**Midterm**

**Tuesday, 1 October 2019@7-10pm**

1. Do bring your laptop for the midterm.
2. Download
3. your individual data (named after your student ID) and
4. d-cdsALL.txt

from the folder

LumiNUS->files->Midterm

Using wrong data will lead to 0 marks.

1. You are required to fill in the following table to summarize your results and place the table on the front page of your answer sheets.
2. Your answers and R code should be properly named following the format

StudentID\_studentName.docx or StudentID\_studentName.pdf

and StudentID\_studentName.R.

1. Your answer sheets and R code should be submitted to the folder

LumiNUS -> files -> Midterm Student Submission

1. **The submission folder will be closed punctually at 10:00pm** (LumiNUS clock). A 10-30% penalty will apply to late submission. Any late submission will get at least a 10% penalty. A late submission received on the next day, i.e. 2 October, will be given a 20% penalty. A submission later than 2 days, i.e. from 3 October onwards, will get a 30% penalty.
2. This is an open-book test. You are allowed to use any teaching materials and internet resources. *HOWEVER, this is an individual test! You shouldn’t collude with any other individual, or plagiarize their work. Suspected collusion or plagiarism will be dealt with according to the NUS’ assessment regulations.*

--- Do not let other people see your answers.

--- If other students ask for help, tell them to ask the lecturer.

|  |  |  |
| --- | --- | --- |
|  | **Student ID** | **A0197899R** |
| Q1. | Type of stationarity | **Non-stationary** |
| Fitted model | **~arma(1,1)+garch(1,1),con.dist=’std’** |
| Q2. | Fitted model | **~arma(1,4)+garch(1,1)** |
| Interval forecasts | **Step1 = [89.00989,89.02330] , step2 = [89.00659,89.02004]** |
| Q3 | Fitted model | **~arma(5,5)+garch(1,1)** |
| Fitted model with dummy | **~arma(0,5)+garch(1,1)** |

**Analyze data and answer the following 3 questions with proper justifications.**

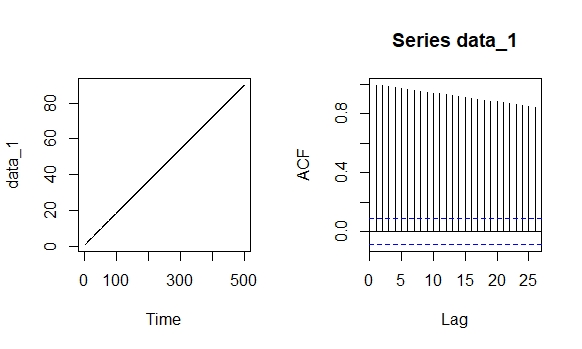
Load your data using the following R code

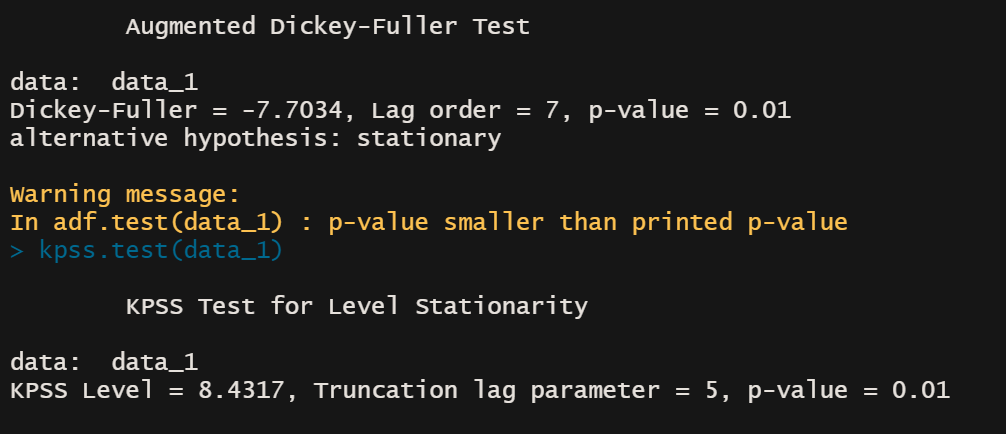
read.table(filename, sep="\t", header=FALSE)

**Question 1.** Consider the first time series (the second column in your data), which is monthly sales of SDA – a FinTech company.

