Forecasting Homework 2

Yuhao Zhao, N17878783

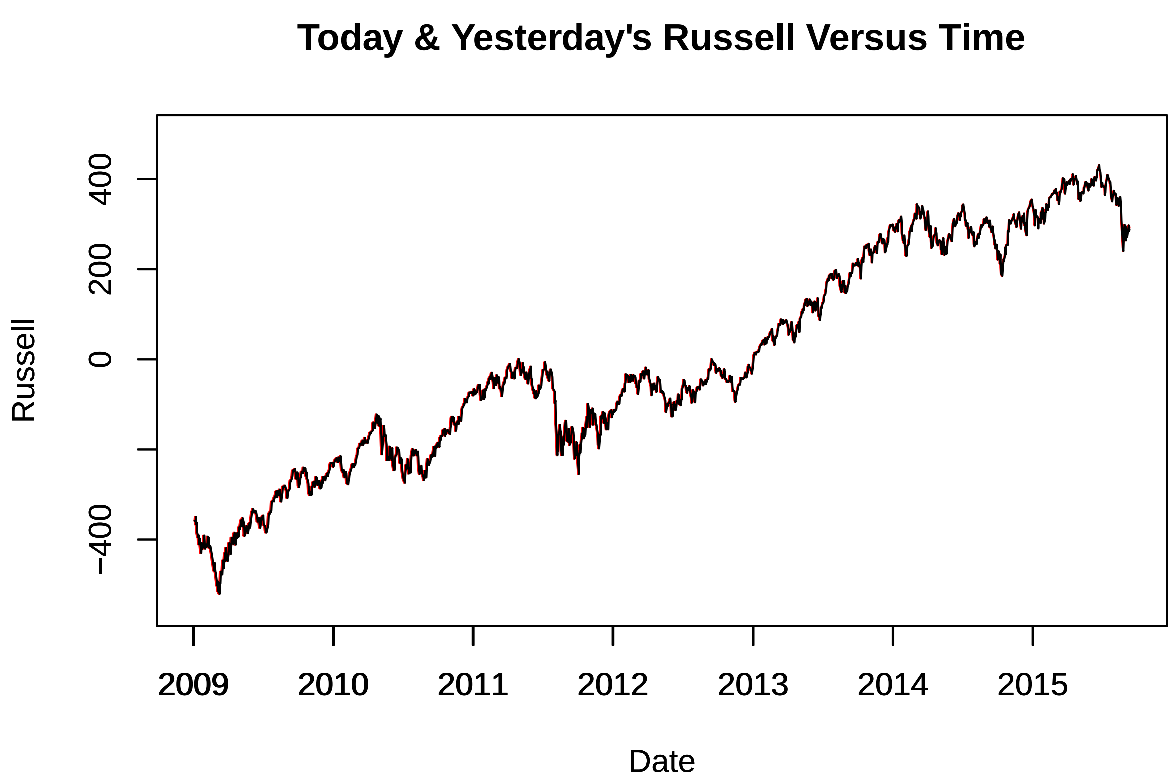
October 7, 2015

Part 1)

