Load libraries and data

```
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
```

```
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
```

```
setwd("~/Desktop/CSC425/hwork4/")
myd = read.table("onlinesales.csv", header = T, sep = ',')
head(myd)
```

```
##         date sales
## 1 1999-10-01  5263
## 2 2000-01-01  5556
## 3 2000-04-01  6062
## 4 2000-07-01  6891
## 5 2000-10-01  9070
## 6 2001-01-01  7874
```

```
tail(myd)
```

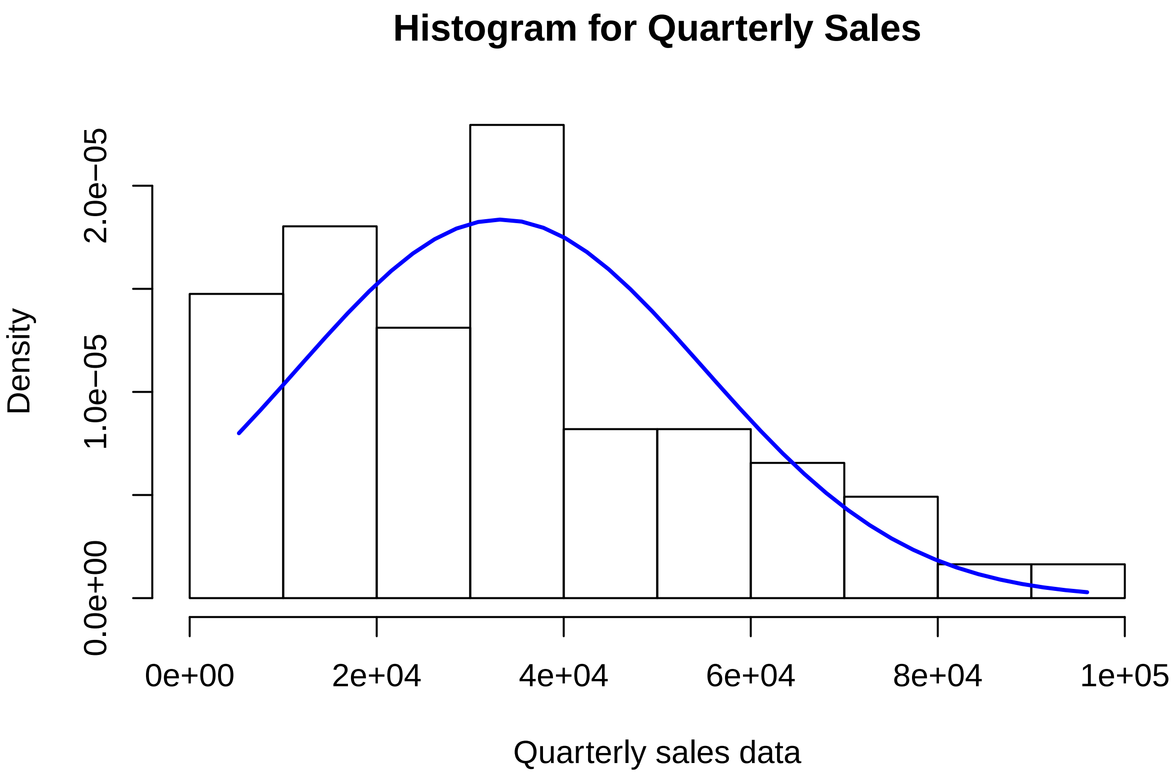
```
##          date sales
## 56 2013-07-01 61857
## 57 2013-10-01 83709
## 58 2014-01-01 66938
## 59 2014-04-01 70134
## 60 2014-07-01 71862
## 61 2014-10-01 95979
```

a) Analyze the distribution of the quarterly sales data. Are the data normally distributed? Provide appropriate statistics, tests and grahps to support your conclusions.

```
x = myd$sales
xts = ts(x,start = c(2000,1), frequency = 4)
```

histogram

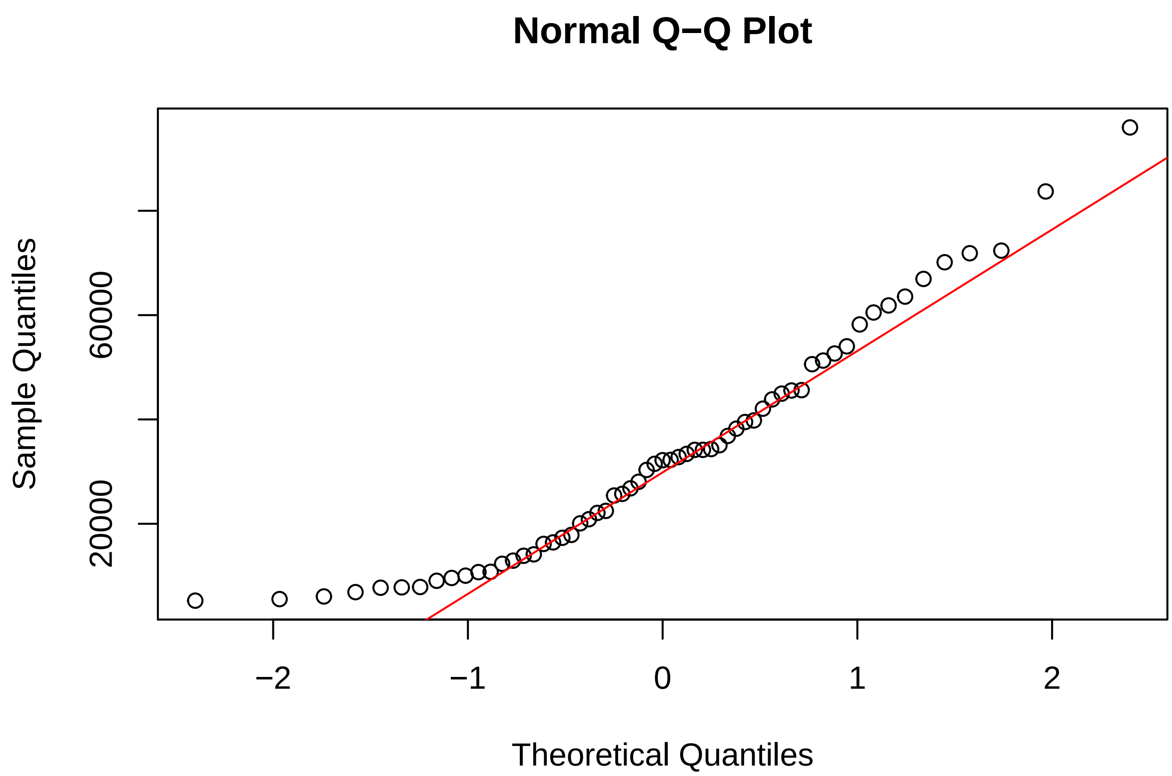
```
hist(x, xlab = 'Quarterly sales data', main = "Histogram for Quarterly Sales", prob = T)
xfit<-seq(min(x), max(x), length = 40)
yfit<-dnorm(xfit, mean = mean(x), sd = sd(x))
lines(xfit, yfit, col = "blue", lwd = 2)
```



