CSC 425 Time Series Analysis, HW1

Akhil Kumar Ramasagaram

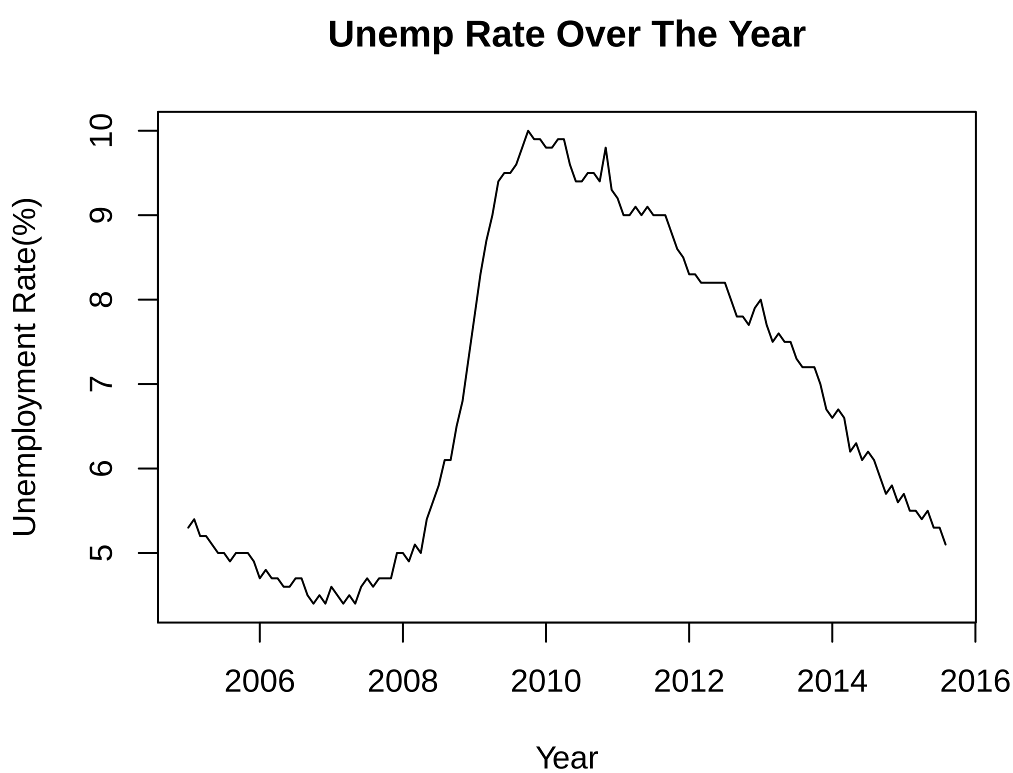
Friday, September 25, 2015

Part I

Consider the monthly US unemployment rates from January 2005 to August 2015. The data file contains dates (date) and unemployment rates (rate).

a) Create a time plot for the observed unemployment rates. Make sure the plot is correctly labeled and titled. Analyze the time trend displayed by the plot, and discuss if data show any striking pattern, such as upward/downward trends or seasonality?

```r
library(fBasics)
library(zoo)
library(tseries)
unemp_data <- read.csv("unemp_rate_m2005_2015.csv")
unemp_rate <- unemp_data$rate
unemp_ts <- ts(unemp_rate, start=c(2005,1), freq=12)
plot(unemp_ts, xlab = "Year", ylab = "Unemployment Rate(%)",
     main = "Unemp Rate Over The Year")
```



According to the graph, it look there is a huge surge and from 2008 to 2010 in the un employment rate followed by a decline trend from 2010 to 2014. This can be the 2008 American economic crisis. We cannot conclude if this is a seasonality or not unless we hypothesize this trend it over entire population.

