CSC425 HW2

Problem 2

load libraries

```
library(tseries)
library(fBasics)
```

```
## Loading required package: timeDate

## Loading required package: timeSeries

##

## Rmetrics Package fBasics

## Analysing Markets and calculating Basic Statistics

## Copyright (C) 2005-2014 Rmetrics Association Zurich

## Educational Software for Financial Engineering and Computational Science

## Rmetrics is free software and comes with ABSOLUTELY NO WARRANTY.

## https://www.rmetrics.org --- Mail to: info@rmetrics.org
```

a) Import the data in R. Use ts() where the starting date is first month of 1980, and frequency is set eqaul to 12

```
setwd("~/Desktop/CSC425/hwork2/")
myd = read.table("NAPM.csv", header = T, sep = ',')
head(myd)
```

```
##       date index
## 1 1/1/1980  46.2
## 2 2/1/1980  50.2
## 3 3/1/1980  43.6
## 4 4/1/1980  37.4
## 5 5/1/1980  29.4
## 6 6/1/1980  30.3
```

```
tail(myd)
```

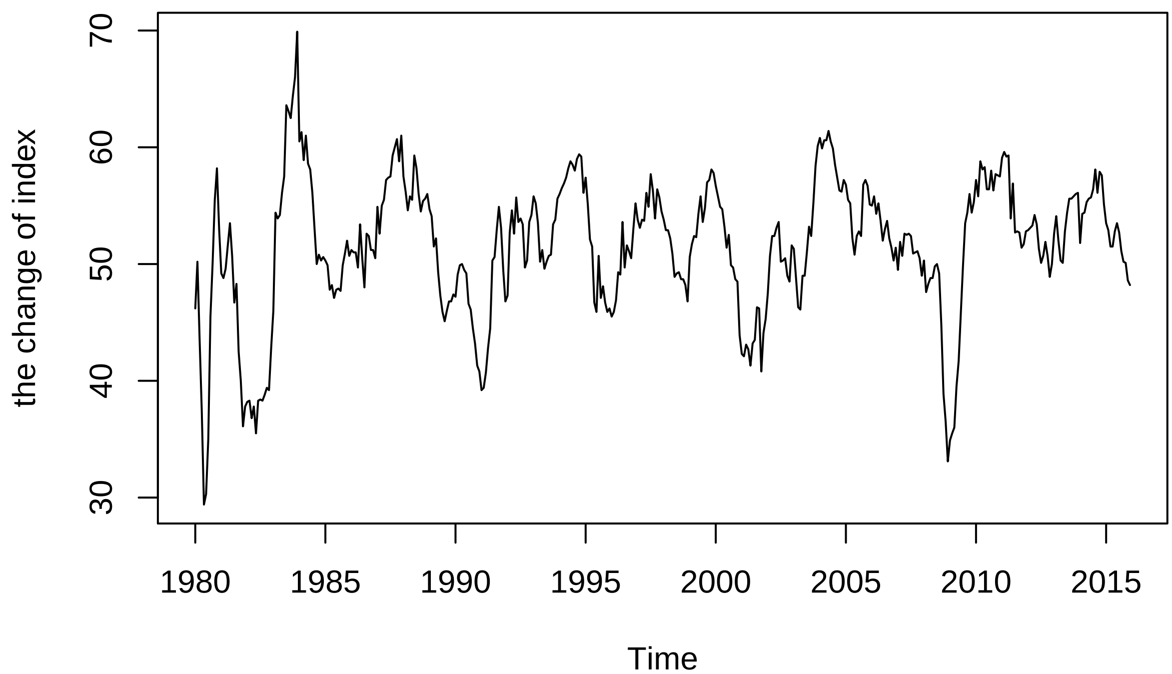
```
##          date index
## 427  7/1/2015  52.7
## 428  8/1/2015  51.1
## 429  9/1/2015  50.2
## 430 10/1/2015  50.1
## 431 11/1/2015  48.6
## 432 12/1/2015  48.2
```

```
TSindex = ts(myd[,2], start = c(1980,1), freq = 12)
TSindex
```

```
##       Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1980 46.2 50.2 43.6 37.4 29.4 30.3 35.0 45.5 50.1 55.5 58.2 53.0
## 1981 49.2 48.8 49.6 51.6 53.5 50.7 46.7 48.3 42.5 40.0 36.1 37.8
## 1982 38.2 38.3 36.8 37.8 35.5 38.3 38.4 38.3 38.8 39.4 39.2 42.8
## 1983 46.0 54.4 53.9 54.2 56.1 57.5 63.6 63.1 62.5 64.4 66.0 69.9
## 1984 60.5 61.3 58.9 61.0 58.6 58.1 56.1 53.0 50.0 50.8 50.3 50.6
## 1985 50.3 49.9 47.8 48.2 47.1 47.8 47.9 47.7 49.9 50.9 52.0 50.7
## 1986 51.2 51.0 51.0 49.7 53.4 50.5 48.0 52.6 52.4 51.2 51.2 50.5
## 1987 54.9 52.6 55.0 55.5 57.2 57.4 57.5 59.3 60.0 60.7 58.8 61.0
## 1988 57.5 56.2 54.6 55.8 55.5 59.3 58.2 56.0 54.5 55.4 