Instructions:
Be verbose. Explain clearly your reasoning, methods, and results in your written work. Write clear code that is well documented. With 99% certainty, you cannot write too many code comments.

Written answers are worth 18 points. Code is worth 2 points. 10 points total.

1. When finished, respond to the question in Canvas as “done.” We will record your grade there.
2. In your code repository, create a folder called “Project01.”
3. In that folder, include
    a. a document (PDF) with your responses.
    b. All code
    c. A README file with instructions for us to run your code

Everything must be checked into your repository by 8am Sunday 2/1. A pull will be done at that time. Documents and code checked in after the instructors pull will not be graded.

Data for problems can be found in CSV files with this document in the class repository.

Problem 1
Given the dataset in problem1.csv
A. Calculate the Mean, Variance, Skewness and Kurtosis of the data
B. Given a choice between a Normal Distribution and a T-Distribution, which one would you choose to model the data? Why?
C. Fit both distributions and prove or disprove your choice in B using methods presented in class.

Problem 2
Given the data in problem2.csv
A. Calculate the pairwise covariance matrix of the data.
B. Is the Matrix at least positive semi-definite? Why?
C. If not, find the nearest positive semi-definite matrix using Higham’s method and the near-psd method of Rebenato and Jackel.
D. Calculate the covariance matrix using only overlapping data.
E. Compare the results of the covariance matrices in C and D. Explain the differences. Note: the generating process is a covariance matrix with 1 on the diagonals and 0.99 elsewhere.

Problem 3

Given the data in problem3.csv

A. Fit a multivariate normal to the data.
B. Given that fit, what is the distribution of X2 given X1=0.6. Use the 2 methods described in class.
C. Given the properties of the Cholesky Root, create a simulation that proves your distribution of X2 | X1=0.6 is correct.

Problem 4

Given the data in problem4.csv

A. Simulate an MA(1), MA(2), and MA(3) process and graph the ACF and PACF of each. What do you notice?
B. Simulate an AR(1), AR(2), and AR(3) process and graph the ACF and PACF of each. What do you notice?
C. Examine the data in problem4.csv. What AR/MA process would you use to model the data? Why?
D. Fit the model of your choice in C along with other AR/MA models. Compare the AICc of each. What is the best fit?

Problem 5

Given the stock return data in DailyReturns.csv.

A. Create a routine for calculating an exponentially weighted covariance matrix. If you have a package that calculates it for you, verify it produces the expected results from the testdata folder.
B. Vary λ. Use PCA and plot the cumulative variance explained of λ in (0,1) by each eigenvalue for each λ chosen.
C. What does this tell us about the values of λ and the effect it has on the covariance matrix?

Problem 6

Implement a multivariate normal simulation using the Cholesky root of a covariance matrix.
Implement a multivariate normal simulation using PCA with percent explained as an input.

Using the covariance matrix found in problem6.csv

A. Simulate 10,000 draws using the Cholesky Root method.
B. Simulate 10,000 draws using PCA with 75% variance
C. Take the covariance of each simulation. Compare the Frobenius norm of these matrices to the original covariance matrix. What do you notice?
D. Compare the cumulative variance explained by each eigenvalue of the 2 simulated covariance matrices along with the input matrix. What do you notice?
E. Compare the time it took to run both simulations.
F. Discuss the tradeoffs between the two methods.