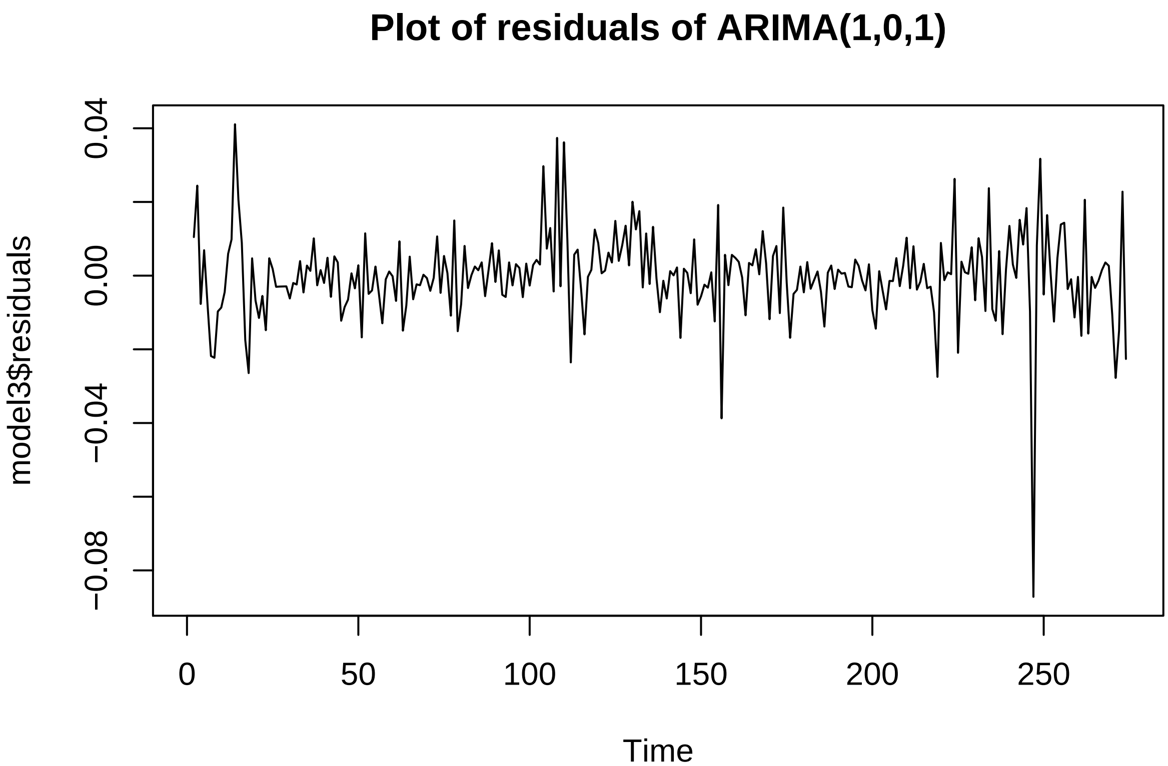
```
##          ar1         ma1   intercept
##  0.81877374 -0.54324190  0.00729667
```

For ARMA model, as ϕ_1 within the unit circle. So this is stationary.

Standard errors

```
##         ar1         ma1   intercept
## 0.072969119 0.106769037 0.001781461
```

Residual plot



Model 4- p=2, q=1

Coefficients

```
##         ar1          ar2          ma1    intercept
##  0.747584706  0.040361900 -0.479360383  0.007306698
```

