Forecasting Project 2

Yuhao Zhao,Yz3085

December 16, 2015

Part A. Introduction

In this project I will model the daily Crude Oil price, Cushing OK WTI Spot Price FOB (Dollars per Barrel), from Jan 02, 1986 to Dec 07, 2015

Part B. Data for analysis

data source: *Spot Prices (Crude Oil in Dollars per Barrel, Products in Dollars per Gallon)*

The data is acquired from U.S Energy Information Administration websiet via *https://www.eia.gov/dnav/pet/pet_pri_spt_s1_d.htm*. It contains the daily Crude Oil spot price from Jan 02, 1986 to Dec 07, 2015.

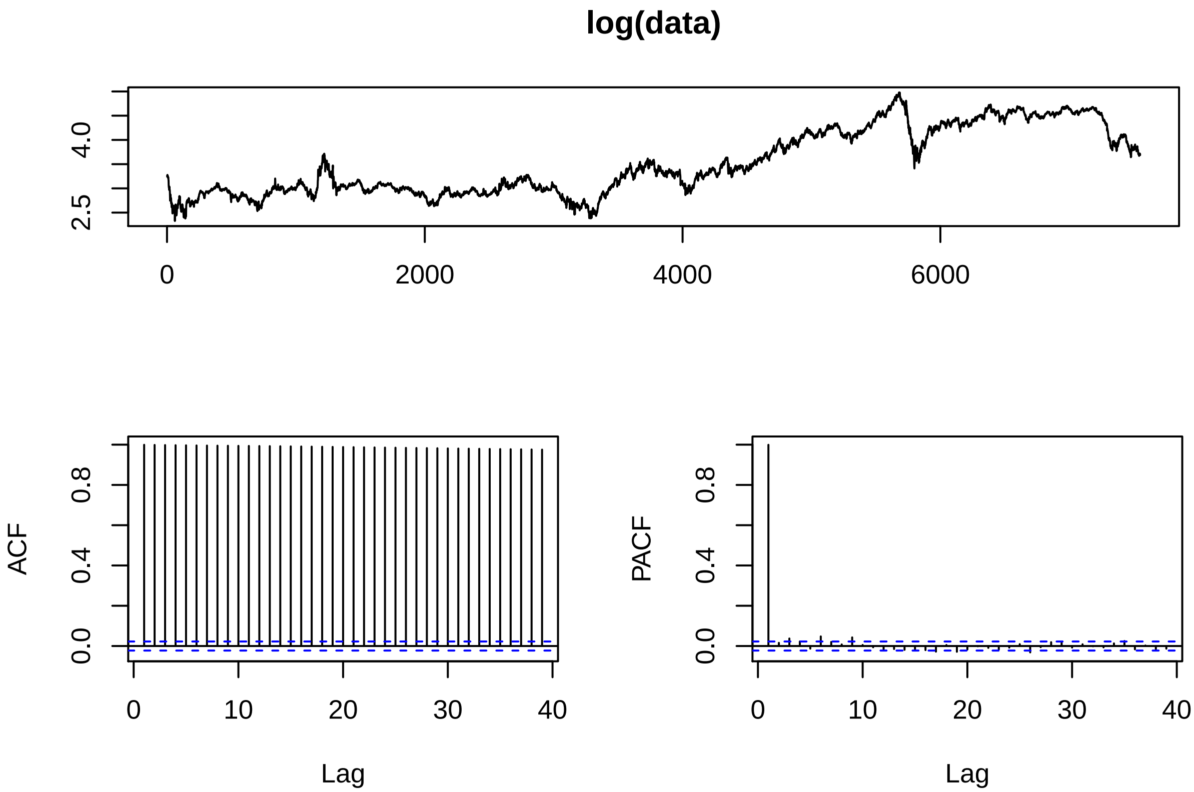
Part C. Data input

```
setwd("/Users/cryan/Desktop/Assignment/STAT-2302/project2")
library('itsmr')
library('forecast')
library('tseries')
data <- read.table("data.txt",fill = TRUE)[,2]
data = as.numeric(na.omit(data))
data_last = tail(data,1)
data = data[1:(length(data)-1)]
```

