FIn600 hw1 solution

```r
# clear the environment
rm(list = ls())
```

Q1

```r
# import data
Q1.data <- read.csv("Rates.csv", header = T)
```

```r
# daily return
usd.ret <- diff(log(Q1.data$USD_by_INR))
gbp.ret <- diff(log(Q1.data$GBP_by_INR))
eur.ret <- diff(log(Q1.data$EUR_by_INR))
```

```r
# sample average
mean(usd.ret)
```

```
## [1] -1.089284e-05
```

```r
mean(gbp.ret)
```

```
## [1] -0.000207834
```

```r
mean(eur.ret)
```

```
## [1] 4.564951e-05
```

```r
# standard deviation
sd(usd.ret)
```

```
## [1] 0.005637217
```

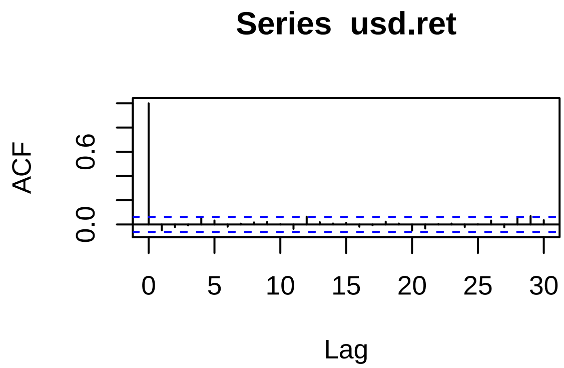
```r
sd(gbp.ret)
```

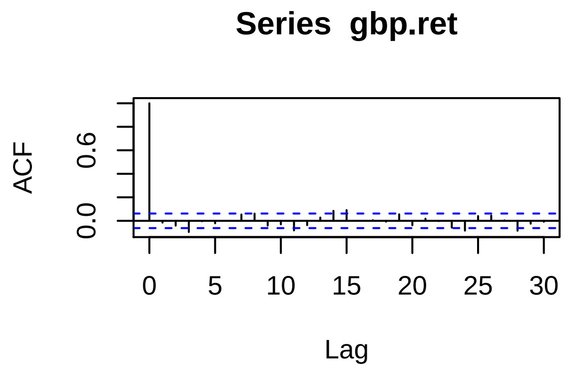
```
## [1] 0.008101245
```

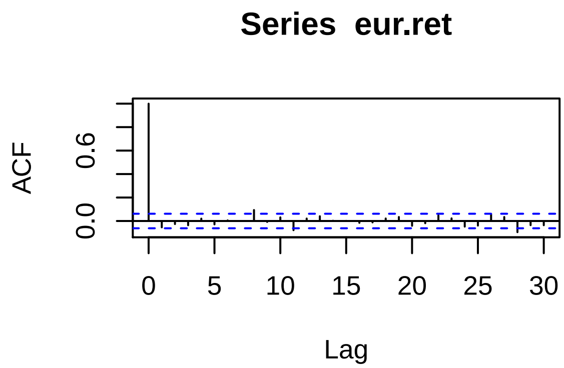
```r
sd(eur.ret)
```

```
## [1] 0.007267491
```

```r
# first order autocorrelation
# use plot to test H0:pho(v) = 0
par(mfrow = c(2,2))
acf(usd.ret,lag.max = 30)
acf(gbp.ret,lag.max = 30)
acf(eur.ret,lag.max = 30)
```








```r
# test H0: mean = 0
t.test(usd.ret, mu=0)
```

```
##
##  One Sample t-test
##
## data:  usd.ret
## t = -0.061135, df = 1000, p-value = 0.9513
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0003605334  0.0003387477
## sample estimates:
##     mean of x
## -1.089284e-05
```

```r
t.test(gbp.ret, mu=0)
```

```
##
##  One Sample t-test
##
## data:  gbp.ret
## t = -0.81167, df = 1000, p-value = 0.4172
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0007103024  0.0002946344
## sample estimates:
##    mean of x
## -0.000207834
```

```r
t.test(eur.ret, mu=0)
```

```
##
##  One Sample t-test
##
## data:  eur.ret
## t = 0.19873, df = 1000, p-value = 0.8425
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0004051065  0.0004964055
## sample estimates:
##    mean of x
## 4.564951e-05
```

All three p-values are greater than zero, so we don't have enough evidence to reject the null hypothesis that mean equals zero.

```r
# Ljung-Box test H0:acf(x,lag = i) = 0
Box.test(usd.ret, lag = 1, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  usd.ret
## X-squared = 2.3155, df = 1, p-value = 0.1281
```

```r
Box.test(gbp.ret, lag = 1, type = "Ljung")
```

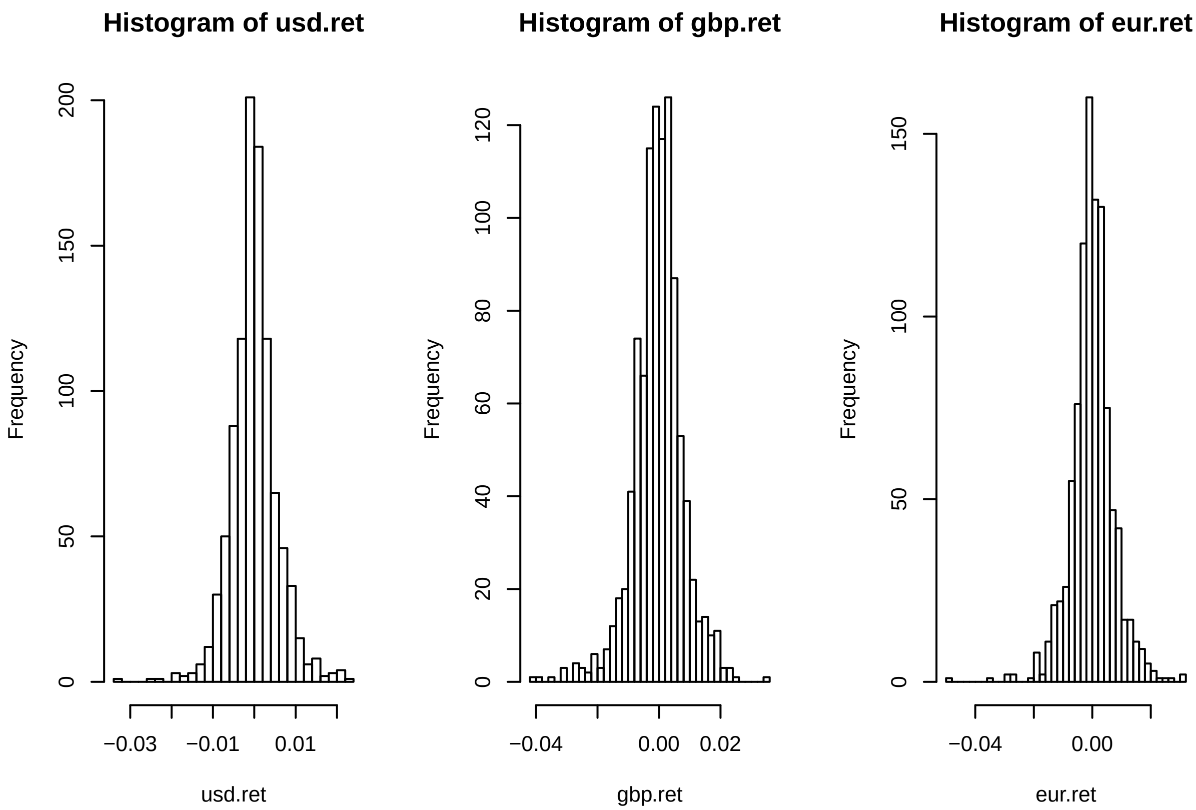
```
##
##  Box-Ljung test
##
## data:  gbp.ret
## X-squared = 0.23845, df = 1, p-value = 0.6253
```

```r
Box.test(eur.ret, lag = 1, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  eur.ret
## X-squared = 2.9365, df = 1, p-value = 0.0866
```