It has a fat tail to the right and hard to say it has a perfect-normal distribution. However, it looks like it is very close to the normal distribution.

qq-line

```
qqnorm(x)
qqline(x, col = 2)
```



Many outliers are detected on the graph. However, it looks like it is very close to the normal distribution.

Perform Jarque-Bera normality test

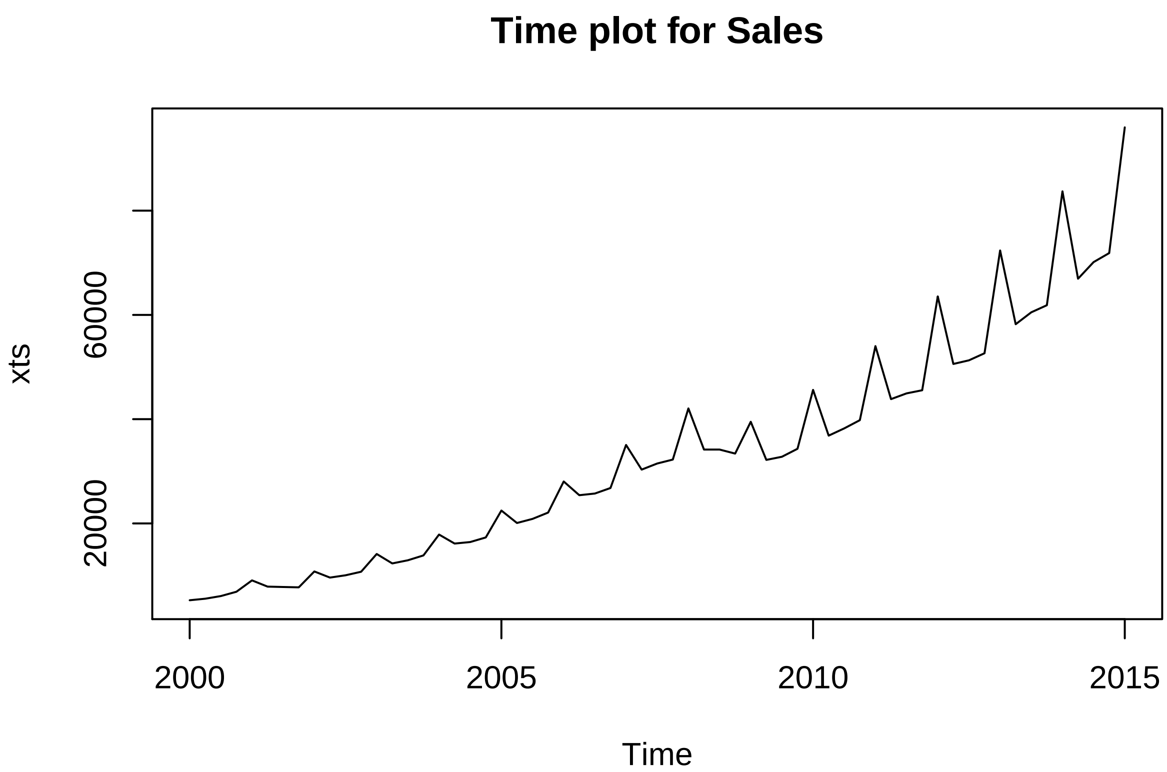
```r
normalTest(xts, method = c("jb"))
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 5.525
##   P VALUE:
##     Asymptotic p Value: 0.06313
##
## Description:
##  Wed Nov  1 00:53:34 2017 by user:
```

The result shows that the p-value is not very significant as of 0.063. Thus, We cannot reject the null hypothesis. Therefore, it is normal.

b) Create a time plot for online sales. Does the plot show variations in variance? Discuss the trends displayed in the plot.

```r
plot(xts, main = 'Time plot for Sales', type = 'l')
```