55.6 56.0
## 1989 54.7 54.1 51.5 52.2 49.3 47.3 45.9 45.1 46.0 46.8 46.8 47.4
## 1990 47.2 49.1 49.9 50.0 49.5 49.2 46.6 46.1 44.5 43.2 41.3 40.8
## 1991 39.2 39.4 40.7 42.8 44.5 50.3 50.6 52.9 54.9 53.1 49.5 46.8
## 1992 47.3 52.7 54.6 52.6 55.7 53.6 53.9 53.4 49.7 50.3 53.6 54.2
## 1993 55.8 55.2 53.5 50.2 51.2 49.6 50.2 50.7 50.8 53.4 53.8 55.6
## 1994 56.0 56.5 56.9 57.4 58.2 58.8 58.5 58.0 59.0 59.4 59.2 56.1
## 1995 57.4 55.1 52.1 51.5 46.7 45.9 50.7 47.1 48.1 46.7 45.9 46.2
## 1996 45.5 45.9 46.9 49.3 49.1 53.6 49.7 51.6 51.1 50.5 53.0 55.2
## 1997 53.8 53.1 53.8 53.7 56.1 54.9 57.7 56.3 53.9 56.4 55.7 54.5
## 1998 53.8 52.9 52.9 52.2 50.9 48.9 49.2 49.3 48.7 48.7 48.2 46.8
## 1999 50.6 51.7 52.4 52.3 54.3 55.8 53.6 54.8 57.0 57.2 58.1 57.8
## 2000 56.7 55.8 54.9 54.7 53.2 51.4 52.5 49.9 49.7 48.7 48.5 43.9
## 2001 42.3 42.1 43.1 42.7 41.3 43.2 43.5 46.3 46.2 40.8 44.1 45.3
## 2002 47.5 50.7 52.4 52.4 53.1 53.6 50.2 50.3 50.5 49.0 48.5 51.6
## 2003 51.3 48.8 46.3 46.1 49.0 49.0 51.0 53.2 52.4 55.2 58.4 60.1
## 2004 60.8 59.9 60.6 60.6 61.4 60.5 59.9 58.5 57.4 56.3 56.2 57.2
## 2005 56.8 55.5 55.2 52.2 50.8 52.4 52.8 52.4 56.8 57.2 56.7 55.1
## 2006 55.0 55.8 54.3 55.2 53.7 52.0 53.0 53.7 52.2 51.4 50.3 51.4
## 2007 49.5 51.9 50.7 52.6 52.5 52.6 52.4 50.9 51.0 51.1 50.5 49.0
## 2008 50.3 47.6 48.3 48.8 48.8 49.8 50.0 49.2 44.8 38.9 36.5 33.1
## 2009 34.9 35.5 36.0 39.5 41.7 45.8 49.9 53.5 54.4 56.0 54.4 55.3
## 2010 57.2 55.8 58.8 58.1 58.3 56.4 56.4 58.0 56.3 57.7 57.6 57.5
## 2011 59.1 59.6 59.2 59.3 53.9 56.9 52.7 52.8 52.7 51.4 51.7 52.8
## 2012 52.9 53.1 53.3 54.2 53.4 51.3 50.1 50.7 51.9 50.7 48.9 50.0
## 2013 52.6 54.1 52.0 50.3 50.1 52.8 54.4 55.6 55.6 55.8 56.0 56.1
## 2014 51.8 54.3 54.4 55.3 55.6 55.7 56.4 58.1 56.1 57.9 57.6 55.1
## 2015 53.5 52.9 51.5 51.5 52.8 53.5 52.7 51.1 50.2 50.1 48.6 48.2
```