1. Plot the time series and ACF plots. Do you think the time series is stationary? Why or why not? What type of non-stationarity can you see in the series and how do you handle this type of non-stationarity? Do you observe any heteroskedasticity in the time series?

**Answer:** the time series and acf plots are as the following. We can see the time series have obvious trend and the acf of it decrease very slowly. So I think it is not stationary. Then I will do the ADF and KPSS test.

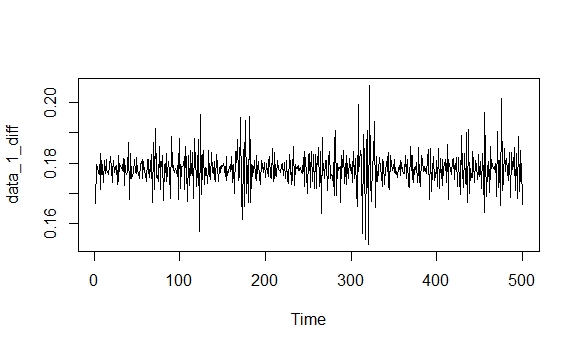




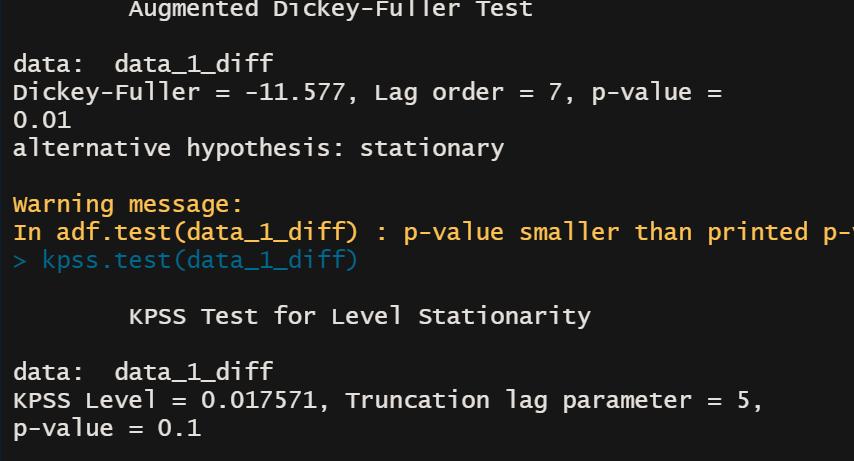
Both ADF and KPSS reject the null hypothesis, to be conservative, and also with the information of the ts plot, let’s assume the original data to be non-stationary.

And it is hard tell

So I choose to do one order difference of the data. And the diff(data) is:



It seems mean-reverting, we do the ADF and KPSS test again:

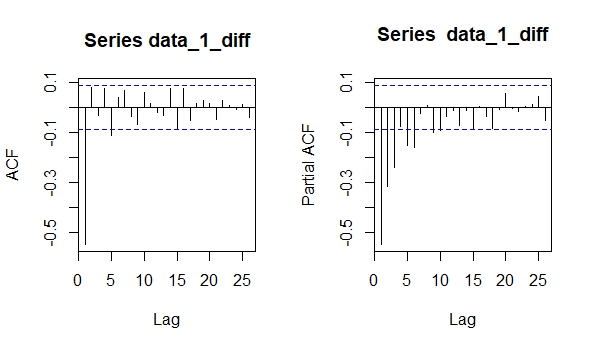


We can see that the ADF and KPSS give the same conclusion that the diff\_ts is stationary.

And Heteroskedastic is hard to tell from these two plots.

1. Propose an appropriate model for the sales data and justify your model assumptions with supporting evidence, including but not necessarily limited to autocorrelations, GARCH effect and distributional assumption.

