Total is 106 points

Problem 1: DO ONLY 5 OF THE 7 discussion / quick questions. 8 points each.
For True/False, you need to explain what is correct or not. Statement with both correct and wrong parts should be marked False, then explain what is correct and false in each statement. If you do more than 5 questions, we will grade 5 random questions, not the 5 best.
You should spend no more than 5 minutes on each question.
a) Write a definition or formula for the semi-variance. Give the main argument in favor. Give the main argument against.

- $SV = \frac{1}{N_{x_i<\bar{x}}} \sum_{x_i<\bar{x}} (x_i - \bar{x})^2$
- Good: only care about "bad" deviations that are below mean
- Bad: If distribution is not very asymmetric, reduces sample size and makes estimate less precise

b) In the SUR system with N seemingly unrelated regressions with T observations, $Y_i = X_i\beta_i + \epsilon_i, i = 1, \ldots N$, define the Σ matrix, write the giant covariance matrix Ω and giant data matrix $\mathbf{X}$ as functions of the Σ and X_i matrices. Write $\widehat{\beta}_{SUR}$ as a function of the giant matrices.

- $X = \begin{pmatrix} X_1 & & 0 \\ & X_2 & \\ 0 & & X_N \end{pmatrix}$ $\quad$ • $\Omega = Cov(\mathcal{E}\mathcal{E}') = \Sigma_N \otimes I_T$
- $\Sigma = \{\sigma_{ij}\}$ covariance of $\varepsilon_i \varepsilon_j$
- $\hat{\beta}_{SUR} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$

c) The only advantage of the SUR when the X matrix is the same for all the N regressions is to get us a more precise estimator of β. True or False. If True, say why, if False explain the advantage of SUR.

When $X_1 = X_2 = \cdots = X_N$ $\quad \hat{\beta}_{SUR} = \begin{pmatrix} \hat{\beta}_{1\,OLS} \\ \vdots \\ \hat{\beta}_{N\,OLS} \end{pmatrix}$

The advantage is that

$V(\hat{\beta}_{SUR}) = \Sigma \otimes (X'X)^{-1}$ gives us cross-equation covariances of $\hat{\beta}$ which we may need for cross equation test

d) After choosing the size of the test (α), we must choose the power so we balance ~~probabilities~~ of type I and type II errors, with not too high (low) power for large (small) sample sizes. True or False

After choosing α, the size of the test, there is nothing left to choose

If we want higher power we need a higher α, (lower power, lower α)

Once α is chosen, only T and data variance determine the power

e) Define p-hacking in your own words, then give one example.

It's another term used by Cam Harvey to talk about data mining to find significant results

If I look at several t-stats to choose the highest (lowest pval) I am p-hacking

f) If we could estimate the index return annual mean (μ) very very very very precisely, with a very very very very long sample, the Arithmetic estimator of long-term return would be efficient and the Geometric estimator would always be biased downward. True or False

• $E(A) = E(V_H) \underbrace{e^{\frac{1}{2}\sigma^2 \frac{H^2}{T}}}$ BIAS

as $T \nearrow \infty$ or $\sigma \downarrow 0$ the bias goes to zero. We are closer to asymptotic situation and A which is MLE is efficient.

Then, as $G < A$, G is biased downward

g) Returns are i.i.d lognormal (μ, σ). You estimated μ with T=60 years of data. Write (no proof) the Geometric and Arithmetic estimators of the long-run expected (compound) return over H future years. For what horizon H is the Geometric estimator also the minimum MSE estimator?