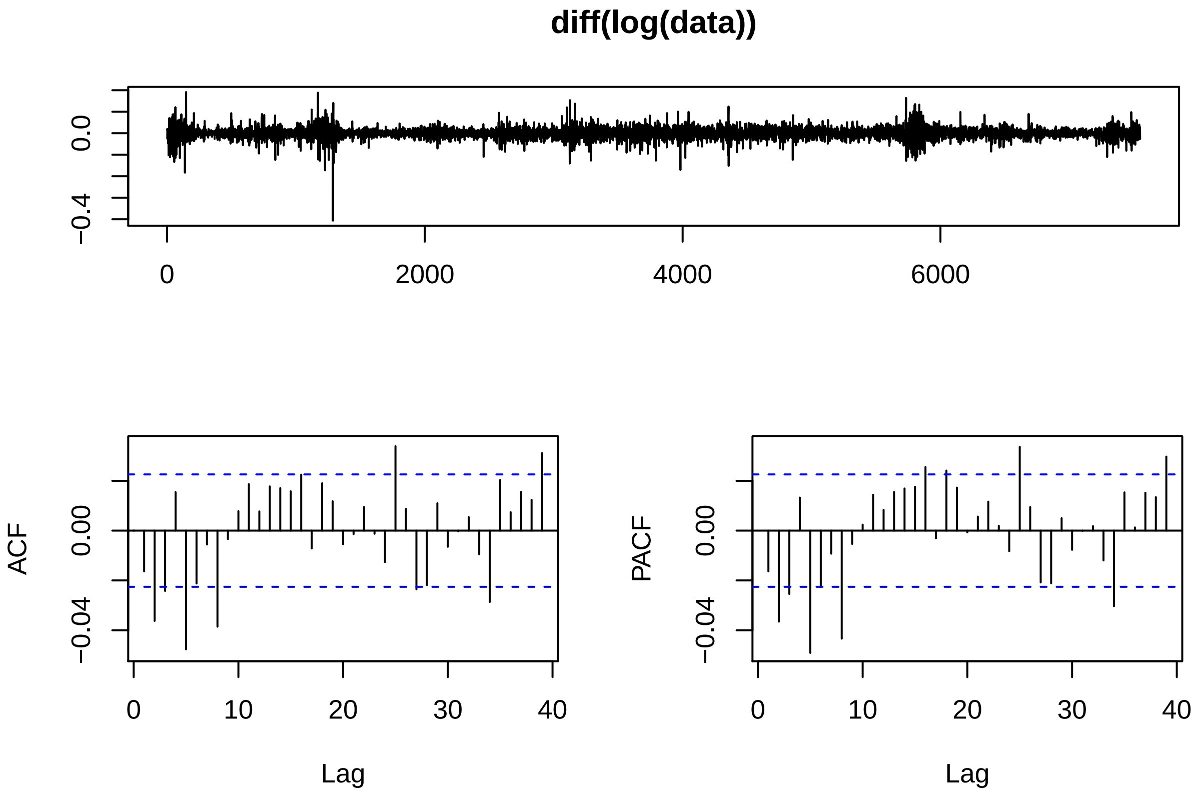
Part D. Data modeling

1)

```
tsdisplay(log(data),points = F)
```




```r
tsdisplay(diff(log(data)),points = F)
```



The plot of logs of data shows that it's non-stationary, since it is not clearly mean-reverting. The differenced logs of Rupee shows strong stationary behavior. It's unclear to identify the ARMA model from the ACF and PACF of differenced logs of data therefore we need to conduct model selection based on the AICc's.