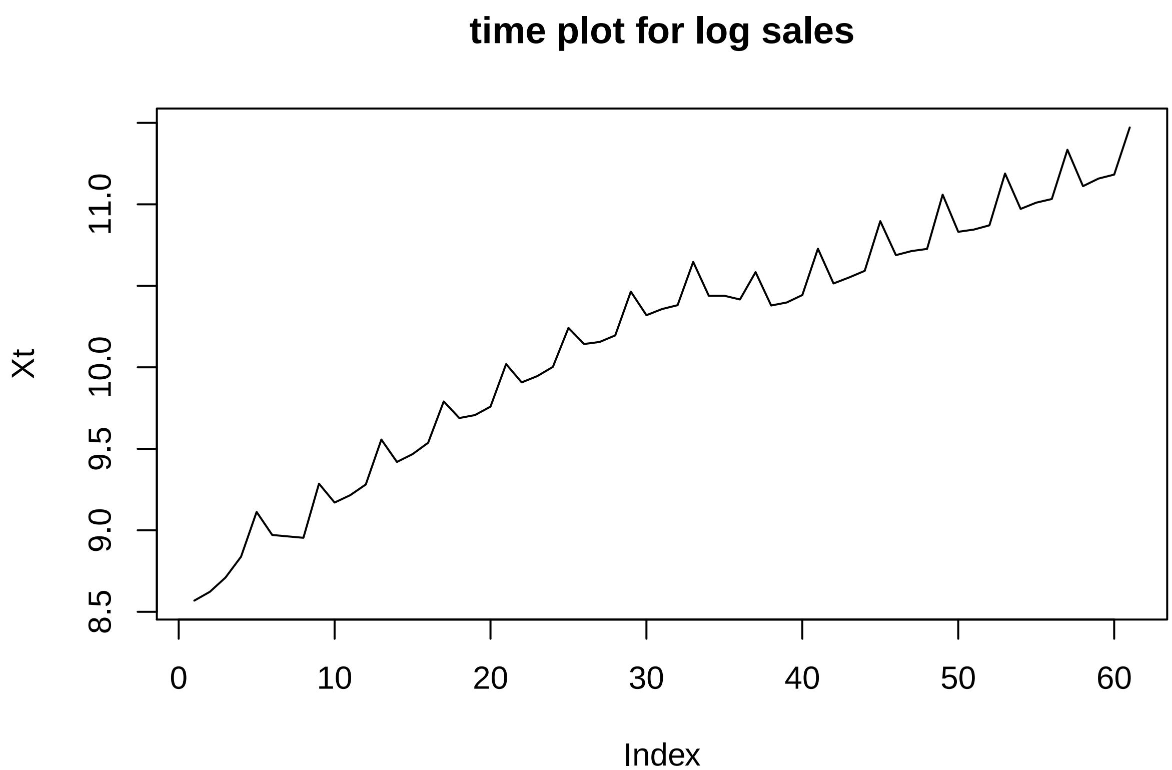


the graph shows that the it has a upward trend as well as variation is not constant over time.

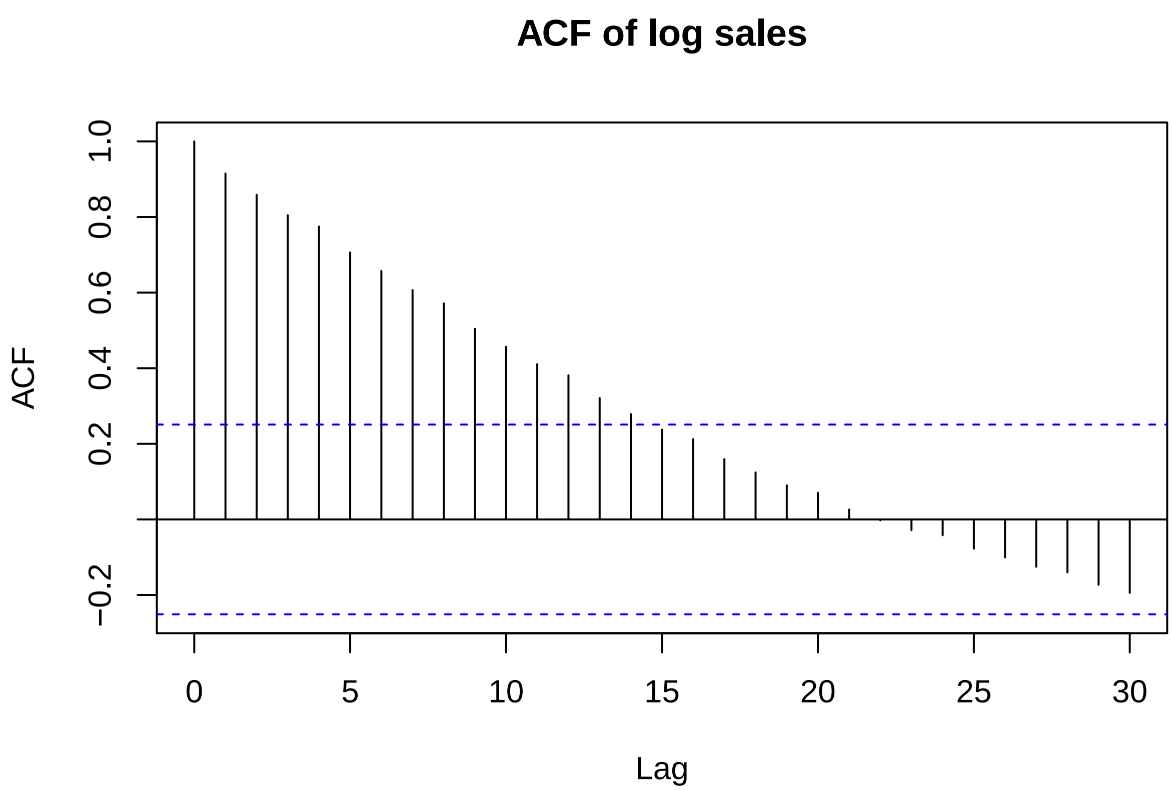
c) Analze the autocorrleation function of Xt and of its first difference. is the differenced time series stationary? Is there any evidence of a seasonal effect in the data?

