Last Name: ____Xuyang_____ **First Name: _______Liu__________**

BU ID: ___U20404938_______

BOSTON UNIVERSITY QUESTROM SCHOOL OF BUSINESS
MF 793 – Fall 2021
Tuesday December 14th, 2021
Final Exam

Write your name NOW – in the space given.

- The exam is online from **8:15am to 11:00am Boston Time.** The last 20 minutes are to give you time to scan your answers and upload your solved exam.

- **At the latest at 10:40am**, 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 8:15am and never leave zoom and must have your video on at all times otherwise points will be taken off.**
  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 !

Problem 1: Short questions 12 points

a) You estimate 16 sector portfolios betas. You find that the correlations between the 16 residuals of these regressions are all about equal to 0.2. By what factor can you reduce the standard deviation of the beta of an equal weighted portfolio of the 16 stocks. Write the formula you use but don't prove it **4 pts**

$$V(\hat{\beta}_{OLS}) = E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] = \sigma^2(X'X)^{-1}X'DX(X'X)^{-1}$$

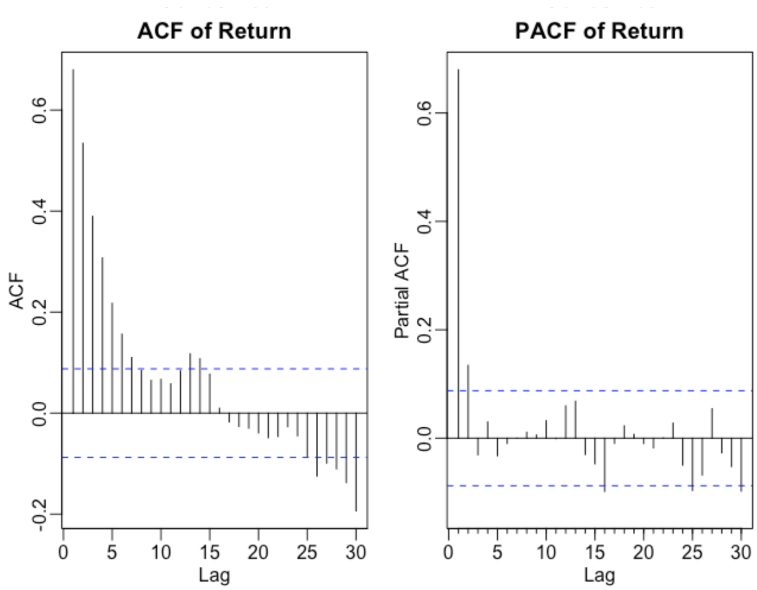
we can estimate $E[X'\varepsilon\varepsilon'X]$

b) Using the arima.sim command in tseries, simulate 500 observations of an AR(2): $\phi_1 = 0.5$, $\phi_2 = 0.2$. Plot the ACF and PACF **next to** each other below on this page. Use mfrow=c(1,2). **8 pts.**

Rcode Here:

```
ret <- rep(0,500);muy <- 0.10/252      ;sigy <- 0.15/sqrt(252)
phi1 <- 0.5;phi2 <- 0.2
alpha <- muy*(1-phi1-phi2)
sige  <- sigy * sqrt(1-phi1^2-phi2^2)
ret[1] <- muy;epst   <- rnorm(500,0,sige)
ret[2] <- alpha + muy*phi1 + epst[i]
for (i in 3:500) { ret[i] <- alpha+ phi1*ret[i-1] + phi2*ret[i-2] + epst[i]}
par(mfrow=c(1,2),mgp=c(1.5,0.5,0),mar=c(3,3,2,0.5))
Acf(ret,lag.max=30);title("ACF of Return")
lines(seq(1,22),phi^seq(1,22),col="red")
Acf(ret,lag.max=30,type="partial")
title("PACF of Return")
```

Plot:



Problem 2: 16 pts We are in January 2022. You use the VR test to compare the variances of the US market return in pre-Covid 2016-2019 included, to the Covid period 2020-2021 included. 12 months and 252 days per year. Your annualized estimate in period 2 is σ2 = 0.17. How small would the pre-covid σ1 have to be to reject the null of equality at the 5% level with monthly, daily return? 2-sided alternitave. Fill in the table and show your computations

	8 pts	**4pts**	**4pts**
	Degrees of Freedom of the F	Cut off of the F	Cut off Value of σ_1
Monthly data	**47,23**	**[0.509,2.147]**	**0.121**
Daily data	**1007 ,503**	**[0.861,1.166]**	**0.158**

Show all your computations below to get credit, included the R code

R code here:

```
cutoff1 <- qf(c(0.025,0.975),47,23)
cutoff2 <- qf(c(0.025,0.975),1007,503)
```

Problem 2. $T = \frac{n_1 S_1^2 (n_2-1) \delta_2^2}{n_2 S_2^2 (n_1-1) \delta_1^2} = \frac{S_1^{*2} \delta_2^2}{S_2^{*2} \delta_2^2} \sim F(n_1-1, n_2-$

Cutoff Value, $\frac{\sigma_1}{\sigma_2} < \sqrt{0.509}$ (Monthly)

$\frac{\sigma_1}{\sigma_2} < \sqrt{0.861}$ (daily)

Problem 3: So, the regression line goes through the means of x and y per the proof on P. 4 of LN10. How about the multiple regression when there are several x variables. Use the exact notation on P. 8 (and the hint!) to prove that the regression plane (call it hyperplane, it sounds more quant.) goes through the data sample means, that is: $\bar{Y} = \bar{x}'\hat{\beta}$. **8pts**

from P.4 of LN 10, in the condition of several x variables

To finish OLS, we need to minimize $(Y-X\beta)'(Y-X\beta)$

Thus, set $\frac{\partial}{\partial\beta}((Y-X\beta)'(Y-X\beta)) = 0 \Rightarrow X'X\hat{\beta} = X'Y$