b) Create the time plot of the index and analyze the time trend displayed by the plot

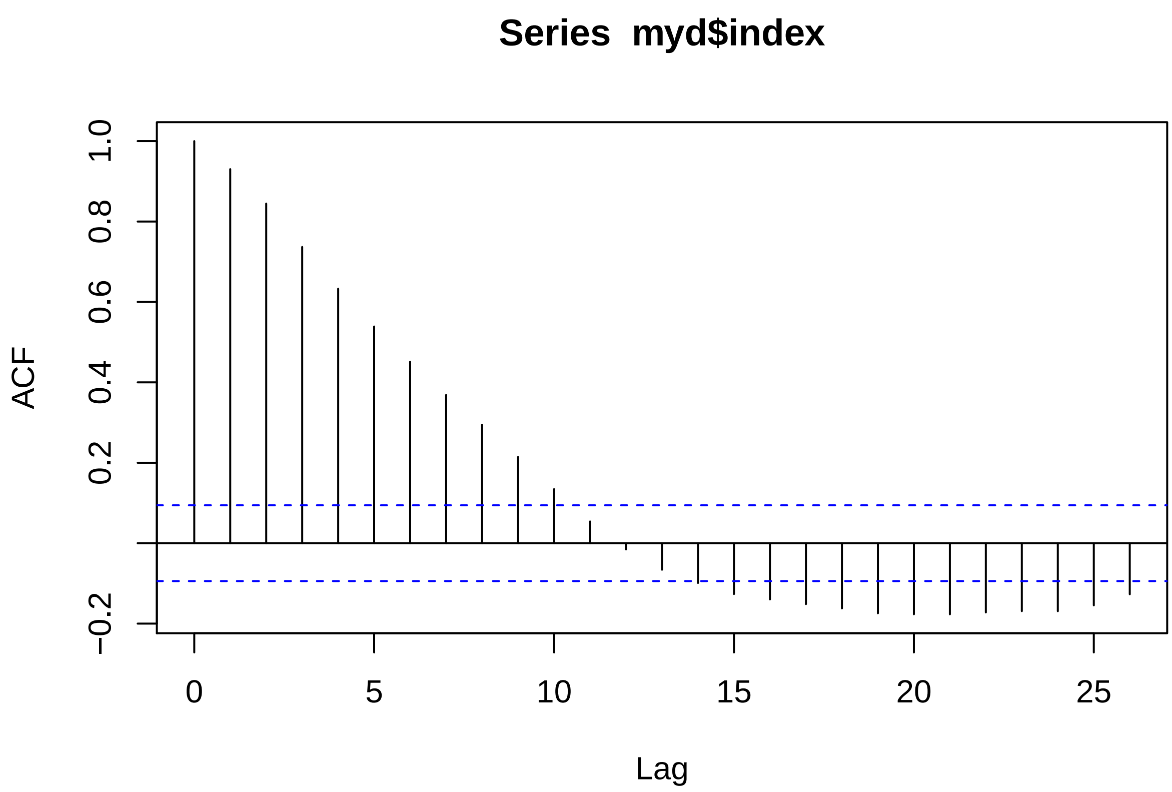
```r
plot(TSindex, ylab = 'the change of index')
```




```
# The plot shows periods of huge fluctuations from 1980 to 1985. The changes of PMI shows relatively st
# Another highlignted poing happend from 2008 to 2010.
```

c) Analyze if the time series is serially correlated using the ACF plot and the Ljung Box test

```r
acf(myd$index)
```



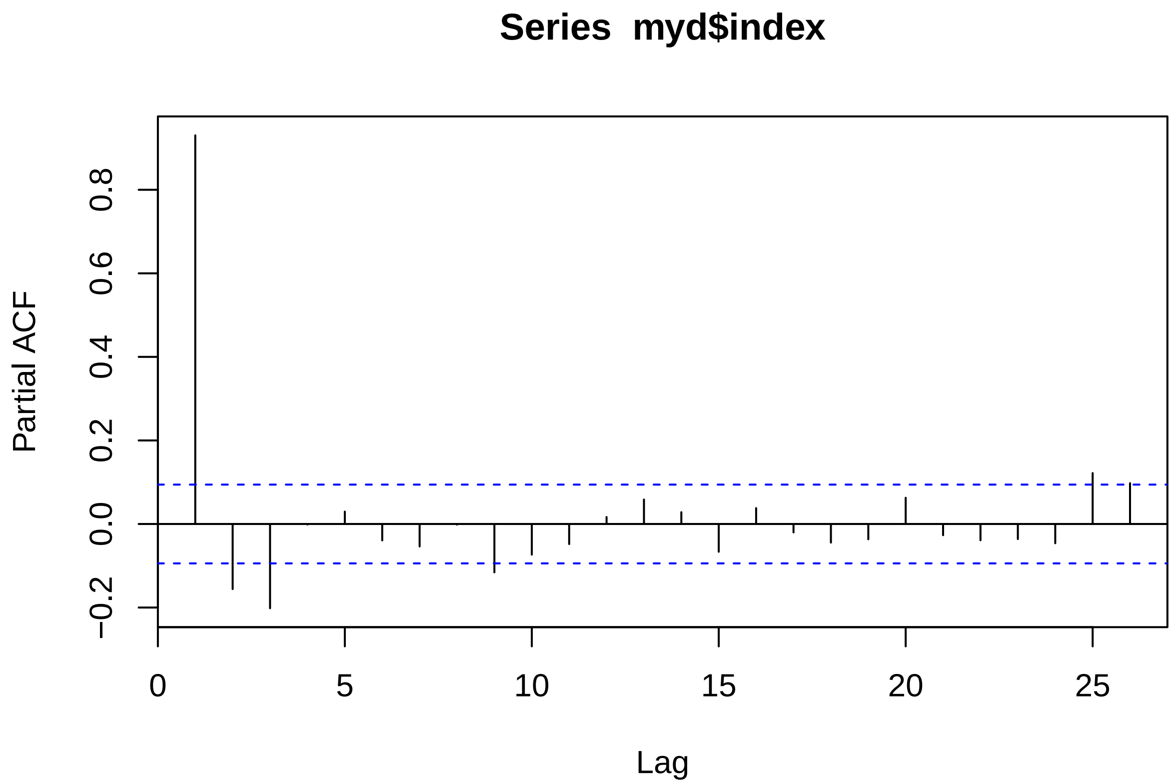
```r
# based on the plot, the serial correlation analysis shows that non-zero sample autocorrelations (signi

Box.test(myd$index, lag = 2, type = 'Ljung')
```

```
##
##  Box-Ljung test
##
## data:  myd$index
## X-squared = 687.59, df = 2, p-value < 2.2e-16
```

```r
# based on the result, it is serial correlation since we can reject the null hypothesis at df 2, and AC
```

