ORIE 4630: Spring Term 2019
Homework #3
Due: Tuesday, February 19, 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.

Recall that a random variable Y is said to have the location-scale Student's t-distribution if

$$Y = a + bT,$$

where $T \sim t_{(\nu)}$, i.e., T has Student's t-distribution with ν degrees of freedom. The parameter a is the location parameter and the parameter b is the scale parameter. Then $E(Y) = a$ (provided $\nu > 1$) and $Var(Y) = b^2 \dfrac{\nu}{\nu - 2}$ (provided $\nu > 2$). The function `fitdistr()` in the package MASS can be used to fit the location-scale Student's t-distribution to returns by the method of maximum likelihood. To fit the location-scale Student's t-distribution to the returns for 3M (MMM), run the following lines:

```
6 install.packages("MASS")
7 library(MASS)
8 fit=fitdistr(Returns$MMM, "t")
9 fit
```

The output from lines 8 and 9 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 maximum likelihood estimate of the degrees of freedom. Standard errors for these maximum likelihood estimates are also provided.

A random variable Y is said to have the standardized Student's t-distribution if

$$Y = \mu + \sigma \sqrt{\frac{\nu - 2}{\nu}}\, T,$$

where $T \sim t_{(\nu)}$, i.e., T has Student's t-distribution with ν degrees of freedom. Then $E(Y) = \mu$, provided $\nu > 1$, and $Var(Y) = \sigma^2$,provided $\nu > 2$. To fit the standardized Student's t-distribution to the returns for 3M (MMM) by maximum likelihood estimation, run the following lines:

```
10 install.packages("fGarch")
11 library(fGarch)
12 ret=Returns$MMM
13 logliksstt=function(beta) sum(-dstd(ret, mean=beta[1], sd=beta[2], nu=beta[3],
14            log=TRUE))
15 start=c(mean(ret), sd(ret), 5)
16 fitstt=optim(start, logliksstt, hessian=TRUE, method="L-BFGS-B",
17            control=list(parscale=c(1e-4, 1e-2, 1)), lower=c(-1,0.001,1))
18 eststt=fitstt$par
19 sestt=sqrt(diag(solve(fitstt$hessian)))
20 parameter=c("mean", "sd", "nu")
21 rbind(parameter, round(eststt, 7), round(sestt, 7))
```

The package `fGarch` is installed in lines 10 and 11 because it contains the function `dsstd()`, which computes the density of the standardized Student's t-distribution that is used in line 13. In line 13, minus the log likelihood function is computed. Note that in line 13, the density of the standardized t-distribution is expressed as a function of the mean (`mean`), the standard deviation (`sd`), and the degrees of freedom (`nu`); the command `log=TRUE` is used so that the logarithm of the density is returned from `dsstd()`. In line 16, minus the log likelihood is minimized by using the function `optim()`; minimizing minus the log likelihood is equivalent to maximizing the log likelihood. Line 15 specifies the starting values for the mean, the standard deviation, and the degrees of freedom that are used for the minimization algorithm implemented by `optim()` in line 16. In line 16, the command `lower=c(-1,0.001,1)` gives lower bounds for the minimization search. In line 16, the results of the minimation algorithm are stored in the object `fitstt`. In particular, the maximum likelihood estimates of the parameters produced by the function `optim()` are contained in the object `fitstt$par` used in line 18. In line 19, the standard errors for the maximum likelihood estimates are computed from the hessian of the log likelihood. Line 21 outputs the maximum likelihood estimates and their standard errors.

Note that the location-scale Student's t-distribution and the standardized Student's t-distribution are both symmetric. To handle the possibility that the returns are skewed, it is useful to have a heavy-tailed distribution that can accommodate skewness. One such skewed distribution is the Fernandez-Steel (F-S) skewed t-distribution. As before, suppose the $T \sim t_{(\nu)}$, and let $f_T(t)$ be probability density of T. Now, consider a random variable U whose probability density function is

$$f_U(u|\xi) = \begin{cases} \frac{2}{(\xi^{-1}+\xi)} f_T(u\xi) & \text{if } u < 0, \\ \frac{2}{(\xi^{-1}+\xi)} f_T(u/\xi) & \text{if } u \geq 0, \end{cases}$$

where ξ (the Greek letter xi) is a positive constant. If $\xi = 1$, then $f_U(u|\xi = 1) = f_T(u)$, so U has the same distribution as T, and hence, U is symmetric. If $\xi > 1$, then U is skewed to the right; if $\xi < 1$, then U is skewed to the left. A random variable Y is said to have the location-scale F-S skewed Student's t-distribution if $Y = a + bU$. Note that the distribution of Y depends on four parameters: the location parameter a, the scale parameter b, the degrees of freedom ν, and the skewness parameter ξ.