Stationary check

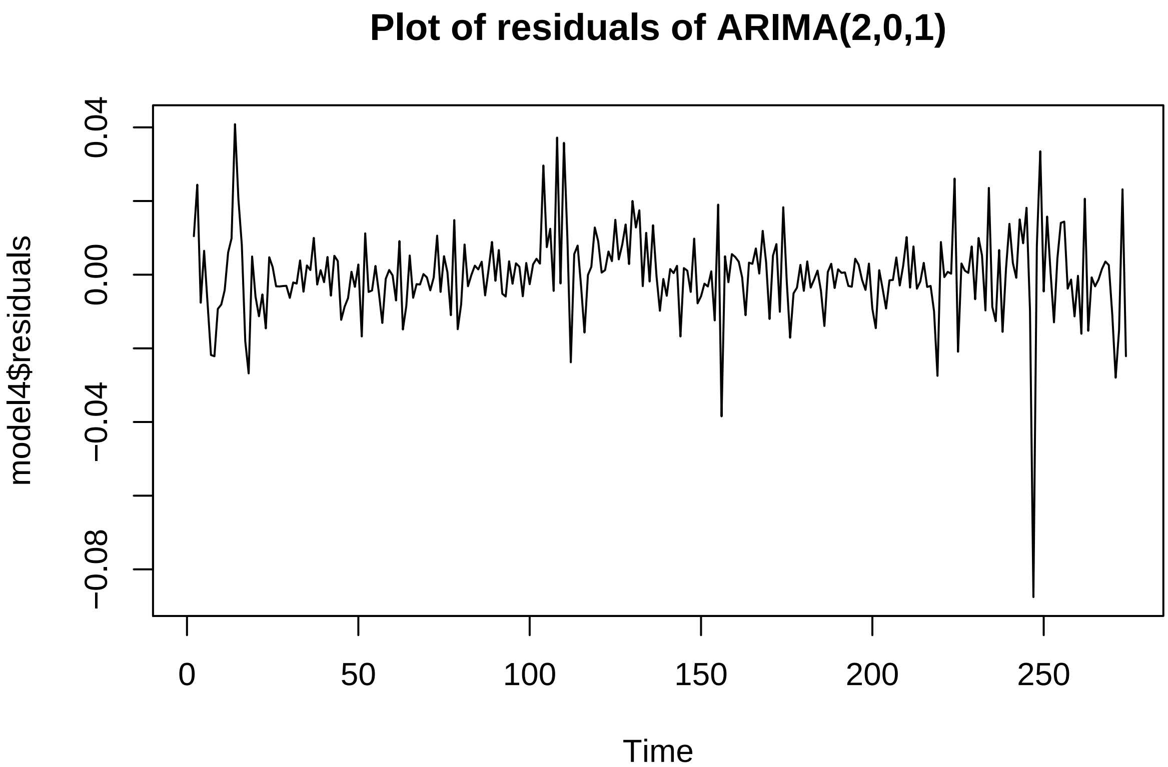
```
## [1] 0.79815378 0.05056908
```

For ARMA model, the characteristic root of AR part is less than 1. So this is stationary.

Standard errors

```
##         ar1         ar2         ma1   intercept
## 0.232533124 0.123413236 0.226047271 0.001737648
```

Residual Plot



Choice between models

The Q statistic, p value for 8 and 12 lags, AIC and BIC values are as below

```
##           8Statistics chi-Sq 8statistics p 12statistics chi-sq
## AR(1)              19.334365     0.0131702            29.90137
## AR(2)              10.201540     0.2511645            18.72061
## AR(1,0,1)           6.927874     0.5444342            14.39146
## AR(2,0,1)           6.722095     0.5668917            14.49738
##           12statistics p       AIC       BIC
## AR(1)        0.002889476 -1627.316 -1616.488
## AR(2)        0.095497802 -1636.589 -1622.151
## AR(1,0,1)    0.276411947 -1640.442 -1626.004
## AR(2,0,1)    0.270079814 -1638.540 -1620.492
```

If we see the values of AIC,BIC, we can see the Model 2 (ARMA(1,0,1)) is the best model as it has the lowest AIC value

6

The Mean square prediction error for the models are as follows

```
## [1] 0.0003307826 0.0003398474 0.0003355703 0.0003398474
```

The mean square method also confirms our choice of model 3 (AR(1,0,1)) as it has the least error.

For a random walk, the mean square prediction error is

```
## [1] 0.0003231982
```

This shows that all our model errors are very close to that of a random walk. This is because we can predict values immediately after the training data (in this case 2006-01-01, 2006-04-01), but our prediction is as bad as a random walk (or sometimes worse) as we try to predict data further away.