d) Analyze the PACF plot and identify the order "p" of the AR(p) model.

```r
pacf(myd$index)
```




```r
# from the plot, the more suitable order for AR model would be AR(3), since the coefficient of correlat
```

e-1) Fit an adequate AR model:

```r
# Examine the significance of the model coefficirents, and discuss which coefficients are significantly
# PACF shows spikes at lag 1,2,3, and ACF shows a tapering patterns.
# so I will check for AR(1), AR(2), AR(3)
library(forecast)
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
## The following object is masked from 'package:timeSeries':
##
##     time<-
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
m1 = Arima(TSindex, order = c(1,0,0))
m1
```

```
## Series: TSindex
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1     mean
##       0.9307  51.2923
## s.e.  0.0172   1.4371
##
## sigma^2 estimated as 4.559:  log likelihood=-940.68
## AIC=1887.36   AICc=1887.42   BIC=1899.57
```

```r
coeftest(m1)
```

```
##
## z test of coefficients:
##
##            Estimate Std. Error z value  Pr(>|z|)
## ar1        0.930734   0.017219  54.053 < 2.2e-16 ***
## intercept 51.292312   1.437119  35.691 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
0.9307/0.0172
```

```
## [1] 54.11047
```

```r
# coefficient is very significant, but I will compare this model to AR(2)

m2 = Arima(TSindex, order = c(2,0,0))
m2
```

```
## Series: TSindex
## ARIMA(2,0,0) with non-zero mean
##
## Coefficients:
##          ar1      ar2     mean
##       1.0884  -0.1688  51.3567
## s.e.  0.0475   0.0475   1.2305
##
## sigma^2 estimated as 4.439:  log likelihood=-934.46
## AIC=1876.93   AICc=1877.02   BIC=1893.2
```

```r
coeftest(m2)
```

```
##
```

```
## z test of coefficients:
##
##            Estimate Std. Error z value  Pr(>|z|)
## ar1        1.088378   0.047524  22.902 < 2.2e-16 ***
## ar2       -0.168805   0.047537  -3.551 0.0003837 ***
## intercept 51.356730   1.230473  41.737 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
-0.1688/0.0475
```

```
## [1] -3.553684
```

```r
# coefficient is also significant, but this model has lower AIC

m3 = Arima(TSindex, order = c(3,0,0))
m3
```

```
## Series: TSindex
## ARIMA(3,0,0) with non-zero mean
##
## Coefficients:
##          ar1     ar2      ar3     mean
##       1.0534  0.0414  -0.1903  51.4142
## s.e.  0.0475  0.0707   0.0480   1.0251
##
## sigma^2 estimated as 4.293:  log likelihood=-926.75
## AIC=1863.51   AICc=1863.65   BIC=1883.85
```

```r
coeftest(m3)
```

```
##
## z test of coefficients:
##
##            Estimate Std. Error z value  Pr(>|z|)
## ar1        1.053415   0.047479 22.1868 < 2.2e-16 ***
## ar2        0.041382   0.070667  0.5856    0.5582
## ar3       -0.190335   0.048032 -3.9627 7.411e-05 ***
## intercept 51.414235   1.025118 50.1545 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
-0.1903/0.0480
```

