MGMT 237E: Empirical Methods in Finance

Homework 4

Yi-Chi Chan Zhaofang Shi
Ahswin Kumar Ashok Kumar George Bonebright

February 16, 2016

1.a. Variance Ratio for Random Walk:

$y_t = \mu + y_t + \epsilon_t$

$= \mu + (\mu + y_{t-2} + \epsilon_{t-1}) + \epsilon_t$

$= 2\mu + y_{t-2} + \epsilon_t + \epsilon_{t-1}$

$= k\mu + y_{t-k} + \sum_{i=0}^{k-1} \epsilon_{t-i}$

$y_t - y_{t-k} = k\mu + \sum_{i=0}^{k-1} \epsilon_{t-i}$

$V[y_t - y_{t-k}] = k\sigma^2$

$\frac{V[y_t - y_{t-k}]}{k\sigma^2} = \frac{k\sigma^2}{k\sigma^2} = 1; k \to \infty$

Variance Ratio for AR(1) Model:

$for : |\phi| < 1$

$y_t = \mu + \phi y_{t-1} + \epsilon_t$

$= \mu + \phi(\mu + \phi y_{t-2} + \epsilon_{t-1}) + \epsilon_t$

$= (1 + \phi)\mu + \phi^2 y_{t-2} + \phi\epsilon_{t-1} + \epsilon_t$

$= (1 + \phi + ... + \phi^{k-1})\mu + \phi^k y_{t-k} + \phi^{k-1}\epsilon_{t-k+1} + ... + \phi\epsilon_{t-1} + \epsilon_t$

$y_t - y_{t-k} = \frac{1-\phi^k}{1-\phi}\mu + (\phi^k - 1)y_{t-k} + \phi^{k-1}\epsilon_{t-k+1} + ... + \phi\epsilon_{t-1} + \epsilon_t$

$V[y_t - y_{t-k}] = (\phi^k - 1)V[y_{t-k}] + \frac{1-\phi^k}{1-\phi}\sigma^2$

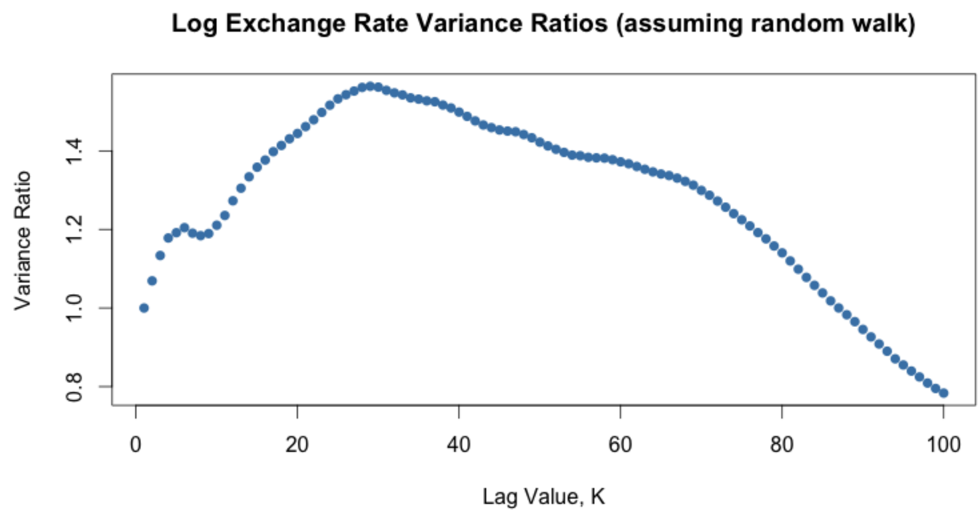
$= \frac{(\phi^k-1)\sigma^2}{1-\phi^2} + \frac{1-\phi^k}{1-\phi}\sigma^2$

