**Boston University Questrom School of Business**

**MF 793 – Fall 2021**

Eric Jacquier

**Problem Set 5**

**Due Sunday November 29th at Noon Boston time**

Problem is in groups of 4 maximum Put all 4 names clearly

* Late Problems are not graded. Do the Problem Set in your usual groups
* All submissions are due on Gradescope. A Gradescope version problem will be created.
* To get a check, you need to answer all the questions including the discussion questions, and have your R code as a printed appendix at the end of the report.
* To get credit, everything which is not a plot, table, or R output **must be hand-written**.
* Time to give readable output to the boss. You will **not** get a check plus if you put irrelevant digits in your Table results. Think before you print.

LAST NAME FIRST NAME

1

2

3

4

**Problem 1:** Was the pre-COVID end of the 2010s more volatile than the early 2010s?

The end of the 2010 decade was turbulent, we can say that! You wonder if it was that bad in terms of stock volatility.

**a)** You check this using the market total return (make sure to add back the risk-free rate for what it matters!) from KF’s data, and the stocks in the files stk-mon-2019 and stk-day-2019 You compare period 1: 2012/1-2016/06 to period 2: 2016/07 – 2019/12.

You use monthly **log-**returns because the well-known fat-tailness of daily log-returns (high kurtosis as is well known) may throw the exact distribution of the Chow test. As your data is the discrete return, make sure to compute log-returns. You compute the Variance Ratio of Period 2 to Period 1.

* Give the theoretical distribution of VRM under H0: σ1 = σ2, given your monthly sample sizes for periods 1 and 2?

The distribution of VRm is a F-distribution: **VRm ~ F(v2=41,v1=53),**

Where the numerator is period 2 and the denominator is period 1.

* What are the cutoffs for a 2-sided 10% test?

The cutoffs is about **[0.608, 1.616]**

* How big must σ2/σ1 be to reject H0 with 10% significance ?

To reject the null by σ2 < σ1, we need **σ2 < 0.608 σ1** that is **σ1 > 1.645 σ2**

* How big must σ1/σ2 be to also reject H0.?

To reject the null by σ1 < σ2, we need **σ2 > 1.616 σ1**

* What?? Why are these 2 ratio cut-offs different? When would they be equal?

Since our sample sizes from period 1 and period 2 are different, the distribution of this F-distribution is asymmetry, which lead to the different ratio cut-offs. Therefore, if we set the sample size (the number of months in both periods) same, the cutoffs would be equal.

**b)** Fill in the first three columns of Table 1. Subscript M for monthly returns. Make sure to report annualized standard deviations.

**Table 1: Volatility for 2012/1-2016/6 and 2016/7-2019/12**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | σ1M (ann.) | σ2M (ann.) | VRM(2/1) | σ1D (ann.) | σ2D (ann.) | VRD |
| Biogen | 0.287 | 0.347 | 1.465 | 0.343 | 0.362 | 1.111 |
| Pfizer | 0.148 | 0.153 | 1.069 | 0.167 | 0.174 | 1.082 |
| Astra-Z | 0.187 | 0.232 | 1.542 | 0.214 | 0.225 | 1.102 |
| Toyota | 0.183 | 0.140 | 0.584 | 0.218 | 0.156 | 0.511 |
| Nike | 0.199 | 0.212 | 1.136 | 0.221 | 0.234 | 1.117 |
| US VW | 0.109 | 0.122 | 0.892 | 0.135 | 0.128 | 0.892 |
| Average |  |  | 1.115 |  |  | 0.970 |

* According to **your** view, how much bigger than σ1 should σ2 be to constitute an **economically** significant deviation from equality.

10% reason?

* Given your view, is the VRM test powerful enough to reject H0, for economically significant deviations from equality?

??

**c)** This lack of statistical power is annoying! You boss asks you to redo the VR test with daily returns. With the added power, you figure you can use a 5% test. Fill in the last 3 columns of Table 1.

* What are the relevant F statistic and the relevant two-sided 5% cut-off values ?

The distribution is about : **VRd ~ F(880,1130)**

And the cutoffs is about : **[0.882,1.132]**

* Conclude: Was the 2nd period stock variance higher than the first period. Statistically, economically?