2)

```
model= auto.arima(data,d = 1,parallel = F,trace = T,ic = "aicc",stepwise = T
                  ,allowmean = F,allowdrift = F,seasonal =F,max.p = 2,max.q=2 )
```

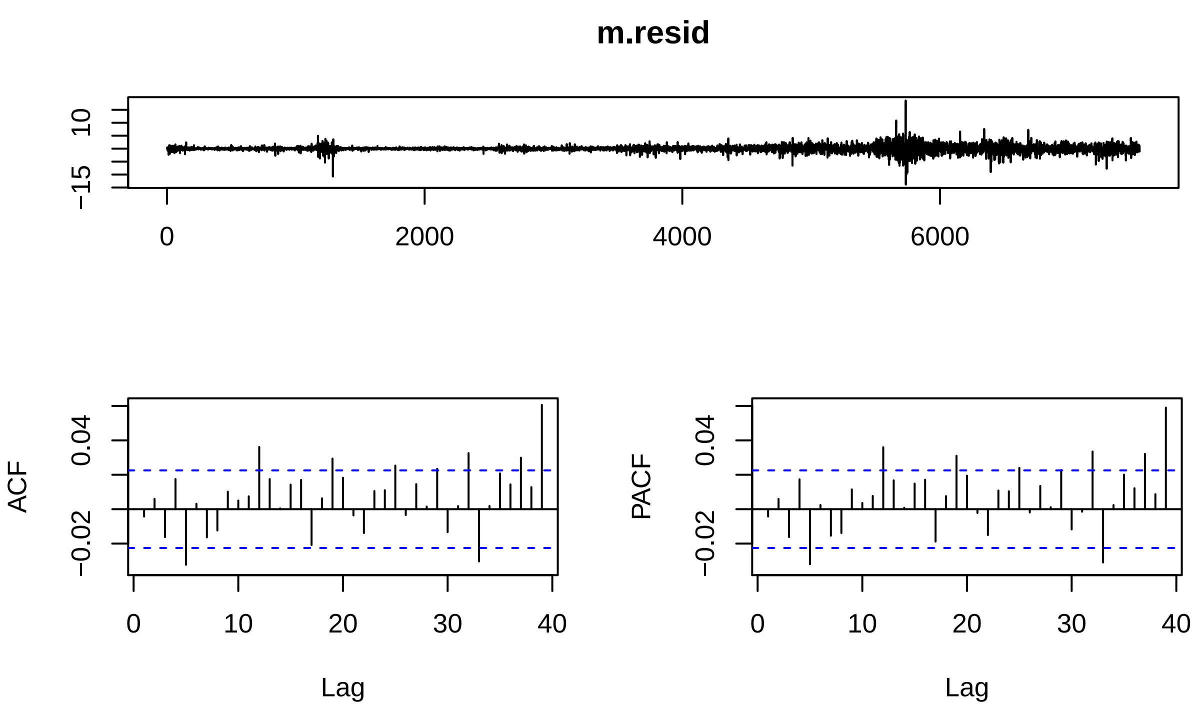
```
##
##  ARIMA(2,1,2)                    : 23540.75
##  ARIMA(0,1,0)                    : 23564.42
##  ARIMA(1,1,0)                    : 23551.29
##  ARIMA(0,1,1)                    : 23549.54
##  ARIMA(1,1,2)                    : 23549.48
##  ARIMA(2,1,1)                    : 23549.36
##  ARIMA(1,1,1)                    : 23550.05
##
##  Best model: ARIMA(2,1,2)
```

```
m.resid = model$residuals
m.fore = forecast::forecast(model,h=1,level = 95)
m.fit = fitted(model)
```