log transformation

```
Xt = log(x)
plot(Xt, main = 'time plot for log sales', type = 'l')
```



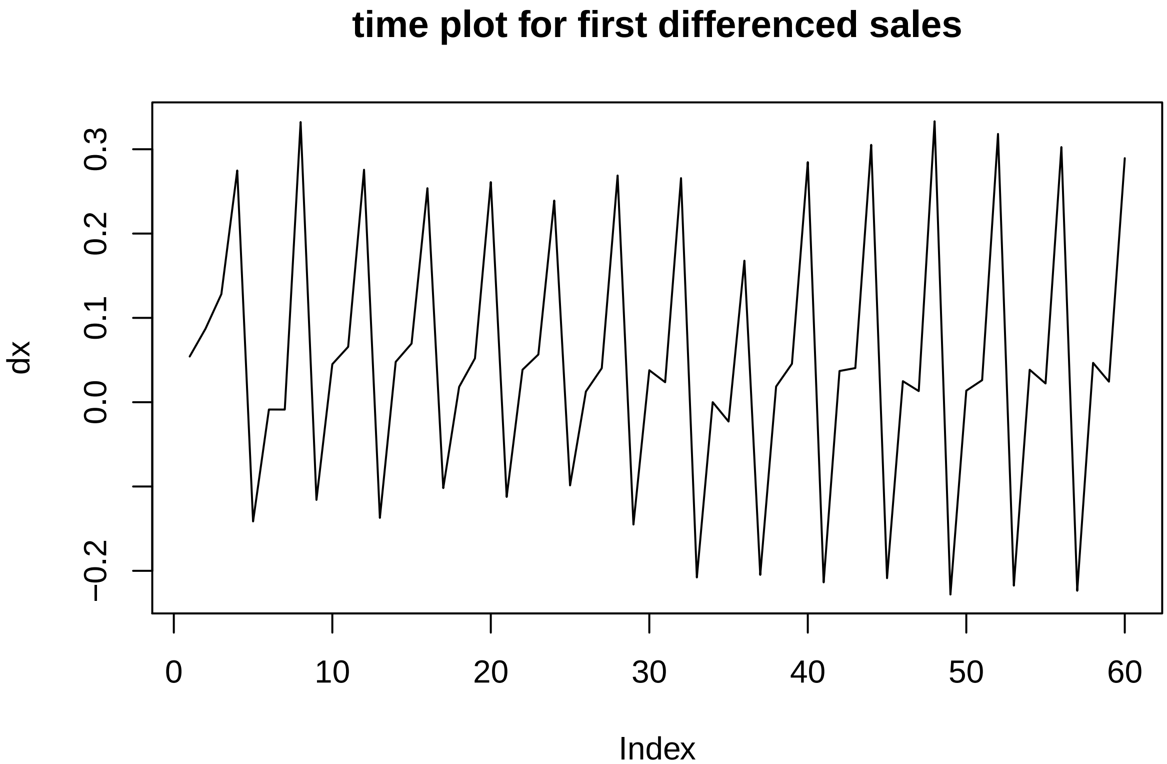

```r
lnts = ts(Xt, start=c(2000,1), frequency = 4)
acf(as.vector(lnts), lag.max= 30, main ="ACF of log sales")
```



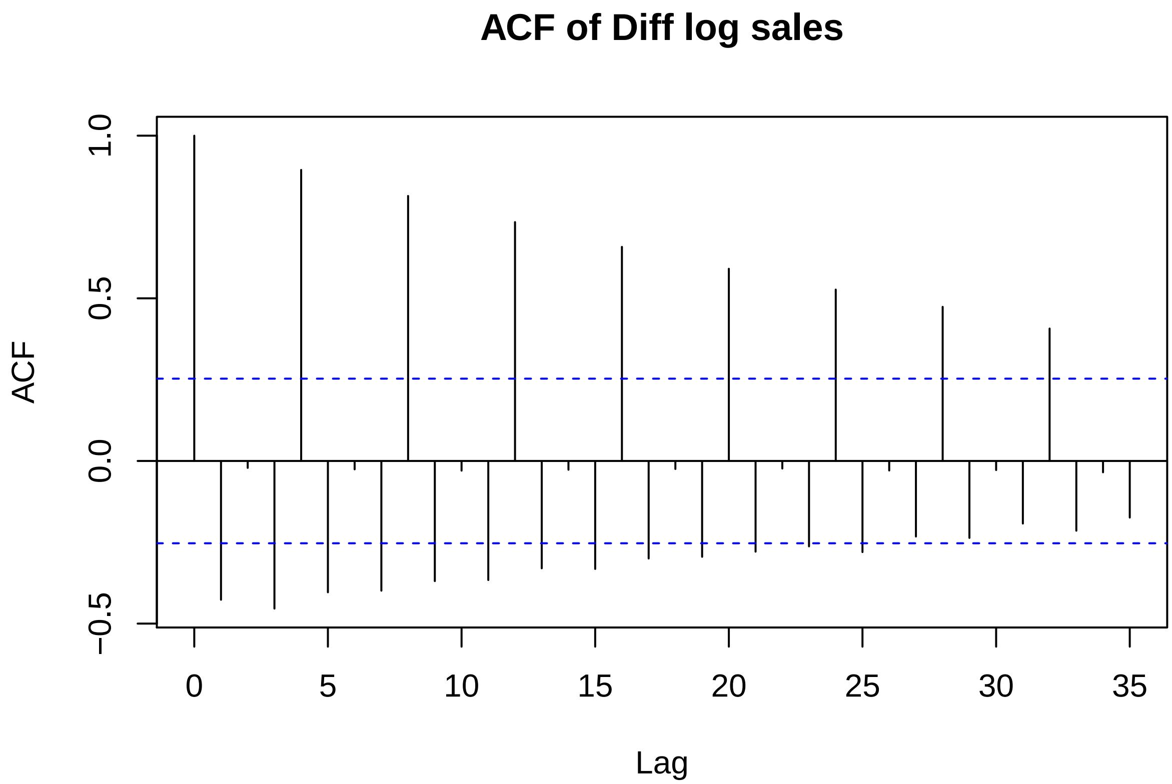
ACF for log sales is decaying slowly. This indicates a non-stationary series. Thus, we need to see the first differnce

