Time Series HW12

18/30

B082040005 高念慈

2023-05-19

1. 8

This problem is concerned with the dynamic (動態) relationship between the spot and futures prices of the S&P 500 index. The data file sp5may.dat hasthree columns: log(futures price), log(spot price), and cost-of-carry (×100). The data were obtained from the Chicago Mercantile Exchange for the S&P 500 stock index in May 1993 and its June futures contract. The time interval is 1 minute (intraday). Several authors used the data to study index futures arbitrage (套利). Here we focus on the first two columns. Let f_t and s_t be the log prices of futures and spot (現貨), respectively. Consider $y_t = f_t - f_{t-1}$ and$ x_t=s_t-s_{t-1}$ Build a regression model with time series errors between $\{y_t\}$ and$\{x_t\}$, with y_t being the dependent variable.

```
sp = read.table('https://faculty.chicagobooth.edu/-/media/faculty/ruey-s-tsay/teaching/fts3/sp5may.da
t',header=T)
head(sp)
```

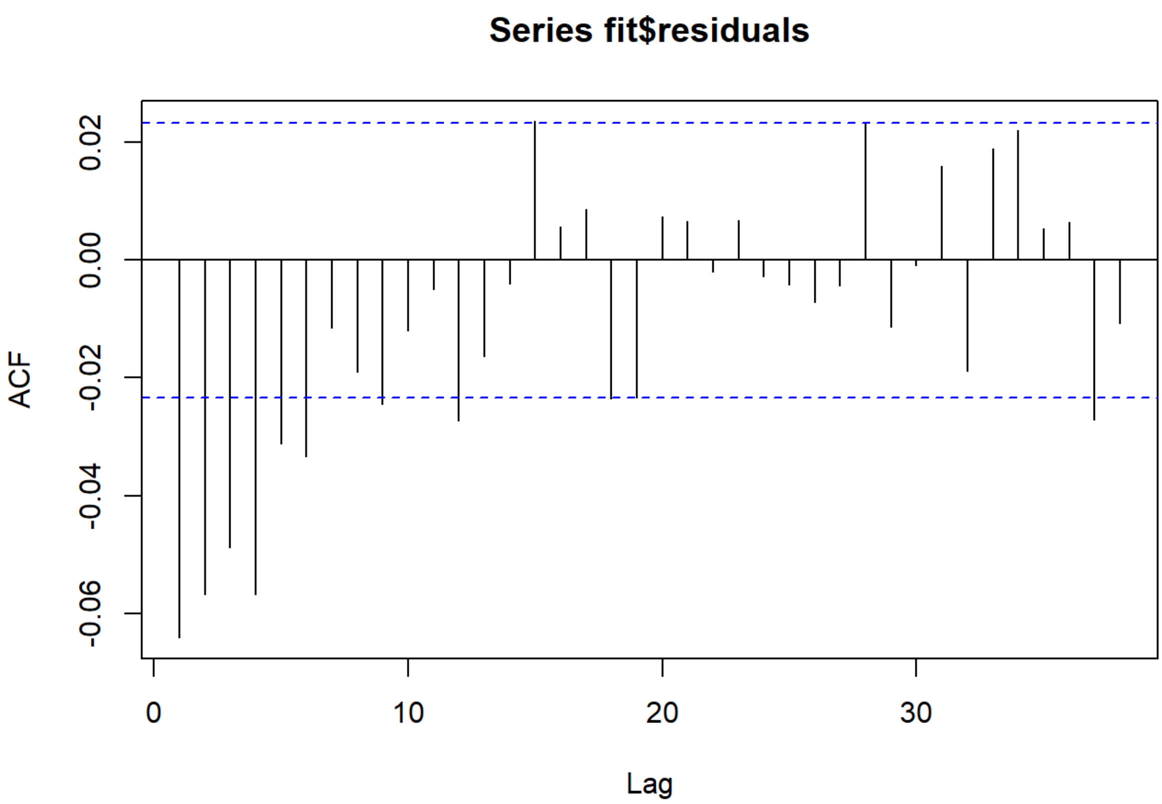
```
##   lnfuture  lnspot     cost
## 1  6.08382 6.08618 -0.16501
## 2  6.08404 6.08623 -0.16501
## 3  6.08473 6.08630 -0.16501
## 4  6.08450 6.08630 -0.16501
## 5  6.08450 6.08623 -0.16501
## 6  6.08439 6.08625 -0.16501
```

```
ft = sp$lnfuture
st = sp$lnspot
yt = diff(ft)
xt = diff(st)

fit = lm(yt ~ xt)
summary(fit)
```

```
##
## Call:
## lm(formula = yt ~ xt)
##
## Residuals:
##        Min         1Q     Median         3Q        Max
## -0.0038484 -0.0001568 -0.0000014  0.0001612  0.0026256
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.354e-06  3.509e-06   0.386      0.7
## xt          6.212e-01  1.754e-02  35.420   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.0002948 on 7058 degrees of freedom
## Multiple R-squared:  0.1509, Adjusted R-squared:  0.1508
## F-statistic:  1255 on 1 and 7058 DF,  p-value: < 2.2e-16
```

```
acf(fit$residuals)
```




```
Box.test(fit$residuals, lag=12, type = "Ljung-Box")
```

```
##
##  Box-Ljung test
##
## data:  fit$residuals
## X-squared = 120.19, df = 12, p-value < 2.2e-16
```