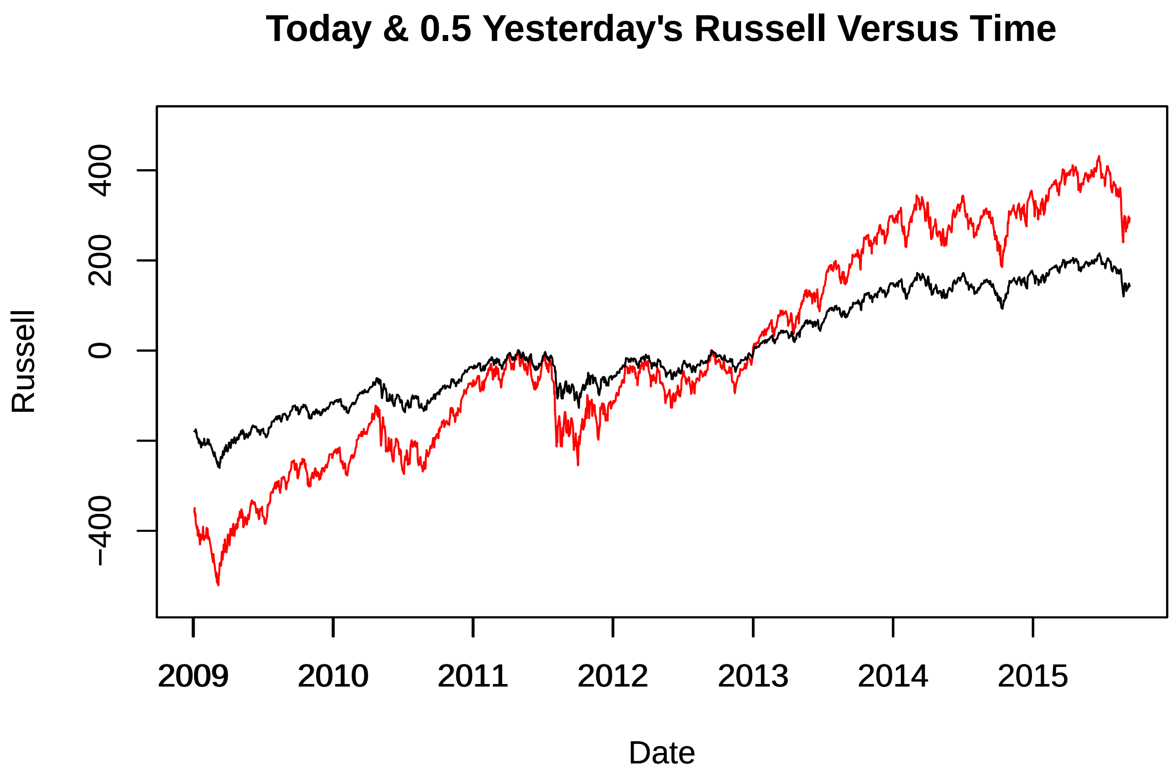
A)

```r
library(itsmr)
original  = read.table('http://people.stern.nyu.edu/churvich/Forecasting/Data/RUSSELL.CSV'
                  ,sep = ",",header = T)
data = original
data[,2] =data[,2] - mean(data[,2])
data[,1] = as.Date(data[,1],"%m/%d/%Y")
n = length(data[,2])
Today = data[2:n,]
Yesterday = data[1:(n-1),]

plot(Today[,2]~Today[,1],type = 'l',ylim = c(-550,500),col = 'red',
     main  = "Today & Yesterday's Russell Versus Time",ylab = 'Russell',xlab = "Date")
par(new = T)
plot(Yesterday[,2]~Yesterday[,1],type = 'l',ylim = c(-550,500),
     ylab = 'Russell',xlab = "Date")
```



```r
plot(Today[,2]~Today[,1],type = 'l',ylim = c(-550,500),col = 'red',
     main  = "Today & 0.5 Yesterday's Russell Versus Time",ylab = 'Russell',xlab = "Date")
par(new = T)
plot(0.5*Yesterday[,2]~Yesterday[,1],type = 'l',
     ylim = c(-550,500),ylab = 'Russell',xlab = "Date")
```



B)

From the plot, we observe that Today's Russell alomst coincide Yesterday's Russell. Therefore, Yesterday's Russell seems to be a bette forecast of Today's Russell.

C)

```r
t1 = mean((Today[,2] - Yesterday[,2])^2);t2 = mean((Today[,2] - 0.5*Yesterday[,2])^2)
MSE = as.matrix(t(c(t1,t2)));colnames(MSE) = c('Using Yesterday','Using 0.5Yesterday')
MSE
```