All three p-values are greater than zero, so we don't have enough evidence to reject the null hypothesis that ACF of lag i equals zero(in this code, $i = 1$).

```r
# (b) histogram of return
par(mfrow = c(1,3))
hist(usd.ret,breaks = 30)
hist(gbp.ret,breaks = 30)
hist(eur.ret,breaks = 30)
```




```r
# Jarque-Bera test, H0:the data are from a normal distribution
library("tseries")
jarque.bera.test(usd.ret)
```

```
##
##  Jarque Bera Test
##
## data:  usd.ret
## X-squared = 400.72, df = 2, p-value < 2.2e-16
```

```r
jarque.bera.test(gbp.ret)
```

```
##
##  Jarque Bera Test
##
## data:  gbp.ret
## X-squared = 331.99, df = 2, p-value < 2.2e-16
```

```r
jarque.bera.test(eur.ret)
```

```
##
##  Jarque Bera Test
##
## data:  eur.ret
## X-squared = 704.77, df = 2, p-value < 2.2e-16
```

All three p-values are less than 0.05, so we should reject the null hypothesis that the data set is following the normal distribution.

(c) aggregate data at weekly level

```r
# change the first column to date form
library(xts)
Date <- as.Date(as.character(Q1.data$Date), "%Y%m%d")
```

```r
# aggregate data into class 'zoo'
zoo.usd <- zoo(Q1.data$USD_by_INR,Date)
zoo.gbp <- zoo(Q1.data$GBP_by_INR,Date)
zoo.eur <- zoo(Q1.data$EUR_by_INR,Date)
```

```r
# change date to weekly data
usd.weekly <- to.weekly(zoo.usd,drop.time = T,name = T,OHLC = F)
gbp.weekly <- to.weekly(zoo.gbp,drop.time = T,name = T,OHLC = F)
eur.weekly <- to.weekly(zoo.eur,drop.time = T,name = T,OHLC = F)
```

```r
# weekly return
usd.ret.weekly <- diff(log(usd.weekly))
gbp.ret.weekly <- diff(log(gbp.weekly))
eur.ret.weekly <- diff(log(eur.weekly))
```

```r
# weekly mean return
mean(usd.ret.weekly)
```

```
## [1] -6.06311e-05
```

```r
mean(gbp.ret.weekly)
```

```
## [1] -0.0009843461
```

```r
mean(eur.ret.weekly)
```

```
## [1] 0.0002289841
```

```r
# weekly return standard deviation
sd(usd.ret.weekly)
```

```
## [1] 0.01153731
```

```r
sd(gbp.ret.weekly)
```

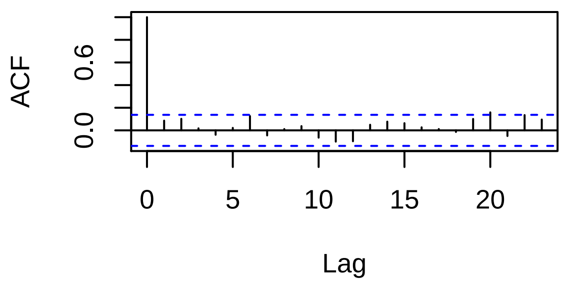
```
## [1] 0.01779926
```

```r
sd(eur.ret.weekly)
```

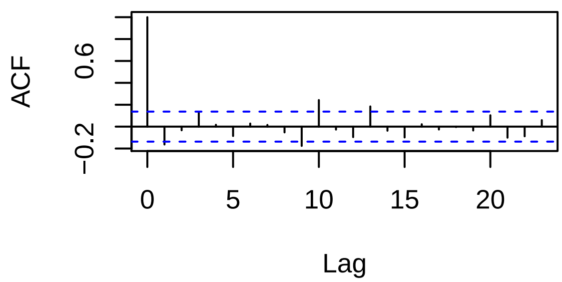
```
## [1] 0.01447418
```

```r
# acf of weekly return
par(mfrow = c(2,2))
acf(coredata(usd.ret.weekly, lag.max = 30))
acf(coredata(gbp.ret.weekly, lag.max = 30))
acf(coredata(eur.ret.weekly, lag.max = 30))
```

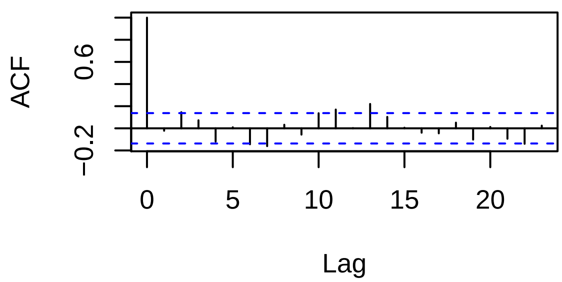
Series coredata(usd.ret.weekly, lag.max =



Series coredata(gbp.ret.weekly, lag.max =



Series coredata(eur.ret.weekly, lag.max =




```r
# t test
t.test(usd.ret.weekly, mu = 0)
```

```
##
##  One Sample t-test
##
## data:  usd.ret.weekly
## t = -0.07506, df = 203, p-value = 0.9402
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.001653333  0.001532070
## sample estimates:
##    mean of x
## -6.06311e-05
```

```r
t.test(gbp.ret.weekly, mu = 0)
```

```
##
##  One Sample t-test
##
## data:  gbp.ret.weekly
## t = -0.78988, df = 203, p-value = 0.4305
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.003441498  0.001472805
## sample estimates:
##     mean of x
## -0.0009843461
```

```r
t.test(eur.ret.weekly, mu = 0)
```

```
##
##  One Sample t-test
##
## data:  eur.ret.weekly
## t = 0.22596, df = 203, p-value = 0.8215
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.001769147  0.002227115
## sample estimates:
##    mean of x
## 0.0002289841
```