- p−value < 0.05 has serial correlation(不好)
- Build a regression model with time series error

```
eacf(fit$residuals)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x x x x x x o o x o o  x  o  o
## 1 x o o o o o o o o o o  o  o  o
## 2 x x o o o o o o o o o  o  o  o
## 3 x x x o o o o o o o o  o  o  o
## 4 x x x x o o o o o o o  o  o  o
## 5 x x x o x o o o o o o  o  o  o
## 6 x x x x x x o o o o o  o  o  o
## 7 x x x x x x x o o o o  o  o  o
```

- AR=1,MA=1

```
fu_sp_model = arima(yt, order = c(1,0,1),xreg = xt, include.mean = F)
fu_sp_model
```

```
##
## Call:
## arima(x = yt, order = c(1, 0, 1), xreg = xt, include.mean = F)
##
## Coefficients:
##          ar1      ma1    xreg
##       0.8204  -0.9359  0.7203
## s.e.  0.0127   0.0083  0.0177
##
## sigma^2 estimated as 8.389e-08:  log likelihood = 47499.05,  aic = -94992.11
```

```
rbind(fu_sp_model$coef-2*sqrt(diag(fu_sp_model$var.coef)),
      fu_sp_model$coef+2*sqrt(diag(fu_sp_model$var.coef)))
```

```
##            ar1        ma1      xreg
## [1,] 0.7950409 -0.9524298 0.6848372
## [2,] 0.8457255 -0.9193937 0.7557746
```

```
Box.test(fu_sp_model$residuals, lag=12, type = "Ljung-Box")
```

fitdf - 2

```
##
##  Box-Ljung test
##
## data:  fu_sp_model$residuals
## X-squared = 8.7632, df = 12, p-value = 0.723
```

- p-value > 0.05 has no serial correlation,(足夠)
- $y_t = 0.7203x_t + e_t$, $e_t = 0.8204e_{t-1} + at - 0.9359a_{t-1}$

6'

2.

The file m-mrk4608.txt contains monthly simple returns of Merck stock from June 1946 to December 2008. The file has two columns denoting (表示) date and simple return. Transform the simple returns to log returns.

```r
data_mrk = read.table('https://faculty.chicagobooth.edu/-/media/faculty/ruey-s-tsay/teaching/fts3/m-mr
k4608.txt',header=T)
head(data_mrk)
```

```
##       date       mrk
## 1 19460628 -0.025926
## 2 19460731 -0.030534
## 3 19460830  0.043307
## 4 19460930 -0.105660
## 5 19461031 -0.008475
## 6 19461130  0.064103
```

```r
logrtn_mrk = log(data_mrk$mrk + 1)
head(logrtn_mrk)
```

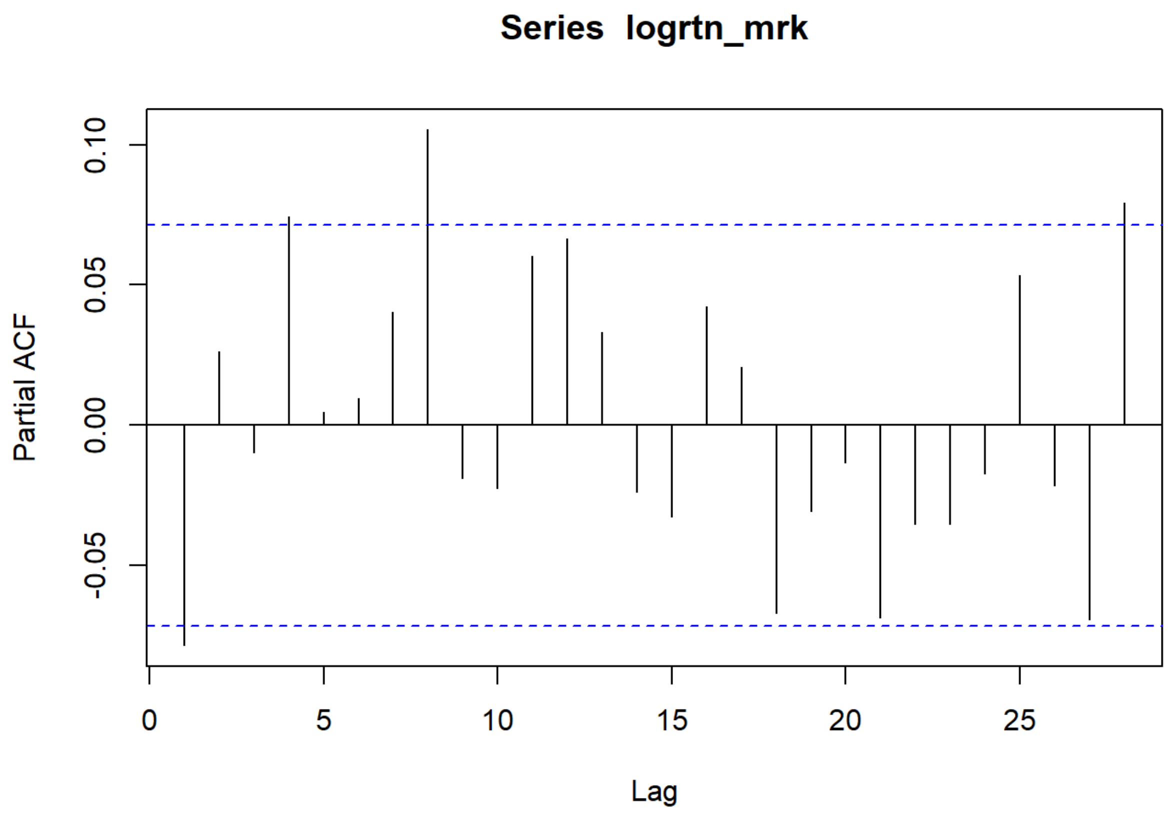
```
## [1] -0.026268003 -0.031009875  0.042395476 -0.111669263 -0.008511117
## [6]  0.062132191
```