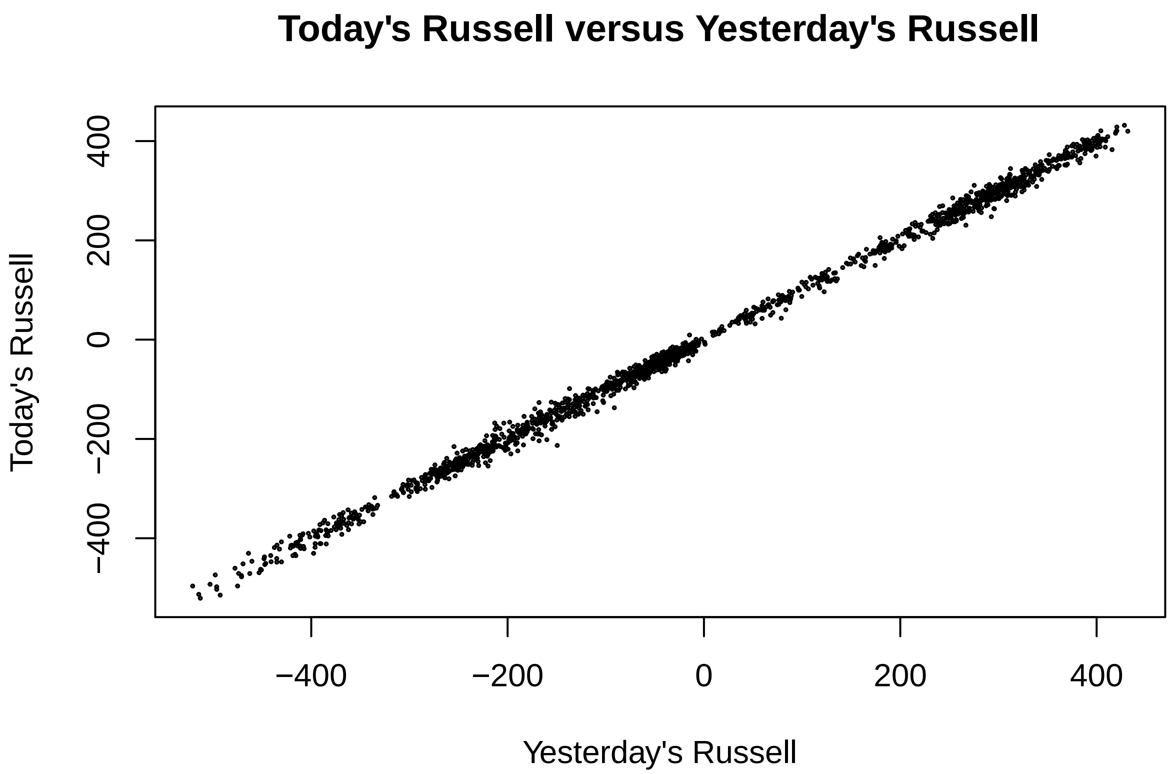
```
##      Using Yesterday Using 0.5Yesterday
## [1,]        126.7687           14315.55
```

The average squared forecast error for using Yesterday is 126.7678, and 14315.55 for using 0.5 Yesterday. The error for using Yesterday is much smaller than that for using 0.5 Yesterday. Therefore, Using Yedsterday is better.

Part 2)

A)

```r
plot(Yesterday[,2],Today[,2],cex = 0.2, ylab = "Today's Russell"
     ,xlab = "Yesterday's Russell",main = "Today's Russell versus Yesterday's Russell")
```



From the plot, it's obvious that there is a linear relation between Today's Russell and Yesterday's Russell.The slope is about 1.

B)

```r
m_2_b = lm(Today[,2] ~ Yesterday[,2])
summary(m_2_b)
```

```
##
## Call:
## lm(formula = Today[, 2] ~ Yesterday[, 2])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -64.255  -5.938   0.554   6.913  44.530
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)    0.386912   0.274261   1.411    0.159
## Yesterday[, 2] 0.998672   0.001148 869.801   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 11.25 on 1682 degrees of freedom
## Multiple R-squared:  0.9978, Adjusted R-squared:  0.9978
## F-statistic: 7.566e+05 on 1 and 1682 DF,  p-value: < 2.2e-16
```

From the summary of the model, the estimated slope is 0.998 and is significant. the intercept is not significantly different from 0. Therefore, the prediction of Today's Russell is $0.386912 + 0.998 \times$ yesterday's Russell. This is consistent with the conclusion we had on Problem 1.