???

**d)** Your boss points out that you ignored the well known fact that daily log-returns are much more fat-tailed than a normal distribution. This is so unfair, you knew that all along!

You must simulate the distribution of the VRD statistic to adjust your confidence cut-offs for non-normality. You do 20000 simulations under H0 of a sample with N1 (period 1) and N2 (period 2) pseudo daily returns, with N1 and N2 the sample sizes in c).

* The mean is irrelevant so you simulate zero mean for H0: σ=σ1=σ2. The standard deviation is irrelevant too since you computing a ratio. So you simulate σ = 1.
* You simulate to fill row 1 of Table 2: Normal log-returns. This is just to check your code is right. Then to simulate fat tails – you use 20000 pseudo daily returns from the Student-t with 5 degrees of freedom.

**Table 2:** Theoretical and simulated Chow Test, 20000 simulations

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Theory with Normal returns | | | | Simulation | | |
|  | ν1 | ν2 | F0.025 | F0.975 | % reject |  | *VR0.975* |
| Normal | 880 | 1130 | 0.882 | 1.132 | 0.0536 | 0.880 | 1.133 |
| t(5) |  |  |  |  | 0.2673 | 0.799 | 1.259 |
| Bootstrap |  |  |  |  | 0.4762 | 0.719 | 1.097 |

% reject is the % times you reject when using the theory F cut-offs.

**e)** Your boss is on a micro-managing rampage these days.

“How do you know stock returns are t(6), what if they are not? Use the Empirical distribution of the data rather than making some likely wrong Student-t assumption! Add a bootstrap row to that table and bring me the results yesterday morning”.

You have so had it you feel like turning your zoom video off so you can make faces.

* To add Row 3, you bootstrap the actual returns. Use the stock market return. That is, you randomly select N1 + N2 returns **with replacement** from your daily market returns series, all N1+N2, daily returns for this. Then you compute VR. You do it 20000 times, and fill in the 3rd row. The R command “sample” will help, don’t forget “replace = T”
* Given Table 2, does the theoretical F reject too much or not enough or about correctly for daily returns?

??

**f )** So what is the actual distribution of VR?In **Figure 1**, show an **F-probability plot** of the 20,000 VRs against the theoretical F, Figure 1a, for the normal monthly case, Figure 1b for the bootstrapped case. How does the distributions departs from the theoretical F. See the help for a F-probability plot at the end.

* Revise, if necessary, your conclusion from Table 1 VRD, by using the proper cutoffs.

**??**

**Problem 2: How precisely can we estimate stock and portfolio betas?**

**a)b)c) handwritten**

**a)** Consider N (i = 1 , N) stocks with market betas βi, and a portfolio of the N stocks with weight vector w. Write (prove) the portfolio beta **βP**, as a function of the N stock betas βI’s and the weight vector w.

**文本, 信件

描述已自动生成**

How about ­as a function of the estimates ­?

**b)** Now consider the standard deviation of each ­. You estimate each βi from a simple regression of stock I on the market return, usual assumptions. You get from each regression output. Assume the N stock beta estimates are uncorrelated. Write the standard deviation of as a function of the N stock and .. N. Assume an equal-weighted portfolio to simplify the result. At what rate does the standard deviation of decrease as the portfolio size increases?

**文本, 信件

描述已自动生成**

**c)** Consider the classic *market regression* Rit = α + β RMt + εit, t= 1, …T. You can run it on monthly or daily returns. Say daily returns have negligible autocorrelation.

Going from Monthly to Daily returns, how do and change?? This is one sentence each!

Going from Monthly to Daily returns, what happens to the standard deviation of and ?

**d)** Reality check. Use the stock files stk-mon-2019 and stk-day-2019 to get the estimates for Table 1, period 2015-2019. For both daily and monthly returns, regress the 4 assets on the market, report the α and β and their standard errors. You can do 5 regressions and save these 4 quantities in one shot by using the method (sapply(?, ’[‘, c(??) ) ) seen in class, see R lecture note and code help below.

Also write in the **annualized** standard deviations for each stock σD and σM.