a. Is there any evidence of serial correlations in the log returns? Use auto-correlations and 5% significance level to answer the question. If yes, remove the serial correlations.

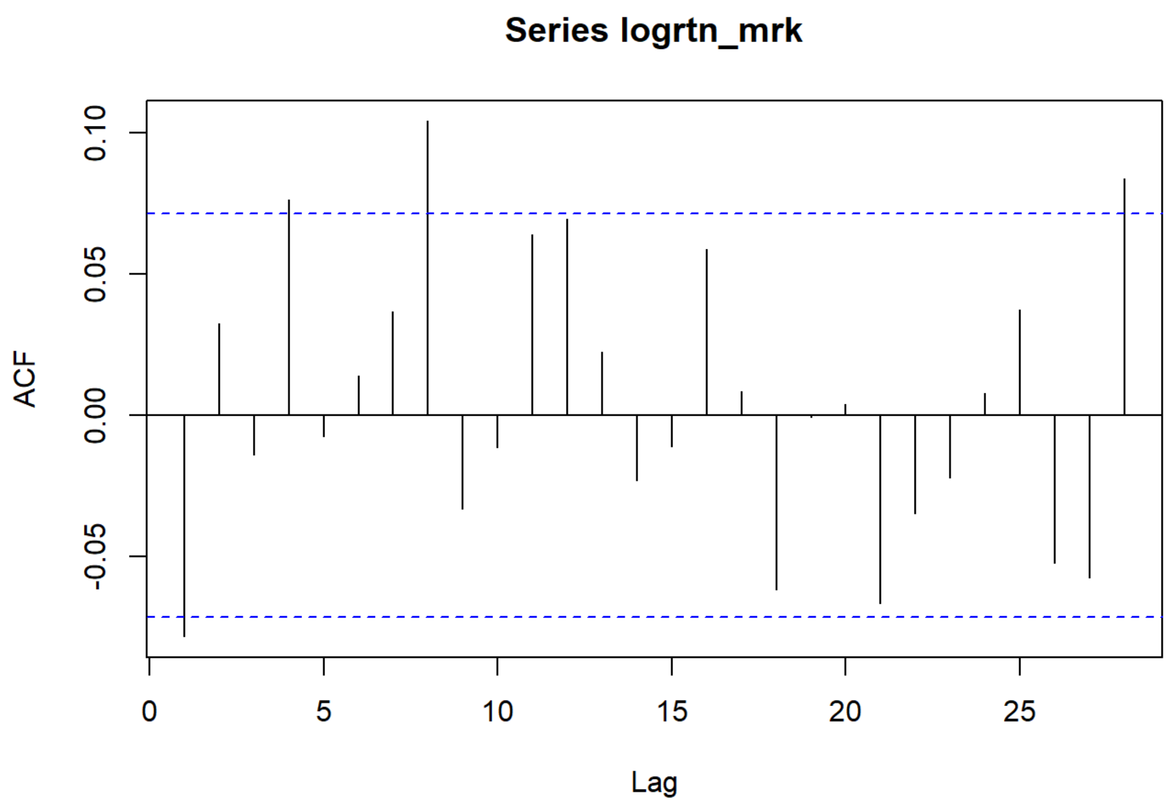
```r
Box.test(logrtn_mrk, lag = 12, type = "Ljung")
```

```
##
##  Box-Ljung test
##
## data:  logrtn_mrk
## X-squared = 27.236, df = 12, p-value = 0.007144
```

```r
pacf(logrtn_mrk)
```




```
acf(logrtn_mrk)
```