**Answer:** From question 1 we know that we accept the one order difference of original data is stationary, so we plot the acf and pacf of the diff(data):

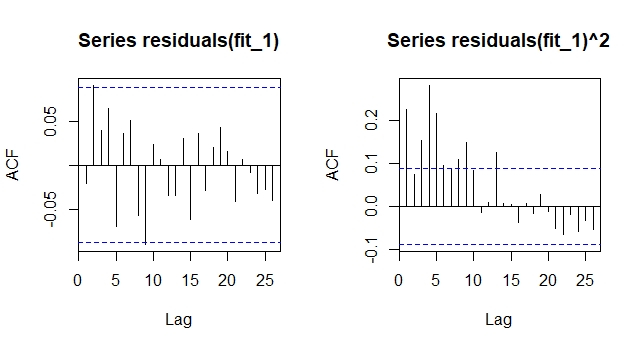


We can see from the ACF that this data may fit MA(1) model, so we use arima(0,0,1) for the diff\_data and we plot the ACF of residuals and residuals square.

fit\_1 = arima(data\_1\_diff,order=c(0,0,1));

acf(residuals(fit\_1))

acf(residuals(fit\_1)^2)



I can see that ARIMA(0,1,1) may fit well for the data, however, the residuals have obvious Garch effect according the acf(residuals^2) plot. Thus I choose use Garch model to fit this data.

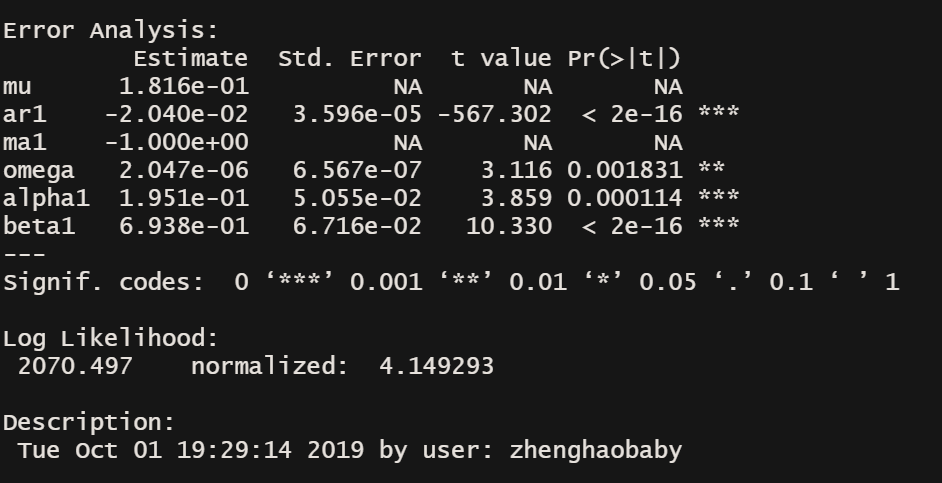
I choose three model as the following:

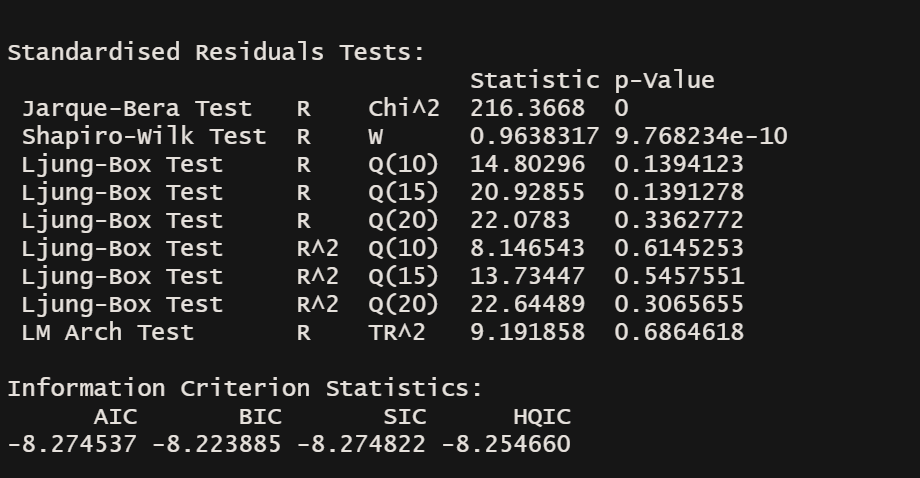
gfit\_1 = garchFit(~arma(0,1)+garch(1,1),data=data\_1\_diff)

gfit\_2 = garchFit(~arma(1,0)+garch(1,1),data=data\_1\_diff)