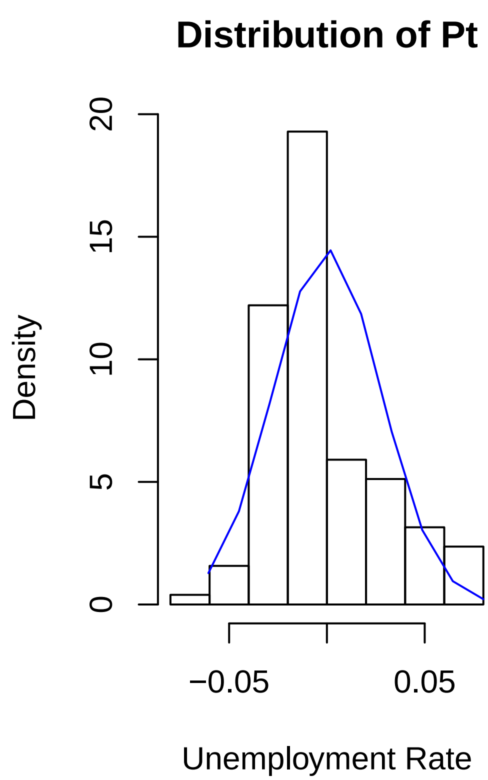
b) Compute the percentage change in unemployment rates using the formula. $P_t = (X_t - X_{t-1})/X_{t-1}$

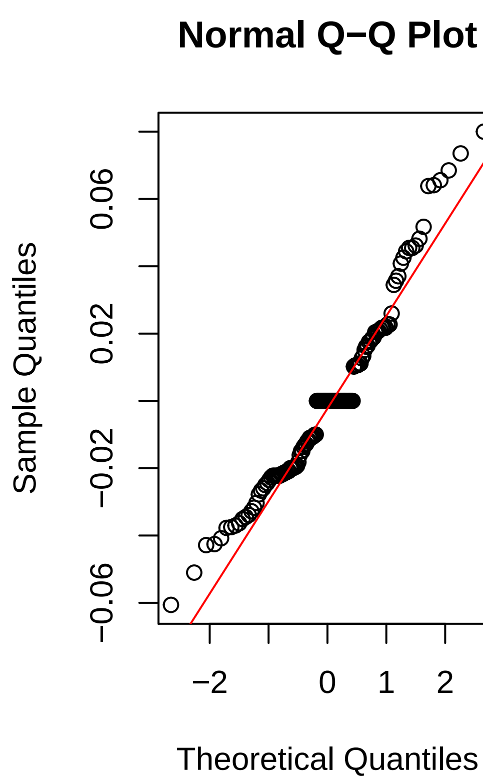
```
unemp_lag <- lag(unemp_ts, k=-1)
pt <- (unemp_ts - unemp_lag)/unemp_lag
head(pt)
```

```
## [1]  0.01886792 -0.03703704  0.00000000 -0.01923077 -0.01960784  0.00000000
```

c) Analyze the distribution of P_t using a histogram and a normal quantile plot. Is the distribution of P_t symmetric? Is it close to a normal distribution?

```
pts <- ts(pt, start=c(2005,1), freq=12)
par(mfrow = c(1,2))
hist(pts, xlab="Unemployment Rate", prob=TRUE,
     main="Distribution of Pt")
x_fit <- seq(min(pts), max(pts), length=10)
y_fit <- dnorm(x_fit, mean=mean(pts), sd=sd(pts))
lines(x_fit, y_fit, col="blue")
qqnorm(pts)
qqline(pts, col = 2)
```