- There are serial correlation of log return of mrk (p-value < 0.05)

```
adfTest(logrtn_mrk,lags=12,type=c("c"))  # p-value < 0.05，不做差分
```

```
##
## Title:
##  Augmented Dickey-Fuller Test
##
## Test Results:
##   PARAMETER:
##     Lag Order: 12
##   STATISTIC:
##     Dickey-Fuller: -5.7098
##   P VALUE:
##     0.01
##
## Description:
##  Fri May 19 06:02:07 2023 by user: user
```

```
eacf(logrtn_mrk)
```

```
## AR/MA
##   0 1 2 3 4 5 6 7 8 9 10 11 12 13
## 0 x o o x o o o x o o o  o  o  o
## 1 x o o x o o o x o o o  o  o  o
## 2 x o o o o o o x o o o  o  o  o
## 3 x x x x o o o x o o o  o  o  o
## 4 o x x x o o o o o o o  o  o  o
## 5 x o o x x o o o o o o  o  o  o
## 6 x x o x x x o o o o o  o  o  o
## 7 x o o x x x x o o o o  o  o  o
```

```
m1 = arima(logrtn_mrk, order=c(2,0,3))
m1
```

```
##
## Call:
## arima(x = logrtn_mrk, order = c(2, 0, 3))
##
## Coefficients:
##           ar1     ar2      ma1      ma2     ma3  intercept
##       -0.0054  0.8207  -0.0786  -0.7662  0.0757     0.0106
## s.e.   0.1406  0.1236   0.1448   0.1445  0.0472     0.0032
##
## sigma^2 estimated as 0.004918:  log likelihood = 930.1,  aic = -1848.2
```

```
rbind(m1$coef-2*sqrt(diag(m1$var.coef)),
      m1$coef+2*sqrt(diag(m1$var.coef)))
```

```
##             ar1      ar2        ma1        ma2         ma3   intercept
## [1,] -0.2865984 0.573523 -0.3682402 -1.0551257 -0.01856149 0.004200501
## [2,]  0.2758634 1.067852  0.2110977 -0.4771746  0.17004015 0.016983852
```

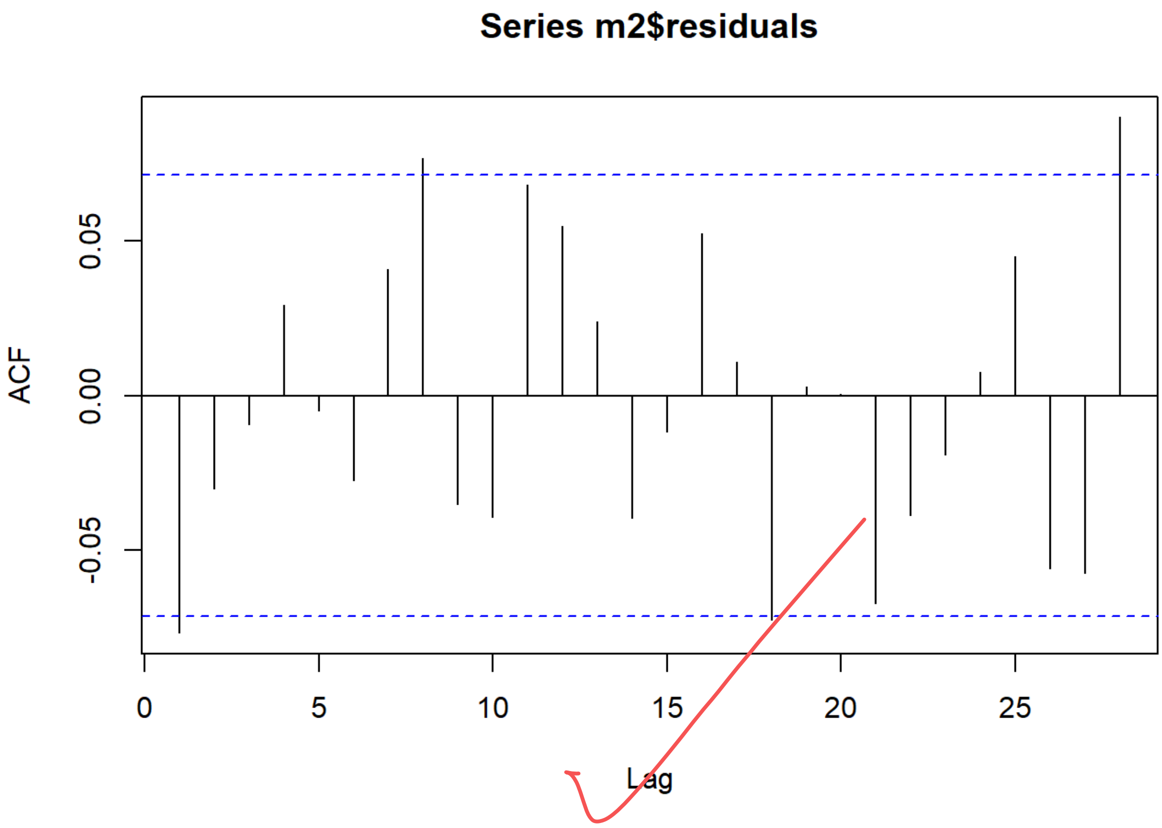
```
m2 = arima(logrtn_mrk,
           order = c(2,0,3),
           fixed = c(0,NA,0,NA,0,NA),
           transform.pars = FALSE)
m2
```

```
##
## Call:
## arima(x = logrtn_mrk, order = c(2, 0, 3), transform.pars = FALSE, fixed = c(0,
##     NA, 0, NA, 0, NA))
##
## Coefficients:
##       ar1     ar2  ma1     ma2  ma3  intercept
##         0  0.8122    0  -0.757    0     0.0106
## s.e.    0  0.1055    0   0.116    0     0.0033
##
## sigma^2 estimated as 0.00495:  log likelihood = 927.61,  aic = -1849.23
```

```
Box.test(m2$residuals, lag = 24, type = "Ljung", fitdf = 5-3)
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 33.271, df = 22, p-value = 0.05817
```

```
acf(m2$residuals)
```