All three p-values are greater than 0.05, so we don't have enough evidence to reject the null hypothesis that mu = 0. This result is the same as daily data.

```r
# Ljung-Box test H0:acf(x,lag = i) = 0
Box.test (coredata(usd.ret.weekly, lag = 1, type = "Ljung"))
```

```
##
##  Box-Pierce test
##
## data:  coredata(usd.ret.weekly, lag = 1, type = "Ljung")
## X-squared = 1.5107, df = 1, p-value = 0.219
```

```r
Box.test (coredata(gbp.ret.weekly, lag = 1, type = "Ljung"))
```

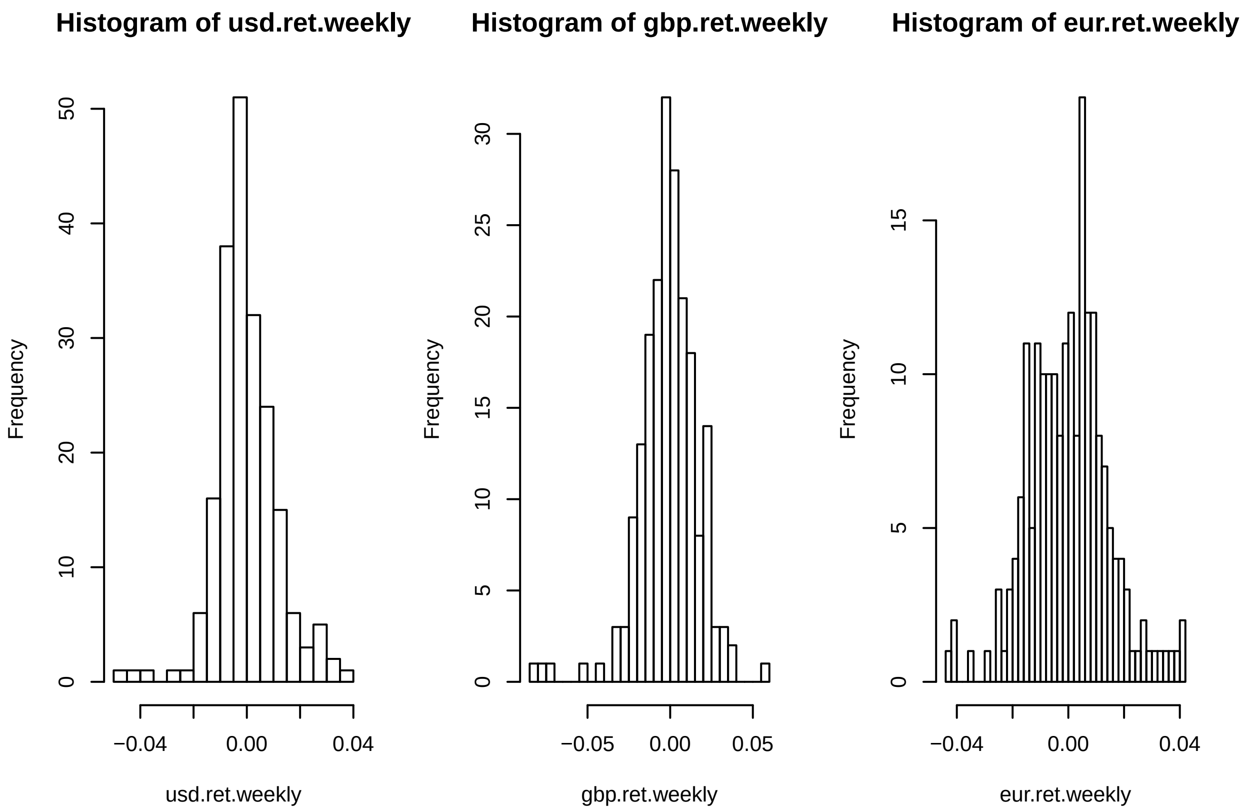
```
##
##  Box-Pierce test
##
## data:  coredata(gbp.ret.weekly, lag = 1, type = "Ljung")
## X-squared = 5.4519, df = 1, p-value = 0.01955
```

```r
Box.test (coredata(eur.ret.weekly, lag = 1, type = "Ljung"))
```

```
##
##  Box-Pierce test
##
## data:  coredata(eur.ret.weekly, lag = 1, type = "Ljung")
## X-squared = 0.10037, df = 1, p-value = 0.7514
```

For BGP weekly data, p-value of Ljung-Box test is less than 0.05, we should reject tha null that ACF of lag 1 = 0, for the rest two data set, p-values are not significant, same result as daily data.

```r
# histogram of weekly return
par(mfrow = c(1,3))
hist(usd.ret.weekly, breaks = 30)
hist(gbp.ret.weekly, breaks = 30)
hist(eur.ret.weekly, breaks = 30)
```




```
# Jarque-Bera test
jarque.bera.test(usd.ret.weekly)
```

```
##
##  Jarque Bera Test
##
## data:  usd.ret.weekly
## X-squared = 72.35, df = 2, p-value = 2.22e-16
```

```
jarque.bera.test(gbp.ret.weekly)
```

```
##
##  Jarque Bera Test
##
## data:  gbp.ret.weekly
## X-squared = 175.54, df = 2, p-value < 2.2e-16
```

```
jarque.bera.test(eur.ret.weekly)
```

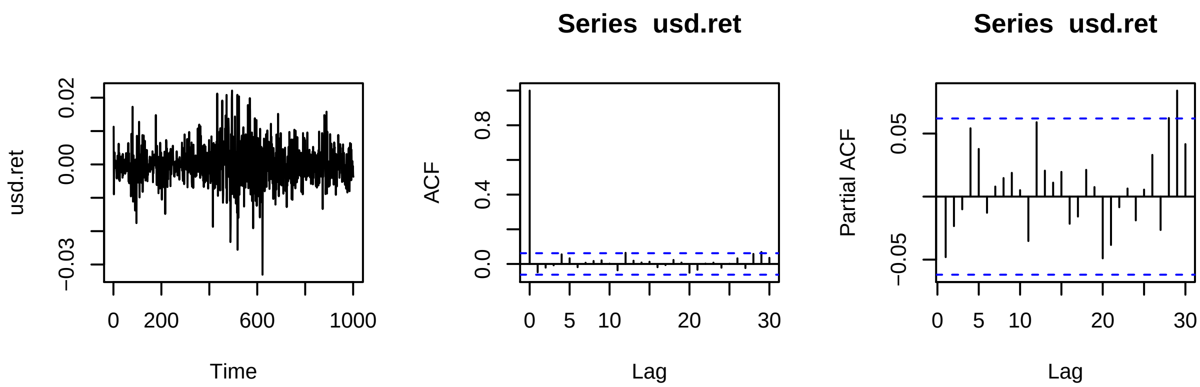
```
##
##  Jarque Bera Test
##
## data:  eur.ret.weekly
## X-squared = 4.8188, df = 2, p-value = 0.08987
```

For USD and GBP data, p-value are significant, we should reject the null hypothesis that data are following normal distribution. But for EUR data, we can not reject the null hypothesis.