first difference

```
dx = diff(Xt)
plot(dx, main = 'time plot for first differenced sales', type = 'l')
```



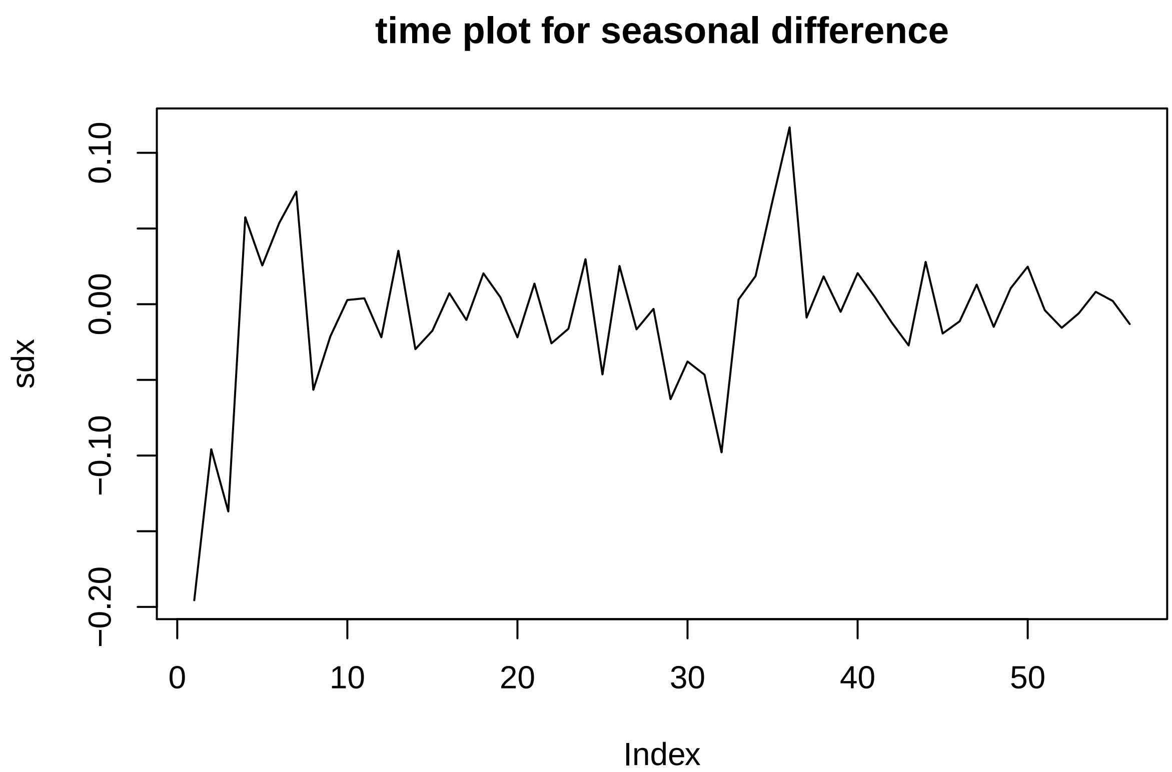

```
dxts= ts(dx, start=c(2000,1), frequency = 4)
acf(as.vector(dxts), lag.max = 35, main = "ACF of Diff log sales")
```



ACF is alternating while it is slowly decaying around near zero. Thus, first difference results of a non-stationary as well.

d) After de-trending and de-seasonalizing the time series, do you obtain a stationary time series?

```
sdx = diff(dx,4)
plot(sdx, main = 'time plot for seasonal difference', type = 'l')
```