- p−value > 0.05 has no serial correlation,(足夠)

b. Is there any evidence of ARCH effects in the log returns? Use the residual series if there are serial correlations in part (a). Use Ljung–Box statistics for the squared returns (or residuals) with 6 and 12 lags of autocorrelations and 5% significance level to answer the question.

```
at = logrtn_mrk - mean(logrtn_mrk)
Box.test(at^2, lag=6, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  at^2
## X-squared = 22.544, df = 6, p-value = 0.0009644
```

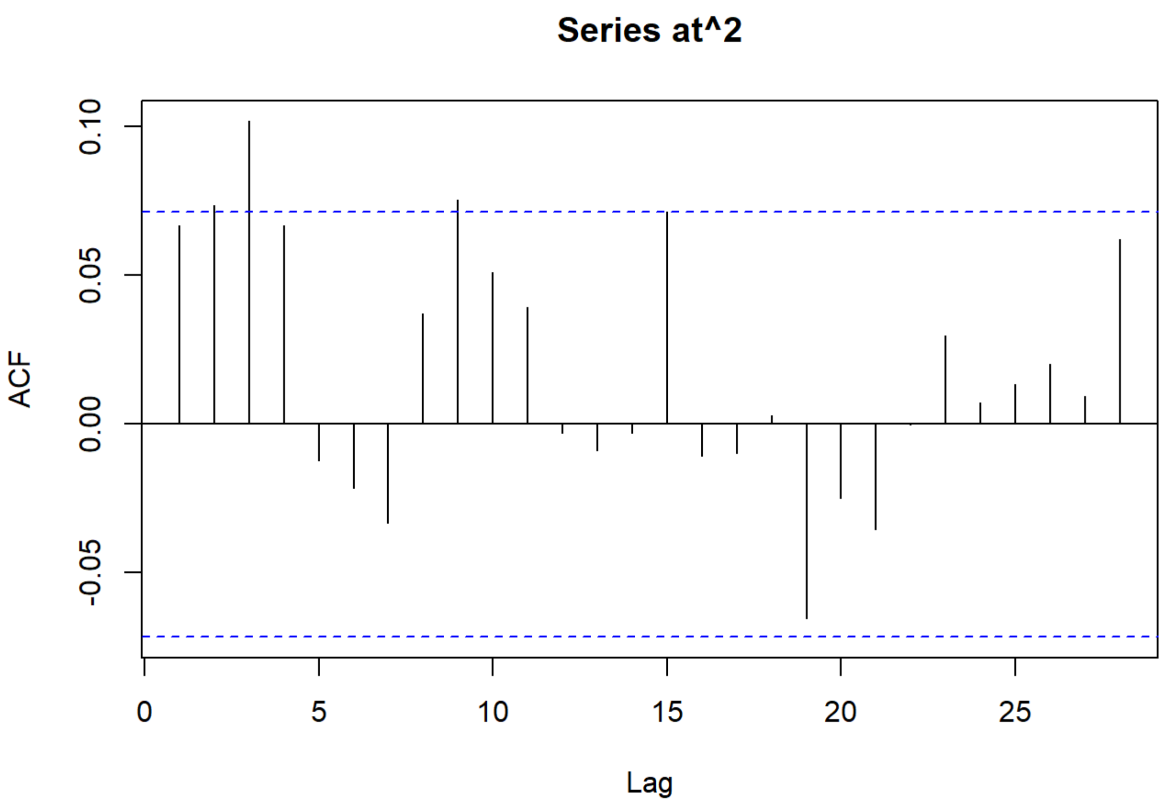
```
Box.test(at^2, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  at^2
## X-squared = 33.013, df = 12, p-value = 0.0009637
```