C)

```
test = (m_2_b$coefficients[2] - 1)/0.001148
p_value = as.numeric(2*pt(test,df = 1682));
p_value
```

```
## [1] 0.247396
```

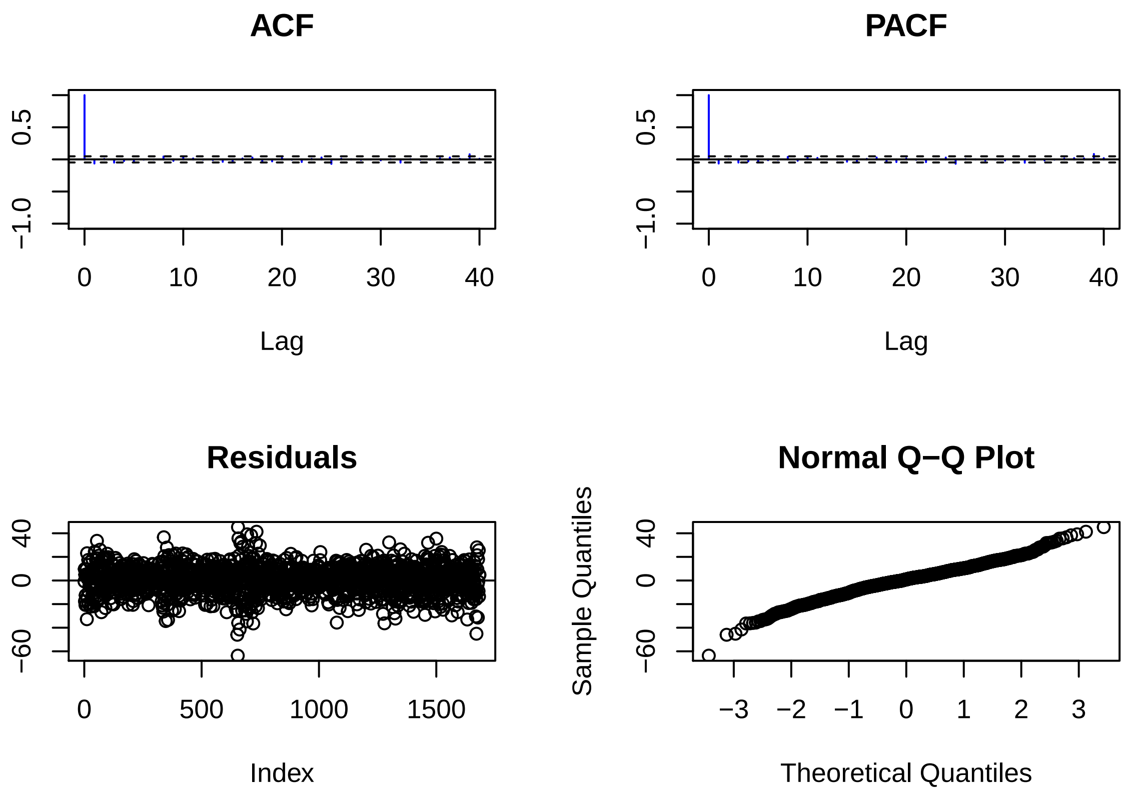
The slope of the fitted regression is 0.998672. After doing a t-test, the p-value is 0.247396. Therefore the slope is not significantly different from 1. The fitted intercept is 0.386912 with p-value 0.159. Therefore the intercept is not significantly different from 0.