The function `dstd()` in the package `fGarch` computes the density of the F-S skewed Student's t-distribution. In the function `dstd()`, the distribution is specified by giving the mean, the standard deviation, the degrees of freedom, and the value of the skewness parameter ξ. Running the following lines produces a plot that compares three densities: the standard normal density; the standardized Student's t-density with mean 0, standard deviation 1, and degrees of freedom 5; and the F-S skewed Student's t density with mean 0, standard deviation 1, degrees of freedom 5, and skewness parameter $\xi = 2$. These lines assume that the package `fGarch` has already been installed.

```
22 x=seq(from=-4, to=4, length=500)
23 normdensity=dnorm(x, mean=0, sd=1)
24 sttdensity=dstd(x, mean=0, sd=1, nu=5)
25 skewtdensity=dsstd(x, mean=0, sd=1, nu=5, xi=2)
26 y=max(c(normdensity, sttdensity, skewtdensity))
27 plot(x, normdensity, type="l", lwd=2, lty=1, col="red", ylim=c(0,y),
28        ylab="density")
29 lines(c(-4,4), c(0,0), lwd=1)
30 lines(x, sttdensity, type="l", lwd=2, col="blue", lty=2)
31 lines(x, skewtdensity, type="l", lwd=2, col="blue", lty=3)
32 legend(-4, y, legend=c("normal", paste("std t") , paste("skew t")),
33        lty=c(1,2,3), lwd=c(2,2,2), col=c("red","blue","blue"), cex=2)
```

The x-values at which the densities are computed for the plot are specified in line 22. For the chosen x-values, the normal density is computed in line 23, the standardized Student's t-density is computed at line 24, and the F-S skewed Student's t-density is computed at line 25.

To fit the F-S skewed Student's t-distribution to the returns for 3M (MMM) by maximum likelihood estimation, run the following lines:

```
34 x=Returns$MMM
35 loglikskt=function(beta) sum(-dsstd(x, mean=beta[1], sd=beta[2], nu=beta[3],
36        xi=beta[4], log=TRUE))
37 start=c(mean(x), sd(x), 5, 1)
38 fitskt=optim(start, loglikskt, hessian=TRUE, method="L-BFGS-B",
39        control=list(parscale=c(1e-4,1e-3,1,1)), lower=c(-0.1,0.01,1.8,0.001))
```

```
40 estskt=fitskt$par
41 seskt=sqrt(diag(solve(fitskt$hessian)))
42 parameter=c("mean", "sd", "nu","xi")
43 rbind(parameter, round(estskt, 7), round(seskt, 7))
```

These lines assume that the package `fGarch` has already been installed. The negative log likelihood function is stored in the object `loglikskt` in line 35. The negative log likelihood is minimized by using the function `optim`, with starting values for the minimization algorithm specified in line 37; the results of the minimization are stored in the object `fitskt`. The maximum likelihood estimates are contained in the object `fitskt$par`, and these are stored inthe object `estskt` in line 40. The standard errors of the maximum likelihood estimates are computed in line 41, and the standard errors are stored in the object `seskt`. Lines 42 and 43 print the estimates and standard errors. The function `round()` in line 43 is used to round the numerical results to 7 decimals.

Linear regression models are useful for discerning how the returns of one asset are related to the returns of another asset. In the simple linear regression model, there is a pair (X, Y) of variables: Y is the the "response" variable, and X is the "predictor" variable. The model stipulates that $Y = \alpha + \beta X + \epsilon$, where ϵ is called the "disturbance" or "error." It is assumed that ϵ is independent of X and that $E(\epsilon|X) = 0$, so $E(Y|X) = \alpha + \beta x$ and $E(\epsilon) = 0$; furthermore, $Var(Y|X) = \sigma_\epsilon^2$, where $Var(\epsilon) = \sigma_\epsilon^2$. Often it is assumed that ϵ is normally distributed, i.e., $\epsilon \sim N(0, \sigma_\epsilon^2)$, so that the conditional distribution of Y given X is $N(\alpha + \beta X, \sigma_\epsilon^2)$. Under the normality assumption, the least squares estimate of (α, β) derived from a sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ is also the maximum likelihood estimate. One goal of linear regression is to reduce variability: generally, $Var(Y|X) = \sigma_\epsilon^2 \le Var(Y)$; and hopefully, $Var(Y|X) = \sigma_\epsilon^2$ is substantially less than $Var(Y)$.