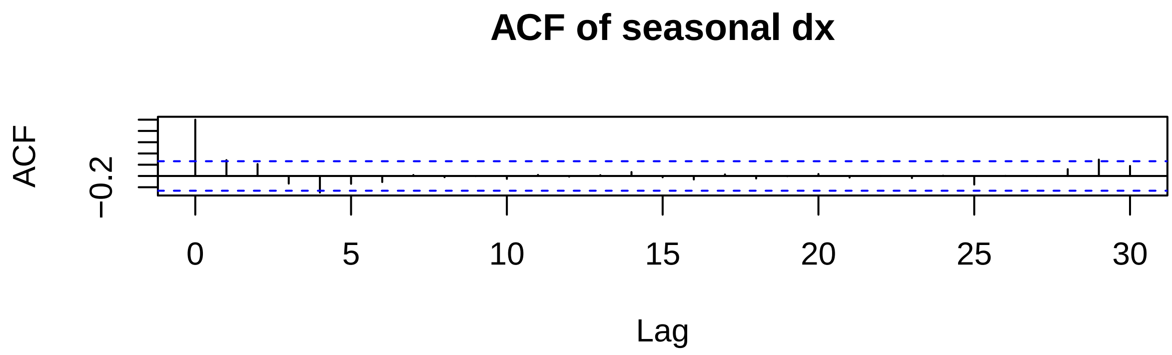


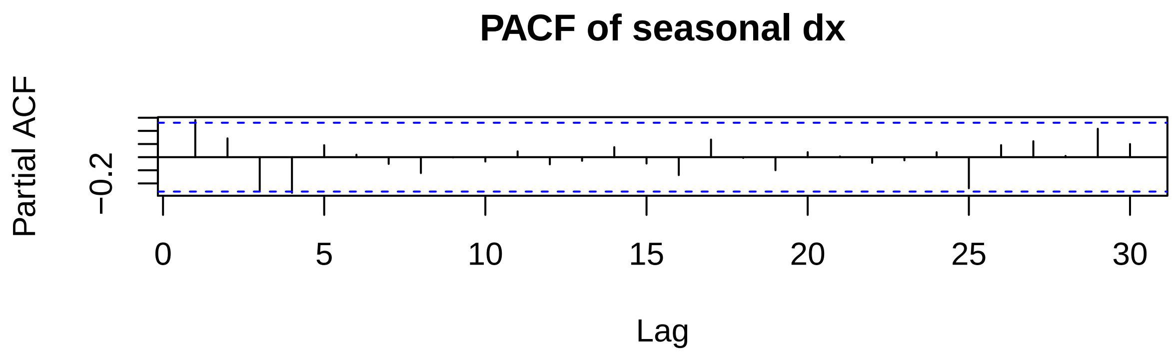

```
sdxts = ts(sdx, start = c(2000,1), frequency = 4)
```

After the seasonal difference, ACF is now stationary.

check ACF and PACF

```
par(mfcol = c(2,1))
acf(as.vector(sdxts), lag.max = 30, main= "ACF of seasonal dx")
pacf(as.vector(sdxts), lag.max = 30, main= "PACF of seasonal dx")
```





e) Identify a seasonal multiplicative model that explains the dynamic behavior of the process Xt. Evaluate the goodness of fit of your model using residual analysis. Explain the steps of your analysis and how you identified the final model

For non-seasonal behavior, the spikes at lag 1 in both ACF and PACF indicates MA(1) and AR(1). For seasonal behavior, the spike at lag 4 in both ACF and PACF indicates seasonal MA(1) and seasonal AR(1). so my seasonal arima model would be (1,1,1)x(1,1,1) period = 4

fit the model

```
m1 = Arima(xts, order = c(1,1,1), seasonal = list(order = c(1,1,1), period = 4), method = 'ML')
m1
```

```
## Series: xts
## ARIMA(1,1,1)(1,1,1)[4]
##
## Coefficients:
##           ar1     ma1    sar1     sma1
##       -0.9654  0.9229  0.9920  -0.9319
## s.e.   0.0704  0.1179  0.0545   0.2383
##
## sigma^2 estimated as 1914582:  log likelihood=-484.52
## AIC=979.03   AICc=980.23   BIC=989.16
```