- There are ARCH effects in log return residuals with lag 6 and lag 12 in mrk. (p-value < 0.05)

c. Identify an ARCH model for the data and fit the identified model. Write down the fitted model.

```
acf(at^2, type="partial")
```



- ARCH(3) (library(fGarch))

```
model1=garchFit(logrtn_mrk~garch(3,0), data=logrtn_mrk, trace=F)
summary(model1)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = logrtn_mrk ~ garch(3, 0), data = logrtn_mrk,
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ garch(3, 0)
## <environment: 0x00000000242a4a70>
##  [data = logrtn_mrk]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##         mu      omega     alpha1     alpha2     alpha3
## 0.0120047  0.0040637  0.0296618  0.0695198  0.0841515
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.0120047   0.0025505    4.707 2.52e-06 ***
## omega   0.0040637   0.0003279   12.393  < 2e-16 ***
## alpha1 0.0296618   0.0391996    0.757   0.4492
## alpha2 0.0695198   0.0372276    1.867   0.0618 .
```

沒 fixed -2

```
## alpha3 0.0841515   0.0391461    2.150   0.0316 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  931.7081    normalized:  1.240623
##
## Description:
##  Fri May 19 06:02:08 2023 by user: user
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  24.8437   4.029569e-06
##  Shapiro-Wilk Test  R    W      0.9943764 0.006934424
##  Ljung-Box Test     R    Q(10)  19.18421  0.03798434
##  Ljung-Box Test     R    Q(15)  28.56444  0.01829135
##  Ljung-Box Test     R    Q(20)  34.90717  0.02060429
```

model 沒过 -2

```
##  Ljung-Box Test     R^2  Q(10)  9.889245  0.4502634
##  Ljung-Box Test     R^2  Q(15)  15.50471  0.4157096
##  Ljung-Box Test     R^2  Q(20)  17.04997  0.6497266
##  LM Arch Test       R    TR^2   12.26932  0.4242995
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.467931 -2.437163 -2.468019 -2.456076
```

- $r_t = 0.0120047 + a_t$, $a_t = \sigma_t \epsilon_t$, $\sigma_t^2 = 0.0040637 + 0.0296618 a_{t-1}^2 + 0.0695198 a_{t-2}^2 + 0.0841515 a_{t-3}^2$

3. 4

The file m-3m4608.txt contains two columns. They are date and the monthly simple return for 3M stock. Transform the returns to log returns.

```
data_3m = read.table('https://faculty.chicagobooth.edu/-/media/faculty/ruey-s-tsay/teaching/fts3/m-3m4
608.txt',header=T)
head(data_3m)
```

```
##       date       rtn
## 1 19460228 -0.077922
## 2 19460330  0.018592
## 3 19460430 -0.100000
## 4 19460531  0.209877
## 5 19460628  0.005128
## 6 19460731  0.076531
```

```
logrtn_3m = log(data_3m$rtn + 1)
head(logrtn_3m)
```

```
## [1] -0.081125460  0.018421282 -0.105360516  0.190518702  0.005114897
## [6]  0.073743834
```

a. Is there any evidence of ARCH effects in the log returns? Use Ljung–Box statistics with 6 and 12 lags of autocorrelations and 5% significance level to answer the question.

```
at_2 = logrtn_3m - mean(logrtn_3m)
Box.test(at_2^2, lag=6, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  at_2^2
## X-squared = 28.116, df = 6, p-value = 8.937e-05
```