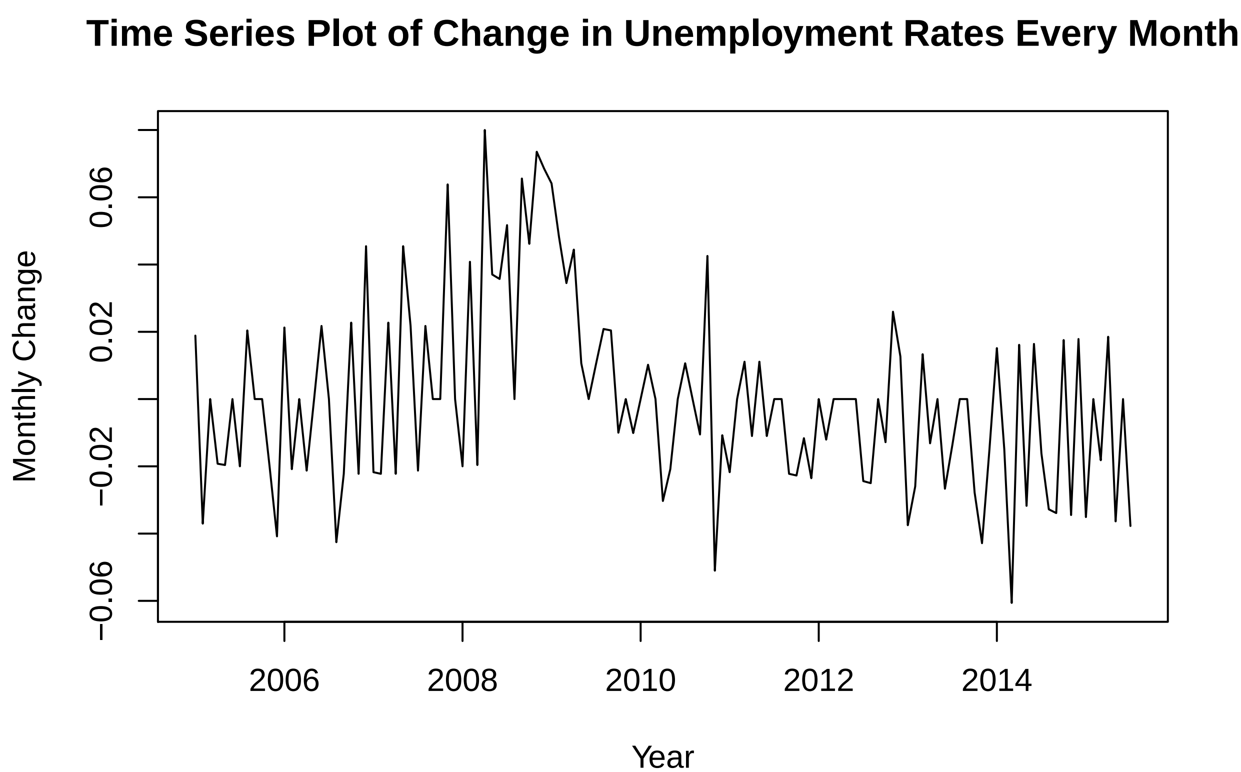




The Distribution is skewed and not normal.

d) Create a time plot for the time series of P_t. Analyze the time trend displayed by the plot, and discuss if data show any striking pattern, such as upward/downward trends or seasonality?

```
plot(pts, xlab = "Year", ylab = "Monthly Change",
     main = "Time Series Plot of Change in Unemployment Rates Every Month")
```



As you can see from the plot, for most part the time series the trend is almost constant and then there is the surge believed to be the 2008 american crisis.

e) Test the null hypothesis of perfect symmetry for the distribution of P_t at 5% significance level.

```
skew_test = skewness(pts)/sqrt(6/length(pts))
2* (1-pnorm(abs(skew_test)))[1]
```

```
## [1] 0.002176524
```

This 5% significance, and p = 0.0021765. We reject the H_o

f) Test the null hypothesis of excess kurtosis equal to zero (normal tails) at 5% significance level.

```
k_stat = kurtosis(pts)/sqrt(24/length(pts))
2*(1-pnorm(abs(k_stat)))[1]
```

```
## [1] 0.5485466
```

With Kurtosis test valued at 0.5999395, we can assume that the tails are a bit thicker. On the other hand with p = 0.5485466 with 5% significance. We reject the H_o

g) Test the hypothesis of normality for the distribution of P_t using the Jarque-Bera test at 5% level.

```
normalTest(pts,method=c("jb"))
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 10.1375
##   P VALUE:
##     Asymptotic p Value: 0.00629
##
## Description:
##  Fri Sep 25 14:26:31 2015 by user: Akhilkumar
```

with 5% significance, we reject the H_o ## Part II

The dataset "oranges.csv" contains monthly prices in US dollars for metric tons of oranges. Data are available at the Data World Bank http://data.worldbank.org/. The dataset contains the variable date and the variable price