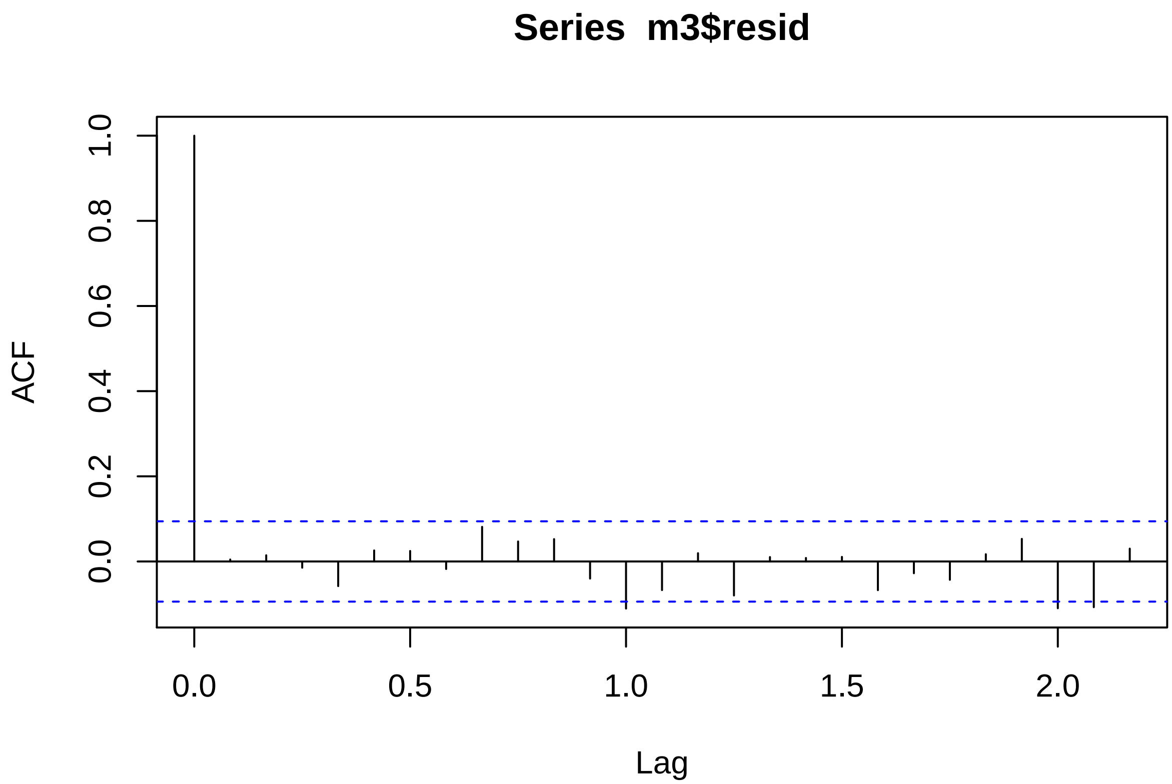
```
## [1] -3.964583
```

```r
# coeefficient is significant too, and this model has lowest AIC amongs the three models. However, t-te
# Since PACF recommended AR(3) is approppreate as shown in the plot, I decided to choose AR(3)
```

e-2). Perform a residual analysis and discuss if the selected model is adequate

i. Compute ACF functions of residuals

```r
acf(m3$resid)
```




```
# m3 has few significant correlations. so that it is very close to independent.
```

ii. Test if residuals are white noise.

```r
Box.test(m3$resid, lag = 3, type = 'Ljung', fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  m3$resid
## X-squared = 0.19705, df = 1, p-value = 0.6571
```

```r
Box.test(m3$resid, lag = 6, type = 'Ljung', fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  m3$resid
## X-squared = 2.2319, df = 4, p-value = 0.6932
```

```r
Box.test(m3$resid, lag = 9, type = 'Ljung', fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  m3$resid
## X-squared = 6.2766, df = 7, p-value = 0.5079
```

```r
Box.test(m3$resid, lag = 12, type = 'Ljung', fitdf = 2)
```

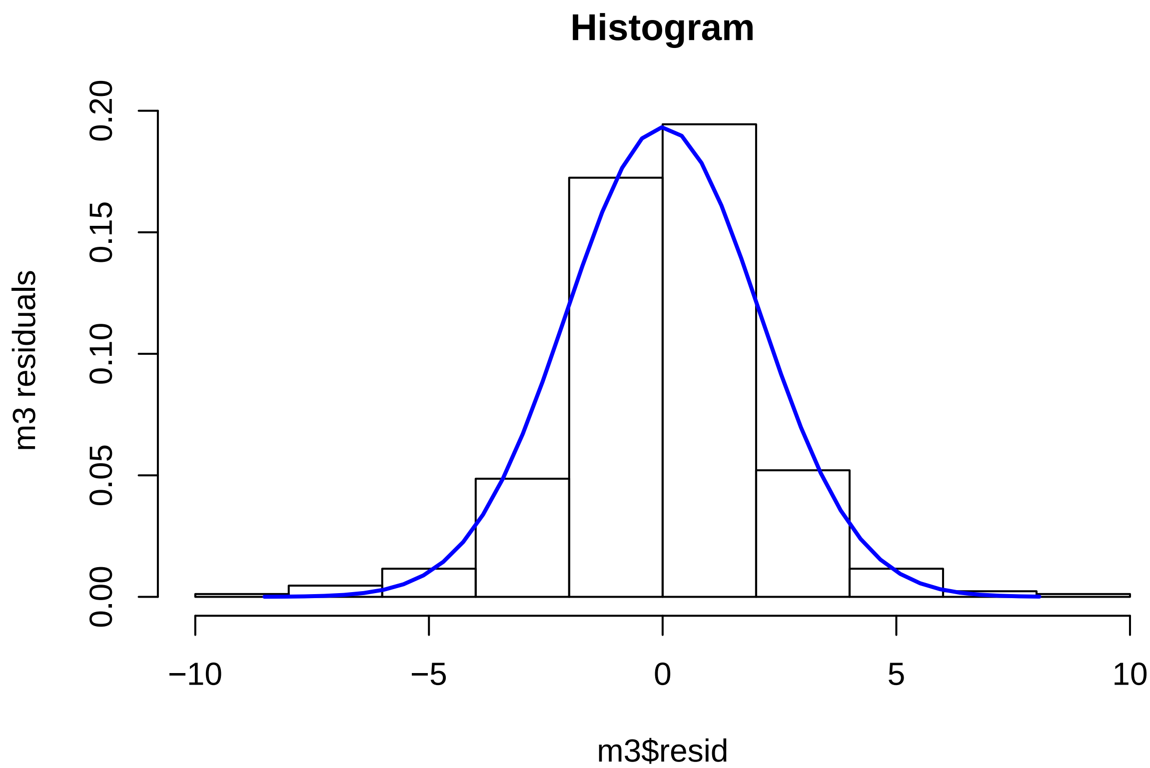
```
##
```

```
##  Box-Ljung test
##
## data:  m3$resid
## X-squared = 13.687, df = 10, p-value = 0.1877
```