Q2 ARMA & Garch model

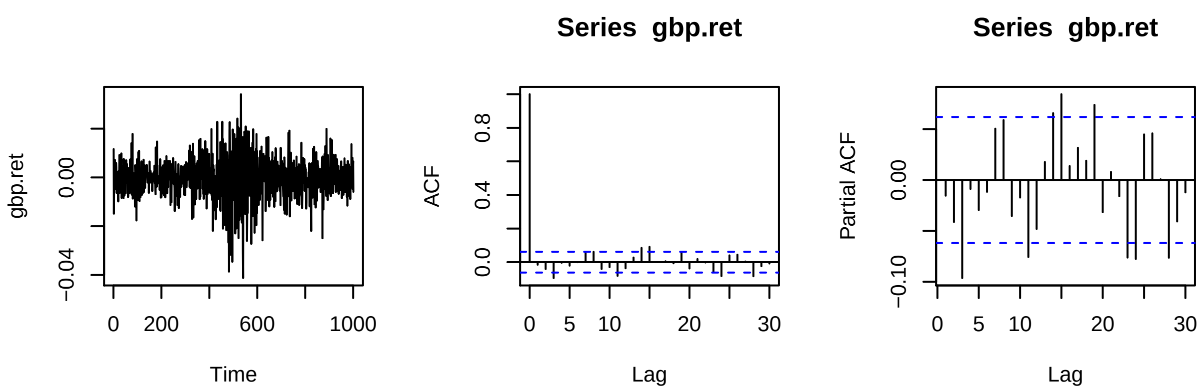
ARMA model

```r
par(mfrow = c(2,3))

ts.plot(usd.ret)
acf(usd.ret)
pacf(usd.ret)
```



For usd.ret, both ACF and PACF are not significat at any lag, it is random walk, so don't need to fit ARMA model.

```r
# plot GBP
par(mfrow = c(2,3))

ts.plot(gbp.ret)
acf(gbp.ret)
pacf(gbp.ret)
```



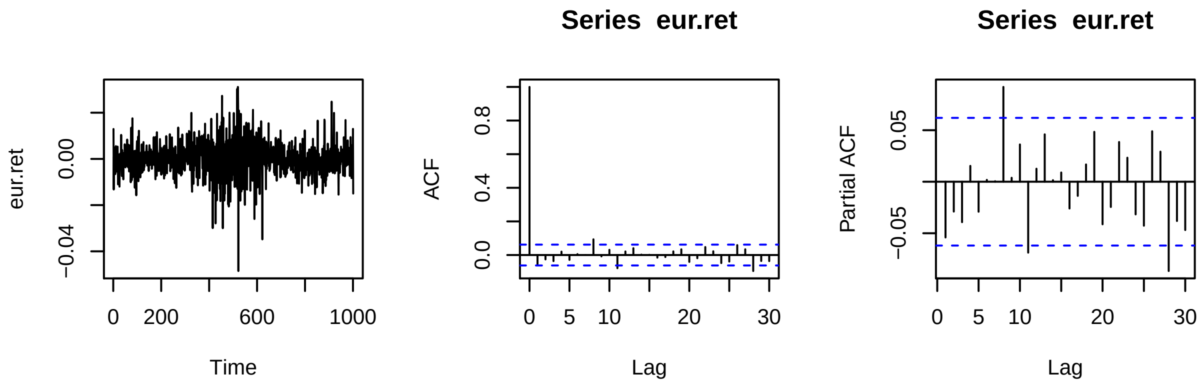
ACF is significant at lag3, and PACF is significant at lag 2, so fit an ARMA(2,3).

```r
# fit ARMA model
arima(gbp.ret, order = c(2,0,3))
```

```
##
## Call:
## arima(x = gbp.ret, order = c(2, 0, 3))
##
## Coefficients:
```

```
##           ar1      ar2      ma1       ma2       ma3  intercept
##       -0.0620  0.0407  0.0415  -0.0866  -0.0997      -2e-04
## s.e.   0.3038  0.2549  0.3022   0.2535   0.0357       2e-04
##
## sigma^2 estimated as 6.479e-05:  log likelihood = 3406.61,  aic = -6799.22
```

```r
# plot EUR
par(mfrow = c(2,3))

ts.plot(eur.ret)
acf(eur.ret)
pacf(eur.ret)
```



For usd.ret, both ACF and PACF are significat at lag 7, but not significant at any other lag, so it is random walk, we don't need to fit ARMA model.

