final

Mingwen Yang

5/6/2020

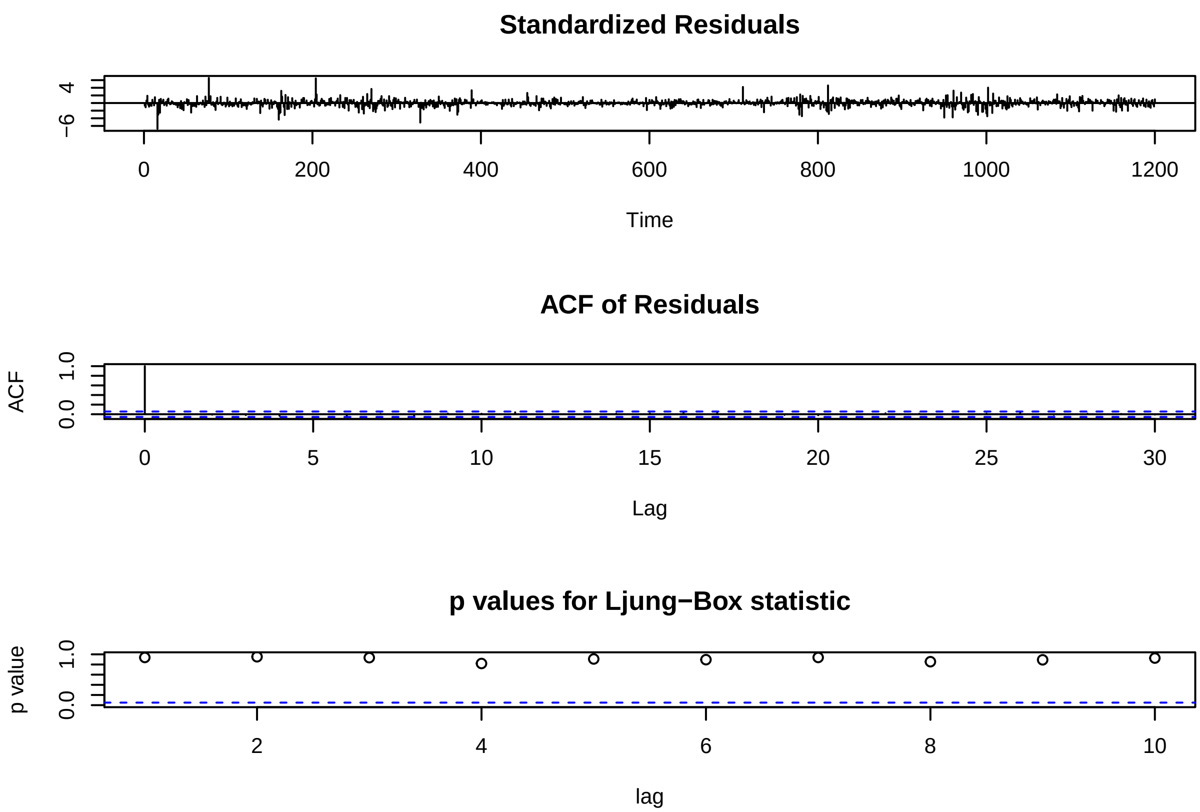
Problem C

Pick up “MSFT” for “Microsoft Corporation”. Log return is stationary due to the small p-value of adf test. So when fit ARIMA model, fix d to 0.

1.

Use the loops to identify the best arima orders. And from the result, we found when p = 3, q = 3, AIC has the smallest value, is -6704.944. So use ARMA(3, 3) to fit the time series.

```
##
## Call:
## arima(x = rt_train, order = c(3, 0, 3))
##
## Coefficients:
##          ar1      ar2     ar3      ma1     ma2      ma3  intercept
##       0.4163  -0.3009  0.7919  -0.4908  0.2736  -0.7575     0.0011
## s.e.  0.1102   0.1084  0.0866   0.1240  0.1277   0.1001     0.0001
##
## sigma^2 estimated as 0.0002163:  log likelihood = 3360.47,  aic = -6704.94
```