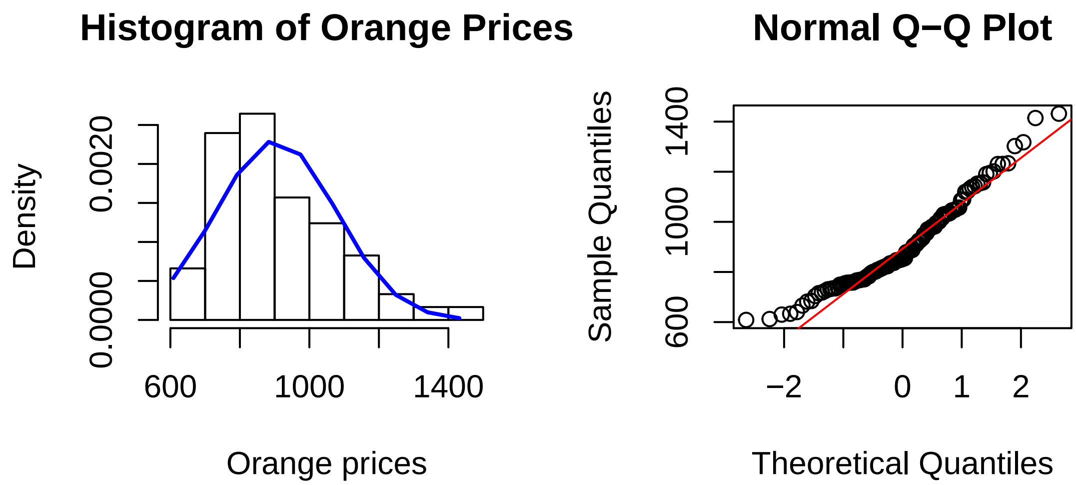
a) Analyze the distribution of monthly prices of oranges using a histogram and a normal quantile plot. Is the distribution of symmetric? Is it close to a normal distribution?

```
orange = read.csv("oranges.csv")
orange_price = orange$price

par(mfrow = c(1,2))

hist(orange_price, xlab="Orange prices", prob=T, main="Histogram of Orange Prices")
x_fit<-seq(min(orange_price),max(orange_price),length=10)
y_fit<-dnorm(x_fit,mean=mean(orange_price),sd=sd(orange_price))
lines(x_fit, y_fit, col="blue", lwd=2)

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



The distribution is a bit skewed towards right.

b) Test the hypothesis of normality for the distribution of prices using the Jarque Bera test at 5% significance level.

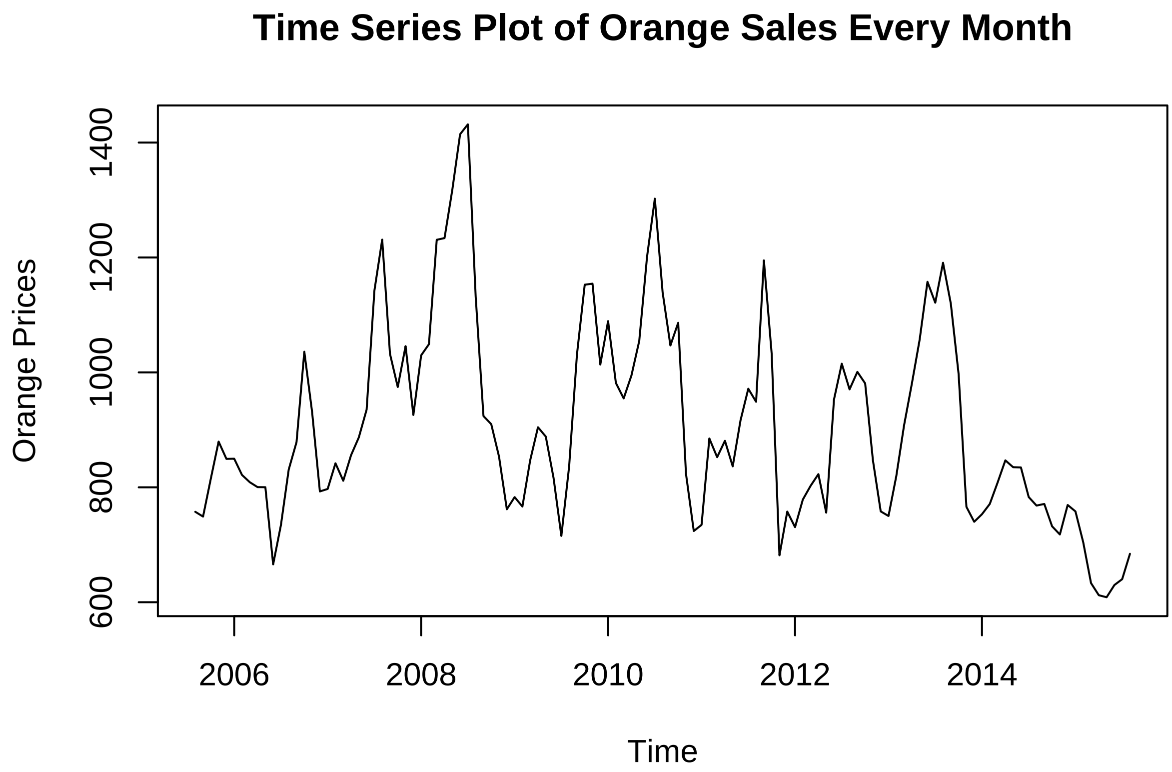
```
normalTest(orange_price,method=c("jb"))
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 12.1672
##   P VALUE:
##     Asymptotic p Value: 0.00228
##
## Description:
##  Fri Sep 25 14:26:31 2015 by user: Akhilkumar
```

with 5% significance, we reject the H_o

c) Create a time plot for the price time series and discuss if data show any striking pattern, such as trends or seasonality?

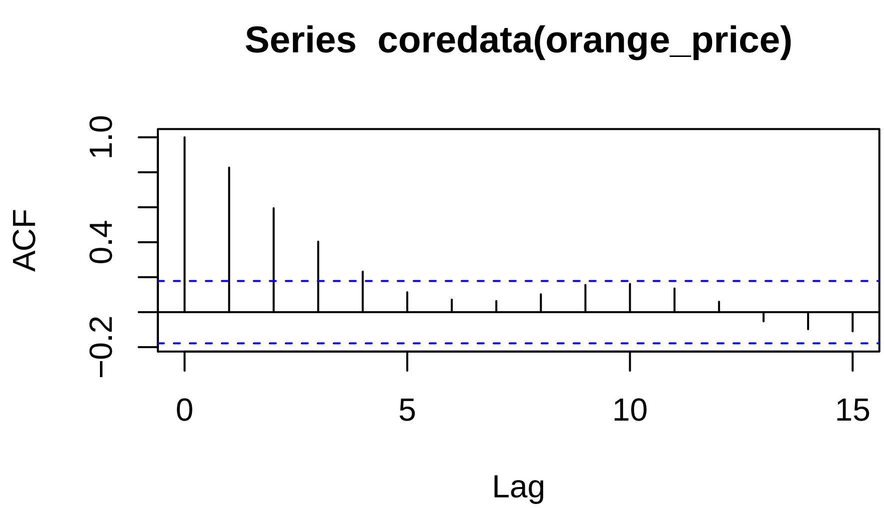
```
ots <- ts(orange_price, start=c(2005,8), freq=12)
plot(ots, type="l", xlab="Time", ylab="Orange Prices",  main = "Time Series Plot of Orange Sales Every
```



There is some level of seasonality.

d) Compute and plot the first 15 lags of the autocorrelation function for the monthly prices and discuss if the series shows evidence of serial correlation.

```
acf(coredata(orange_price), plot=T, lag=15)
```



According to the plot there is a serial correlation and with lag = 4. e) Use the Ljung Box test to hypothesis that the monthly prices have significant serial correlation.

```
Box.test(ots,lag=15,type='Ljung')
```

```
##
##  Box-Ljung test
##
## data:  ots
## X-squared = 173.6655, df = 15, p-value < 2.2e-16
```

The p-value is very low which shows weak serial correlation.

f) Discuss in general the importance of weak stationarity for time series analysis, and explain the methods that are used to analyze whether a TS is stationary.

Weak stationarity is a process whose joint probability distribution does not change when shifted in time. Consequently, parameters such as the mean and variance, also do not change over time and do not follow any trends. For example, economic data are often seasonal and/or dependent on a non-stationary price level.

There are two methods to analyze if a Time Series is stationary. * Auto-Correlation function series. * Ljung Box test.