(b) Garch Model

```r
# fit GARCH model for daily usd return
garch.model1 <- garch(usd.ret^2, order = c(1,1))
```

```
##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     4.628785e-09     1.000e+00
##      2     5.000000e-02     1.000e+00
##      3     5.000000e-02     1.000e+00
##
##     IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##      0    1 -9.002e+03
##      1   26 -9.002e+03  6.40e-06  3.04e-04  5.1e-09  1.0e+19  5.1e-10  1.59e+15
##      2   27 -9.003e+03  6.54e-05  5.75e-05  2.3e-09  2.0e+00  2.6e-10  9.32e+00
##      3   28 -9.003e+03  7.05e-07  7.86e-07  2.5e-09  2.0e+00  2.6e-10  9.40e+00
##      4   29 -9.003e+03  8.08e-09  8.20e-09  2.6e-09  2.0e+00  2.6e-10  9.39e+00
##      5   44 -9.016e+03  1.39e-03  2.27e-03  3.3e-01  2.0e+00  5.0e-02  9.38e+00
##      6   47 -9.031e+03  1.74e-03  2.78e-04  8.4e-01  0.0e+00  5.4e-01  2.78e-04
##      7   49 -9.091e+03  6.60e-03  1.79e-03  1.5e-01  2.0e+00  2.2e-01  2.47e+01
##      8   70 -9.095e+03  3.86e-04  2.63e-03  2.2e-10  2.1e+00  3.5e-10  2.44e+05
```

```
##     9   71 -9.103e+03  9.29e-04  3.19e-03  1.6e-10  2.0e+00  3.5e-10  2.93e+05
##    10   93 -9.138e+03  3.84e-03  2.94e-03  1.7e-02  2.0e+00  2.7e-02  4.20e+05
##    11   95 -9.148e+03  1.01e-03  9.49e-04  3.2e-03  2.0e+00  5.4e-03  5.96e+07
##    12   97 -9.162e+03  1.61e-03  1.98e-03  6.4e-03  2.0e+00  1.1e-02  7.10e+08
##    13   99 -9.196e+03  3.63e-03  5.63e-03  2.1e-02  2.0e+00  3.7e-02  1.08e+07
##    14  126 -9.207e+03  1.25e-03  3.87e-03  2.6e-11  3.5e+00  4.6e-11  3.29e+06
##    15  147 -9.213e+03  6.19e-04  1.00e-03  7.7e-03  2.0e+00  1.4e-02  3.97e+05
##    16  157 -9.214e+03  1.02e-04  2.69e-04  7.7e-12  5.6e+01  1.4e-11  5.86e+03
##    17  158 -9.214e+03  1.07e-05  1.39e-05  7.4e-12  2.0e+00  1.4e-11  4.10e+00
##    18  159 -9.214e+03  5.68e-07  6.12e-07  7.6e-12  2.0e+00  1.4e-11  5.39e+00
##    19  166 -9.214e+03 -3.73e-10  1.55e-09  8.4e-15  4.5e+00  1.5e-14  5.49e+00
##
##  ***** FALSE CONVERGENCE *****
##
##  FUNCTION    -9.213949e+03   RELDX        8.359e-15
##  FUNC. EVALS      166          GRAD. EVALS      19
##  PRELDF       1.545e-09      NPRELDF      5.495e+00
##
##      I      FINAL X(I)        D(I)          G(I)
##
##      1    5.724062e-11     1.000e+00     1.106e+09
##      2    1.453742e-01     1.000e+00     4.535e+01
##      3    8.971469e-01     1.000e+00    -7.902e+01
```

```
# fit GARCH model, use daily GBP return squred
garch.model2 <- garch(gbp.ret^2, order = c(1,1))
```

```
##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     1.825848e-08     1.000e+00
##      2     5.000000e-02     1.000e+00
##      3     5.000000e-02     1.000e+00
##
##    IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##     0    1 -8.332e+03
##     1   12 -8.332e+03  1.75e-06  3.25e-06  1.4e-09  1.4e+18  1.4e-10  2.25e+12
##     2   13 -8.332e+03  6.41e-09  4.89e-09  1.4e-09  2.0e+00  1.4e-10  2.77e+01
##     3   20 -8.332e+03 -7.12e-14  7.50e-15  3.9e-15  2.0e+00  3.9e-16 -8.14e-02
##
##  ***** FALSE CONVERGENCE *****
##
##  FUNCTION    -8.332294e+03   RELDX        3.859e-15
##  FUNC. EVALS       20          GRAD. EVALS       3
##  PRELDF       7.503e-15      NPRELDF     -8.141e-02
##
##      I      FINAL X(I)        D(I)          G(I)
##
##      1    1.811074e-08     1.000e+00    -1.622e+05
##      2    5.000000e-02     1.000e+00    -6.708e+02
##      3    5.000000e-02     1.000e+00    -1.042e+02
```