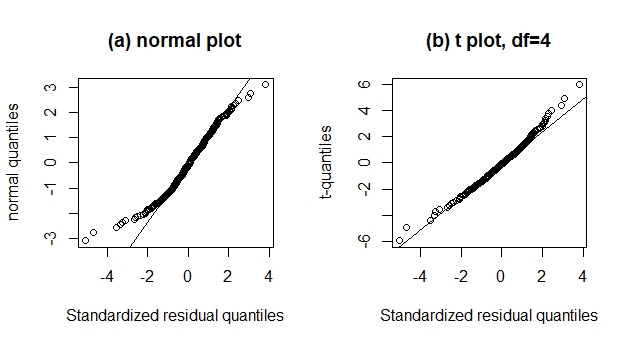
gfit\_3 = garchFit(~arma(1,1)+garch(1,1),data=data\_1\_diff)

according to the AIC and BIC of each model, the gfit\_3 model fit the best.





Then we should see the residuals of this model.



We can see that the residuals may more fit t-distribution rather than normal distribution. Thus we should change our assumption about the distribution.

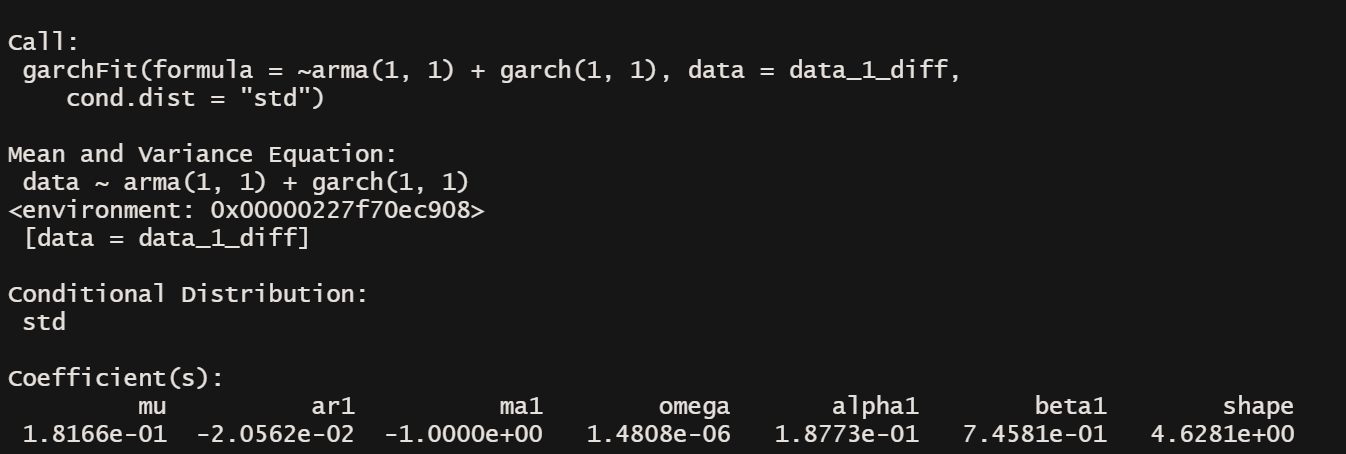
Model:

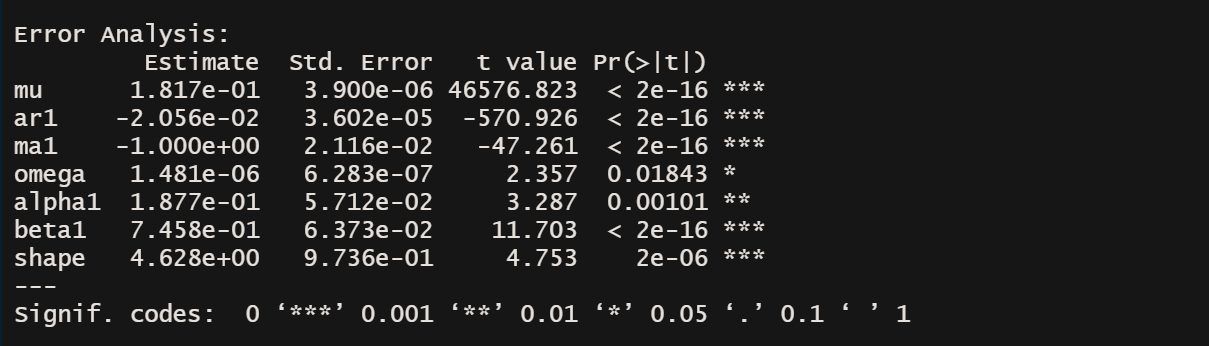
gfit\_4 = garchFit(~arma(1,1)+garch(1,1),data=data\_1\_diff,cond.dist = 'std')