R Code

```
suppressMessages(library(xts))
suppressMessages(library(readxl))
setwd("C:/_UCLA/237E_Empirical/Assigments/Assignment7")

PPI.data <- read_excel("PPIFGS.xls")
PPI <- xts(PPI.data$VALUE,as.Date(PPI.data$DATE))

#1

par(mfrow=c(2,2))

plot(PPI,main="PPI levels")
abline(h=mean(PPI))

plot(diff(PPI),main="difference between PPI levels")
abline(h = mean(diff(PPI),na.rm=T))

plot(log(PPI), main="log of PPI levels")
abline(h = mean(log(PPI),na.rm=T))

plot(diff(log(PPI)),main="difference in log of PPI levels")
abline(h = mean(diff(log(PPI)),na.rm=T))
par(mfrow=c(1,1))


#3
par(mfrow=c(1,2))
acf(diff(log(PPI))[-1],lag.max = 12)
par(mfrow=c(1,1))

#4
pacf(diff(log(PPI))[-1],lag.max = 12)

##Model 1 - p=1, q=0
###Coefficients and stationary check
model1 <- arima(diff(log(PPI)),order=c(1,0,0),method = "ML")
model1$coef
```

```
###Standard errors
sqrt(diag(model1$var.coef))

###Residual Plot
plot(model1$residuals,main="Plot of residuals of AR(1)")


choose.factor <- matrix(nrow=4,ncol=6)
colnames(choose.factor) <- c("8Statistics chi-Sq","8statistics p","12statistics chi-sq","12statistics p
row.names(choose.factor) <- c("AR(1)","AR(2)","AR(1,0,1)","AR(2,0,1)")

test8 <- Box.test(model1$residuals,8,"Ljung-Box")
test12 <- Box.test(model1$residuals,12,"Ljung-Box")

choose.factor[1,] <- c(test8$statistic,test8$p.value,test12$statistic,test12$p.value,model1$aic,BIC(mode

##Model 2 - p=2, q=0

###Coefficients
model2 <- arima(diff(log(PPI)),order=c(2,0,0),method = "ML")
coefs2 <- model2$coef
coefs2

###Stationary check
roots <- polyroot(c(1,-coefs2[1],-coefs2[2]))
characteristic <- 1/roots
Mod(characteristic)

###Standard errors
sqrt(diag(model2$var.coef))

###Residual plot
plot(model2$residuals,main="Plot of residuals of AR(2)")

test8.2 <- Box.test(model2$residuals,8,"Ljung-Box")
test12.2 <- Box.test(model2$residuals,12,"Ljung-Box")

choose.factor[2,] <- c(test8.2$statistic,test8.2$p.value,test12.2$statistic,test12.2$p.value,model2$aic

##Model 3 - p=1, q=1
###Coefficients and stationary proof
model3 <- arima(diff(log(PPI)),order=c(1,0,1),method = "ML")
model3$coef

###Standard errors
sqrt(diag(model3$var.coef))

###Residual plot
plot(model3$residuals,main="Plot of residuals of ARIMA(1,0,1)")

test8.3 <- Box.test(model3$residuals,8,"Ljung-Box")
test12.3 <- Box.test(model3$residuals,12,"Ljung-Box")
```

```
choose.factor[3,] <- c(test8.3$statistic,test8.3$p.value,test12.3$statistic,test12.3$p.value,model3$aic

##Model 4- p=2, q=1
###Coefficients
model4 <- arima(diff(log(PPI)),order=c(2,0,1),method = "ML")
coefs4 <- model4$coef
coefs4

###Stationary check
roots <- polyroot(c(1,-coefs4[1],-coefs4[2]))
characteristic <- 1/roots
Mod(characteristic)

###Standard errors
sqrt(diag(model4$var.coef))

###Residual Plot
plot(model4$residuals,main="Plot of residuals of ARIMA(2,0,1)")

test8.4 <- Box.test(model4$residuals,8,"Ljung-Box")
test12.4 <- Box.test(model4$residuals,12,"Ljung-Box")

choose.factor[4,] <- c(test8.4$statistic,test8.4$p.value,test12.4$statistic,test12.4$p.value,model4$aic


#6

suppressMessages(library("forecast"))
model1.reest <- Arima(diff(log(PPI))[index(diff(log(PPI)))<"2006-01-01",],order=c(1,0,0),method = "ML")
model2.reest <- Arima(diff(log(PPI))[index(diff(log(PPI)))<"2006-01-01",],order=c(2,0,0),method = "ML")
model3.reest <- Arima(diff(log(PPI))[index(diff(log(PPI)))<"2006-01-01",],order=c(1,0,1),method = "ML")
model4.reest <- Arima(diff(log(PPI))[index(diff(log(PPI)))<"2006-01-01",],order=c(2,0,0),method = "ML")

true.value <- diff(log(PPI))[index(diff(log(PPI)))>="2006-01-01",]

model1.forecasted.value <- forecast(model1.reest,nrow(true.value))$mean
model2.forecasted.value <- forecast(model2.reest,length(true.value))$mean
model3.forecasted.value <- forecast(model3.reest,length(true.value))$mean
model4.forecasted.value <- forecast(model4.reest,length(true.value))$mean

mspes <- c()
mspes[1] <- sum((coredata(model1.forecasted.value) - true.value)^2)/(length(true.value))
mspes[2] <- sum((coredata(model2.forecasted.value) - true.value)^2)/(length(true.value))
mspes[3] <- sum((coredata(model3.forecasted.value) - true.value)^2)/(length(true.value))
mspes[4] <- sum((coredata(model4.forecasted.value) - true.value)^2)/(length(true.value))
mspes

#Random walk- forecast value is same as last value
randomWalk.forecasted.value <- rep(diff(log(PPI))["2005-10-01"],length(true.value))
mspe.randomWalk <- sum((randomWalk.forecasted.value - true.value)^2)/length(true.value)
mspe.randomWalk
```