Last Name: __Liu__________ **First Name: ______Xuyang____________**

BU ID: ___U20404938_____

BOSTON UNIVERSITY QUESTROM SCHOOL OF BUSINESS
MF 793 – Fall 2021
Monday October 25th, 2021
Midterm Exam

Write your name NOW – in the space given.

- The exam is online from **8:30am to 11:00am Boston Time.** The last 15 minutes are to give you time to scan your answers and upload your solved exam.

- **At the latest at 10:45am**, stop working on your exam and start uploading. There will be no added time after 11:00am.

- Upload your completed exam on Gradescope.

- You may consult anything posted on the MF793 Questrom Tools site as well as your own notes. Use R for computations. **You cannot use the "*internet*":** anything not in line with the notation or the assumptions of the course will be considered wrong with no recourse. If your answer appears to have been copied from an internet web site, you will get a zero.

- The exam is zoom proctored: **You must be on zoom starting at 8:30am and never leave zoom and must have your video on at all times otherwise points will be taken off your total grade.**
  If we don't see you all the time, you will lose points.

- You **cannot communicate with anybody by any means** during the exam. If you are at home, make it clear that you cannot be disturbed or talked to during the exam.

- You can not use any device to call, email, text, chat, zoom, etc.. with anybody. Put your phone in Do Not Disturb mode during the duration of the exam. Do not use email.

- You must be alone in a physical room during the exam.

- You must report cheating if you are aware of it

Violation of these rules will result in disciplinary action

- Word and Algebra / proofs must be handwritten in order to get points
- R code must be shown in the indicated space or you will get no points for your answers even if correct.
- The exam has a number of independent questions: algebraic or number calculations as in class or homework, or a bit different, discussion questions involving a short but complete justification of the answer to check that you understood the discussions in class.
- Discussion questions may have a True / False feature: If part of a statement is correct and part is False, you must label the statement as False. Then in your discussion, write clearly what is correct, what is false. and why. If you say True, explain why the entire statement is true. Saying True with no explanation gets zero point even if it is correct.
- Only answers in the space provided for each question count. Answers to a question in the space of another get zero credit
- Correct numerical answers without justification or starting theoretical formula get zero point.
- Be neat and show your work: Answers without work or motivation get no credit. Numerical answers without first writing the formula used get no credit. Wrong final answers with correct initial work get partial credit.
- Be concise: Incorrect statements cost points even if you also write the correct answer next to them. Correct statements unrelated to the question get no credit.

Take a couple minutes to read problems. Check what looks obvious to you and do it first.
GOOD LUCK !

Did you write your name on the front page ?

Problem 1: short questions. 15 pts

a) You find that the standard deviation of weekly log-returns is 2.6 times the standard deviation of daily log-returns for your data. Explain briefly, using the proper formula, what feature of the portfolio return may have caused this result. Your data has 252 days per year and 52 weeks per year. **6 pts**

log return usually can be time-aggretion.

$Log V_T = \log(1+R_1)(1+R_2)\cdots(1+R_T) = \sum_t \log(1+R_t) = \sum_t r_t$

Then. $V(Log V_N) \equiv \sigma_N^2 = N\sigma^2 \Rightarrow \sigma_N = \sqrt{N}\,\sigma$

Since we're talking about daily and weekly data. N should be $\frac{252}{52} \approx 4.8$

$\therefore \frac{\sigma_N}{\sigma} = \sqrt{N} = \sqrt{4.8} \approx \cancel{2.6}\ 2.2$, However, we get $\sqrt{N} \approx 2.6$

$\Rightarrow$ Think about $\sigma_m^2 = V(\Sigma r_{Dt}) = \Sigma\sigma^2 + \sum_t \sum_{s \neq t} Cov(r_{Dt}\, r_{Ds}) = N\sigma_D^2 + STUFF$.

we usually ~~think~~ assume no returns autocorrelation.

But the truth seems that the returns have autocorrelation, and

$\sum_t \sum_{s \neq t} Cov(r_{Dt}\, r_{Ds}) \neq 0 \Rightarrow$ That's why we get $2.6 > \sqrt{\frac{252}{52}}$.

b) *"So if your test comes back positive, your true chance of having the disease is actually 1 out of 51".*

You are fascinated by Dr Morgan's article which you read several times. He does know Bayes. But he does not mention one of the inputs needed in his numerical example.

Define the relevant outcomes, write Bayes Theorem and state which input he doesn't give us. **6 pts**

What should that probability be for Dr Morgan's result (1 in 51) to be correct **3 pts**

2. Define B is Positive. and A is having the disease.

According to Bayes formular:

$P(B) = P(B|A)\cdot P(A) + P(B|A^x)\cdot P(A^x)$

Then we got $P(A|B) = \frac{P(B|A)\cdot P(A)}{P(B)}$, from the article. we must know the probability of having disease and the correct rate of testing.

When. correct rate of testing is about 95%, then, the reselut tends to be $\frac{1}{51}$.