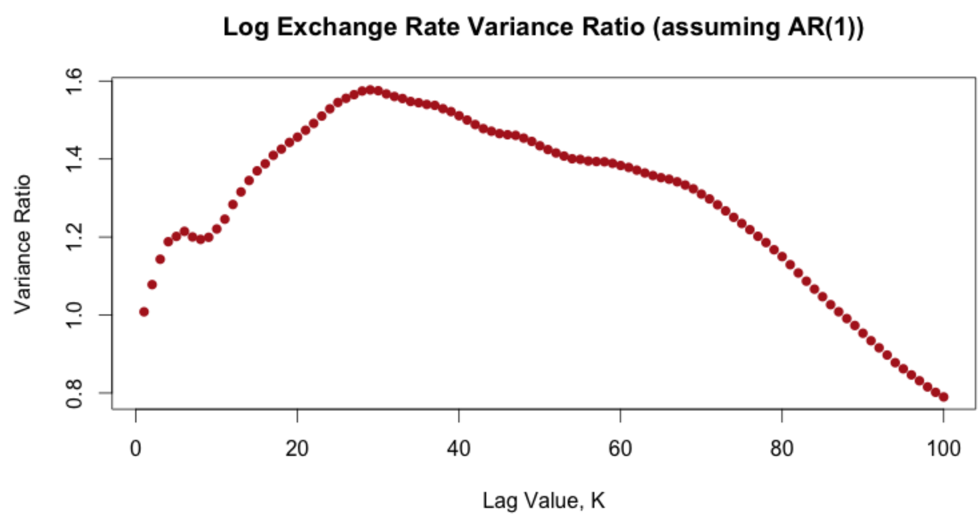
$= \frac{\sigma(1-\phi^k)}{1-\phi^2}\sigma^2$

$\frac{V[y_t - y_{t-k}]}{k\sigma^2} = \frac{\phi(1-\sigma^k)}{k(1-\phi^2)}$

$\frac{V[y_t - y_{t-k}]}{k\sigma^2} \to 0; k \to \infty$

1.b. Variance Ratio Plots:

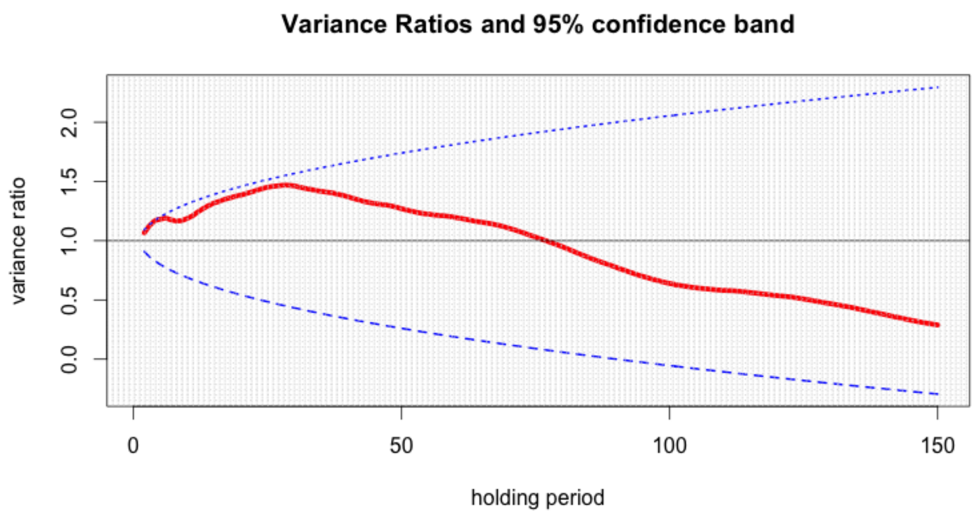




Variance ratio approaches 0 as lags increase, so the series may not follow a random walk, but a time trend. It should also be noted that the sample size must be significantly greater than the number of lags (k) in order to have a reasonable variance estimate. Otherwise, the ratio will always trend to zero as the number of samples decreases and the k value increases.

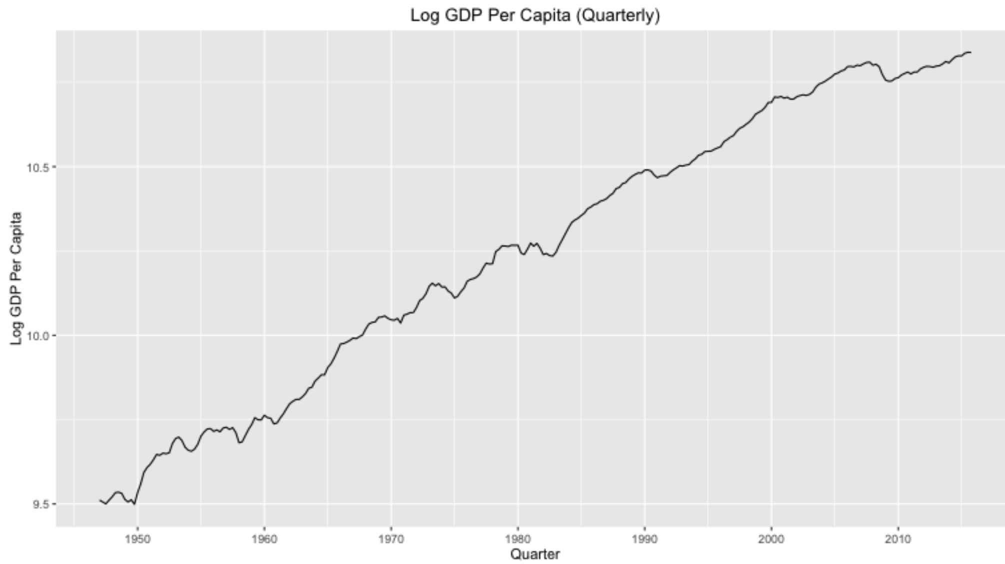
1.c. Variance Ratio test (vrtest):