D)

From part C) we already know that Today's Russell has a linear relationship with Yesterday's Russell with slope 1. However,to decide whether Russell is a random walk, we also need to test whether the residuals are iid white noise.

```
itsmr::test(Today[,2]- Yesterday[,2])
```

```
## Null hypothesis: Residuals are iid noise.
## Test                        Distribution Statistic   p-value
## Ljung-Box Q                Q ~ chisq(20)     35.98    0.0155 *
## McLeod-Li Q                Q ~ chisq(20)    562.82         0 *
## Turning points T(T-1121.3)/17.3 ~ N(0,1)      1127    0.7432
## Diff signs S      (S-841.5)/11.8 ~ N(0,1)       826    0.1909
## Rank P       (P-708543)/11522.7 ~ N(0,1)    710457    0.8681
```



From the R out put, the Ljung-Box test gives a 0.0155 p-value which shows strong evidence that the residuals are not iid white noise process. In particular from the Residual plot and the Normal Q-Q Plot, the residuals are not normally distributed and are not uncorrelated. Therefore the Ruessell is not a random walk.

E)

```
sqrt(0.9978)
```

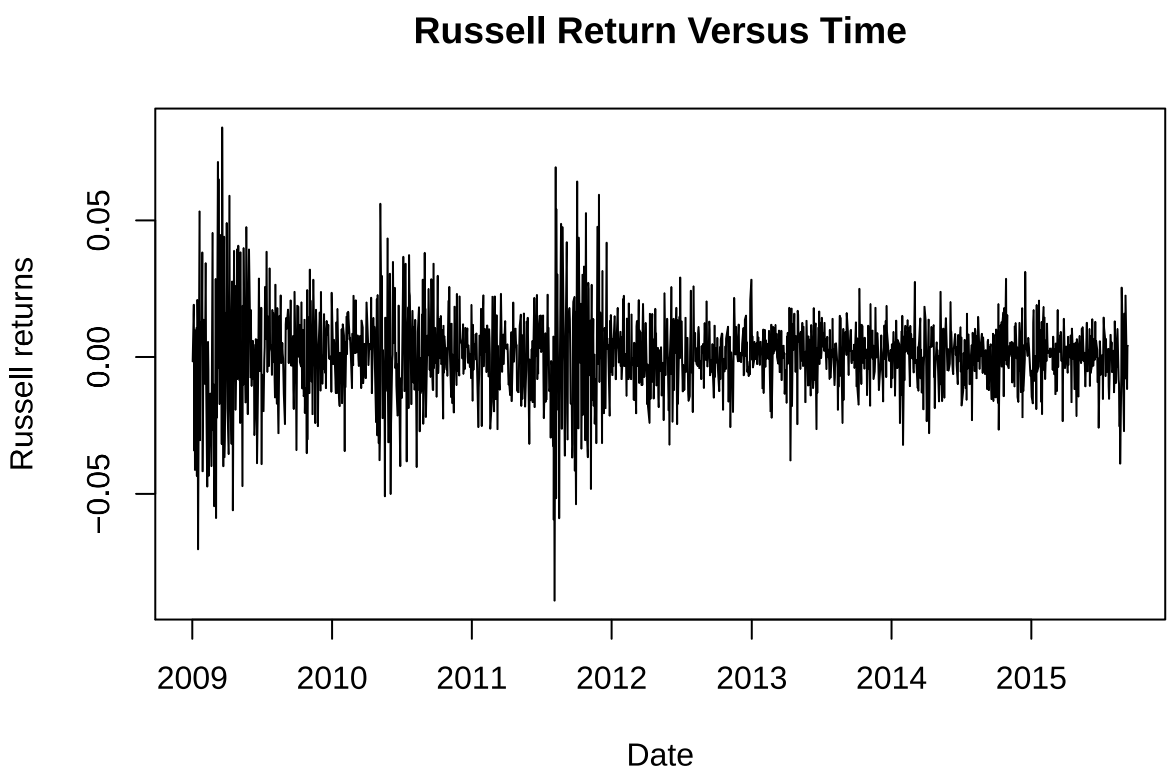
```
## [1] 0.9988994
```

The correlation coefficient between Today's Ruessell and Yesterday's Russell is 0.9988994. This means there is a extremely strong linear association between Today's Ruessell and Yesterday's Russell.