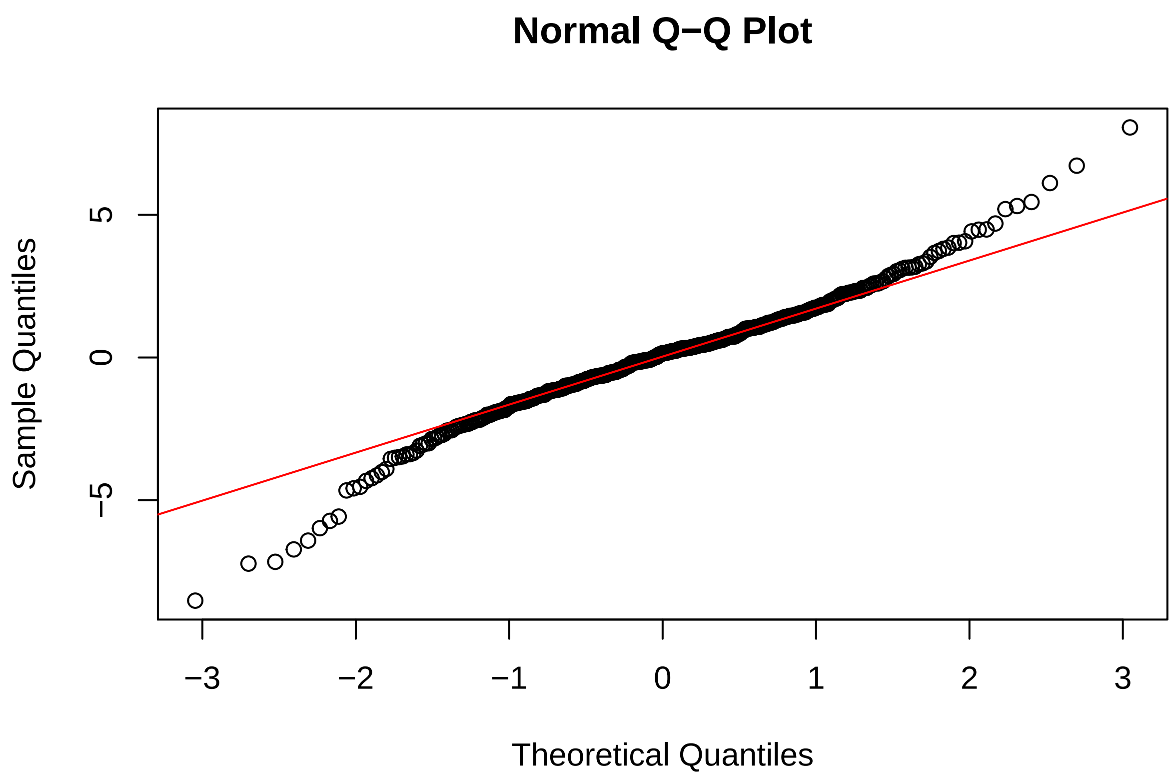
```
# m3 residuals are white noise since the Ljung box test show that p-values for tests with m = 3,6,9,12,
# indicating that there is no evidence of serial correlation in the residuals.
```

iii. plot histogram and normal quantile plots of residuals

```
# using these methodes, I can clarify which model is more suitable.
# histogram for m3
hist(m3$resid, ylab = 'm3 residuals', prob = T, main = 'Histogram')
xfit <- seq(min(m3$resid), max(m3$resid), length = 40)
yfit <- dnorm(xfit, mean = mean(m3$resid), sd = sd(m3$resid))
lines(xfit, yfit, col ="blue", lwd = 2)
```




```
# qq-plot for m3
qqnorm(m3$resid)
qqline(m3$resid, col =2)
```




```
# histogram for m3 show that the residuals are normal.However, qq-plot for m3 is very closer to a norma
```

