ORIE 4630: Spring Term 2019
Homework #2
Due: Tuesday, February 12, 2019

Students are required to work independently on homework. You should not give or receive help from other students. You should also not receive help from students or former students who took this course in previous years and who may have solutions to similar problems. The solutions you submit should be your own work and not copied from elsewhere.

Homework is due at the end of lecture (12:55pm) on the due date. You will usually have one week to do the assignments. Please don't wait until the homework is nearly due to start. Late homework is not accepted. Also, homework is not accepted by email. You can submit your assignment in lecture or in the drop box in Rhodes Hall.

Please print your name on the front of your homework so that it is legible.

Include your `R` code, output, graphs, and other work with your homework. This will allow the grader to find any errors you make and to give partial credit.

This assignment concerns daily returns of certain stocks from Jan 4, 2006 to Aug 18, 2017. The log gross returns for the stocks under consideration are contained in a comma separated values (csv) file named `returns.csv`. You should download this file from the course Blackboard site and put it into your `R` or `Rstudio` working directory. The file has 35 columns. The first column shows the date (Date), and the next 30 columns are for the stocks that are the components of the Dow Jones Industrial average (DOW). The final four columns are for the Dow Jones Industrial average index (DOW), the NASDAQ composite index (NASD), the NASDAQ 100 index (NASD100), and the S&P 500 index (SP500). There are 2927 days of returns. The log gross returns are calculated from adjusted closing prices downloaded from Yahoo.

Start `R` or `Rstudio` and run the following lines:

```
1 Returns = read.csv("returns.csv")
2 names(Returns)
3 class(Returns$Date)
4 Returns$Date = as.Date(Returns$Date, format="%m/%d/%Y")
5 class(Returns$Date)
```

Line 1 reads the data into a data frame named `Returns`. If `returns.csv` is not in your working directory, then you need to give a complete path to that file in line 1. Line 2 prints the names of the columns (variables) of `Returns`. Line 3 verifies that the class of `Returns$Date` is `factor`; line 4 changes the class of `Returns$Date` to `Date`, as is verified in line 5.

Normal probability plots can be obtained by using the function `qqnorm()`. To create a normal probability plot of the returns for Caterpillar (CAT), run the following lines:

```
6 qqnorm(Returns$CAT, datax=TRUE)
7 qqline(Returns$CAT, datax=TRUE, col="blue", lwd=2)
```

In line 6, the command `datax=TRUE` puts the sample quantiles on the horizontal axis, so the normal quantiles are on the vertical axis. The function `qqline()` puts a straight line through the pair of first quartiles and the pair of third quartiles in the normal probability plot. In line 7, the command `col="blue"` makes the line blue. The command `lwd=2` controls the width of the line; the larger the specified number, the wider the line.

The sample standardized skewness and the sample standardized kurtosis can be obtained by using the functions `skewness()` and `kurtosis()` in the `moments` package. To install the `moments` package and find the skewness and kurtosis of the returns for Caterpillar (CAT), run the following lines:

```
8 install.packages("moments")
9 library(moments)
10 skewness(Returns$CAT)
11 kurtosis(Returns$CAT)
```

The Shapiro-Wilk and Jarque-Bera tests for normality can be conducted by using the functions `shapiro.test()` and `jarque.test()` in the `moments` package. To conduct the Shapiro-Wilk test and the test for the returns for Caterpillar (CAT), run the following lines:

```
12 shapiro.test(Returns$CAT)
13 jarque.test(Returns$CAT)
```

In addition to the normal probability plot, another graphical way to assess whether the returns are normally distributed is to plot a frequency histogram of the returns with a "matching" normal density curve superimposed. (The "matching" normal density has the same mean and standard deviation as the returns, and it is suitably scaled so that it has the same area as the frequency histogram.) If the returns are indeed normally distributed, then the frequency histogram and the matching normal density curve should be close. To prepare the enhanced histogram for the returns for Caterpillar (CAT), run the following lines:

```
14 x <- Returns$CAT
15 h<-hist(x, breaks=100, col="red")
16 xfit<-seq(min(x),max(x),length=200)
17 yfit<-dnorm(xfit,mean=mean(x),sd=sd(x))
18 yfit <- yfit*diff(h$mids[1:2])*length(x)
19 lines(xfit, yfit, col="blue", lwd=2)
```

The function `hist()` produces the frequency histogram; in line 15, the number of bins is set to 50, and the color red is specified for the histogram. Values of the normal density are computed by using the function `dnorm()` in line 17. The values of the normal density curve are scaled in line 18 so that the area under the density curve is the same as the area of the frequency histogram, which is the number of returns times the width of the bins. In line 19, the density curve is plotted as a blue curve superimposed on the histogram by using the function `lines()`.

The function `fitdistr()` in the package `MASS` can be used to fit Student's t-distribution to returns by the method of maximum likelihood. To fit Student's t-distribution to the returns for Caterpillar (CAT), run the following lines:

```
20 install.packages("MASS")
21 library(MASS)
22 fit=fitdistr(Returns$CAT, "t")
23 fit
```

In line 22, the declaration "t" specifies that a location-scale Student's t-distribution is fit to the returns. The output from lines 20 to 23 give: i) m, the maximum likelihood estimate of the location parameter; ii) s, the maximum likelihood estimate of the scale parameter; and iii) df, the degrees of freedom of the fitted t-distribution. From these estimates, the estimated mean of the returns is m, and the estimated standard deviation of the returns is $\mathtt{s} \times \sqrt{\mathtt{df}/(\mathtt{df}-2)}$