1. Write down the fitted model for sales. Interpret the meaning of each coefficient and discuss significance.

**Answer:**

+0.74581



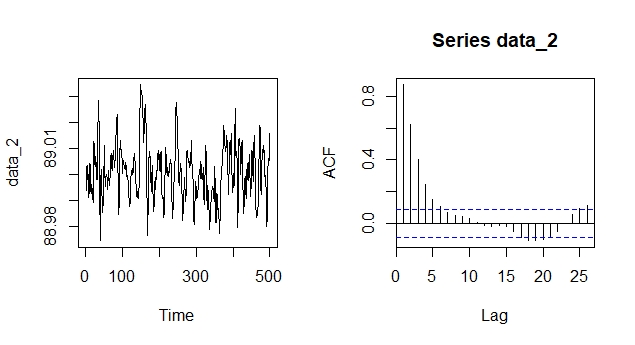


The mu, ar1, ma1 are the coefficients for conditional mean and the omega, alpha1, beta1 are the coefficients for conditional variance. And the shape is the df of t-distribution. We can see from the picture that all the coefficients are significant according to 5% standard.

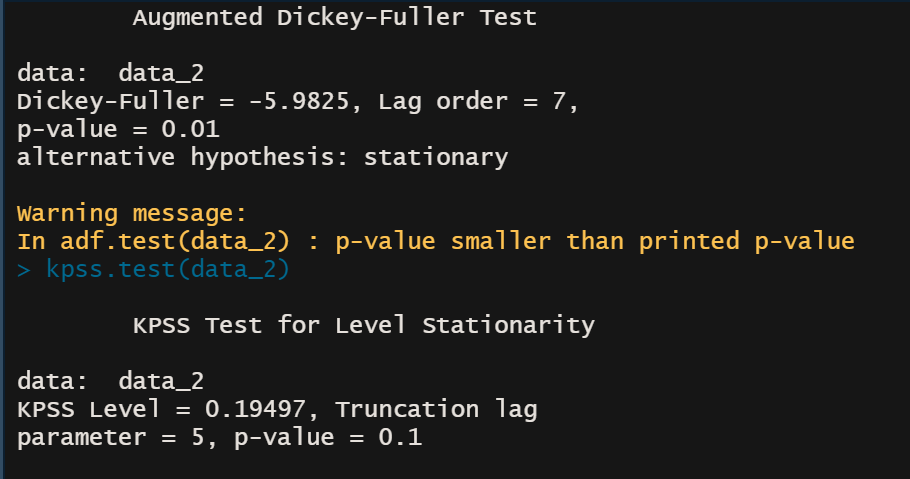
**Question 2.** Over years, SDA made gains in many forms of investing in real estate, bonds, etc. as displayed in the third column of your data.

1. Plot the time series and ACF plots. Do you think the time series is stationary? If yes, what type of non-stationarity can you see in the series and how do you handle this type of non-stationarity? Do you observe any heteroskedasticity in the time series?

**Answer:** First I plot the time-series and ACF.



It seems that the time series are mean-reverting, I choose to do the ADF and KPSS test.:

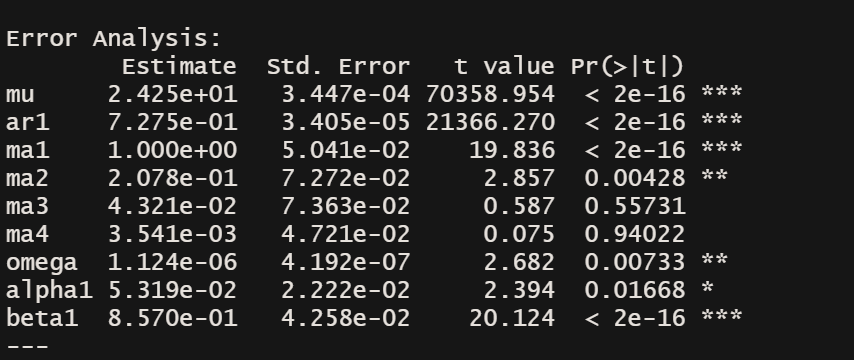


Both ADF and KPSS test conclude that the time series are stationary. From the time series plots I think there exists **heteroskedasticity** because some time it has large variance.