Problem 2: Conditional Densities 12pts

a) x~U(0,1) and y~U(0,x). Compute p(y|x), p(x,y), p(y). **6 pts**

a) $f_X(x) = \begin{cases} 1 & 0<x<1 \\ 0, & \text{otherwise.} \end{cases} \Rightarrow \underset{f_{Y|X}(y|x)}{\overset{p(y|x)}{\shortparallel}} = \begin{cases} \frac{1}{x}, & 0<y<x<1 \\ 0, & \text{otherwise} \end{cases}$

Then $p(x,y) = p(x) \cdot p(y|x) = \begin{cases} \frac{1}{x}, & 0<y<x<1 \\ 0, & \text{otherwise.} \end{cases}$

$p(y) = \int_{y}^{1} \frac{1}{x}\, dx = \ln x \Big|_y^1 = -\ln y$

b) Now compute E(y) in a simple manner. Name which result you use and show the steps. **6 pts**

b) $E(y) = \int \frac{0+x}{2} = \frac{x}{2}$ where $0<y<x<1$

Problem 3: .. Fund Performance 10 pts

You track the performance of N=420 funds over 2 periods. In period 1, X_1=126 funds beat the market, in period 2, only X_2=105 beat the market. Assume all fund returns are iid . Call p1 and p2 the probability of beating the market in periods 1 and 2. You test the null hypothesis that average fund performance stayed constant over the two periods: H_0: p1 = p2.

a) What is the relevant statistic to test H0? **2pts**

b) What is the asymptotic mean of that statistic? **2pts**

c) What is the asymptotic variance of that statistic? **2pts**

d) Write a **98%** CI (not 95 !) for the statistic. **2pts**

e) Given that interval, do you reject H_0 at the 2% level: **2pts**

Problem 3.

a) It's student-t test, so called t-statistic :

$$\cancel{t = \frac{\bar{X}-\mu}{\frac{s_X}{\sqrt{n}}}} \qquad t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

b) $\bar{X}_1 = \frac{\sum_{i=1}^{126} X_i}{126}$. $\bar{X}_2 = \sum_{i=1}^{10}$

c)

Problem 4: 12pts You are star, you are on TV! You play the three doors game, doors A, B, C . Behind one door is a Tesla Z. The other two have nothing. You announce to the hysteric crowd that you will open door A, but do not open it yet. To help you (ummh!?), the host opens door B, of course nothing behind ! Crowd howls. She now allows you to change your mind. Crowd gasps. Will you open A or will you switch and open C ? The host knows where the car is and will never open that door. Initially the car has the same chance of being behind any door. Let's denote outcomes. For (say) door C, we call C: the car is behind C, we call "OC": the host opens door C. Same for door B.

a) Compute p(OB). You will need it for Bayes theorem **4pts**

Problem 4:

a) According to Bayes formular

$$P(OB) = P(OB|A)\cdot P(A) + P(OB|B)\cdot P(B) + P(OB|C)\cdot P(C)$$

$$= \frac{1}{3}\times\frac{1}{2} + \frac{1}{3}\times 0 + \frac{1}{3}\times 1 = \frac{1}{2}$$

b) Compute p(A | OB) by Bayes theorem. **4pts**

b) $$P(A|OB) = \frac{P(OB|A)\cdot P(A)}{P(OB)} = \frac{1}{3}$$

c) Compute p(C | OB) by Bayes theorem (The answer 1-p(A|OB) gets zero points !) **4pts**

Do you switch? 0 pts.

c) $$P(C|OB) = \cancel{P}\, 1 - P(A|OB) = \frac{P(OB|C)\cdot P(C)}{P(OB)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

Because now I only have 2 choices.

$$\cancel{\Rightarrow P(C|OB) = 1 - \frac{1}{3} = \frac{2}{3}}$$

Problem 5 12pts You have T observations of i.i.d. normally distributed returns. Derive (prove) the pdf of $\hat{h}$ the estimate of the precision h. Precision is the inverse of variance ($h = 1/\sigma^2$). What density do you start with, explain all the steps of the change of variable method.

2pts What is the starting Random Variable which pdf you know: $x =$

2pts Write $\hat{h}$ as a function of x : $\hat{h} =$

2pts Write p(x)

2pts Inverse transform:

2pts Inverse transform derivative:

2pts Final result : $p(\hat{h}) =$

Problem 6 Short things. Choose 3 of the 4 questions below. 18 pts

a) Short algebra. $X \sim N(\mu , \sigma)$. $\mathbf{Y = e^X}$. Derive (prove) the formula for $E(Y^2)$. **6pts**

φ $x = \frac{\log M}{2}$

Problem . 6

a) $X \sim N(\mu,\sigma)$. $Y = e^X \to Y^2 = e^{2X} \to 2X = \ln Y^2$. $M = Y^2$

PDF of $Y^2 = e^{2X} \Rightarrow P_Y(y_0) = P_X(g^{-1}(y_0)) \times \left|\frac{d}{dy} g^{-1}(y_0)\right| = \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-\frac{(\frac{\log M}{2}-\mu)^2}{2\sigma^2}} \cdot \frac{1}{2M} = \frac{1}{2\sqrt{2\pi}\sigma y^2} e^{-\frac{(\log y-}{2\sigma}}$

