Instructions:
Be verbose. Explain clearly your reasoning, methods, and results in your written work.

No code is necessary but including it in your answer could result in partial credit.

Written answers are worth the amounts stated.

Total points available on this exam are 110. That means there are 10 points of extra credit available.

1. Answers should be formatted as a PDF. You may convert your Python notebook, if you use one, directly to PDF.
2. Restate the question along with the question number before each answer
3. When finished, upload your PDF to Canvas along with your code (optional).
4. Do not check code or answers into your repository until after the exam is completed by all.

Data for problems are available in the repository named by the question number.

You may use your notes and the internet for **coding syntax help only**. You may not use an LLM – turn off any coding helper you may have installed. You may not work with other students – all work must be your own.

All students will be held to the Duke Community Standard

Duke's Community Standard:

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity.
To uphold the Duke Community Standard:

1. I will not lie, cheat, or steal in my academic endeavors,
2. I will conduct myself honorably in all my endeavors; and
3. I will act if the Standard is compromised.

1. (10 pts) Explain the difference in thinking between data modeling for risk analysis vs data modeling for forecasting.

2. (20 pts) Using problem2.csv
    a. Calculate the Mean, Variance, Skewness and Kurtosis of the data (8)
    b. Given a choice between a normal distribution and a t-distribution, which one would you choose to model the data and why based on part a alone (4)?
    c. Fit both distributions and prove or disprove your choice in b. (8)

3. (20 pts) Using problem2.csv and your fitted models. These are returns of a stock
    a. Calculate the VaR (5% alpha) as distance from 0 for both models (8)
    b. Calculate the ES (5% alpha) as the distance from 0 for both models (8)
    c. Discuss the results. What do you notice? Why is that? (4)

4. (10 pts) Using problem4.csv
    a. Calculate the exponentially weighted correlation matrix with lambda = 0.94 (3)
    b. Calculate the exponentially weighted variances with lambda=0.97 (3)
    c. Combine A and B to form a covariance matrix (3)
    d. Why would you do something like this in practice? (1)

5. (30 pts) Using the data in problem5.csv. These data contain missing values
    a. Calculate the pairwise covariance of the data (10)
    b. Is your matrix Positive Definite, Positive Semi-definite, or Non Definite? (5)
    c. If the matrix is non definite, use Higham's method to fix the matrix. (10)
    d. For each principal component, list the variance explained and the cumulative variance explained, sorted from largest to smallest variance explained. (5)

6. (20 pts) Using problem6.csv. These data are prices 3 stocks. You own 100 shares of each stock. Using arithmetic returns:
    a. De-mean the return series so that the mean of each is 0. Fit a Student T model for each stock. Report the fit values. (3)
    b. Simulate the system using a Gaussian Copula. Report the correlation matrix you used in the copula. (3)
    c. What is the VaR and ES at the 5% alpha level for each stock expressed in $? (7)
    d. What is the VaR and ES at the 5% alpha level for the total portfolio expressed in $? (7)