Once the degrees of freedom for the fitted Student's t-distribution is obtained, a t-probability plot can be prepared to see how well the fitted Student's t-distribution matches the distribution of the returns. For purpose of the discussion here, suppose that the estimate df of the degrees of freedom for the Caterpillar returns is 3.06. To obtain a t-probability plot for the returns for Caterpillar, run the following lines:

```
24 degf=3.06
25 n=length(Returns$CAT)
26 t_quantiles=qt((1:n)/(n+1), df=degf)
27 qqplot(Returns$CAT, t_quantiles)
28 line=lm(qt(c(0.25, 0.75), df=degf)~quantile(Returns$CAT, c(0.25, 0.75)))
29 abline(line, col="blue", lwd=2)
```

In line 26, the function qt() computes the $\frac{i}{n+1}$-quantile ($i = 1, \ldots, n$) from the relevant Student's t-distribution (with degrees of freedom 3.06), where n is the number of returns, declared in line 25. In line 27, the function qqplot() is used to produce the t-probability plot; note that the returns are on the horizontal axis and the t-quantiles are on the vertical axis. In line 28, the function lm() computes the line that runs through through the pair of first quartiles and the pair of third quartiles in the t-probability plot. The function abline() in line 29 adds this line to the plot; note that the color blue is used for the line.

Another way to assess graphically whether the distribution of the returns follows the fitted location-scale Student's t-distribution is to plot a frequency histogram of the returns with the fitted Student's t-density curve superimposed, where the fitted Student's t-density is suitably scaled so that it has the same area as the frequency histogram. If the distribution of the returns is indeed close to the fitted location-scale Student's t-distribution, then the frequency histogram and the scaled fitted Student's t-density curve should be close. In addition to supposing that the estimate df of the degrees of freedom for the returns for Caterpillar is 3.06, suppose that the estimate m of the location parameter is 0.000565, and suppose that the estimate s of the scale parameter is 0.0130. To prepare the enhanced histogram for the returns for Caterpillar (CAT), run the following lines:

```
30 x <- Returns$CAT
31 degf=3.06
32 m=0.000565
33 s=0.0130
34 h<-hist(x, breaks=50, col="red")
35 xfit<-seq(min(x),max(x),length=200)
36 yfit<-dt((xfit-m)/s,degf)/s
37 yfit <- yfit*diff(h$mids[1:2])*length(x)
38 lines(xfit, yfit, col="blue", lwd=2)
```

The function hist() in line 34 produces the histogram with 50 bins and in the color red. The fitted location-scale Student's t-density is computed using the function dt() in line 36; it is scaled to have the same area as the frequency histogram in line 37; and it is superimposed on the frequency histogram by using the function lines() in line 38.

Sometimes we are interested in determining the dates of extreme returns. To find the dates of the four smallest returns for Caterpillar (CAT), run the following lines:

```
39 ind=(Returns$CAT<=sort(Returns$CAT)[4])
40 Returns$Date[ind]
```

In line 39, the function `sort(Returns$CAT)` orders the returns in `Returns$CAT` from smallest to largest; the variable `sort(Returns$CAT)[4]` is the fourth smallest return. The variable `ind` is an indicator that specifies which entries in `Returns$CAT` are less than or equal to the fourth smallest. In line 40, the entries in `Returns$Date` corresponding to the entries indicated by `ind` are printed. Before running line 40, make sure that you have run line 4, so that `Returns$Date` is of class `Date`.

Questions:

1. [10 points] Run lines 1 to 5. Adapt lines 6 and 7 to produce a normal probability plot for Goldman Sachs (GS). Submit your normal probability plot. What does the plot suggest about the distribution of the returns compared to the normal distribution. Justify your answer briefly.

2. [10 points] By adapting lines 8 to 11, find the sample standardized skewness and the sample standardized kurtosis for the returns for Goldman Sachs (GS). Submit your output. What do these moments suggest about the distribution of the returns compared to the normal distribution? Justify your answer briefly.

3. [10 points] Perform the Shapiro-Wilk test and the Jarque-Bera test for the returns for Goldman Sachs (GS) by adapting lines 12 and 13. Submit your output. What are the p-values from the test? What do you conclude from the tests? Justify your answer briefly.

4. [15 points] By adapting lines 14 to 19, prepare a histogram for the returns for Goldman Sachs (GS) with a matching normal density curve superimposed. Submit your histogram. What is the mean of the matching normal distribution? What is the standard deviation of the matching normal distribution?

5. [20 points] By adapting lines 20 to 23, fit a location-scale Student's t-distribuion to the returns for Goldman Sachs (GS). Submit your output.
i) What is the maximum likelihood estimate of the location parameter?
ii) What is the maximum likelihood estimate of the scale parameter?
iii) What is the maximum likelihood estimate of the degrees of freedom?
iv) What is the standard error for estimate of the degrees of freedom?
v) Give a 95% confidence interval for the degrees of freedom.
vi) What is the mean of the fitted location-scale Student's t-distribution?
vii) What is the standard deviation of the fitted location-scale Student's t-distribution?

6. [10 points] By adapting lines 24 to 29 and using your results from question 5, prepare a t-probability plot for the returns for Goldman Sachs (GS). Submit your plot. Based on the plot, do the returns appear to follow a location-scale Student's t-distribution? Justify your answer briefly.

7. [10 points] By adapting lines 30 to 38, prepare a histogram of the returns from Goldman Sachs (GS) with the fitted location-scale Student's t-distribution from your answer to question 5 superimposed. Submit your histogram.

8. [15 points] By adapting lines 39 and 40, find the dates of the two smallest and the three largest returns for Goldman Sachs (GS). Submit your code and your output. State the dates and the returns.