Part 3

A)

```
data = original
returns = diff(data[,2])/data[1:(n-1),2]
date =as.Date(data[,1],"%m/%d/%Y")[1:(n-1)]
plot(returns~date,type = 'l',
     main  = "Russell Return Versus Time",ylab = 'Russell returns',xlab = "Date")
```




```r
res = cbind(mean(returns),sd(returns));colnames(res) = c('mean','std')
res
```

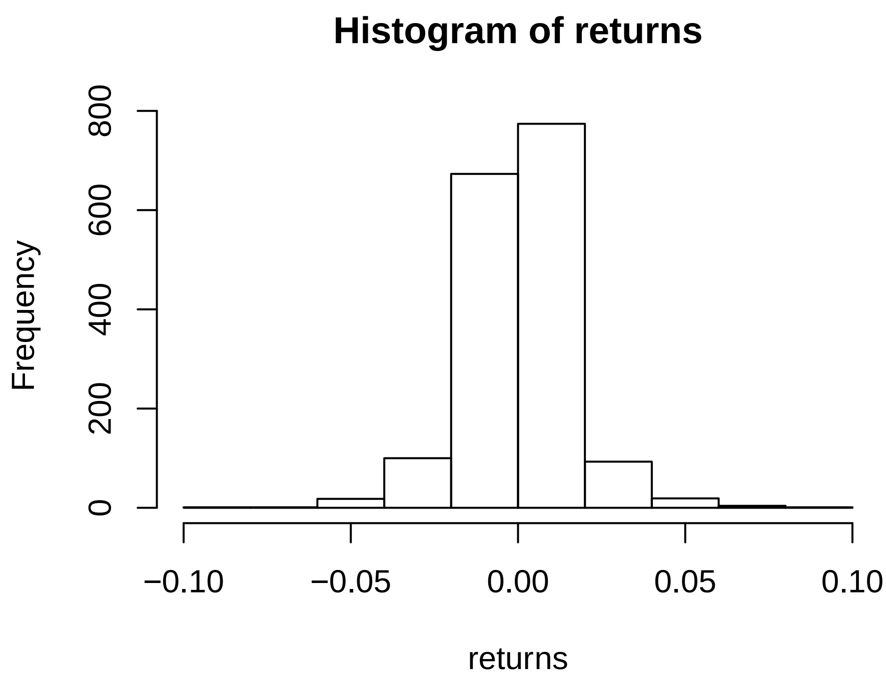
```
##              mean        std
## [1,] 0.0006098695 0.01536515
```

```r
t.test(returns,mu = 0,alternative = "two.sided",conf.level = 0.95)
```

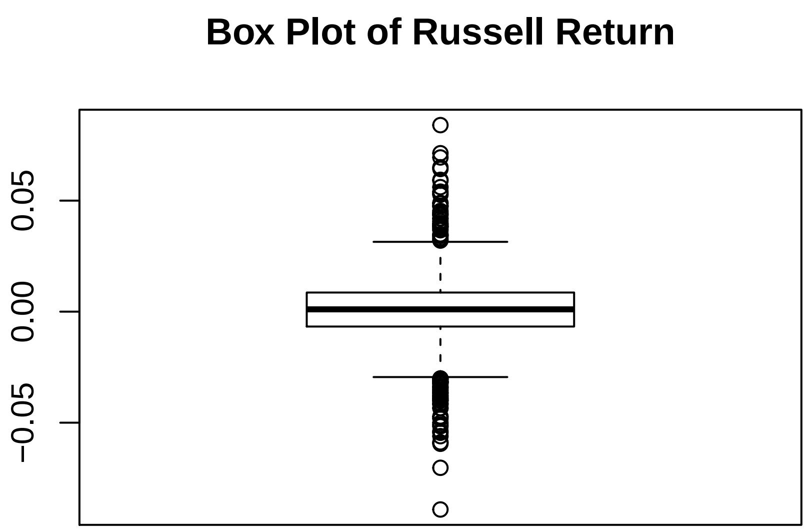
```
##
##  One Sample t-test
##
## data:  returns
## t = 1.6288, df = 1683, p-value = 0.1035
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0001245197  0.0013442587
## sample estimates:
##    mean of x
## 0.0006098695
```