Make sure to *appropriately* transform the α’s so they are comparable between daily and monthly regression and explain what α you are writing

***Answer:*** *I annualized the alpha*

Table 1: Market regressions from monthly and daily returns. Period 2015-2019

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Monthly Regression on USVW | | | |  | Daily Regression on USVW | | | |
|  | σD | α | tα | β | sβ | σM | α | tα | β | sβ |
| Biogen | 0.38 | -0.1259 | -0.81 | 1.06 | 0.07 | 0.34 | -0.1132 | -0.92 | 1.13 | 0.33 |
| Pfizer | 0.18 | -0.0010 | -0.02 | 0.76 | 0.03 | 0.16 | 0.0139 | 0.21 | 0.62 | 0.15 |
| Astra-Z | 0.23 | 0.0350 | 0.38 | 0.70 | 0.04 | 0.21 | 0.0657 | 0.70 | 0.42 | 0.21 |
| Nike | 0.23 | 0.0552 | 0.64 | 0.97 | 0.04 | 0.20 | 0.0732 | 0.92 | 0.81 | 0.18 |
| EW Port |  |  |  | 0.87 | 0.03 |  |  |  | 0.75 | 0.12 |

It looks like this below is a lot, but these are one sentence or one or two word answers, as we walk you through the story step by step. Enjoy.

**d1)** Compare the 4 monthly and daily βs.

* Are the monthly and daily βs very different? **No**
* Are their standard errors very different? **Yes**
* So, which frequency seems better as an estimations strategy?

**Answer :** daily seems better because these data have less standard error, which leads to more precise results.

**d2)** Compare the 4 monthly and daily α’s.

* Are the monthly and daily very different (after appropriate transformations)?

**No**

* What “appropriate” transformation did you use?

**Answer:** I use annualization for appropriate transformation. Alpha is the excess mean return of the stock after explained by beta. Thus, we can annualize it to make it comparable.

* Are the αs estimated more precisely with daily than with monthly returns? Use the t-statistic of α to answer **No, t-statics seems not too much change**.
* Explain why or why not in one sentence:

**Answer:** means can’t be more precisely estimated by using higher frequency returns, and alpha is a part of excess mean. So it can’t be more precise with larger frequency.

**d3)** Portfolio vs stock regression precision. Make an equal weighted portfolio “EW Port” of the four stocks and regress it on the US market return. Fill in the β estimates and std. deviations in Table 1

* What is the average sβi for the stock regressions? **Daily:0.05, Monthly: 0.22**
* What is the sβP for theEW Port regression? **Daily:0.03, Monthly: 0.12**
* Given your result in b), do the cross-correlation of the 4 look close to 0?

**Answer:** Yes, since the average beta for the four regression and the bate of EW Port are nearly same, we can conclude that the cross-correlation of 4 look close to 0

* Get the 4 residual vectors (of the 4 stock regressions) in one data matrix. Compute the 4x4 correlation matrix, and compute the average of the 6 cross-correlations. Make sure to remove the ones! What are they (for monthly and daily regressions)?

**For daily data:**

The correlation matrix:

文本, 信件

描述已自动生成

And the average of 6 cross-correlations: **0.0592**

**For monthly data:**

The correlation matrix:

文本, 信件

描述已自动生成

And the average of 6 cross-correlations: **0.0606**

**Problem 3:** Autocorrelation and heteroskedasticity in daily regressions

Pick the **Pfizer** regression and use the methods shown in class and in the R lecture note to compute OLS robust standard errors and confidence intervals. Fill in Table 1.

Table 1: Sandwich estimates of the slope standard error for the Pfizer daily regression

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | OLS |  | 5% | 95% |
| OLS – iid | 0.7612 | 0.0300 | 0.7023 | 0.8201 |
| HC – White | 0.7612 | 0.0327 | 0.6970 | 0.8254 |
| HAC - Andrews | 0.7612 | 0.0338 | 0.6948 | 0.8276 |

* From these results, does there seem to be a lot of heteroskedasticity in the daily residuals? a lot of autocorrelation?