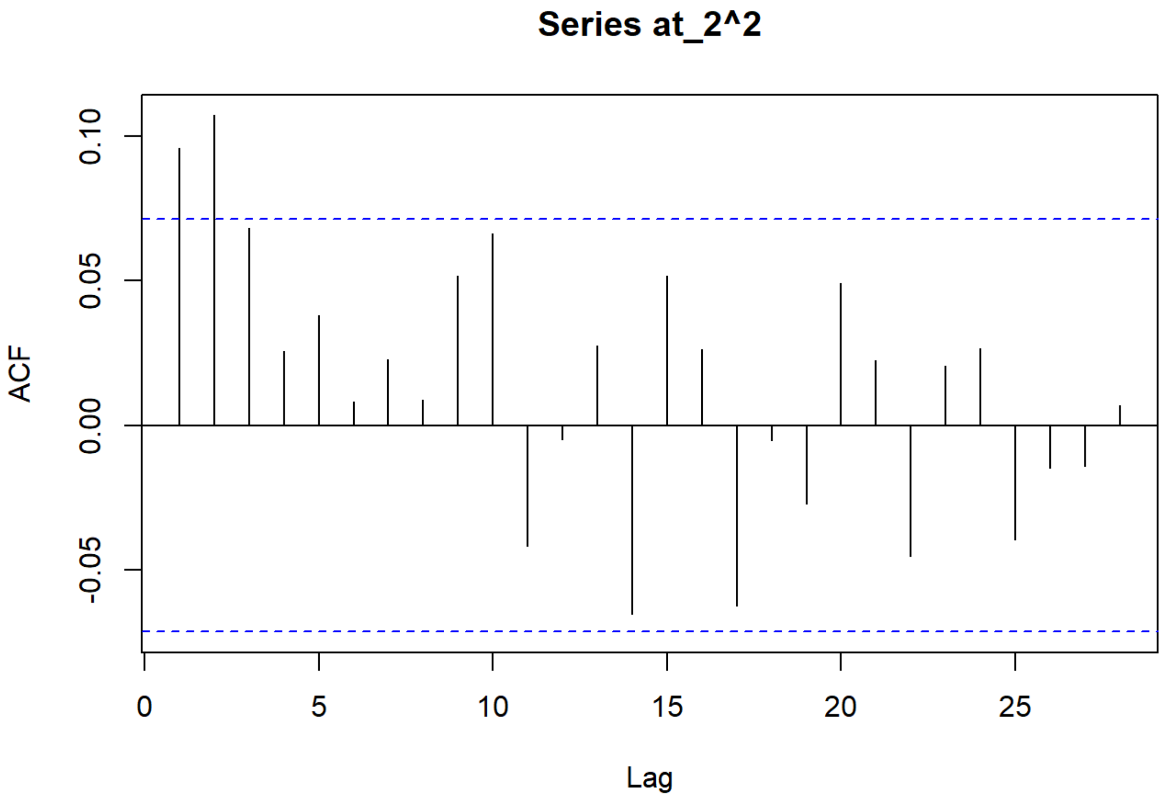
```
Box.test(at_2^2, lag=12, type="Ljung")
```

```
##
##  Box-Ljung test
##
## data:  at_2^2
## X-squared = 38.761, df = 12, p-value = 0.0001152
```

- There are ARCH effects in log return residuals with lag 6 and lag 12 in 3m. (p-value < 0.05)

b. Use the PACF of the squared returns to identify an ARCH model. What is the fitted model?

```
acf(at_2^2,type = "partial")
```



- ARCH(2)

```
model2 = garchFit(logrtn_3m ~ garch(2,0), data=logrtn_3m, trace=F)
summary(model2)
```

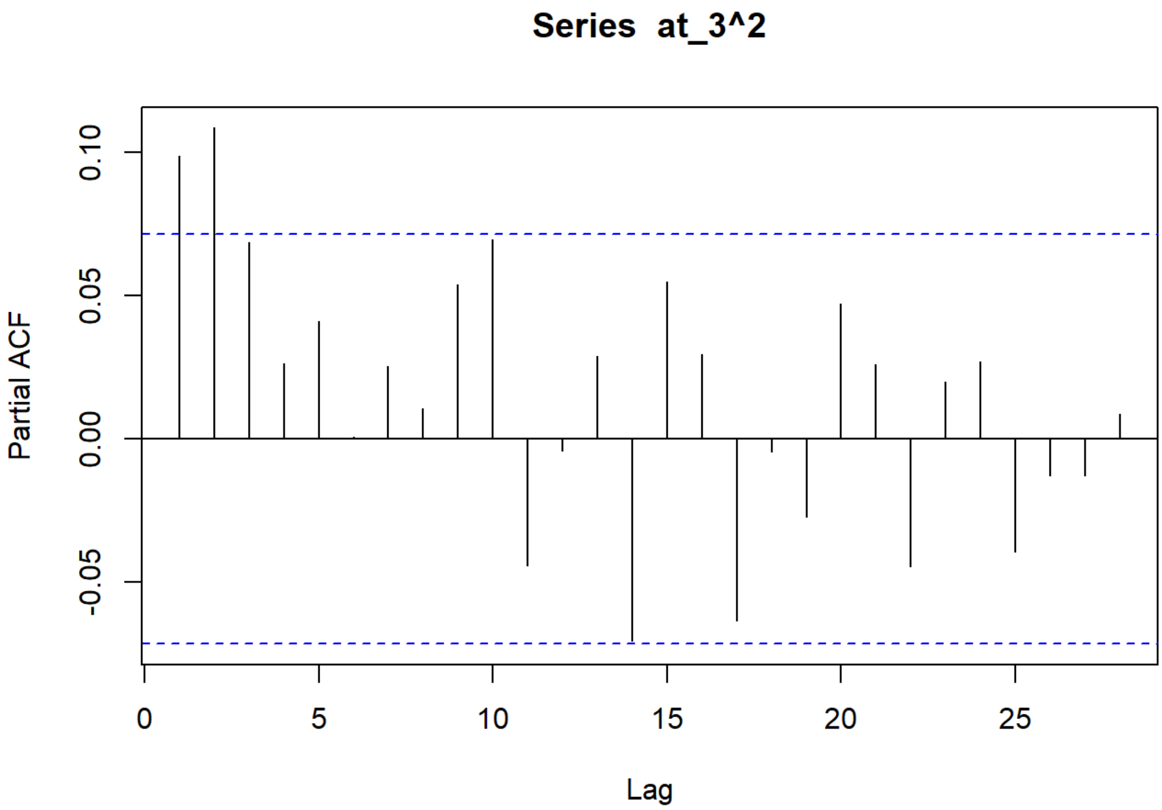
```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = logrtn_3m ~ garch(2, 0), data = logrtn_3m,
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ garch(2, 0)
## <environment: 0x000000003e8c31f0>
##  [data = logrtn_3m]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##       mu     omega    alpha1    alpha2
## 0.010615  0.003228  0.078122  0.128041
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.010615    0.002205    4.814 1.48e-06 ***
## omega   0.003228    0.000256   12.609  < 2e-16 ***
## alpha1  0.078122    0.044993    1.736   0.0825 .
## alpha2  0.128041    0.053228    2.406   0.0162 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  1017.364    normalized:  1.347502
##
## Description:
##  Fri May 19 06:02:09 2023 by user: user
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  41.06987  1.207233e-09
##  Shapiro-Wilk Test  R    W      0.9917295 0.0003164345
##  Ljung-Box Test     R    Q(10)  19.81877  0.0310142
##  Ljung-Box Test     R    Q(15)  29.37285  0.01439508
##  Ljung-Box Test     R    Q(20)  32.61738  0.03714796
##  Ljung-Box Test     R^2  Q(10)  9.053344  0.5270487
##  Ljung-Box Test     R^2  Q(15)  16.91694  0.3238542
##  Ljung-Box Test     R^2  Q(20)  24.24292  0.2319419
##  LM Arch Test       R    TR^2   10.14719  0.6030498
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.684407 -2.659895 -2.684463 -2.674965
```