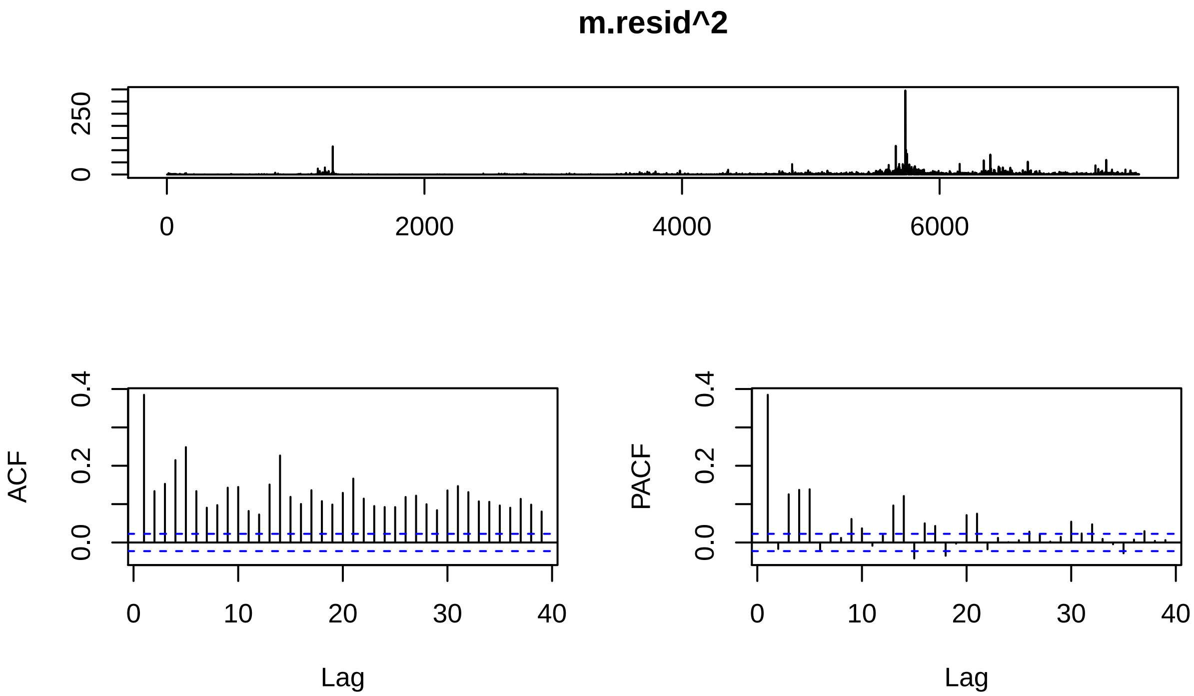
The model I fit with lowest aicc is ARIMA(2,1,2)

3)

```
tsdisplay(m.resid,points = F)
```



```r
tsdisplay(m.resid^2,points = F)
```



The ACF and PACF plot of the residuals show that there is some evidence of autocorralation with some certain lags. Therefore the residuals are not uncorrelated thus not independent. In particular, the residuals shows strong conditional heteroscedasticity.

4)

```r
AICc_arch <- function(data,order)
{
  N = length(data)-1
  res = c()
  for (i in order[1]:order[2])
  {
    if (i == 0) {
        loglik = -0.5*N*(1 + log(2*pi*mean(data^2)))
        aicc = -2*loglik+2*(i +1)*N/(N-i-2)
        res = rbind(res,c(i,loglik,aicc) )}
    else{
      loglik= logLik(garch(data,c(0,i),control = garch.control(trace = F)))
      aicc = -2*loglik+2*(i +1)*N/(N-i-2)
      res = rbind(res,c(i,loglik,aicc ))
    }
  }
  colnames(res) = c('order q','log lik','AICc')
  res
}
res = AICc_arch(m.resid,c(0,10));res
```

```
##       order q    log lik     AICc
##  [1,]       0 -11764.360 23530.72
##  [2,]       1 -10696.387 21396.78
##  [3,]       2  -9827.145 19660.29
##  [4,]       3  -9349.810 18707.63
##  [5,]       4  -9048.623 18107.25
##  [6,]       5  -8851.816 17715.64
##  [7,]       6  -8681.323 17376.66
##  [8,]       7  -8623.566 17263.15
##  [9,]       8  -8596.387 17210.80
## [10,]       9  -8549.282 17118.59
## [11,]      10  -8512.050 17046.13
```

```
m.garch = garch(m.resid,c(1,1),control =garch.control(trace = F))
N = length(m.resid)-1
m.garch.loglik = logLik(m.garch)
m.garch.aicc = -2*m.garch.loglik+2*(2 +1)*N/(N-2-2);m.garch.aicc[1]
```

```
## [1] 16702.75
```

```
summary(m.garch);m.garch.loglik
```

```
##
## Call:
## garch(x = m.resid, order = c(1, 1), control = garch.control(trace = F))
##
## Model:
## GARCH(1,1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.97245 -0.55835  0.02552  0.61106  6.10679
##
## Coefficient(s):
##     Estimate  Std. Error  t value Pr(>|t|)
## a0 0.0009123   0.0001734     5.26 1.44e-07 ***
## a1 0.0843859   0.0032953    25.61  < 2e-16 ***
## b1 0.9217269   0.0028066   328.42  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Diagnostic Tests:
##  Jarque Bera Test
##
## data:  Residuals
## X-squared = 4144.8, df = 2, p-value < 2.2e-16
##
##
##  Box-Ljung test
##
## data:  Squared.Residuals
## X-squared = 1.0212, df = 1, p-value = 0.3122
```

```
## 'log Lik.' -8348.372 (df=3)
```