If (X, Y) is a pair of contemporaneous returns for two assets, it is likely that the disturbance ϵ is not normally distributed. A more realistic model would be that ϵ has the standardized Student's t-distribution with mean 0, standard deviation σ_ϵ, and degrees of freedom ν; in this case, the conditional distribution of Y given X is the standardized Student's t-distribution with mean $\alpha + \beta X$, standard distribution σ_ϵ, and degrees of freedom ν. In the following `R` code, the least squares estimates of α and β, which are the maximum likelihood estimates under the normal model, are computed, and they are compared with the maximum likelihood estimates of α and β derived from using a standardized Student's t-distribution for the conditional distribution of Y given X.

Suppose that Y is a return from 3M (MMM) and X is the contemporaneous return from the Dow Jones Industrial average (DOW). To estimate the linear regression of Y on X based on the sample returns in the objects `Returns$MMM` and `Returns$DOW` by the method of least squares, run the following lines:

```
44 fitLSE=lm(Returns$MMM~Returns$DOW)
45 fitLSE
46 summary(fitLSE)
47 round(confint(fitLSE, level=0.99), 7)
48 round(head(fitLSE$residuals), 7)
```

The function `lm()` in line 44 fits the linear regression model by least squares and stores the results in the object `fitLSE`. The output that results from the call to `fitSLE` in line 45 shows the maximum likelihood estimate $(\hat{\alpha}, \hat{\beta})$ and the standard errors of the estimates. A more extensive summary of the results from the linear regression analysis are obtained from the function `summary()` used in line 46. Note that the "residual standard error," which is an unbiased estimate of σ_ϵ, is reported in this summary. The function `confint()` used in line 47 produces confidence intervals for α and β; the command `level=0.99` in line 47 results in 99% confidence intervals being shown. Line 48 causes the first six regression residuals to be shown. The residual for observation (X_i, Y_i) is $\hat{\epsilon}_i = Y_i - (\hat{\alpha} + \hat{\beta} X_i)$; according to the simple linear regression model, the observation (X_i, Y_i) has $Y_i = \alpha + \beta X_i + \epsilon_i$, so the residual $\hat{\epsilon}_i$ is an estimate of the disturbance ϵ_i.

Again suppose that Y is a return from 3M (MMM) and X is the contemporaneous return from the Dow Jones Industrial average (DOW); now suppose that the conditional distribution of Y given X is a standardized Student's t-distribution with mean $\alpha + \beta X$. Thus, $Y = \alpha + \beta X + \epsilon$, where the disturbance ϵ has the standardized Student's t-distribution with ν degrees of freedom; ϵ has mean 0, provided $\nu > 1$, and standard deviation σ_ϵ, provided $\nu > 2$. To estimate the linear regression of Y on X based on the sample returns in the objects `Returns$MMM` and `Returns$DOW` by maximum likelihood, run the following lines:

```
49 y=Returns$MMM
50 x=Returns$DOW
51 fitLSE=lm(y~x)
52 loglik=function(beta) sum(-dstd(y, mean=beta[1]+beta[2]*x, sd=beta[3],
53        nu=beta[4], log=TRUE))
54 start=c(as.numeric(fitLSE$coefficient), sd(fitLSE$residuals), 5)
55 fit=optim(start, loglik, hessian=TRUE, method="L-BFGS-B",
56        control=list(parscale=c(1e-4, 1, 1e-3, 1)), lower=c(-10,0.001,0.001,1))
57 mle=fit$par
58 se=sqrt(diag(solve(fit$hessian)))
59 lowerCL=round(mle+qnorm(0.005)*se, 7)
60 upperCL=round(mle+qnorm(0.995)*se, 7)
61 parameter=c("alpha", "beta", "sd", "nu")
62 rbind(parameter, round(mle, 7), round(se, 7))
63 cbind(parameter, lowerCL, upperCL)
```