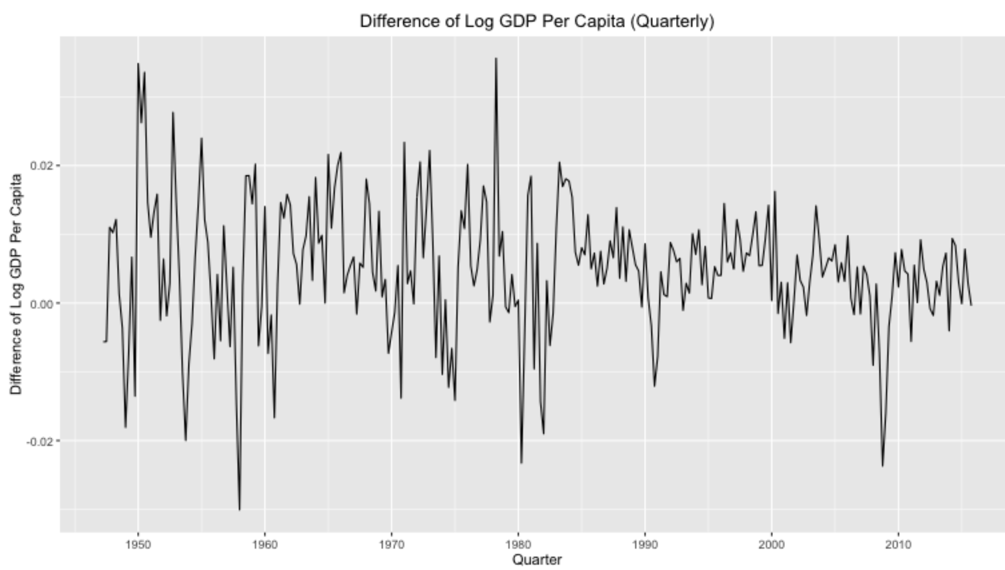
```
$stat
[1] 1.256834

$sum
[1] 1.122364
```



Variance ratio initially increases above 1, but decreases thereafter.

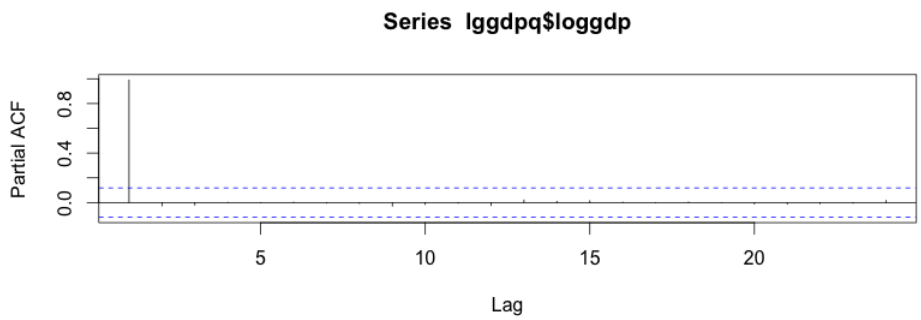
2.a. Per Capita GDP Plots:

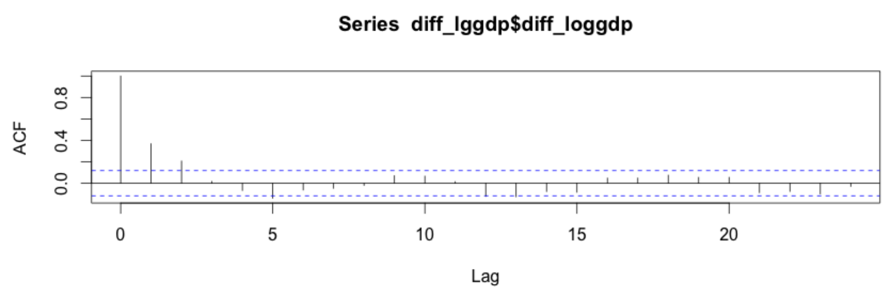




The plot of log GDP shows an upward trend. The first difference of log GDP seems to vary around about a constant mean, although the variation appears to change over time. The AR model of log/difference series is significant at lag 3, and the sample PACF of the log/difference series shown significant lag of 12. Therefore, Lag =13 that will take this into account in the augmented Dicky-Fuller unit root test. With growth of GDP over time, if the process has a unit root, it is expected to have a drift. If it is trend stationary, it is reasonable to expect a time trend.

Plots of ACF/PACF:





2.b. We will calculate two augmented Dicky-Fuller tests, one with a constant/drift, and the other one with a constant and a time trend.

ADF test with constant:
$H_0 : \phi_1 = 1; H_1 : |\phi_1| < 1$
$NullModel : loggdp_t = \phi_0 + loggdp_{t-1} + \epsilon_t$
$AlternativeModel : loggdp_t = \phi_0 + \phi_1 * loggdp_{t-1} + \epsilon_t$

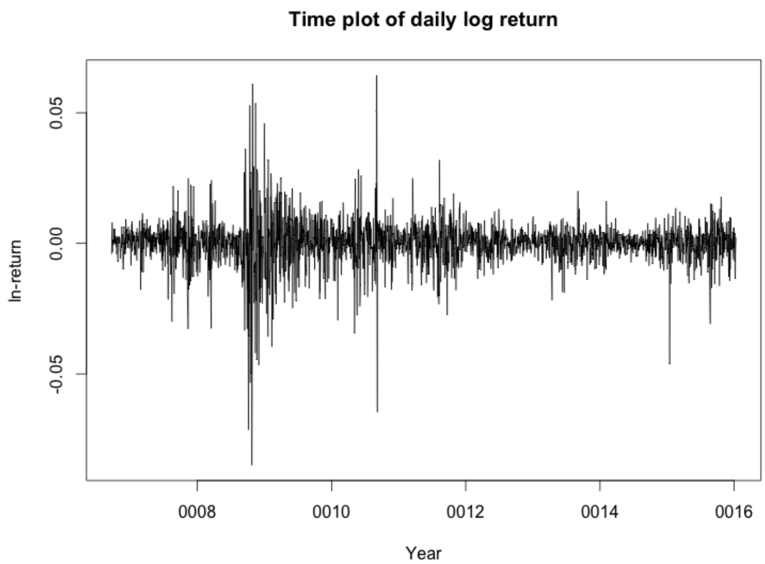
ADF test with a time trend and a constant:
$H_0 : \phi_1 = 1; H1 : |\phi_1| < 1$
$NullModel : loggdp_t = \phi_0 + loggdp_{t-1} + \vartheta * t + \epsilon_t$
$AlternativeModel : loggdp_t = \phi_0 + \phi_1 * loggdp_{t-1} + \vartheta * t + \epsilon_t$

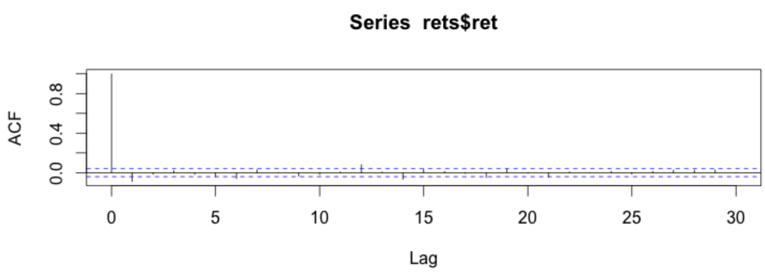
2.c. ADF Test Results:

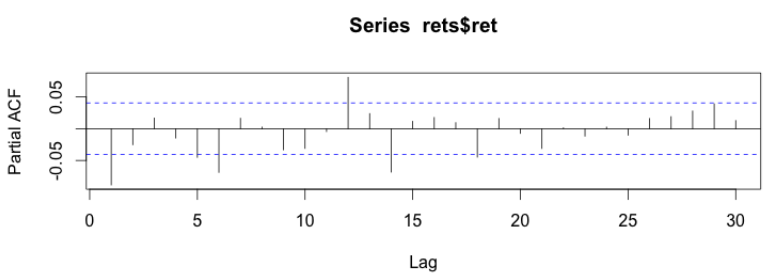
```
> adfTest (lggdpq$loggdp,lags=13,type=c("c"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 13
  STATISTIC:
    Dickey-Fuller: -1.7758
  P VALUE:
    0.3957

> adfTest (lggdpq$loggdp,lags=13,type=c("ct"))

Title:
 Augmented Dickey-Fuller Test

Test Results:
  PARAMETER:
    Lag Order: 13
  STATISTIC:
    Dickey-Fuller: -0.8176
  P VALUE:
    0.9592
```