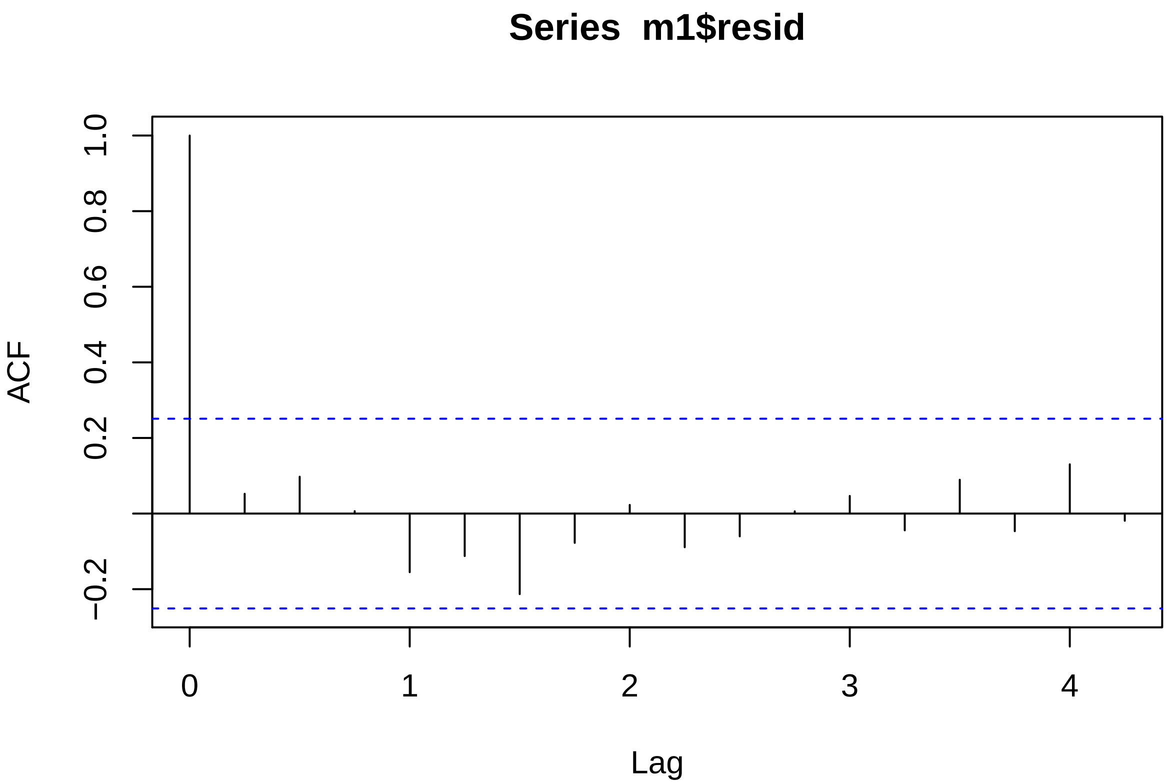
```
coeftest(m1)
```

```
##
## z test of coefficients:
##
##       Estimate Std. Error  z value  Pr(>|z|)
## ar1  -0.965368   0.070377 -13.7171 < 2.2e-16 ***
## ma1   0.922927   0.117914   7.8271 4.991e-15 ***
## sar1  0.992008   0.054537  18.1895 < 2.2e-16 ***
## sma1 -0.931935   0.238290  -3.9109 9.194e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

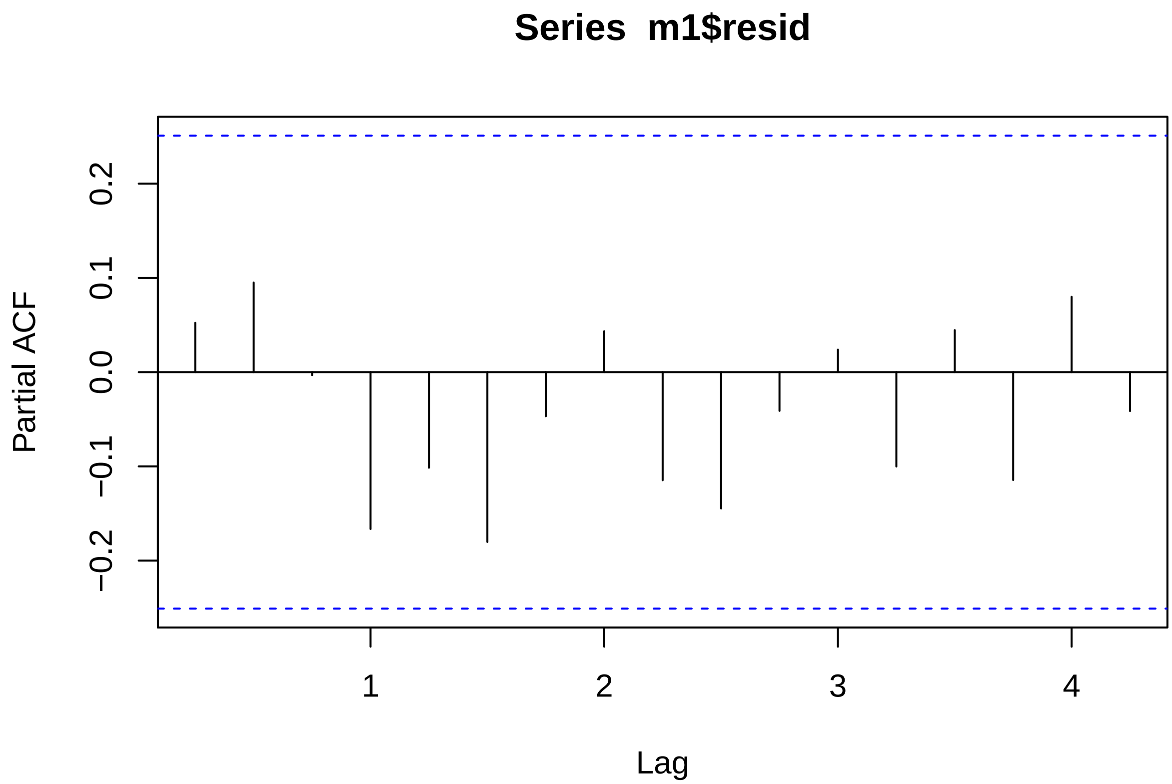
ma1 is not significant at all so it can be removed. Therefore, I will revise model 1 by get rid of MA(1)

residuals analysis

```r
acf(m1$resid)
```




```r
pacf(m1$resid)
```



ACF and PACF show a White Noise

Ljung box test on residuals

```r
Box.test(m1$residuals, 6, "Ljung-Box", fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 6.4622, df = 4, p-value = 0.1672
```