From the outout we see that the lowest aicc provided by arch model is 17046.13, and the garch(1,1) model reterns a 16702 aicc. Therefore we use garch(1,1) to model the residuals. The p-values for the parameter values of garch(1,1) are all significant. The complete form of the model is

$$\epsilon_t|\psi_{t-1} \sim N(0, h_t), h_t = 0.0009123 + 0.0843859 \times \epsilon_{t-1}^2 + 0.9217269 \times h_{t-1}$$

5)

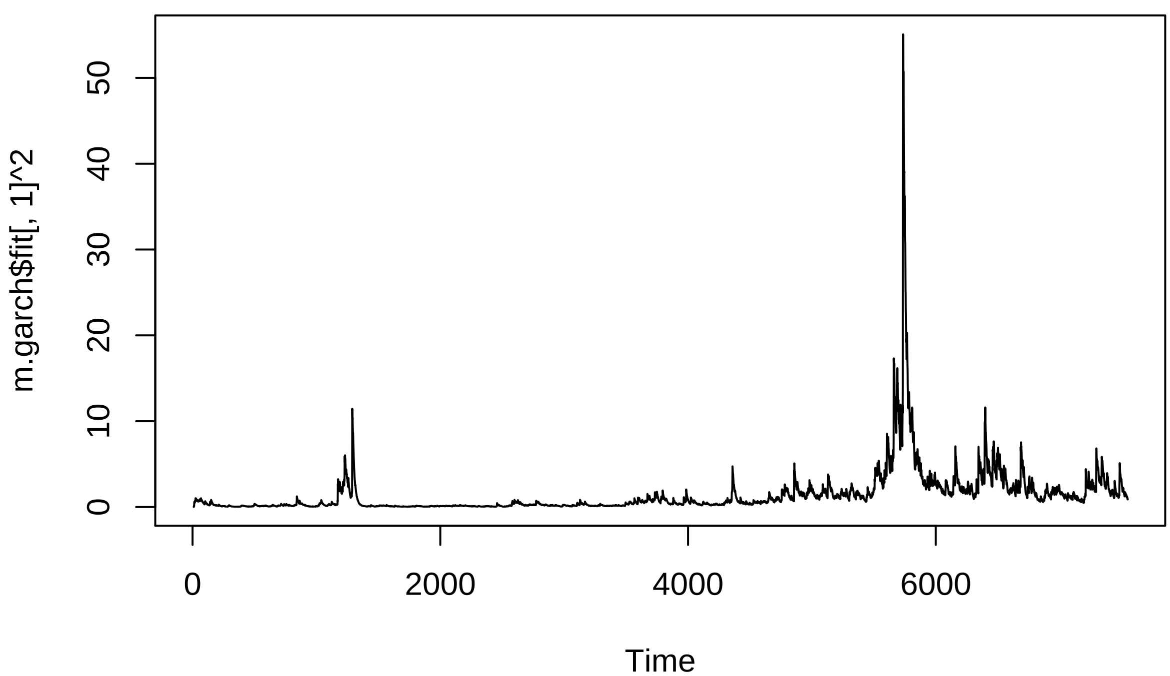
```
cond_var = tail(m.garch$fit[,1],1)
h_f1 = m.garch$coef[1]+m.garch$coef[2]* tail(m.resid,1)^2 +
          m.garch$coef[3]* tail(m.garch$fit[,1]^2,1);
m.garch.fore= m.fore$mean[1] + c(0,-qnorm(0.975)*sqrt(h_f1),qnorm(0.975)*sqrt(h_f1));
colnames(m.garch.fore) = colnames(m.fore)
m.fore;m.garch.fore
```

```
##      Point Forecast    Lo 95   Hi 95
## 7551       39.98567 37.73224 42.2391

##                  a0       a0
## 39.98567 38.07564 41.89571
```