Both tests fail to reject the null hypothesis of unit root in log GDP. It indicates that log GDP is non-stationary process.

3.a. Time Series Plot of Currency Data:





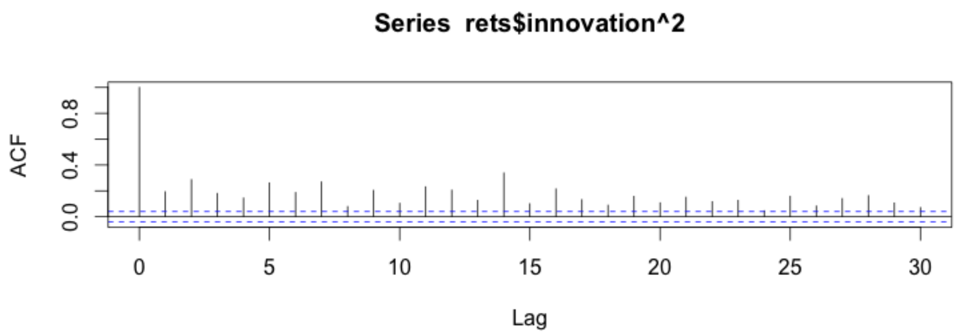


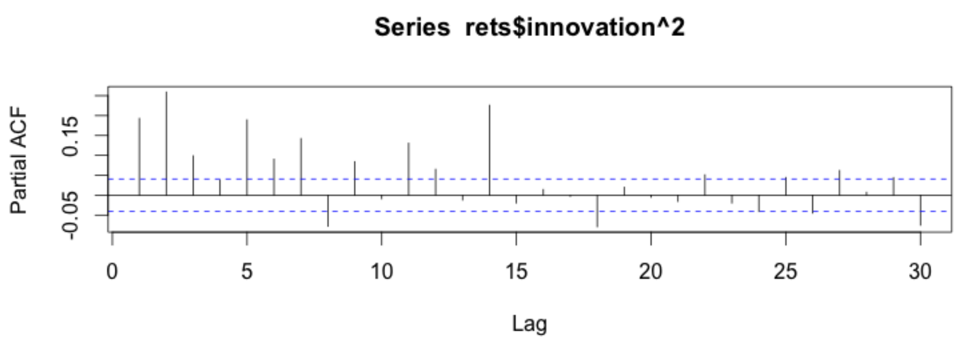
Log return seems to be a constant plus innovation. Abstract conditional mean from variance.

Ljung-Box Q-test rejects the null hypothesis of no autocorrelation in volatility.

```
        Box-Ljung test

data:  rets$innovation^2
X-squared = 2118.4, df = 30, p-value < 2.2e-16
```

ACF/PACF of squared-innovations:



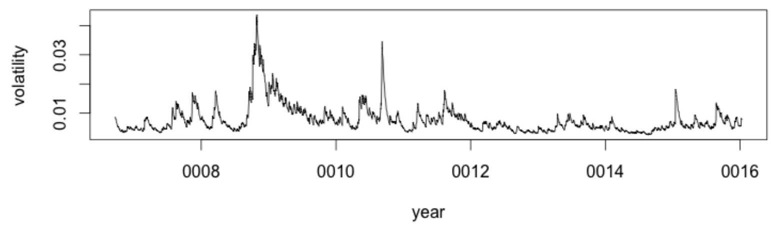


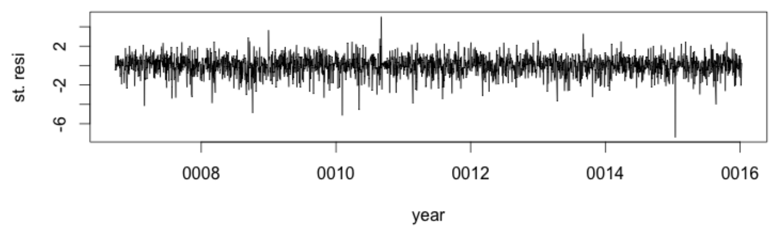
Squared-innovations show significant PACF for lags 1 to 12 as well as lag 18 and lag 30.
Try GARCH (1,1) with Gaussian innovations.