**YOUR RCODE:**

*# Problem 1*

*# a)*

*monthret <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM5/stk-mon-2019.csv",header=T)*

*dailyret <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM5/stk-day-2019.csv",header=T)*

*logmonret <- cbind(monthret[1],log(1 + monthret[,2:7]))*

*logmonret1 <- logmonret[logmonret$date<="20160630"&logmonret$date>="20120101",]*

*logmonret2 <- logmonret[logmonret$date<="20191231"&logmonret$date>="20160701",]*

*length(logmonret1[,1])#54*

*length(logmonret2[,1])#42*

*# the distribution : F(v1=53,v2=41)*

*cutoff <- qf(c(0.05,0.95),41,53)*

*cutoff*

*# b)*

*#compute USVW*

*ff<- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM5/FF-Factors-day.csv",header=T)*

*ffm <-read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM5/FF-Factors-mon.csv",header=T)*

*mkt <- cbind(ff[1],log(1+(ff[2]+ff[5])/100))*

*mkt1 <- mkt[mkt$X<="20160630"&mkt$X>="20120101",][,2]*

*mkt2 <- mkt[mkt$X<="20191231"&mkt$X>="20160701",][,2]*

*mkt1sd <- sd(mkt1)\*sqrt(252)*

*mkt2sd <- sd(mkt2)\*sqrt(252)*

*(mkt2sd/mkt1sd)^2*

*mktm <- cbind(ffm[1],log(1+(ffm[2]+ffm[5])/100))*

*mktm1 <- mktm[mktm$X<="20160630"&mktm$X>="20120101",][,2]*

*mktm2 <- mktm[mktm$X<="20191231"&mktm$X>="20160701",][,2]*

*mktm1sd <- sd(mktm1)\*sqrt(12)*

*mktm2sd <- sd(mktm2)\*sqrt(12)*

*(mkt2sd/mkt1sd)^2*

*#compute stocks*

*sd1 <- apply(logmonret1[,2:7],2,sd)*

*sd2 <- apply(logmonret2[,2:7],2,sd)*

*annsd1 <- sd1 \* sqrt(12)*

*annsd2 <- sd2 \* sqrt(12)*

*(annsd2/annsd1)^2*

*mean((annsd2/annsd1)^2)*

*# then compute the daily data*

*logdayret <- cbind(dailyret[1],log(1 + dailyret[,2:7]))*

*logdayret1 <- logdayret[logdayret$date<="20160630"&logdayret$date>="20120101",]*

*logdayret2 <- logdayret[logdayret$date<="20191231"&logdayret$date>="20160701",]*

*sdd1 <- apply(logdayret1[,2:7],2,sd)*

*sdd2 <- apply(logdayret2[,2:7],2,sd)*

*annsdd1 <- sdd1 \* sqrt(252)*

*annsdd2 <- sdd2 \* sqrt(252)*

*(annsdd2/annsdd1)^2*

*mean((annsdd2/annsdd1)^2)*

*# c*

*length(logdayret1[,1])#1131*

*length(logdayret2[,1])#881*

*# the distribution : F(880,1130)*

*cutoff2 <- qf(c(0.025,0.975),880,1130)*

*cutoff2*

*# d)*

*#row1 (normal)*

*number <- length(logdayret1[,1]) + length(logdayret2[,1])*

*simday <- matrix(rnorm(number\*20000,mean=0,sd=1),ncol=20000)*

*simday1 <- simday[1:1131,]*

*simday2 <- simday[1132:2012,]*

*simsd1 <- apply(simday1,2,sd)*

*simsd2 <- apply(simday2,2,sd)*

*simvr <- (simsd2/simsd1)^2*

*c(quantile(simvr,0.025),quantile(simvr,0.975))*

*rej1 <- (length(simvr[simvr > qf(0.975,880,1130)]) + length(simvr[simvr < qf(0.025,880,1130)]))/length(simvr)*

