Homework 7

1.
Download dataset from the following case study:
PROBLEM 1: 10-fold Cross-validation (10 in-sample data and 10 out-of-sample data) for Standard Deviation minimization
http://uryasev.ams.stonybrook.edu/index.php/research/testproblems/financial_engineering/mortgage-pipeline-hedging/ .
The provided code and data is for 10-fold cross validation with the whole dataset. Modify the code and input data to perform 5-fold cross-validation.

2.
In the downloaded dataset pick 2 factors (out of 3 factors). Students with last name starting with letters A, B, C, D, E, F, G should pick variables x1, x2; starting with H, I, J, K, L, M, N, O, P, Q, R should pick variables x2, x3; starting with S, T, U, V, W, X, Y, Z should pick variables x1, x3.

Replace standard deviation by CVaR deviation with confidence level 0.90 in PROBLEM 1 (function: cvar_dev(0.90, matrix_...).

I. Randomly select 100 rows from the 1000 rows as testing set
II. Randomly select 3 rows from the remaining 900 rows as training set
III. Fit the two-factor model and three-factor model with the training set; obtain in-sample CVaR deviation
IV. Calculate the out-of-sample CVaR deviation of the two models on the testing set
V. Repeat steps I to IV at least 20 times and calculate the average of CVaR deviations for both in-sample and out-of-sample values
VI. Repeat steps I to V with different number of rows in step II; the suggested numbers are 20,40,60…,880,900.

3.
Build a graph using the results computed in the previous question similar to “Figure 4: Performance of models with different flexibility” in Module 7 Notes.

4.
Find the break-even point of the number of rows in the design matrix when you should switch to a more flexible model. Provide a succinct report containing the graph and explanations. Submit the code and numerical results.