```r
Box.test(m1$residuals, 12, "Ljung-Box", fitdf = 2)
```

```
##
##  Box-Ljung test
##
## data:  m1$residuals
## X-squared = 7.9634, df = 10, p-value = 0.6324
```

by Ljung-box test, residuals cannot reject hyphothesis of white noise.

f)

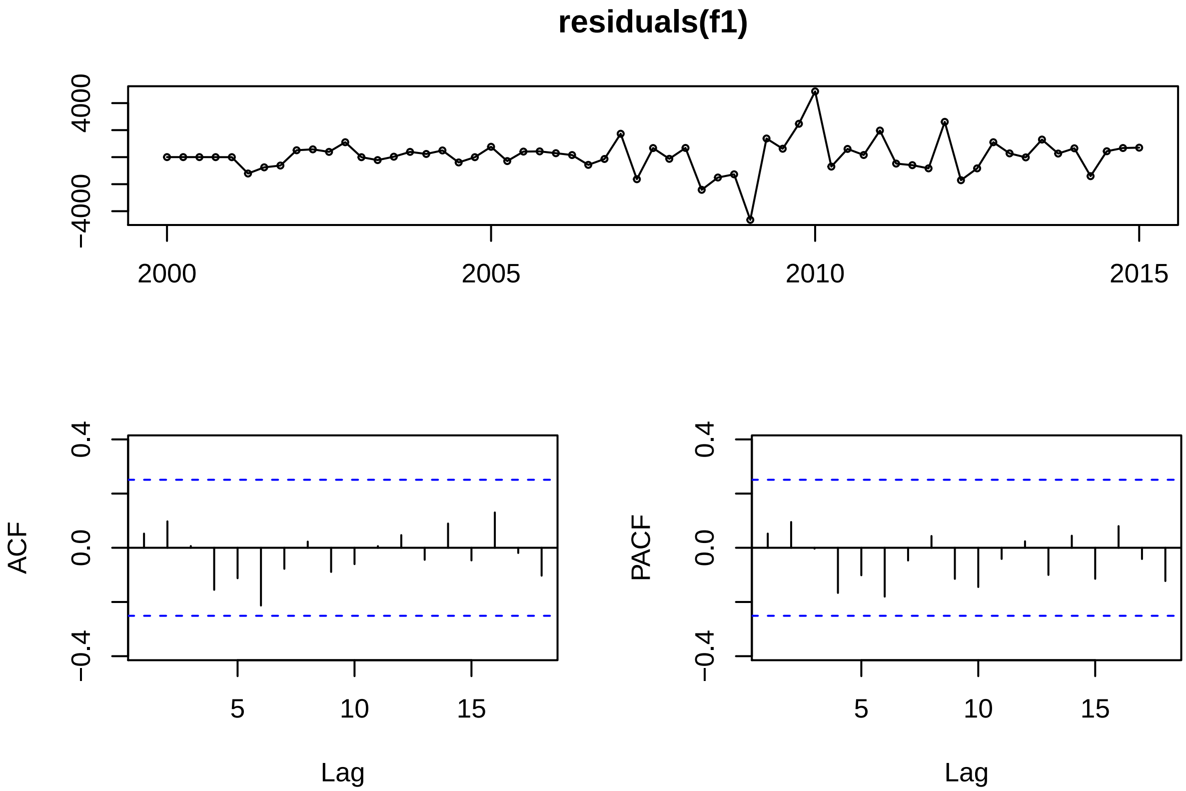
(1+0.9654B) * (1-0.9920B^4) * (1-B) * (1-B^4) Xt = (1+0.9229B) * (1-0.9319B^4) at

g) Compute forecasts for sales for the next 4 quarters.

```
f1 = forecast(m1, h=4)
f1
```

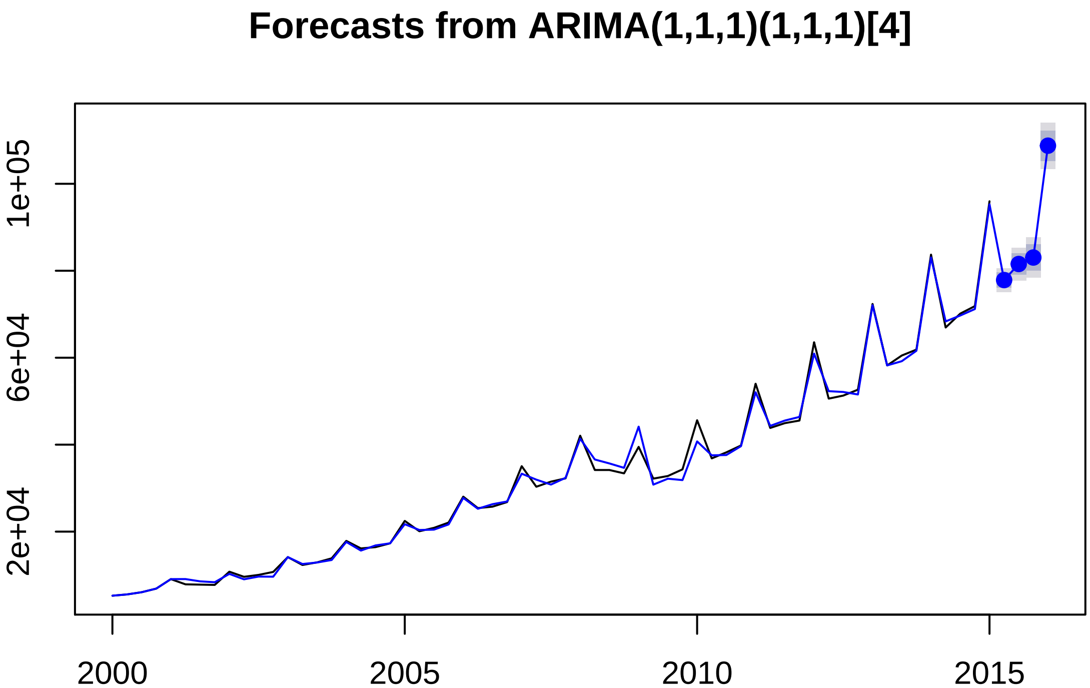
```
##         Point Forecast     Lo 80     Hi 80     Lo 95     Hi 95
## 2015 Q2       77855.11  76072.60  79637.63  75128.99  80581.23
## 2015 Q3       81549.86  79079.22  84020.51  77771.33  85328.40
## 2015 Q4       83040.16  79995.26  86085.05  78383.39  87696.92
## 2016 Q1      108762.32 105268.39 112256.25 103418.82 114105.83
```

```
tsdisplay(residuals(f1))
```



h)

```
plot(f1, include = 100)
lines(ts(c(f1$fitted, f1$mean), frequency = 4, start = c(2000,1)), col = "blue")
```



forecasts are consistent since it fluctuates around zero, and it slightly has a upward trend.