*#row2 (t)*

*simday2 <- matrix(rt(number\*20000,5),ncol=20000)*

*sim2day1 <- simday2[1:1131,]*

*sim2day2 <- simday2[1132:2012,]*

*sim2sd1 <- apply(sim2day1,2,sd)*

*sim2sd2 <- apply(sim2day2,2,sd)*

*sim2vr <- (sim2sd2/sim2sd1)^2*

*c(quantile(sim2vr,0.025),quantile(sim2vr,0.975))*

*rej2 <- (length(sim2vr[sim2vr > qf(0.975,880,1130)]) + length(sim2vr[sim2vr < qf(0.025,880,1130)]))/length(sim2vr)*

*#row3 (bootstrap)*

*mktv1 <- matrix(sample(size=1131\*20000,replace=T,x=mkt1),ncol=20000)*

*mktv2 <- matrix(sample(size=881\*20000,replace=T,x=mkt2),ncol=20000)*

*mksd1 <- apply(mktv1,2,sd)*

*mksd2 <- apply(mktv2,2,sd)*

*mktvr <- (mksd2/mksd1)^2*

*c(quantile(mktvr,0.025),quantile(mktvr,0.975))*

*rej3 <- (length(mktvr[mktvr > qf(0.975,880,1130)]) + length(mktvr[mktvr < qf(0.025,880,1130)]))/length(mktvr)*

*# e)*

*qqplot(qf(ppoints(20000),880,1130),simvr,col="red",main="20,000 normal return samples against the theoretical F",xlab="F(880,1130)",ylab="Emperical Ratio")*

*abline(0,1)*

*qqplot(qf(ppoints(20000),880,1130),mktvr,col="red",main=" 20,000 bootstrap return samples against the theoretical F",xlab="F(880,1130)",ylab="Emperical Ratio")*

*abline(0,1)*

*#Problem 2*

*# d)*

*dstockret <- logdayret[logdayret$date<="20191231"&logdayret$date>="20150101",]*

*dusvw <- mkt[mkt$X<="20191231"&mkt$X>="20150101",][,2]*

*dbio <- dstockret[,6]*

*dpf <- dstockret[,2]*

*daz <- dstockret[,7]*

*dnik <- dstockret[,3]*

*#compiute the sigma d*

*sd(dbio)\* sqrt(252)*

*sd(dpf)\* sqrt(252)*

*sd(daz)\* sqrt(252)*

*sd(dnik)\* sqrt(252)*

*reg1<- lm(dbio ~ dusvw)*

*summary(reg1)*

*-0.0004995877 \* 252 #annulize?*

*reg2 <- lm(dpf ~ dusvw)*

*summary(reg2)*

*-4.044e-06 \* 252*

*reg3 <- lm(daz ~dusvw)*

*summary(reg3)*

*0.0001390 \* 252*

*reg4 <- lm(dnik ~ dusvw)*

*summary(reg4)*

*0.0002192 \* 252*

*#now compute the month data*

*dstockretm <- logmonret[logmonret$date<="20191231"&logmonret$date>="20150101",]*

*dusvwm <- mktm[mktm$X<="20191231"&mktm$X>="201501",][,2]*

*dbiom <- dstockretm[,6]*

*dpfm <- dstockretm[,2]*

*dazm <- dstockretm[,7]*

*dnikm <- dstockretm[,3]*

*# sigma month*

*sd(dbiom)\* sqrt(12)*

*sd(dpfm)\* sqrt(12)*

*sd(dazm)\* sqrt(12)*

*sd(dnikm)\* sqrt(12)*

*regm1<- lm(dbiom ~ dusvwm)*

*summary(regm1)*

*-0.01110 \* 12 #annulize?*

*regm2 <- lm(dpfm ~ dusvwm)*

*summary(regm2)*

*0.001157 \* 12*

*regm3 <- lm(dazm ~dusvwm)*

*summary(regm3)*

*0.005475 \* 12*

*regm4 <- lm(dnikm ~ dusvwm)*

*summary(regm4)*

*0.006096 \* 12*

*# Make an equal weighted portfolio “EW Port” of the four stocks and regress it on the US market return*