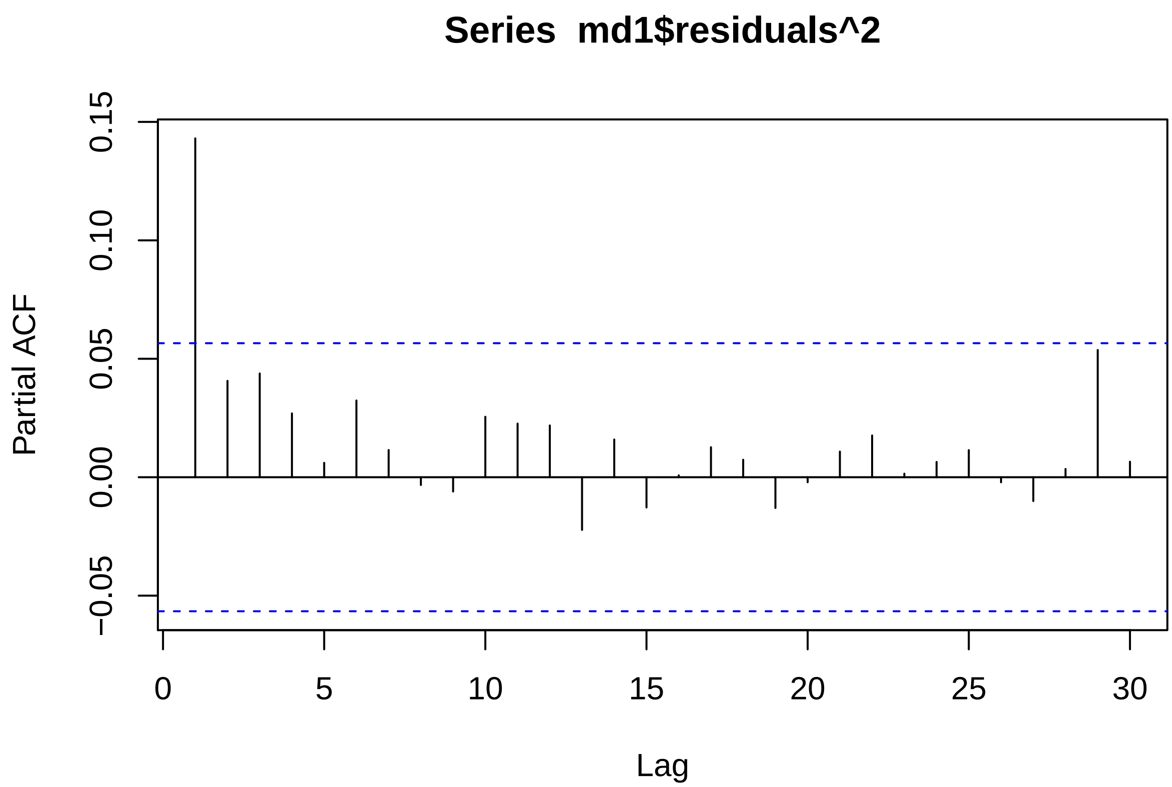
After fitting the model, perform model diagnosis on it. We could see from the result of model, all coefficients are quite significant. And from the diagnosis result, we could see there is no significant correlation in residuals from the plot of residuals. ACF of residuals cuts off at 0, and the p-value of box test is large enough. So this model is relatively adequate.

2.

Use ARMA(3, 3) to predict the log return's direction. Keep fitting model with 1200 days' data, then predict one day ahead. After checking the direction with all 56 days' data in test set, I found 25 out of 56 predicted data sharing the same direction with real data.

3.

```
##
##  Box-Ljung test
##
## data:  md1$residuals^2
## X-squared = 39.291, df = 10, p-value = 2.257e-05
```



Apply box-test on the square of the residual of ARMA model, then it could be found that p-value is small enough. So we can conclude that there is ARCH effect. And to plot the PACF of the square of the residual of ARMA model, the order of ARCH model would be 1. So I fitted ARMA + GARCH model to train set next.