fixed -2

model 沒過. -2

c. There are 755 data points. Refit the model using the first 750 observations and use the fitted model to predict the volatilities for t from 751 to 755 (the forecast origin is 750)

```
data_3m_2 = data_3m[1:750,2]
logrtn_3m_2 = log(data_3m_2 + 1)
at_3 = logrtn_3m_2 - mean(logrtn_3m_2)
pacf(at_3^2)
```




```
model3 = garchFit(logrtn_3m_2 ~ garch(2,0), data=logrtn_3m_2, trace=F)
summary(model3)
```

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = logrtn_3m_2 ~ garch(2, 0), data = logrtn_3m_2,
##     trace = F)
##
## Mean and Variance Equation:
##  data ~ garch(2, 0)
## <environment: 0x0000000026f816c0>
##  [data = logrtn_3m_2]
##
## Conditional Distribution:
##  norm
##
## Coefficient(s):
##        mu      omega     alpha1     alpha2
## 0.0109216  0.0032036  0.0800948  0.1264017
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##         Estimate  Std. Error  t value Pr(>|t|)
## mu      0.0109216   0.0022035    4.956 7.18e-07 ***
## omega   0.0032036   0.0002542   12.602  < 2e-16 ***
## alpha1  0.0800948   0.0449762    1.781   0.0749 .
## alpha2  0.1264017   0.0525631    2.405   0.0162 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  1012.827    normalized:  1.350436
```

-2

```
##
## Description:
##  Fri May 19 06:02:09 2023 by user: user
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  42.29061  6.557086e-10
##  Shapiro-Wilk Test  R    W      0.9914578 0.0002502714
##  Ljung-Box Test     R    Q(10)  20.25332  0.02694571
##  Ljung-Box Test     R    Q(15)  29.55338  0.01363788
##  Ljung-Box Test     R    Q(20)  32.82521  0.03526079
```

model 沒过

```
##  Ljung-Box Test     R^2  Q(10)  9.654786  0.4712839
##  Ljung-Box Test     R^2  Q(15)  18.49898  0.2373414
##  Ljung-Box Test     R^2  Q(20)  25.76902  0.1735767
##  LM Arch Test       R    TR^2   10.73783  0.5515106
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -2.690206 -2.665566 -2.690262 -2.680711
```

```
forecast = predict(model3, n.ahead = 5)
forecast
```

```
##   meanForecast  meanError standardDeviation
## 1   0.01092165 0.07074067        0.07074067
## 2   0.01092165 0.06003729        0.06003729
## 3   0.01092165 0.06422518        0.06422518
## 4   0.01092165 0.06316346        0.06316346
## 5   0.01092165 0.06359692        0.06359692
```