*ewpt <- 0.25 \* (dbio + dpf +daz + dnik)*

*regw<- lm(ewpt ~ dusvw)*

*summary(regw)*

*ewptm <- 0.25 \* (dbiom + dpfm +dazm + dnikm)*

*regwm<- lm(ewptm ~ dusvwm)*

*summary(regwm)*

*# residual cross-correlation*

*resmatd <- matrix(0,ncol=4, nrow=1258)*

*resmatd[,1]<- reg1$residuals*

*resmatd[,2]<- reg2$residuals*

*resmatd[,3]<- reg3$residuals*

*resmatd[,4]<- reg4$residuals*

*cor(resmatd)*

*resmatm <- matrix(0,ncol=4, nrow=60)*

*resmatm[,1]<- regm1$residuals*

*resmatm[,2]<- regm2$residuals*

*resmatm[,3]<- regm3$residuals*

*resmatm[,4]<- regm4$residuals*

*cor(resmatm)*

*# Problem 3*

*library(sandwich) # Contains corrections for cov matrices*

*library(lmtest) # contains command coeftest*

*vcov(reg2)*

*vcov(summary(reg2))*

*coeftest(reg2) # OLS*

*coefci(reg2)*

*confint(reg2)*

*coeftest(reg2,vcov=vcovHC) # Hal White Heteroskedasticity only*

*vcovHC(reg2)*

*coefci(reg2,vcov=vcovHC,level=0.95)*

*coeftest(reg2,vcov=vcovHAC) # Hal White Heteroskedasticity only*

*vcovHAC(reg2)*

*coefci(reg2,vcov=vcovHAC,level=0.95)*

**Construct an F-probability plot**

To do Figure 2, generalizing qqnorm to any distribution not just a normal: Combine the commands ppoints and qqplot

**qqplot(x1,x2)** sorts x1 and x2 and does a scatter plot, it is comparison of two empirical densities, x1 and x2 have the same length.

**qqnorm(x2)**, is just like qqplot(x1,x2) where x1 is the theoretical quantiles of the normal for as many points as x2.

To generalize it, we can use qqplot and replace x1 by the quantiles of the theoretical distribution we want to compare with x2. The command ppoints will help.

**ppoints(nsim)** will generate evenly spaced points on [0,1], like nsim evenly spread quantiles of a density. Recall the inverse transform theorem: The inverse CDF gives us the theoretical values that generated these quantiles:

qf(ppoints(1000),dof1,dof2)) gives 1000 sorted values according to the F.

**qqplot(qf(ppoints(1000),dof1,dof2), mydata)** # for a 1000 long mydata

# plots a F-probability plot of mydata

**Run several regressions (several y variables) on the same X variable**

Do not loop around lm for several regressions. You do them like with:

mymod <- lm( stkrets~rm)

# creates the N regressions if stkrets is a matrix of the N stocks returns

modsum<-summary(mymod)

# creates summary object with all the good stuff in it.

coefficients(mymod)

# also works to print the output with estimates and std.errs.

To run in multivariate mode (several regressions), lm wants a matrix or ts object. When we read data by read.csv they appear as a list to lm. This needs to be modified. I just do this:

zstocks<-read.csv(“zfile.csv”,header=T)

strkets<-as.matrix(zstocks) # step to select the correct columns for the y variables

# remove other variables, dates, etc… of course

mymod and modsum are now **multivariate** linear model object,

lists where each regression is an item in the list,

then each regression is itself a list in the list.

It makes it hard to retrieve in a vector for example, all the standard errors. To see the problem, do names(mymod), names(modsum).

Some commands now don’t work. You cannot do things like: confint(mymod)

What we want (estimates and their stddevs, maybe the ts) is now 2 steps down in the list hierarchy.

You can extract it with the **lapply** (list apply) and **sapply** commands, similar to **apply** but extract components of lists. In coefficients(mymod), you see in what order the output is. Now try this for example, you will see what it does:

*lapply(modsum, coefficients)*

is exactly the same as coefficients(modsum). But is is now a list. Look at it.

Then this below takes what we need from it:

sapply(lapply(modsum,coefficients), '[', c(1,3) )

coefficients output for all ‘[‘ says to c(1,3) takes elements 1 and 3

regressions go down a level

R, like all languages stores column wise, so find the order of the intercept and slope, and standard errors or whatever you want in coefficients(modsum), and use that in c(?,?,?...)