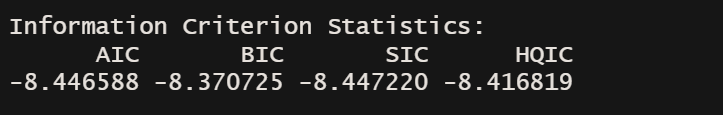
1. Propose an appropriate model for the gains data. Write down the fitted model and report its AIC. Interpret the meaning of each coefficient and discuss significance.

**Answer:** fit\_2\_g = garchFit(~arma(1,4)+garch(1,1),data=data\_2)

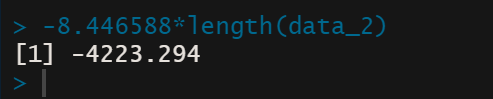
**Model arma(1,4)+garch(1,1)**



Standard AIC is



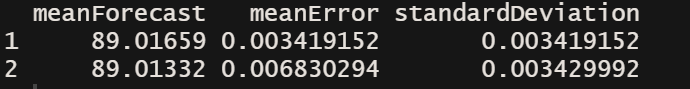
So the AIC IS



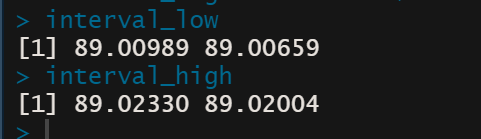
The mu, ar1,ma1,ma2,ma3,ma4 are coefficients for conditional mean, and the omega, alpha1,beta1 are coefficients for conditional variance. The **ma3 and ma4** can is not significant according to the %5 standard and other coefficients are all significant.

1. Compute 1- and 2-step ahead forecasts and show their confidence interval.

**Answer:**



From the picture we can see the 1and 2 step ahead forecast and the interval is as the following:



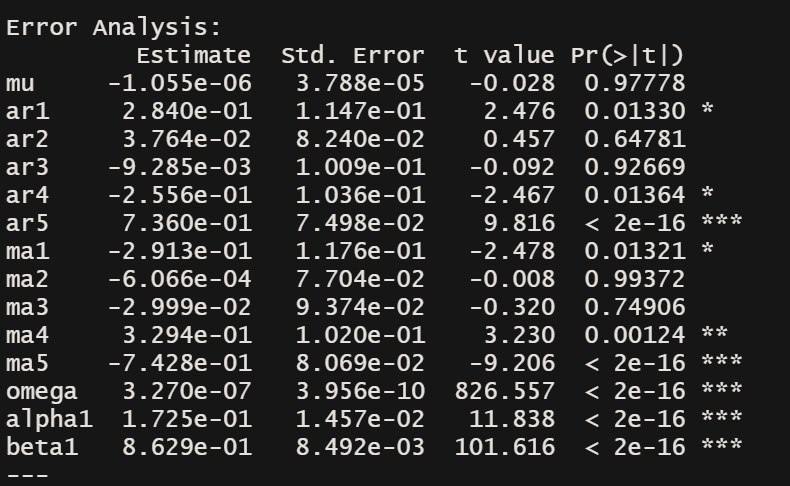
Step1 = [89.00989,89.02330] , step2 = [89.00659,89.02004]

**Question 3.** Consider the daily CDS spreads (3-year maturity) of Allstate Insurance from January 01, 2004 to September 19, 2014. The period includes the financial crisis of 2008 so that the CDS spreads vary substantially. The data are in the file d-cdsALL.txt (column 2). Since the spreads are small, we consider the time series = 100 × (spread3y). In addition, sample ACF of shows strong persistence in the serial dependence. Therefore, we analyze the differenced series .

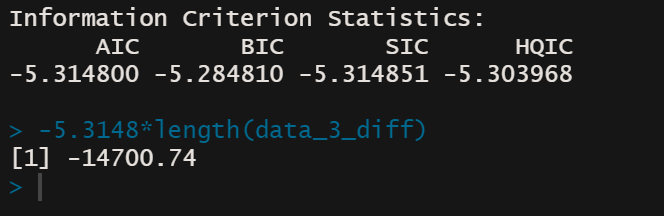
1. Build a time series model for . Write down the fitted model. [You may start with a model suggested by the auto.arima command in forecast package.]