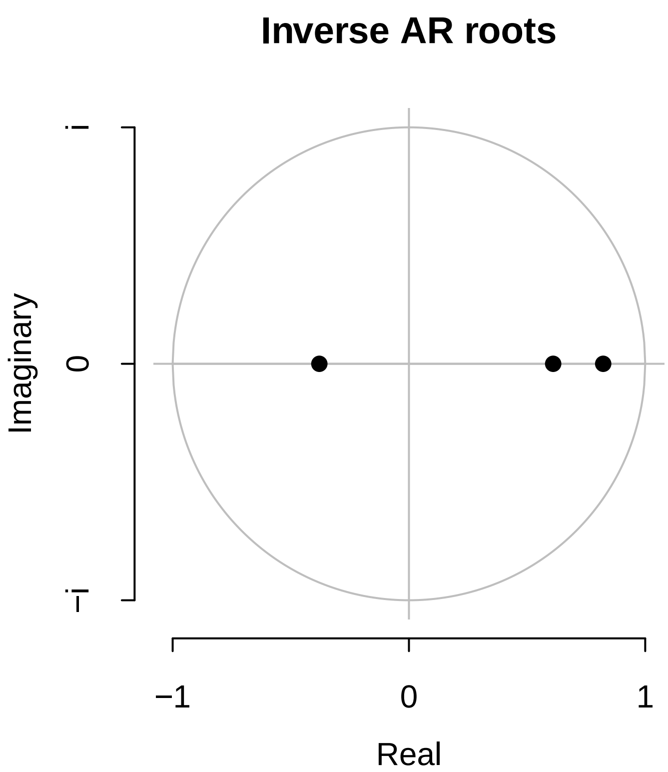
f) Discuss the result of your residual analysis, and draw conclusions on whether the selected model is appropriate to describe the time behavior of the PMI index

According to the results, I selected an appropriate model. Firstly, m3resid plot shows that there are few autocorrelations. qq-norm and qq-line indicate that residuals are normal. Based on these reasons, I can conclude that m3 model is pretty good for my analysis.

g) write down the expression of the estimated AR(p) model, and check if the AR model represents a stationary process.

X_t - 51.414235 = 1.053415(x_(t-1) - 51.414235) - 0.041382(x_(t-2) - 51.414235) - 0.190335(x_(t-3) - 51.414235) a_t1. By using wolframalpha's solutions of 1 - 1.053415x - 0.041382x^2 + 0.190335x^3 = 0, I got x = -2.63697, x = 1.21629, and x = 1.6381 Thus, characteristic roots are w1 = 1/-2.63697 = -0.3792231, w2 = 1/1.21629 = 0.8221723, and w3 = 1/1.6381 = 0.6104633. All three characteristic roots are less than 1. Therefore, AR(3) represents a stationary process.

```
plot(m3)
```



h) Compute up to 5-step ahead forecasts with origin at the end of the data, i.e. 12/01/2015. Write down the forecasts and their 95% prediction intervals

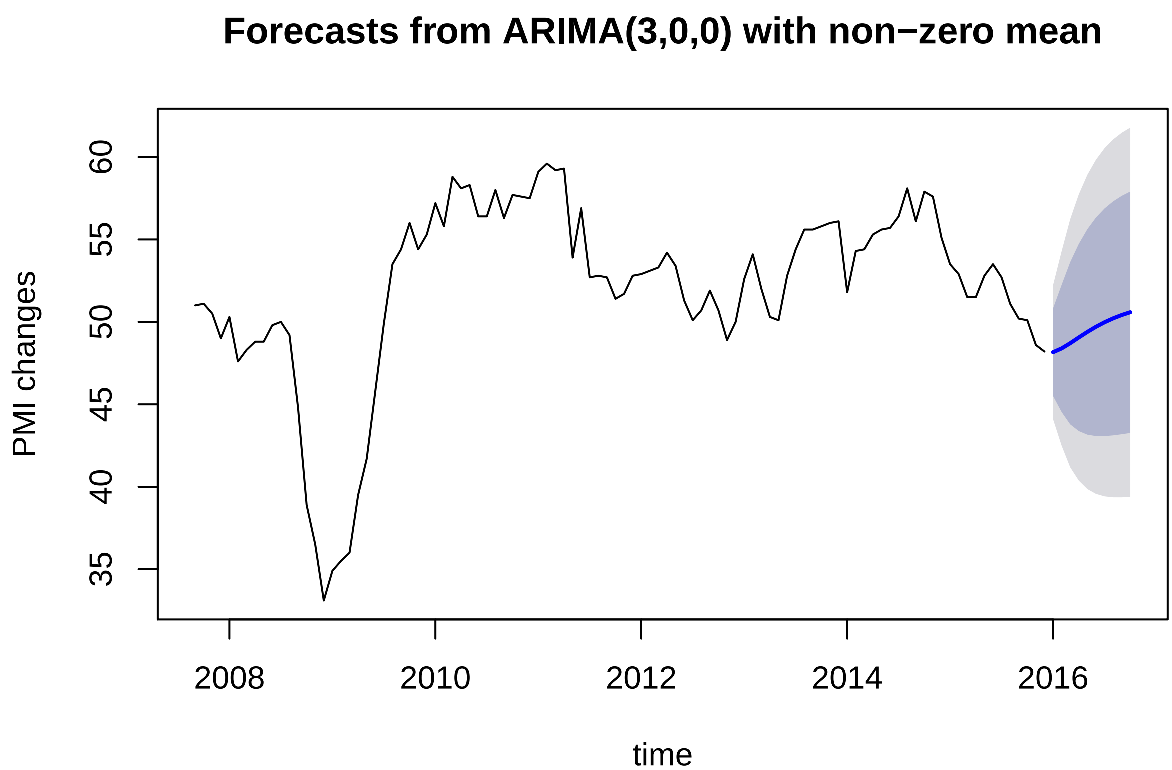
```r
# forecasts with h = 5.
# I keep seeing this messages (Error: could not find function "forecast.Arima"). After some googling, I
# forecast() produces forecasts and 80%~90% prediction intervals.
f = forecast(m3, h=5)
f
```

```
##          Point Forecast    Lo 80    Hi 80    Lo 95    Hi 95
## Jan 2016       48.16200 45.50679 50.81720 44.10121 52.22278
## Feb 2016       48.39092 44.53430 52.24753 42.49273 54.28910
## Mar 2016       48.70662 43.78580 53.62745 41.18087 56.23238
## Apr 2016       49.05590 43.37937 54.73244 40.37439 57.73742
## May 2016       49.39333 43.15998 55.62669 39.86024 58.92642
```

```r
#forecast() provides that 48.16200 in Jan 2016, 48.39092 in Feb 2016, 48.70662 in Mar 2016, 49.05590 in
```