To be precise:

$$X' = \begin{pmatrix} 1 & \cdots & 1 \\ x_1^1 & \cdots & x_m^1 \\ \vdots & & \\ x_1^n & \cdots & x_m^n \end{pmatrix} \quad \text{And} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & x_1^1 & \cdots & x_1^n \\ \vdots & & & \\ 1 & x_m^1 & \cdots & x_m^n \end{pmatrix}$$

we got first row of $X'Y = \sum_{i=1}^{m} y_i = m \cdot \bar{y}$

And the first row of $X'X = m(\Sigma x_i^1, \cdots \Sigma x_i^n) = m(1, \bar{x}_1, \cdots \bar{x}_n)$

which means $\bar{Y} = \bar{x}'\hat{\beta}$

$\therefore$ The regression plane goes the data sample means.

Problem 4: 12 points

a) First post-COVID Holiday party. You forgot how annoying this Donald guy can be. Here he is pompously explaining to Kayla the new Questrom hire, that "the mean can't become negligible in estimating variance since it shrinks at the same rate N as variance itself if we sample N times more often". With a basic formula written on a napkin and a couple sentences Carla explains to him what's really going on. What did she say? **6pts**

Donald can't be right.
Actually, we can ignore the mean when computing high-frequency data.
$S_t^2 = n\sum_{i=0}^{n-1}(R_{t-\frac{i}{n}} - \hat{u}_t)^2/n \approx \sum_{i=0}^{n-1} R^2_{t-\frac{i}{n}}$, So we can just use the sample mean of x^2 to get variance.

b) Your sloppy intern estimated an AR(2) on monthly VIX: $VIX_t = \phi 0 + 0.7\ VIX_{t-1} + 0.1\ VIX_{t-2} + \epsilon_t$. She forgot the intercept!! Next summer, you will take someone from Questrom. Well, the average VIX for the period is 18 and, and talking about sloppy, you have an easy way to find $\boldsymbol{\phi_0}$ without going back to the data. Show you computation and result. **6 pts**

$$\lim_{k\to\infty} E_t(Y_{t+k} \mid Y_t) = \frac{\phi_0}{1-\phi_1-\phi_2} = 18$$

$$\Rightarrow \phi_0 = 18\cdot(1-0.7-0.1) = \cancel{3.6}\ 3.6$$

Problem 5 20pts

a) You ran an AR(1) on monthly Log(VIX), you found: $\phi_0 = 0.7$ $\phi_1 = 0.84$. $\sigma_\varepsilon = 0.23$. You are confident that the <u>noise is normally distributed</u>. What are the <u>unconditional</u> mean and standard deviation of Log(VIX) implied by these parameters? Show work and formula **6 pts**

unconditional mean: $\lim_{k\to\infty} E_t[Y_{t+k}|Y_t] = \frac{\phi_0}{1-\phi_1} = 4.375$

unconditional Std: $\lim_{k\to\infty} Sd(Y_{t+k}|Y_t) = \sqrt{\frac{\sigma^2}{1-\phi_1^2}} = 0.424$

b) VIX is now $VIX_T = 43$. Use your model to give a one-step ahead forecast of $\text{Log}(VIX_{T+1})$ and its standard deviation. Show work and formula **6pts**

① $Log(VIX_{t+1}) = 0.7 + 0.84\, log(VIX_t) + \varepsilon_{t+1}$

$\Rightarrow log(VIX_{t+1}) = 0.7 + 0.84 \cdot log(43) = 3.859$

② $V_t(Y_{t+1}|Y_t) = V_t(a + \phi Y_t + \varepsilon_{t+1}|Y_t) = \sigma_\varepsilon^2$

$\Rightarrow std(Y_{t+1}|Y_t) = \sigma_\varepsilon = 0.23$.

c) You proudly explain to Kayla that you diagnosed heteroskedasticity in the "naïve" linear model, so you use a Log-linear model. "Interesting" she says "but traders don't care about forecasting Log(VIX), they want to know VIX". Oops! But your knowledge of the lognormal density allows you to resolve this easily. Use your log-AR, properties of the lognormal, and your numbers in b), to compute a one-step forecast of VIX_{T+1} and its standard deviation. Write all theoretical formulas you use. **8pts**

$E[Y_{t+1}|Y_t] = e^{logVIX_{t+1}} = e^{3.859} = 47.42$

$Std = e^{0.23} = 1.2586$

Problem 6

"*Even it the conditional daily distribution of returns is normal with a different variance every day, the unconditional distribution is not norma*l" You must confess you still "don't get it". You decide to simulate it. Simulate 10,000 i.i.d. $\varepsilon \sim N(0,1)$. Then simulate 10000 independent numbers around 1 that will be the standard deviations. Multiply your ε's by these numbers and you will have 10,000 normals each with a different standard deviation. What to pick to simulate the random σ's? You want something which is 1 on average but can be above or below 1. Bingo, you remember, a chi-square divided by its degrees of freedom!

a) Find ν the DOF so that the standard deviation of $\chi 2(\nu)/\nu$ is 0.5 **4pts**

ν = 8

b) Simulate 10000 independent normals distributed as $R \sim N(0, \sigma \sim \chi^2(\nu)/\nu)$. Write your R code – 2 lines at most **8 pts**

R code here:

```
std <- rchisq(10000,8) / 8
rand <- rnorm(10000,sd(std))
```

c) Show a normal probability plot of the 10,000 random and the qqline. Estimate the kurtosis. **10pts**

Kurtosis = 6.325 **2pts**