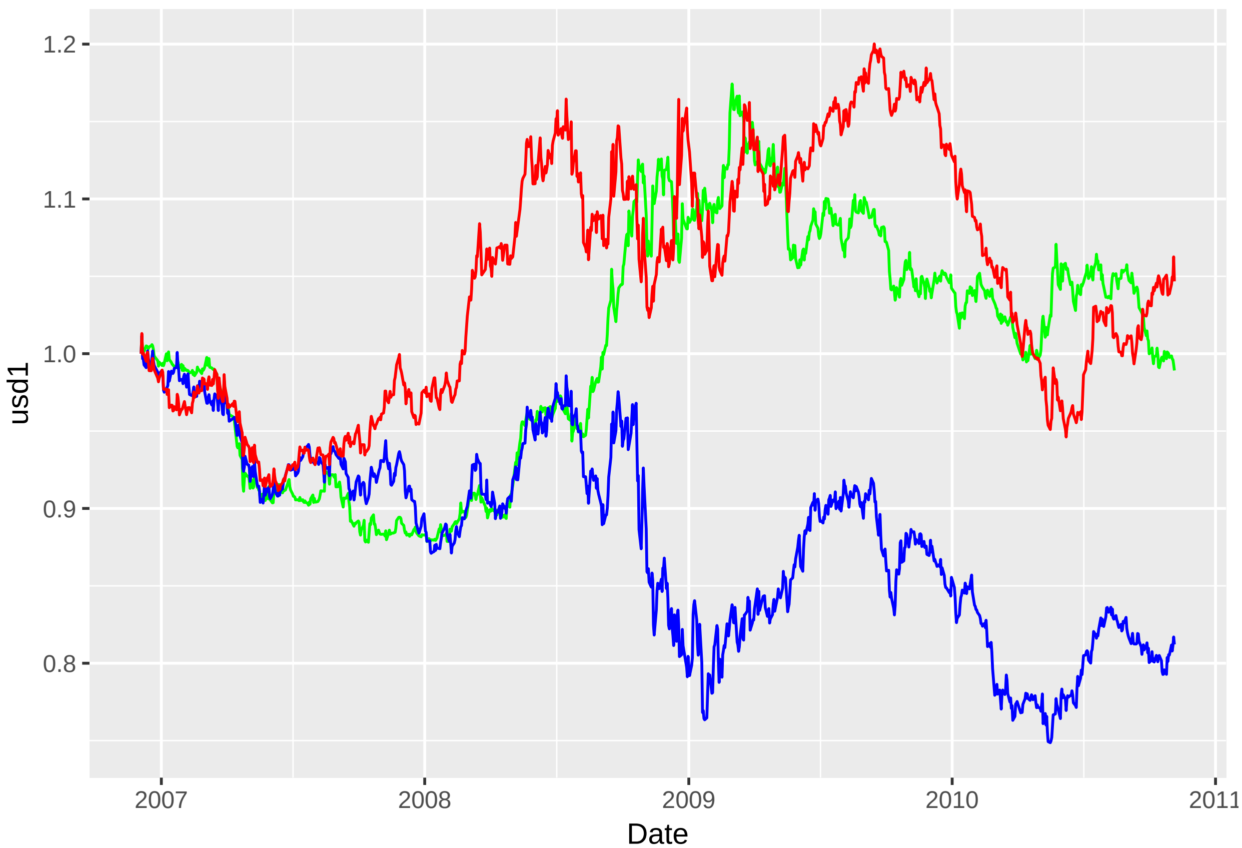
```r
# fit GARCH model for daily usd return
garch.model3 <- garch(eur.ret^2, order = c(1,1))
```

```
##
##  ***** ESTIMATION WITH ANALYTICAL GRADIENT *****
##
##
##      I     INITIAL X(I)        D(I)
##
##      1     1.517955e-08     1.000e+00
##      2     5.000000e-02     1.000e+00
##      3     5.000000e-02     1.000e+00
##
##     IT   NF      F         RELDF    PRELDF    RELDX   STPPAR   D*STEP   NPRELDF
##      0    1 -8.464e+03
##      1   11 -8.468e+03  4.41e-04  6.07e-04  1.4e-08  2.8e+18  1.4e-09  8.36e+14
##      2   12 -8.469e+03  6.09e-05  8.96e-05  9.7e-09  2.0e+00  1.4e-09  9.28e+01
##      3   13 -8.469e+03  8.44e-06  9.62e-06  1.3e-08  2.0e+00  1.4e-09  9.06e+01
##      4   14 -8.469e+03  1.01e-07  9.26e-08  1.4e-08  2.0e+00  1.4e-09  9.08e+01
##      5   28 -8.524e+03  6.54e-03  1.15e-02  4.4e-01  2.0e+00  7.9e-02  9.02e+01
##      6   29 -8.540e+03  1.88e-03  2.20e-03  3.1e-01  2.0e+00  7.9e-02  5.46e-01
##      7   32 -8.576e+03  4.17e-03  4.44e-03  5.3e-01  2.0e+00  3.2e-01  9.82e-01
##      8   34 -8.584e+03  9.27e-04  9.19e-04  6.5e-02  2.0e+00  6.3e-02  3.78e-01
##      9   36 -8.602e+03  2.09e-03  1.87e-03  1.1e-01  2.0e+00  1.3e-01  2.37e+00
##     10   38 -8.658e+03  6.50e-03  4.63e-03  1.6e-01  2.0e+00  2.5e-01  1.87e+02
##     11   40 -8.756e+03  1.12e-02  3.69e-03  2.7e-02  2.0e+00  5.0e-02  2.72e+04
##     12   50 -8.784e+03  3.15e-03  3.50e-03  5.3e-11  3.2e+00  1.0e-10  8.03e+07
##     13   56 -8.784e+03  4.17e-09  1.98e-08  6.5e-14  4.5e+00  1.2e-13  7.11e+04
##     14   58 -8.784e+03 -1.01e-09  1.87e-10  8.8e-15  2.0e+00  1.7e-14  7.81e+04
##
##  ***** FALSE CONVERGENCE *****
##
##  FUNCTION    -8.783957e+03   RELDX        8.807e-15
##  FUNC. EVALS      58         GRAD. EVALS      14
##  PRELDF       1.870e-10      NPRELDF      7.805e+04
##
##      I      FINAL X(I)        D(I)          G(I)
##
##      1    4.347006e-11     1.000e+00     1.972e+08
##      2    6.051868e-02     1.000e+00    -5.329e+02
##      3    9.456852e-01     1.000e+00     5.756e+01
```

(c) is there a co-movement among them

```r
# convert the starting points of the series unity
usd1 <- Q1.data$USD_by_INR / Q1.data$USD_by_INR[1]
gbp1 <- Q1.data$GBP_by_INR / Q1.data$GBP_by_INR[1]
eur1 <- Q1.data$EUR_by_INR / Q1.data$EUR_by_INR[1]

Q1.df <- data.frame(Date,usd1, gbp1, eur1)

library("ggplot2")
```

```r
ggplot(Q1.df, aes(x = Date, y = usd1)) +
  geom_line(aes(y = Q1.df$usd1), col = "green")+
  geom_line(aes(y = Q1.df$gbp1), col="blue") +
  geom_line(aes(y = Q1.df$eur1), col = "Red")
```



Does the co-movement vary over different time regimes? Yes, before end of 2008, all three series move in the similar way, but then, GBP rates went low and USD and EUR went high, before 2010, all three rates went in the same direction again.

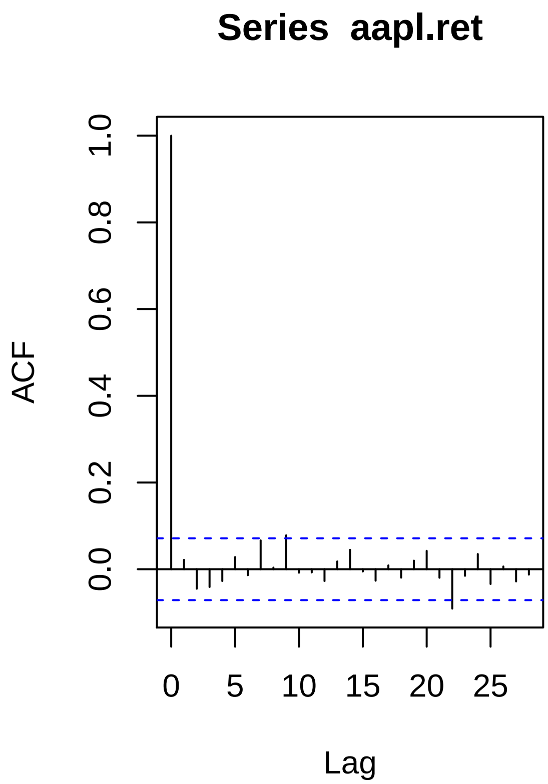
Q3

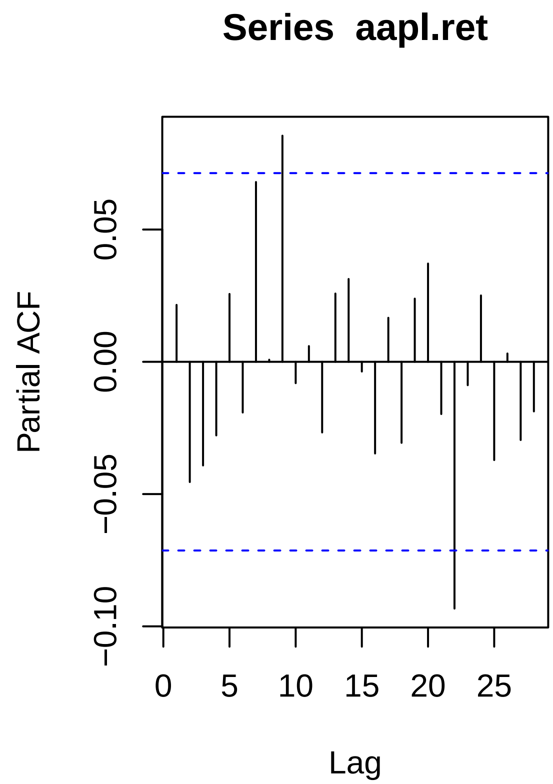
```r
# import data
Q3.data <- read.csv("AAPL.csv", header = T)
```