The forecast interval for ARIMA-GARCH is [38.07564,41.89571] which is narrower than that for the ARIMA model. The last data is 37.64, either of the interval covered the data. This means both of the ARIMA and ARIMA-GARCH interval are too narrow.

6)

```
plot(m.garch$fit[,1]^2)
```

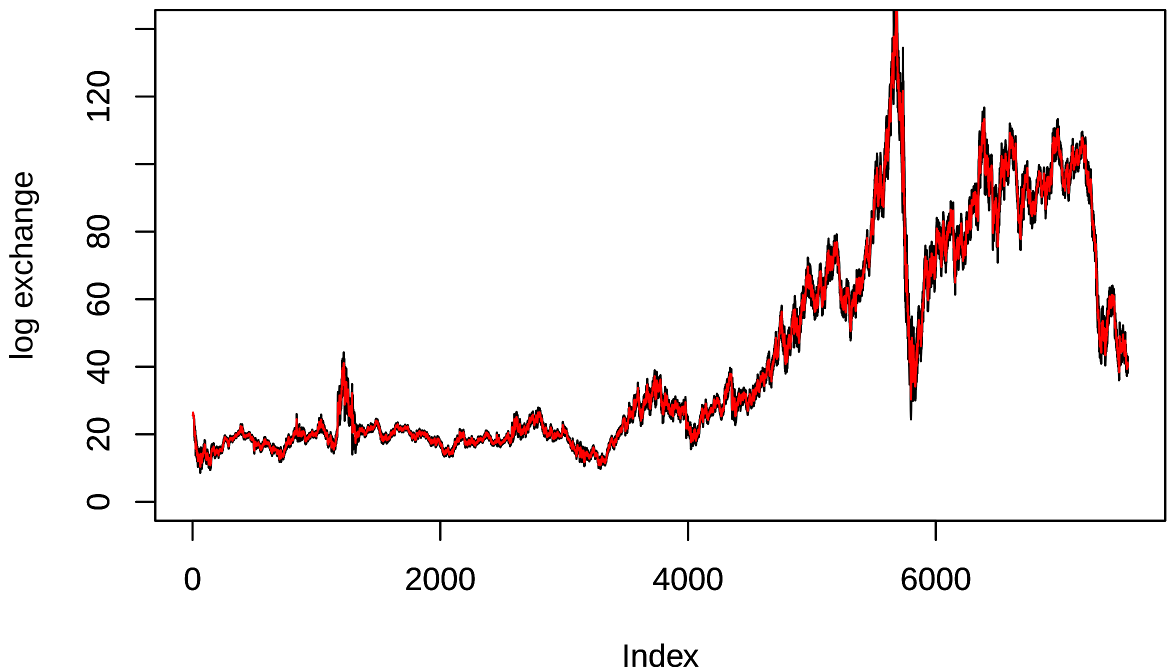


```
ht = m.garch$fit[,1]^2
```

The plot shows high volatility during time index 1000 to 1200, and index 5000 to 6000.It corresponds to the year 1987-1990 and 2006-2009. This highly volatile period agree with the log oil price itself.