i) Plot the 10-step ahead forecasts and discuss whether the forecasts exhigit a trand that is consistent with the observed dynamic behavior of the process

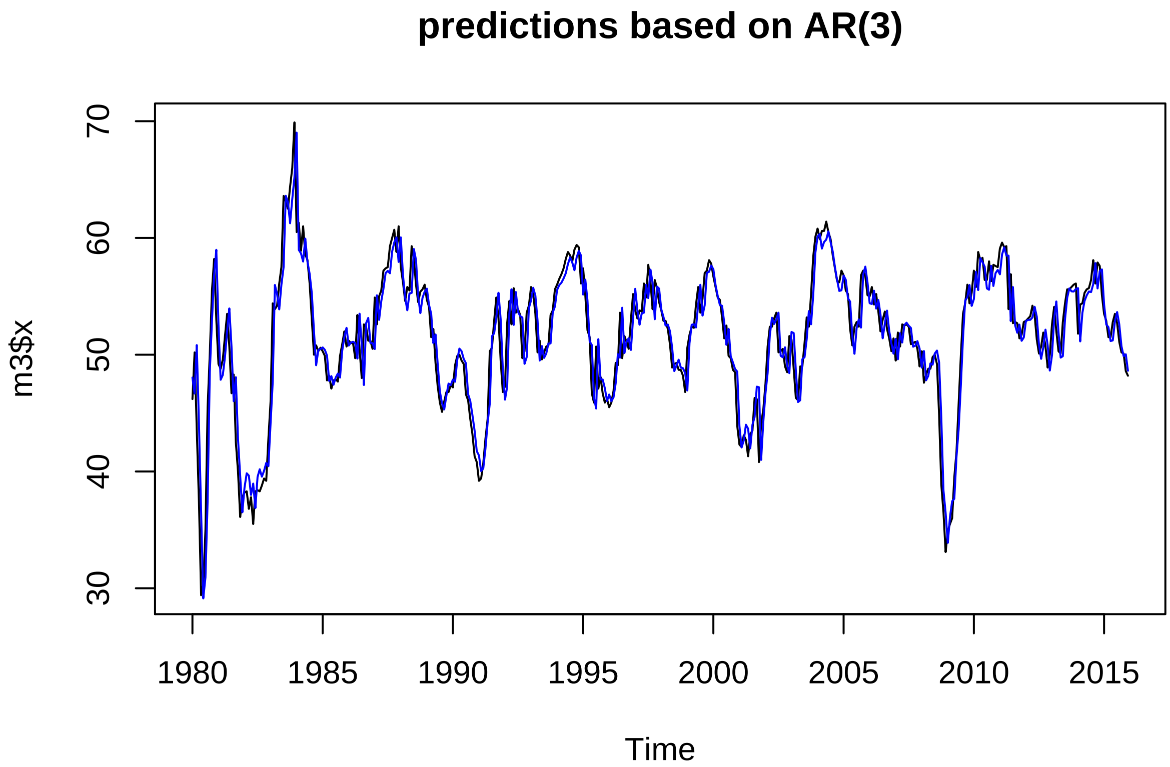
```r
plot(forecast(m3, h=10), include = 100, xlab = "time", ylab = "PMI changes")
```




```
# the plot shows that out-of-sample predictions based on AR(3) model
```

j) A PMI reading above 50% indicates that the manufacturing economy is generally expanding; below 50% that it is generally declining. Do the model forecasts predict that manufacturing economy is generally expanding or contracting?

```
plot(m3$x, main = "predictions based on AR(3)")
lines(fitted(m3), col = "blue")
```




```
# Based on the graph, there has been a huge decrease about every 10 years. After 2015, I anticipate tha
```

k) what do the model forecasts converge to?

The model converges to 50 based on the plot above