• $A = e^{H(\hat{\mu} + \frac{1}{2}\sigma^2)}$

• $G = e^{H\hat{\mu}}$

• $MMSE = e^{H(\hat{\mu} + \frac{K}{2}\sigma^2)}$

with $K = 1 - \frac{3H}{T}$

$K = 0$ when $H = \frac{T}{3}$

Problem 2: 24 points
Consider the regression $Y = X\beta + \epsilon$, with $\epsilon \sim N(0, \Omega)$, where Ω is a **known** covariance matris. You collect T observations of X and Y. Since Ω is known, you can do everything *given* Ω, no need to estimate it.

a) 4 points Write (no proof) the joint density of the error vector $p(\epsilon)$

$$p(\varepsilon) = \frac{1}{\sqrt{2\pi}^{T} |\Omega|^{T/2}} \exp\left\{-\frac{1}{2} \varepsilon' \Omega^{-1} \varepsilon\right\}$$

b) 6 points Given $p(\epsilon)$, write the joint density of the data $p(Y)$. Write exactly the known result you use to go from ϵ to Y, and show how you apply it.

- $Y = X\beta + \varepsilon$ • Need to use change of variable

$$P_Y(Y) = P_X(\varepsilon(Y)) \left|\frac{\partial \varepsilon}{\partial Y}\right|$$

$$\left|\frac{\partial \varepsilon}{\partial Y}\right| = 1 \qquad p(Y) = \frac{1}{\sqrt{2\pi}^{T} |\Omega|^{T/2}} \exp\left\{\frac{-1}{2}(Y - X\beta)' \Omega^{-1} (Y - X\beta)\right\}$$

c) 6 points Find $\widehat{\beta}_{MLE}$

$$\text{Log}\,\ell = -\frac{1}{2}(Y - X\beta)' \Omega^{-1} (Y - X\beta) - xxx$$

$$\frac{\partial \text{Log}\,\ell}{\partial \beta} = \frac{\partial}{\partial \beta}\left[\frac{-1}{2}\left[\cancel{2}\beta' X' \Omega^{-1} Y + \beta' X' \Omega^{-1} X \beta + Y' \Omega^{-1} Y\right]\right]$$

$$= X' \Omega^{-1} Y - X' \Omega^{-1} X \beta \qquad = 0$$

$$\boxed{\widehat{\beta}_{MLE} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y}$$

d) 8 points Write the general result for the asymptotic covariance matrix of an MLE estimator $\widehat{\theta}$ of a parameter θ. Use it to find the variance of $\hat{\beta}_{MLE}$

Take second derivative

$$\frac{\partial^2 \text{Log} P}{\partial \beta} = - X' \Omega^{-1} X$$

- $$V(\hat{\beta}) = - \left[E\left(\frac{\partial^2 \text{Log} P}{\partial \beta} \right)^{-1} \right]$$ general result

- $$V(\hat{\beta}) = - - (X' \Omega^{-1} X)^{-1} = \boxed{(X' \Omega^{-1} X)^{-1}}$$

Problem 3: 18 points

Your assistant, an intern from the Questrom MSMF program, estimated a two-regression SUR system, of a small and a large firm portfolio return on the market return. You want to test some hypotheses on small vs large firms betas. Each regression noise taken separately is i.i.d normal.. Here are some of the results.

Estimates	S(mall)	L(arge)
β	1.06	0.98
$s(\beta)$	0.03	0.02

The intern did not print the covariance matrix of the residuals or the market variance. (What do they teach at Questrom, you will talk to the director about that !)

But he reports that the correlation between $\widehat{\beta_S}$ and $\widehat{\beta_L}$ is **0.5**. You will have to trust him on that!

a) 6 points Write the covariance matrix of the estimates of the vector (β_S, β_L). Then compute its inverse. (numbers)

$$C(\hat{\beta}) = \begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix} \qquad V(\hat{\beta}) = \begin{pmatrix} .03 & 0 \\ 0 & .02 \end{pmatrix}\begin{pmatrix} 1 & .5 \\ .5 & 1 \end{pmatrix}\begin{pmatrix} .03 & 0 \\ 0 & .02 \end{pmatrix}$$