$E(x) = \int_{-\infty}^{\infty} \frac{1}{2\sqrt{2\pi}\sigma} e^{-\frac{(\log y-\mu)^2}{2\sigma^2}} dy = \frac{1}{2} \cdot \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{[\ln y-\mu]^2}{2\sigma^2}} dy$ Define $t = \frac{\ln(y)-\mu}{\sqrt{2}\sigma}$

$\Rightarrow \frac{1}{2}\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} \cdot e^{-t^2} \, d e^{\sqrt{2}\sigma t+\mu} = \frac{1}{2} e^{\mu+\frac{\sigma^2}{2}}$.

b) Write(prove) the formula for the 25th percentile of Y : Y_0 / $\Pr(Y < Y_0) = 0.25$ as an exact function of μ and σ. Write the R command which you use to find the (only!) number in the formula. **6pts**

c) Your assistant checked normality for a few funds, you find something wrong in his table. You don't know **wha**t is wrong but something **is** wrong, explain to him in a few lines! **6pts**

	Kurt
VSXYZ	4.70 [4.25 , 5.38]

c). The upper boud difference is about $5.38 - 4.70 \approx 0.68$,
while the lowerbound difference is about $4.70 - 4.25 \approx 0.45$
When we compute CI $\Rightarrow [4.7 + z \cdot \frac{24}{T}, 4.7 - z \cdot \frac{24}{T}]$. which means the deviation from 4.70 should be in same degree. But Here we got the quite different result.

d) You look again at these GE results. It's infuriating one cannot reject the null $\mu_M=0$ with 96 months of data ! Using the numbers below (no need to go back to the data). How many months would we need with those estimates of μ and σ to *just about* reject H_0? I guess some people would call that to have a p-value just equal to 0.05. Two-sided alternative for this one. **6 pts**

μ_M	σ_M
0.42 [-0.85 ,1.69]	6.33 [5.43,7.22]

Problem 7: Use your data to compute the log-returns on APPLE from 2011-2015 included, daily and monthly. Fill the table below σ estimates. Assume that i.i.d. normal log-returns and write the **exact (not approximate)** CIs. Show you R code and your theoretical starting formulas below. Approximate intervals get zero credit. **18 pts**

	T=	T=
	Daily data	Monthly Data
	σ_D	σ_M
Raw Data	0.0205 [0.0198,0.0214]	0.0772 [0.0654,0.0942]
Annualized	0.33 [0.196, 0.941]	0.27 [0.160, 0.769]

Prove formula for an exact CI on s (sample σ) given sample size T and normal iid log-returns: **(6 pts)**

For $X \sim (\mu, \sigma) \longrightarrow \frac{(T-1)s^2}{\sigma^2} \sim \chi^2(T-1)$ given $1-\alpha$.

$$\Rightarrow P\left\{ \chi^2_{1-\frac{\alpha}{2}}(T-1) < \frac{(T-1)s^2}{\sigma^2} < \chi^2_{\frac{\alpha}{2}}(T-1) \right\} = 1-\alpha.$$

$$\Rightarrow P\left\{ \frac{(T-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}(T-1)} < \sigma^2 < \frac{(T-1)s^2}{\chi^2_{\frac{\alpha}{2}}(T-1)} \right\} = 1-\alpha$$

$$\therefore \text{CI of } \sigma: \left[s\sqrt{\frac{T-1}{\chi^2_{1-\frac{\alpha}{2}}(T-1)}},\ s\sqrt{\frac{T-1}{\chi^2_{\frac{\alpha}{2}}(T-1)}} \right]$$

One more page =>

R code for Table **6pts**

Problem 7

```
daily<-read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM3/stk-11-day-2010-2017.csv",header =T)
monthly <- read.csv("/Users/liuxuyang/Desktop/BU\ FALL\ 2021/MF793/HM3/stk-11-mon-2010-2017.csv",header =T)
logdaily <- log(1+daily[,2:13])[,2]
logdaily <- logdaily[253:1510]
n1 <- length(logdaily)

logmonth <- log(1+monthly[,2:13])[,2]
logmonth <- logmonth[13:72]
ud = mean(logdaily)
ud
sdd = sd(logdaily)
sdd
sdm <- sd(logmonth)
sdm
n2 <- length(logmonth)
c(sdd * sqrt((n1-1)/qchisq(0.975,n1-1)) , sdd * sqrt((n1-1)/qchisq(0.025,n1-1)))
c(sdm * sqrt((n2-1)/qchisq(0.975,n2-1)) , sdm * sqrt((n2-1)/qchisq(0.025,n2-1)))

sda<-sdd*sqrt(252)
sma<-sdm*sqrt(12)

c(sda * sqrt((n1/252-1)/qchisq(0.975,n1/252-1)) , sda * sqrt((n1/252-1)/qchisq(0.025,n1/252-1)))
c(sma * sqrt((n2/12-1)/qchisq(0.975,n2/12-1)) , sma * sqrt((n2/12-1)/qchisq(0.025,n2/12-1)))
```