And choose ARMA(3, 3) + GARCH(1, 1) with t-distribution from ARMA(3, 3) + ARCH(1) models and ARMA(3, 3) + GARCH(1, 1) models with different error distribution, i.e. normal, t, skewed t distribution. All coefficients in it are significant and has the smallest AIC, -5.892. And from the results of all the box test of residual, p-value is large enough, so there is no correlation among residual, indicating it is a adequate model.
We got the model

$$
\begin{aligned}
r_t &= 4.09 \times 10^{-4} + 0.41 r_{t-1} - 0.39 r_{t-2} + 0.67 r_{t-3} + a_t - 0.51 a_{t-1} + 0.39 a_{t-2} - 0.7 a_{t-3} \\
a_t &= \sigma_t \epsilon_t \\
\sigma^2 &= 1.53 \times 10^{-5} + 0.16 a_{t-1}^2 + 0.8 \sigma_{t-1}^2
\end{aligned}
$$

```
##
## Title:
##  GARCH Modelling
##
## Call:
##  garchFit(formula = ~arma(3, 3) + garch(1, 1), data = rt_train,
##     cond.dist = c("std"), trace = F)
##
## Mean and Variance Equation:
##  data ~ arma(3, 3) + garch(1, 1)
## <environment: 0x000000002aa3a168>
##  [data = rt_train]
##
```

```
## Conditional Distribution:
##  std
##
## Coefficient(s):
##          mu          ar1          ar2          ar3          ma1          ma2
##  4.0867e-04   4.0649e-01  -3.8870e-01   6.7449e-01  -5.0693e-01   3.8893e-01
##         ma3        omega       alpha1        beta1        shape
## -7.0043e-01   1.5252e-05   1.6451e-01   7.9604e-01   3.5069e+00
##
## Std. Errors:
##  based on Hessian
##
## Error Analysis:
##          Estimate  Std. Error  t value Pr(>|t|)
## mu      4.087e-04   1.997e-04    2.047 0.040668 *
## ar1     4.065e-01   1.425e-01    2.853 0.004333 **
## ar2    -3.887e-01   1.105e-01   -3.517 0.000437 ***
## ar3     6.745e-01   8.317e-02    8.110 4.44e-16 ***
## ma1    -5.069e-01   1.446e-01   -3.505 0.000457 ***
## ma2     3.889e-01   1.228e-01    3.167 0.001543 **
## ma3    -7.004e-01   8.422e-02   -8.316  < 2e-16 ***
## omega   1.525e-05   6.956e-06    2.193 0.028335 *
## alpha1  1.645e-01   5.071e-02    3.244 0.001179 **
## beta1   7.960e-01   6.002e-02   13.263  < 2e-16 ***
## shape   3.507e+00   3.966e-01    8.842  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log Likelihood:
##  3546.287    normalized:  2.955239
##
## Description:
##  Thu May 07 17:43:32 2020 by user: yangm
##
##
## Standardised Residuals Tests:
##                                 Statistic p-Value
##  Jarque-Bera Test   R    Chi^2  6389.009  0
##  Shapiro-Wilk Test  R    W      0.9013267 0
##  Ljung-Box Test     R    Q(10)  4.845864  0.9012336
##  Ljung-Box Test     R    Q(15)  10.69661  0.7737897
##  Ljung-Box Test     R    Q(20)  13.52703  0.8536448
##  Ljung-Box Test     R^2  Q(10)  3.396785  0.9704873
##  Ljung-Box Test     R^2  Q(15)  4.475002  0.9957154
##  Ljung-Box Test     R^2  Q(20)  5.49566   0.9994272
##  LM Arch Test       R    TR^2   3.828325  0.9863387
##
## Information Criterion Statistics:
##       AIC       BIC       SIC      HQIC
## -5.892145 -5.845486 -5.892311 -5.874568
```

4.

Like the prediction method in question 2, perform the rolling forecast for the last 56 days from 2019-10-9. And compute the total predicted profit, which is 7050.8 in this situation. So at the end of process, the value of portfolio would be $107050.8.

5.

Just as the method of prediction like before, with ARMA(3, 3) + GARCH(1, 1) this time, and at the end of process, the value is $108520.6.

6.

Because we need not to consider transaction costs, in addition to long the stock when prediction is positve, short the stock when prediction is negative. And we could get the value of $114217.9.

We could set the mean value of the last 10 days before as the middel line, adding the predicted volatility as the upper line, while minus the predicted volatility as the down line. Then if the predicted return over upper line, buy with all value; if between the upper line and middel line, buy with 70% money; if between middle and down line, sell 70%; if lower than down line, sell all. Then at the end of process, we will get $108501.