Answer: garchFit(~arma(5,5)+garch(1,1),data=data\_3\_diff)

**Model:** **arma(5,5)+garch(1,1)**

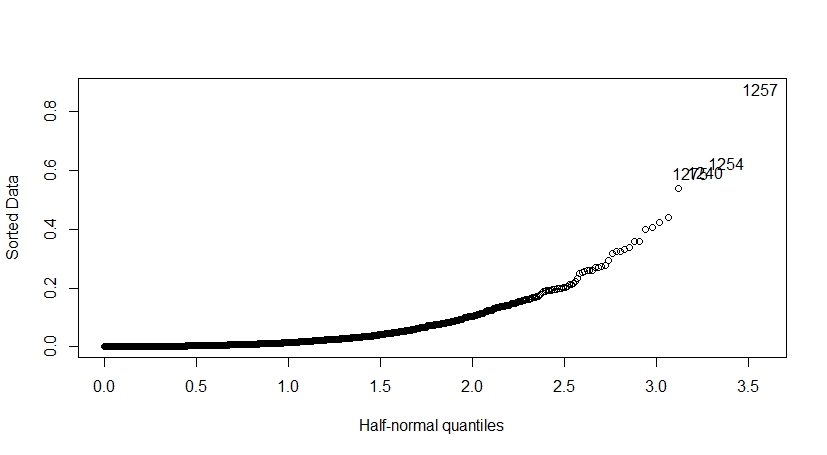


AIC:



1. To improve the fit, identify sequentially the largest four outliers of the fitted model. Write down the fitted model with the four largest outliers included as dummy.

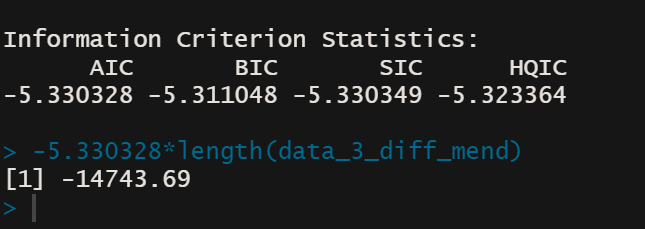
Answer: First, use the halfnorm function to find the outliners



From this picture I can know the index of the largest four outliners is 1257,1254,1240,1275, Thus make this four to be dummy. Then we fit the data again and get the model as the following:

fit\_3\_g\_mend = garchFit(~arma(0,5)+garch(1,1),data=data\_3\_diff\_mend

**Model:** arma(0,5)+garch(1,1)



And we can see the AIC significantly decrease by setting the largest four as dummy.

1. Let be the residuals of the model in question (8) and be the lag- ACF of . Test versus Ha : for some . Draw your conclusion.

**Answer:**

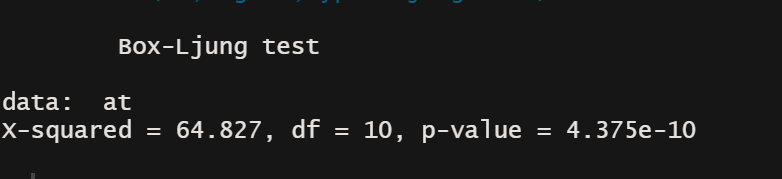
at = residuals(fit\_3\_g)

acf(at)

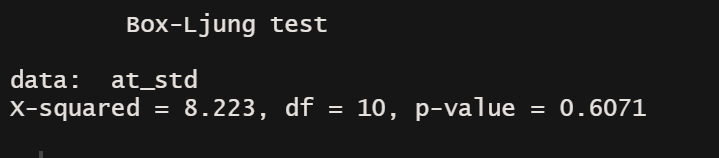
Box.test(at,lag=10,type="Ljung-Box")

at\_std = residuals(fit\_3\_g)/fit\_3\_g@sigma.t

acf(at\_std)

Box.test(at\_std,lag=10,type="Ljung-Box")

We can see for residuals, the p-value is very small and we can reject the H0.



However, for the standard residuals, we accept the H0