From the ordinary t-test, the p-value is 0.1035. Therefore we don't have enough evidence to reject Null hypothesis. We are in favor of the Null hypothesis, which means we are in favor of that the mean return is zero. Therefore the mean is not significantly different from zero. B) —

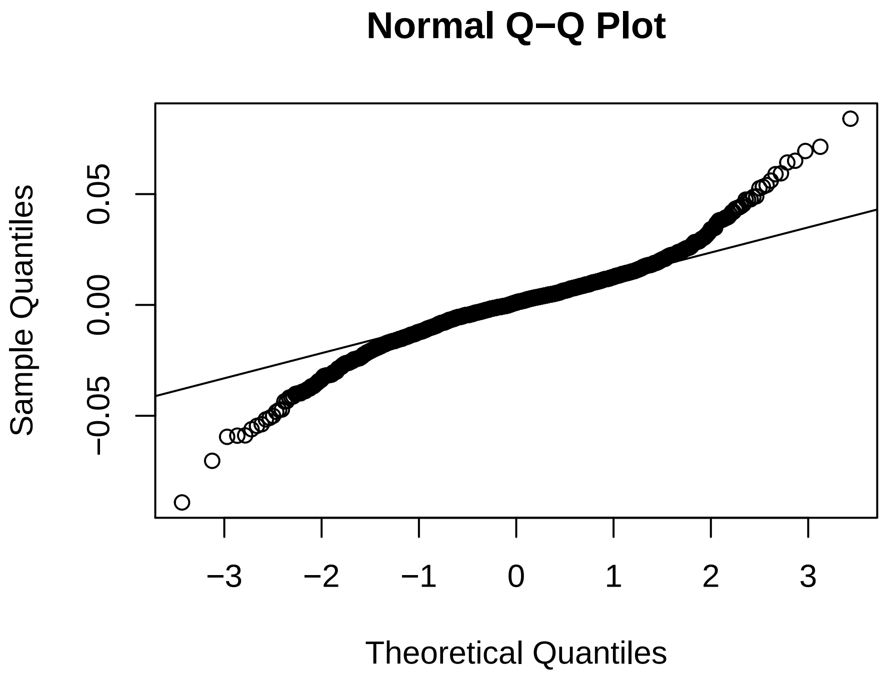
```r
hist(returns,breaks = 10)
```




```r
boxplot(returns,main = 'Box Plot of Russell Return')
```



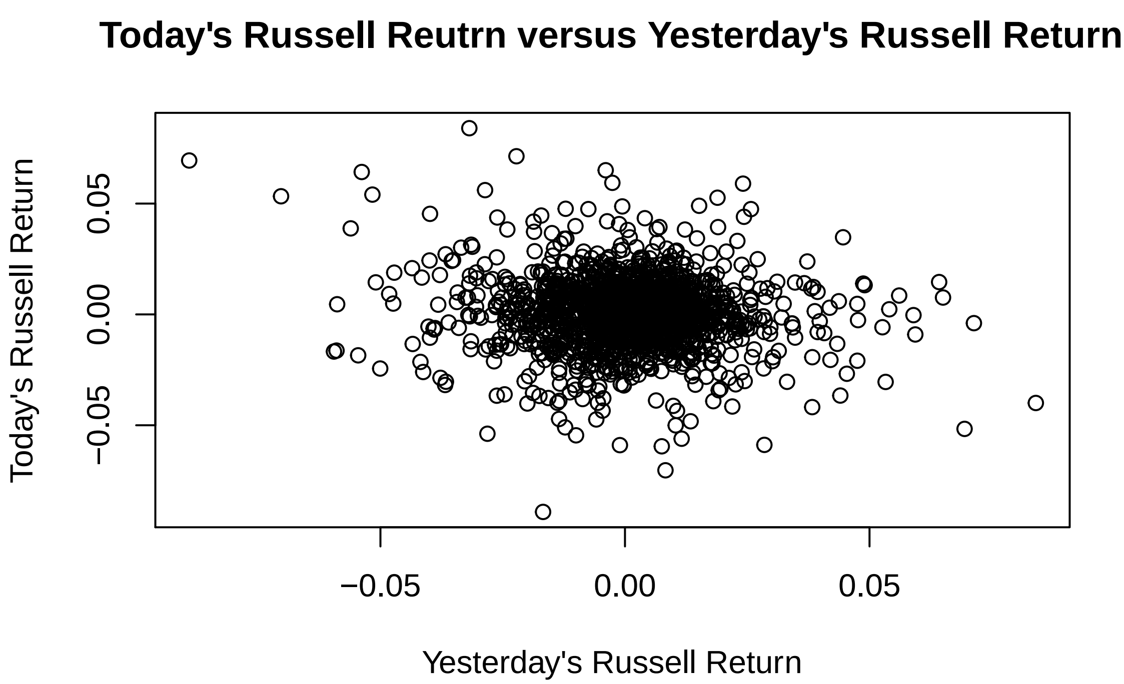

```r
qqnorm(returns);qqline(returns)
```