7)

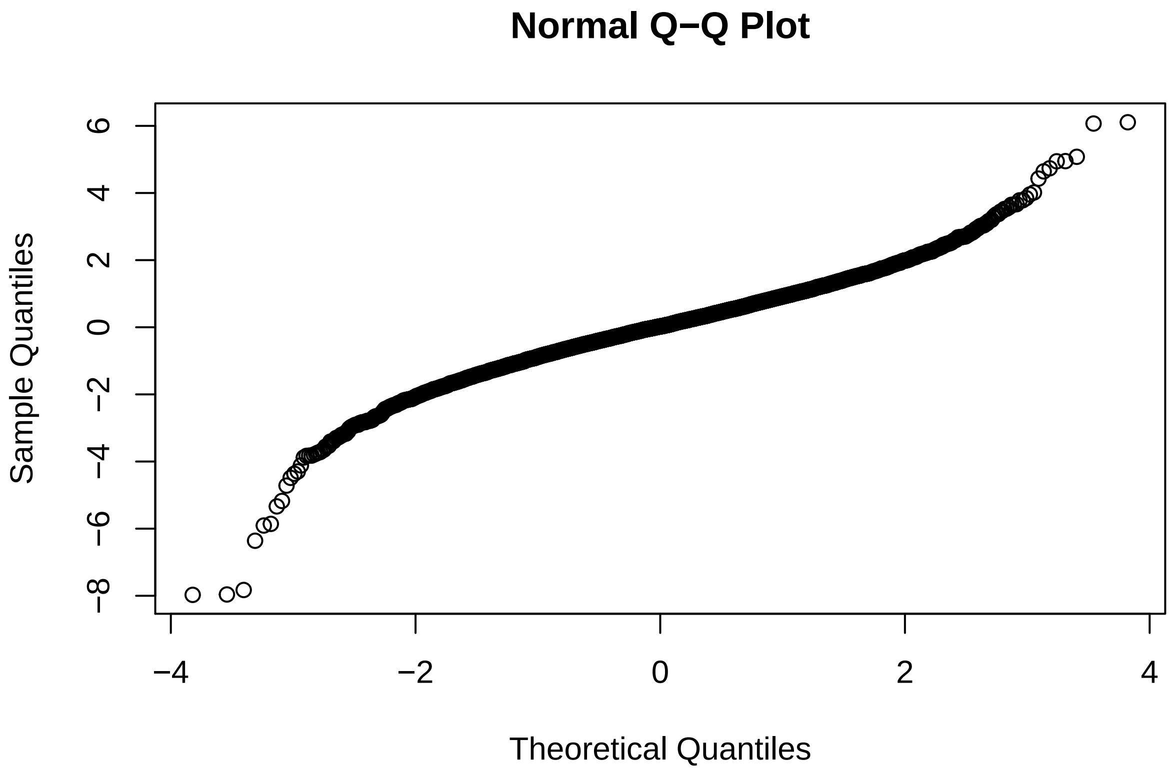
```
conf_int = matrix(0, 7550, 3)
conf_int[,1] = m.fit
conf_int[,2:3] = m.fit + cbind(-qnorm(0.975)*sqrt(ht),qnorm(0.975)*sqrt(ht))

plot(conf_int[,2],type = 'l',ylim = c(0,140),ylab = c(''));par(new = T)
plot(conf_int[,3],type = 'l',ylim = c(0,140),ylab =c(''));par(new = T)
polygon(c(1:7550,rev(1:7550)),c(conf_int[,3],rev(conf_int[,2])),col="grey");par(new = T)
plot(conf_int[,1],type = 'l',ylim = c(0,140),col = 'red',ylab = c('log exchange'))
```



The one-step-ahead forecast covers the true log oil price most of the times. The confidence interval is wide when the previous log oil price volatility is high and narrow when it's low. The model seems to have great predictive power in terms of the high accuracy. However we are using the model fitted by the whole data and explaining the same data at the same time. This forecast interval may be too optimistic.

8)

```
m.garch.resid = m.garch$residuals
qqnorm(na.omit(m.garch.resid))
```



From the Normal Q-Q plot, the residuals still shows clear long-tailedness. Thus the garch model is not adequante to describe the leptokurtosis in the data.

9)

```
cur = na.omit(m.garch.resid)
sum(abs(cur)>1.96)/length(cur)
```

```
## [1] 0.05091488
```

The interval fails 0.05091488 of the time. The last day oil price is 37.64. The forecast interval for ARIMA-GARCH is [38.07564,41.89571] which is narrower than that for the ARIMA model. Either of the interval covered the data. This means both of the ARIMA and ARIMA-GARCH interval are too narrow