$$= \begin{pmatrix} .03^2 & .02 \times .03 \times .5 \\ .02 \times .03 \times .5 & .02^2 \end{pmatrix} = \begin{pmatrix} .03^2 & .0003 \\ .0003 & .02^2 \end{pmatrix}$$

$$V(\hat{\beta})^{-1} = \frac{1}{.03^2\, .02^2 - .0003^2}\begin{pmatrix} .02^2 & -.0003 \\ -.0003 & .03^2 \end{pmatrix} = \begin{pmatrix} 1481 & -1111 \\ -1111 & 3333 \end{pmatrix}$$

b) 6 points Write the theoretical (letters) formula of the Wald test of $H_0 : \beta_S = \beta_L = 1$. Then compute the test value(numbers) using your numerical inputs.

$$W = (\hat{\beta} - \beta_0)' \, V(\hat{\beta})^{-1} (\hat{\beta} - \beta_0)$$

$$= (.06, -.02)\begin{pmatrix} 1481 & -1111 \\ -1111 & 3333 \end{pmatrix}\begin{pmatrix} .06 \\ -.02 \end{pmatrix} = (.06, -.02)\begin{pmatrix} 66.6 \\ -133 \end{pmatrix}$$

$$\boxed{= 6.6}$$

c) **6 points** What is the asymptotic distribution of your test (no proof)? Use the attached table to conclude at the 5% significance level.

$W \overset{\bullet}{\sim} \chi^2(2)$ $\chi^2_{.95}(2) = 6$

$6.6 > 6$ Reject at 5%

Problem 4: 24 points

Your assistant (that intern again !) estimated several market regressions, assuming (it's OK) i.i.d normally distributed noise for each. He tested the null hypothesis $H_0 : \beta = 1$ at the 5% level. Given the large sample size, it's OK to assume that the t-stat is normally distributed and use a 1.96 cutoff.

Estimates	P_1	P_2	P_3	P_4	P_5
β	0.920	0.930	1.030	1.051	1.104
$s(\beta)$	0.040	0.035	0.025	0.030	0.042
t-statistic	-2.0*	-2.0*	1.2	1.7	2.48*

a) **8 points** For question a), assume all these regressions are **truly** unrelated. Under the null that all βs are equal to 1, what is the probability of finding at least one significant t-stat at the 5% size when searching from 5? What is the actual size of such the test that data mines for a significant t-stat by looking at 5 of them?

If regressions are unrelated, $\hat{\beta}$ and tests of $\hat{\beta}$ are unrelated

$\Pr(\text{at least 1 reject}) = 1 - P(\text{No reject})$

$= 1 - .95^5 = 0.22$

That IS the actual size of the data mining test

b) 8 points Define the Bonferroni correction for data mining? How many of the 5 t-stats are now significant if you apply this correction? What can you say of the size of the data-mining test when corrected by Bonferroni?

- The correction: to get a mining test of size $\leq .05$ we need to use a size $\frac{.05}{m} = \frac{.05}{1} = .01$ on each individual test
- The $t_{.01/2} = -2.58$ $t_{.995} = 2.58$

We reject NO NULL

- The size of the Bonferroni test is $\leq .05$ not sure that it is exactly equal

c) 8 points Consider the first test on P_1 by itself, no data mining issues. Write the probability of rejecting the null, under the alternative: $H_1 : \beta_1 = 1.1$ Use standard R notation pnorm(x, mu, sigma) to write the needed CDFs (no numerical result needed).

Under H_1 $\hat{\beta}_1 \sim N(1.1, s_{\hat{\beta}} = .04)$

Power $= Pr(\hat{\beta}_1 < 1 - 1.96 \times 0.04)$
$+ Pr(\hat{\beta}_1 > 1 + 1.96 \times 0.04)$

$= \text{pnorm}(\underset{x}{1 - 1.96 \times 0.04}, \underset{\mu}{1.1}, \underset{\sigma}{0.04})$
$+ 1 - \text{pnorm}(1 + 1.96 \times 0.04, 1.1, 0.04)$