From the Q-Q plot, the quantile of Ruessell Returns doesn't show linear relationship with the theoretical normal quantiles.Therefore, it's clear that the Russell returns are not normally distributed.

C)

```
n = length(returns)
Today = returns[2:n];Yesterday = returns[1:(n-1)]
plot(Yesterday,Today,ylab = "Today's Russell Return",xlab = "Yesterday's Russell Return",
     main = "Today's Russell Reutrn versus Yesterday's Russell Return")
```



The plot of today's Russell Reutrn versus yesterday's Russell Return is very different from the Russell plot. In this plot, data points are clusterd over the original. The today's Russell are easier to predict than today's Russell Return.

D)

```
m_3_d  = lm(Today~Yesterday)
summary(m_3_d)
```

```
##
## Call:
## lm(formula = Today ~ Yesterday)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.091414 -0.007206  0.000450  0.008162  0.080199
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.0006710  0.0003732   1.798   0.0724 .
## Yesterday   -0.0984406  0.0242720  -4.056 5.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0153 on 1681 degrees of freedom
## Multiple R-squared:  0.00969,    Adjusted R-squared:  0.009101
## F-statistic: 16.45 on 1 and 1681 DF,  p-value: 5.227e-05
```

The prediction of Today's return is 0.0006710 -0.0984406 $\times$ Yesterday's return. The slope is significantly different from 0. The intercept is not.

Part 4

Since X_t is stationary and $E(X_t) = 0$. We want to find the best liner predictor of X_t given X_{t-1}. The target is to mininize MSE w.r.t the coefficients a and b. Let $X = X_{t-1}, Y = X_t, \hat{Y} = a + bX$

MSE $= E[(Y - \hat{Y})^2] = E[(Y - (a + bX))^2] = E(Y^2) + E(a^2 + b^2X^2 + 2abX) - 2E(aY + bXY)$

$\frac{\partial MSE}{\partial a} = 2a + 2bE(X) - 2E(Y)①$
$\frac{\partial MSE}{\partial b} = 2bE(X^2) + 2aE(X) - 2E(XY)②$

Since X_t is stationary, $E(X) = E(X_t) = E(X_{t-1}) = E(Y) = 0, Var(X_t) = Var(X_{t-1})$
① $\rightarrow 2a = 0$, therefore a = 0
② $\rightarrow 2bE(X^2) - 2E(XY) = 0$, therefore b = $\frac{E(XY)}{E(X^2)}$
$Since Var(X) = E(X^2) - E(X)^2 = E(X^2)$, and $E(X^2) = E(Y^2).b = \frac{E(XY) - 0 \times 0}{\sqrt{E(X^2)E(X^2)}} = \frac{E(XY) - E(X)E(Y)}{\sqrt{var(X)var(Y)}} = \rho_1$

Therefore, the best linear predictor of X_t based on X_{t-1} is $\rho_1 X_{t-1}$