(a)

```r
# Daily log return
aapl.ret<- diff(log(Q3.data$Adj.Close))
```

```r
# ACF and PACF
par(mfrow=c(1,2))
acf(aapl.ret)
pacf(aapl.ret)
```





ACF and PACF are not significant when lag < 10, there is no serial correlation in the daily log returns.

```
# Ljung-Box Test
Box.test (aapl.ret, lag = 1, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  aapl.ret
## X-squared = 0.35099, df = 1, p-value = 0.5536
```

p-value is greater than 0.05, so we can not reject th null hypothesis that there is no correlation at lag 1.

(b)

```
# pivot1 = (high + low )/2
pivot1 = (Q3.data$High + Q3.data$Low)/2

# pivot2 = (high + low + close) /3
pivot2 = (Q3.data$High + Q3.data$Low + Q3.data$Close)/3
```

```
# pivot return
pivot1.ret <- diff(log(pivot1))
pivot2.ret <- diff(log(pivot2))
```

```
# test for serial correlation
Box.test (pivot1.ret, lag = 1, type = "Ljung")
```

```
##
##  Box-Ljung test
```

```
##
## data:  pivot1.ret
## X-squared = 12.485, df = 1, p-value = 0.0004103
```

P-value less than 0.05, we should reject the null hypothesis that there is no correlation at lag 1.

```
Box.test (pivot2.ret, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  pivot2.ret
## X-squared = 22.455, df = 1, p-value = 2.152e-06
```

P-value less than 0.05, we should reject the null hypothesis that there is no correlation at lag 1.

(c)

```
# unit root test of log price : augmented Dicky-Fuller test
adf.test(log(Q3.data$Adj.Close))
```

```
##
##  Augmented Dickey-Fuller Test
##
## data:  log(Q3.data$Adj.Close)
## Dickey-Fuller = -1.3346, Lag order = 9, p-value = 0.86
## alternative hypothesis: stationary
```

p-value is greater than 0.05, we cannot reject the null hypothesis that unit root exists.

Q4

Compute various measures of variance computed

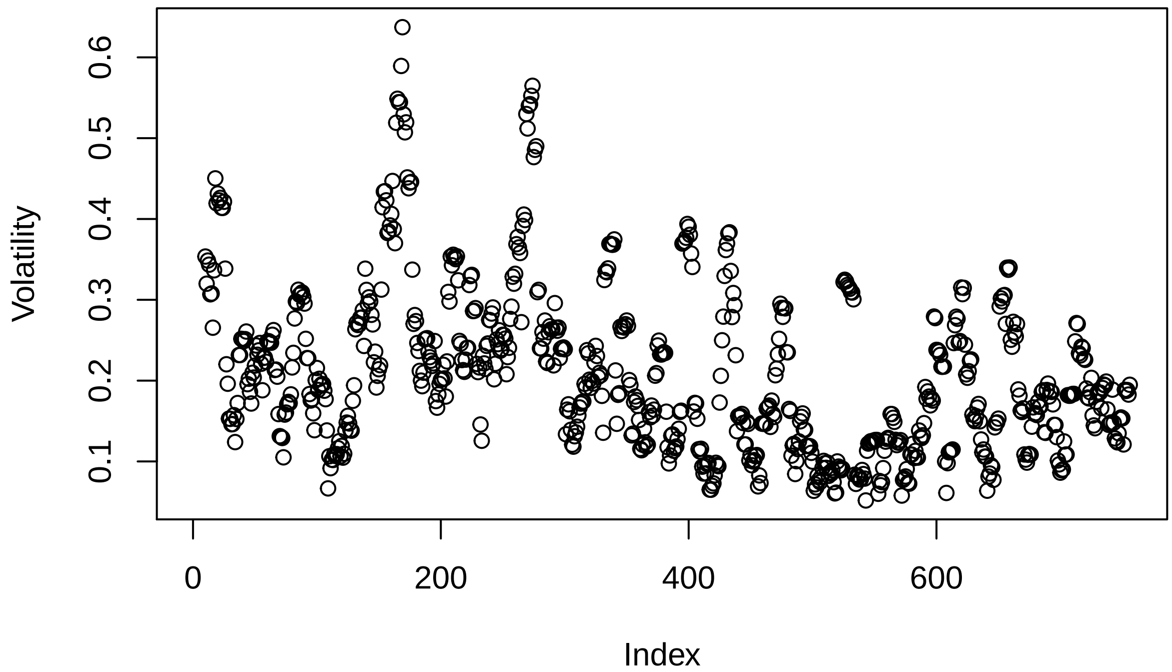
from the entries of the price bars.

Comment on their correlation with log volume.

```
# Using realized volatility in textbook page 108
# other answers are fine as well as it is reasonable
library("TTR")
```

```
## Warning: package 'TTR' was built under R version 3.4.3
```

```
Volatility <- volatility(Q3.data$Adj.Close,n=10)
plot(Volatility)
```