The log likelihood function is computed in line 52. In line 54, starting values are specified for the minimization conducted using the `optim()` function in line 55: the starting values for (α, β) are the least squares estimates; the starting value for the standard deviation is the sample standard deviation of the residuals obtained from least squares; and the starting value for the degrees of freedom is 5. The maximum likelihood estimates of $(\alpha, \beta, \sigma_\epsilon, \nu)$ are stored in the object `mle` in line 57, and the standard errors are stored in the object `se` line 58. Lower and upper endpoints of 99% confidence intervals are computed in lines 59 and 60. The output from line 62 presents the maximum likelihood estimates with their standard errors; the output from line 63 presents the confidence intervals.

Questions:

1. [15 points] Run lines 1 to 5. Adapt lines 6 to 9 to fit a location-scale Student's t-distribution for the returns from Goldman Sachs (GS). Submit your output.
i) What is the maximum likelihood estimate of the degrees of freedom?
ii) What is the standard error for the scale parameter?
iii) Test the null hypothesis that the mean of the returns is 0 against the alternative hypothesis that the mean exceeds 0. Use the 5% significance level for your test. Would you conclude that the mean is positive?

2. [15 points] Adapt and run lines 10 to 21 to fit the standardized Student's t-distribution to the returns for Goldman Sachs (GS) by maximum likelihood. Submit your output.

i) What is the maximum likelihood estimate of the degrees of freedom?
ii) Observe that the maximum likelihood estimate of the standard deviation is different from the maximum likelihood estimate of the scale parameter in Question 1. Are these results inconsistent? Justify your answer.

3. [15 points] Recall that the F-S skewed Student's t-density is right-skewed if $\xi > 1$ and is left-skewed if $\xi < 1$.
i) By adapting and running lines 22 to 33, create a plot that compares three densities: the standard normal density; the standardized Student's t-density with mean 0, standard deviation 1, and degrees of freedom 3; and the F-S skewed Student's t density with mean 0, standard deviation 1, degrees of freedom 3, and skewness parameter $\xi = 2.5$. Submit your plot.
ii) In your plot from Part i), is the F-S skewed Student's t-density skewed to the right or skewed to the left?
iii) Repeat Part i) using skewness parameter $\xi = 1/2.5$. Submit your plot.
iv) In your plot from Part iii), is the F-S skewed Student's t-density skewed to the right or skewed to the left?

4. [15 points] By adapting and running lines 34 to 43, fit the F-S skewed Student's t-distribution to the returns for Goldman Sachs (GS) by the method of maximum likelihood. Submit your output.
i) What is the maximum likelihood estimate $\hat{\xi}$ of ξ? Does this estimate suggest that the returns for Goldman Sachs (GS) are skewed to the right or skewed to the left?
ii) What is the standard error of $\hat{\xi}$?
iii) Test the null hypothesis $H_0 : \xi = 1$ against the alternative hypothesis $H_A : \xi \neq 1$ using the 5% level of significance. Is there evidence that the returns for Goldman Sachs (GS) are skewed?

5. [20 points] By adapting and running lines 44 to 48, estimate the linear regression of the returns of Goldman Sachs (GS) on the returns of the S&P 500 index (SP500) by the method of least squares. Provide 95% confidence intervals. Submit your output.
i) What are the least squares estimates of α and β?
ii) What is the standard error of the estimate of β?
iii) What is the 95% confidence interval for β?
iv) Based on your confidence interval in Part iii), would you reject the null hypothesis $H_0 : \beta = 1$ in favor of the alternative hypothesis $H_A : \beta \neq 1$ in a test of level 5%?
v) What is the estimate of σ_ϵ?

6. [20 points] By adapting and running lines 49 to 63, use maximum likelihood to estimate the linear regression of the returns of Goldman Sachs (GS) on the returns of the S&P 500 index (SP500) under the assumption that the model disturbances have the standardized Student's t-distribibuion. Provide 95% confidence intervals. Submit your output.
i) What are the maximum likelihood estimates of α and β?
ii) What is the standard error of the estimate of β?
iii) What is the 95% confidence interval for β?
iv) Based on your confidence interval in Part iii), would you reject the null hypothesis $H_0 : \beta = 1$ in favor of the alternative hypothesis $H_A : \beta \neq 1$ in a test of level 5%?
v) What is the estimate of σ_ϵ?