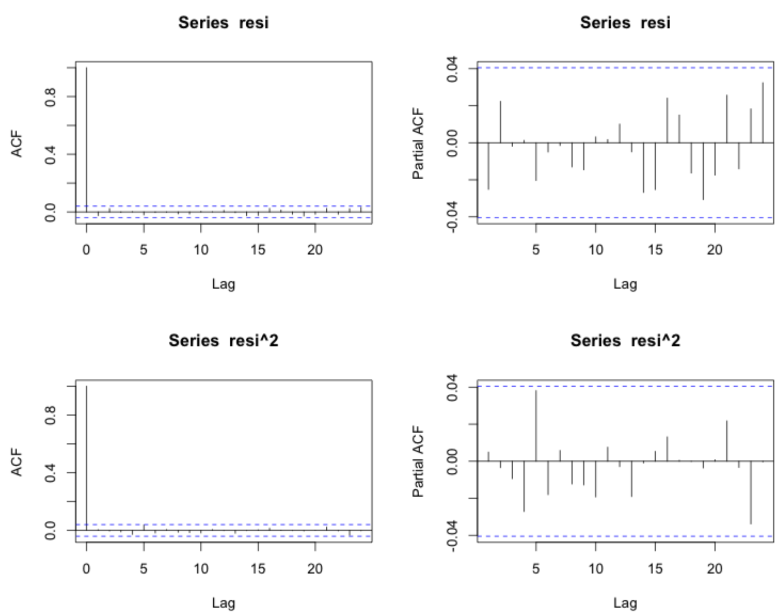
GARCH(1,1) Model Summary:

```
Title:
 GARCH Modelling

Call:
 garchFit(formula = ~1 + garch(1, 1), data = rets$ret, trace = F)

Mean and Variance Equation:
 data ~ 1 + garch(1, 1)
<environment: 0x10f886120>
 [data = rets$ret]

Conditional Distribution:
 norm

Coefficient(s):
        mu       omega      alpha1       beta1
1.5208e-04  5.9516e-07  1.3766e-01  8.6636e-01

Std. Errors:
 based on Hessian

Error Analysis:
        Estimate  Std. Error  t value Pr(>|t|)
mu     1.521e-04   1.103e-04    1.379 0.168042
omega  5.952e-07   1.700e-07    3.500 0.000464 ***
alpha1 1.377e-01   1.331e-02   10.341  < 2e-16 ***
beta1  8.664e-01   1.121e-02   77.290  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log Likelihood:
 8389.85    normalized:  3.582344

Standardised Residuals Tests:
                                Statistic p-Value
 Jarque-Bera Test   R    Chi^2  892.9103  0
 Shapiro-Wilk Test  R    W      0.9718357 0
 Ljung-Box Test     R    Q(10)  4.683022  0.9113224
 Ljung-Box Test     R    Q(15)  7.879335  0.9285291
 Ljung-Box Test     R    Q(20)  12.68464  0.8905009
 Ljung-Box Test     R^2  Q(10)  7.705103  0.6576169
 Ljung-Box Test     R^2  Q(15)  8.692776  0.8930256
 Ljung-Box Test     R^2  Q(20)  9.236635  0.9800008
 LM Arch Test       R    TR^2   8.01036   0.7843202

Information Criterion Statistics:
      AIC       BIC       SIC      HQIC
-7.161273 -7.151437 -7.161279 -7.157690
```

Volatility and ACF/PACF Plots:







No significant residuals; it appears that the GARCH(1,1) model is a good fit for the data.

3.b. 20 day Prediction:

	meanForecast	meanError	standardDeviation
1	0.0001520827	0.007362577	0.007362577
2	0.0001520827	0.007417575	0.007417575
3	0.0001520827	0.007472386	0.007472386
4	0.0001520827	0.007527016	0.007527016
5	0.0001520827	0.007581469	0.007581469
6	0.0001520827	0.007635751	0.007635751
7	0.0001520827	0.007689865	0.007689865
8	0.0001520827	0.007743816	0.007743816
9	0.0001520827	0.007797608	0.007797608
10	0.0001520827	0.007851245	0.007851245
11	0.0001520827	0.007904732	0.007904732
12	0.0001520827	0.007958072	0.007958072
13	0.0001520827	0.008011268	0.008011268
14	0.0001520827	0.008064326	0.008064326
15	0.0001520827	0.008117247	0.008117247
16	0.0001520827	0.008170037	0.008170037
17	0.0001520827	0.008222697	0.008222697
18	0.0001520827	0.008275232	0.008275232
19	0.0001520827	0.008327644	0.008327644
20	0.0001520827	0.008379937	0.008379937

The 20 trading day return volatility (standard deviation) is 0.03525.