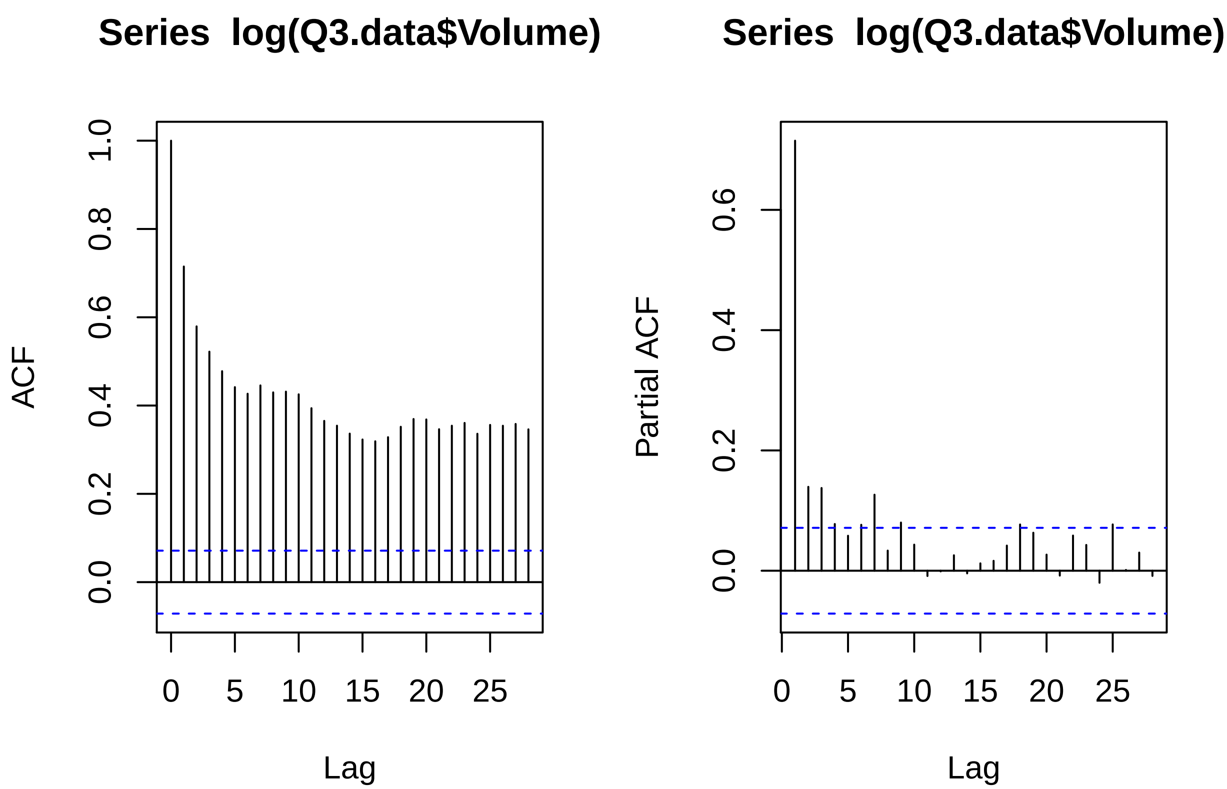

```r
# correlation with log volume
Volatility <- Volatility[!is.na(Volatility)]
log.Vol <- log(Q3.data$Volume)[-(1:9)]
cor(Volatility, log.Vol)
```

```
## [1] 0.5364175
```

It is relatively high.

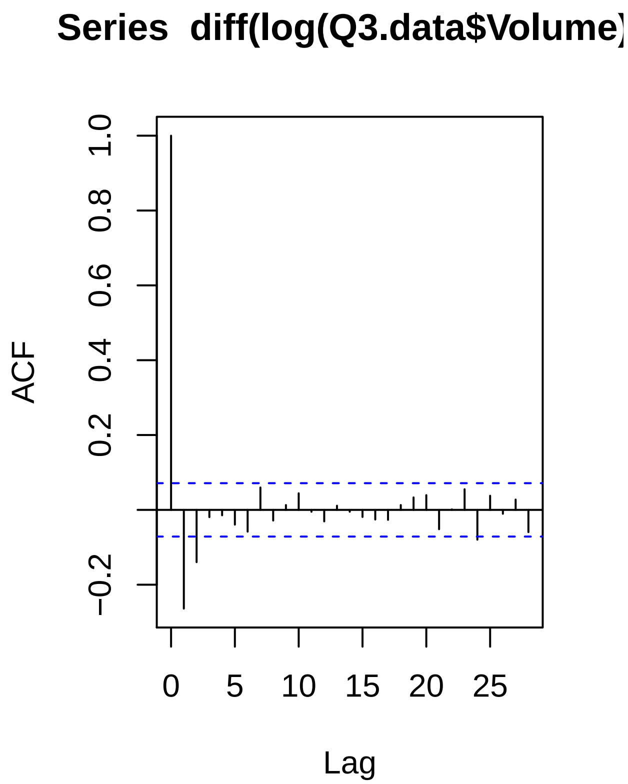
Use the ARIMA modeling to come up with a parsimonions model for log volume. ##Comment on the model accuracy by setting aside a validation data set.

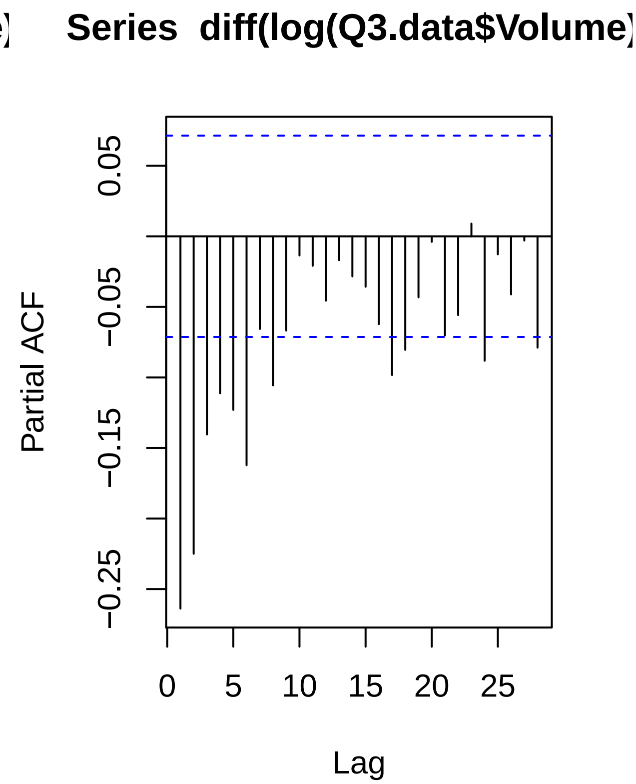
```r
par(mfrow = c(1,2))
acf(log(Q3.data$Volume))
pacf(log(Q3.data$Volume))
```



ACF is tail-off after many lags, so autocorrelation exist. PACF is significant at lag 1, lag2, lag3. Difference the data now.

```
par(mfrow = c(1,2))
acf(diff(log(Q3.data$Volume)))
pacf(diff(log(Q3.data$Volume)))
```





After difference once, ACF is significant at lag 1 and lag2. try an MA(2) model.

```r
arima(diff(log(Q3.data$Volume)), order = c(0,0,2))
```

```
##
## Call:
## arima(x = diff(log(Q3.data$Volume)), order = c(0, 0, 2))
##
## Coefficients:
##           ma1      ma2  intercept
##       -0.4754  -0.3070    -0.0012
## s.e.   0.0351   0.0402     0.0024
##
## sigma^2 estimated as 0.09068:  log likelihood = -165.56,  aic = 339.12
```

Both coefficients are significant.