3.c. Value at Risk at 5% on $2billion long position in currency is $115.95 million.
The m1 model predicts a constant mean of 1.521e-04, but it is not significantly different from 0, given its p-value of 0.168¿.05. So we assume mean of 20-day log return is 0.

```r
# MGMT237e HW5
##setwd("C:/Users/SallyShi/Desktop/MGMT237E-Empirical Methods in Finance/HW4")
##install.packages("vrtest")
##install.packages("fGarch")

library(lubridate)
library(xts)
library(ggplot2)
library(dplyr)
library(vrtest)
library(fUnitRoots)
library(zoo)
library(fGarch)

# question 1 part b
# get monthly exchange rate from 1970 to 2008
usdi<-read.csv("dollar_broadindex_tradeweighted.csv")
usdi$Date<-as.Date(usdi$Date,format("%m/%d/%Y"))
usdi<-usdi[which(usdi$Date=="1971-01-31"):which(usdi$Date=="2008-11-30"),]
usdi<-select(usdi,Date,Close)
usdi$log<-log(usdi$Close)
#calculate and plot variance ratio for log exchange rate
sigma_sq1=var(diff(usdi$log))
ar1<-arima(usdi$log,order=c(1,0,0))
sigma_sq2=ar1$sigma2

par(mfrow=c(1,1))
vr1_usdi<- sapply(1:100,function(x){
  var(diff(usdi$log,lag=x))/(x*sigma_sq1)
})
plot(vr1_usdi,main="Log Exchange Rate Variance Ratios (assuming random walk)", xlab="Lag Value, K", ylab="Variance Ratio",pch=16, col="steelblue")
vr2_usdi<- sapply(1:100,function(x){
  var(diff(usdi$log,lag=x))/(x*sigma_sq2)
})
plot(vr2_usdi,main="Log Exchange Rate Variance Ratio (assuming AR(1))", xlab="Lag Value, K", ylab="Variance Ratio",pch=16, col="firebrick")
# variance ratio approximate 0 as lags goes larger, so the series may not follow a random walk,
# but a time trend.

# question 1 part c
y <- usdi$log
nob <- length(y)
r <- diff(y,lag=1)
Auto.VR(r)

kvec <- c(2,5,10,30,50,100,150)
VR.plot(r,kvec)
# variance ratio increase above 1 first, and decrease below 1 as holding period gets gets longer.
# It also shows that the series is not unit root.

```

```r
# question 2 part a
gdpq<-read.csv("rgdppc.csv")
colnames (gdpq) <-c("Date","GDP")
gdpq$Date<-as.yearqtr(gdpq$Date,format = "%Y-%m-%d")
gdpq$Date <-as.Date(gdpq$Date, format="%Y%q")
gdpq$loggdp<-log(gdpq$GDP)
lggdpq<-gdpq[,c("Date","loggdp")]
# get difference for quarterly gdp
diff_lggdp <-diff(gdpq$loggdp)
diff_lggdp <-data.frame(gdpq$Date[2:length(gdpq$Date)],diff_lggdp)
colnames (diff_lggdp)<-c("Date","diff_loggdp")
# plot Log GDP and its changes
p1<-ggplot(lggdpq)+geom_line(aes(lggdpq$Date,lggdpq$loggdp)) + xlab("Quarter") + ylab("Log GDP Per Capita") + ggtitle("Log GDP Per Capita (Quarterly)")
p1
p2<-ggplot(diff_lggdp)+geom_line(aes(diff_lggdp$Date, diff_lggdp$diff_loggdp)) + xlab("Quarter") + ylab("Difference of Log GDP Per Capita") + ggtitle("Difference of Log GDP Per Capita (Quarterly)")
p2

par(mfrow=c(2,1))
acf(lggdpq$loggdp)
pacf(lggdpq$loggdp)

par(mfrow=c(2,1))
acf(diff_lggdp$diff_loggdp)
pacf(diff_lggdp$diff_loggdp)

m1=ar(diff_lggdp$diff_loggdp,method = 'mle')
m1$order

# The plot of log GDP shows an upward trend. The first difference of log GDP seems to vary around a
# fixed mean leve, although the variabilty appears to be smaller in recent years.
# AR model of difference series is significant at lag 3, and the sample PACF of the differenced
# series shown significant lag of 12. Therefore, Lag =13 that will take this into account in the
# augmented Dicky-Fuller unit root test.

# With growth of GDP over time, if the process has a unit root, it is expected to have a drift.
# If it is trend stationary, it is reasonable to expect a time trend.

# question 2 part b

# Therefore, we will do two augmented Dicky-Fuller test, one with a constant/drift,
# and the other one with a constant and a time trend.

# ADF test with constant
# H0: phi_1=1 H1:|phi_1|<1
# null model: loggdp_t=phi_0+loggdp_t-1+error_t
```

```
# alternative model: loggdp_t=phi_0+phi_1*loggdp_t-1+error_t

# ADF test with a time trend and a constant
# H0: phi_1=1 H1:|phi_1|<1
# null model: loggdp_t = phi_0 + loggdp_t-1 + omega*t + error_t
# alternative model: loggdp_t = phi_0 + phi_1*loggdp_t-1 + omega*t + error_t

# question 2 part c
# ADF test with constant
adfTest (lggdpq$loggdp,lags=13,type=c("c"))
# ADF test with a time trend and a constant
adfTest (lggdpq$loggdp,lags=13,type=c("ct"))

# Both test fail to reject the null hypothesis of unit root in log GDP. It indicates that log GDP
# is non-stationary process.

# question 3 part a
data <-read.csv("currency_fund_prices.csv")
data$Date <- as.Date(data$Date,format("%m/%d/%Y"))
# calculate daily log return
ccy<-select(data,Date,Adj.Close)
ccy$ret[2:length(ccy$Adj.Close)] <-diff(log(ccy$Adj.Close))
rets <- select(ccy,Date,ret)
rets <- rets[2:length(ccy$ret),]
par(mfrow=c(1,1))
plot(rets,type="l",xlab = 'Year', ylab = 'ln-return', main ='Time plot of daily log return')
par(mfrow=c(2,1))
acf(rets$ret, lag=30)
pacf(rets$ret, lag=30)

# seems log return is a constant plus innovation. abstract conditional mean from variation
rets$innovation <-rets$ret-mean(rets$ret)
Box.test(rets$innovation^2,lag=30,type='Ljung')
# LB Q-test rejects the null of no autocorrelation in volatility
par(mfrow=c(2,1))
acf(rets$innovation^2, lag=30)
pacf(rets$innovation^2, lag=30)

# innovation squared shows significant PACF for 1 to 12 lags as well as lag 18 and 30.
# one can employ more parsimonious GARCH model
# Try GARCH (1,1) with Gaussian innovations
m1 <- garchFit(~1+garch(1,1),data=rets$ret,trace=F)
summary(m1)
# GARCH(1,1) model: r_t = 1.521e-04 + a_t, a_t = sigma_t * e_t
# sigma_t^2 = 5.952e-07 + 1.377e-01 * a_(t-1)^2 + 8.664e-01 * sigma_(t-1)^2
# AIC = -7.161273
v1 = volatility(m1) # obtain volatility
resi <- residuals(m1,standardize=T) # standard residuals
vol <- data.frame(rets$Date,v1)
res <- data.frame(rets$Date,resi)
par(mfrow=c(2,1)) # show volatility and residuals
```

```
plot (vol,xlab='year',ylab='volatility', type='l')
plot (res,xlab='year',ylab='st. resi', type='l')
par(mfrow=c(2,2)) # obtain ACF & PACF
acf(resi, lag=24)
pacf(resi, lag=24)
acf(resi^2, lag=24)
pacf(resi^2, lag=24)
# no significant ACF/PACF for GARCH residuals
# model is good fit for the data

# question 3 part b
# prediction of log return level for next 20 periods
pred <- predict(m1,20)
# log return volatility over 20 days, assuming iid daily log return, variance over 20 day period
# equals sum of standard error squared on each day.
std <- sqrt(sum(pred$standardDeviation^2))
# the 20-trading-day return volatility (standard deviation) is 0.03525

# question 3 part c
# VaR at 5% on $2bn long position in currency is $115.95 million,
# m1 model predict constant mean is 1.521e-04, but it is not significant different from 0, given
# its p-value is 0.168>.05. So we assume mean of 20-day log return approximates 0.
2000*abs(qnorm(.05,mean=0,sd=std))
```