Problem set # 4

Due: Wednesday, February 19, by 8 am

Chang Deng U71725017 dengc@bu.edu

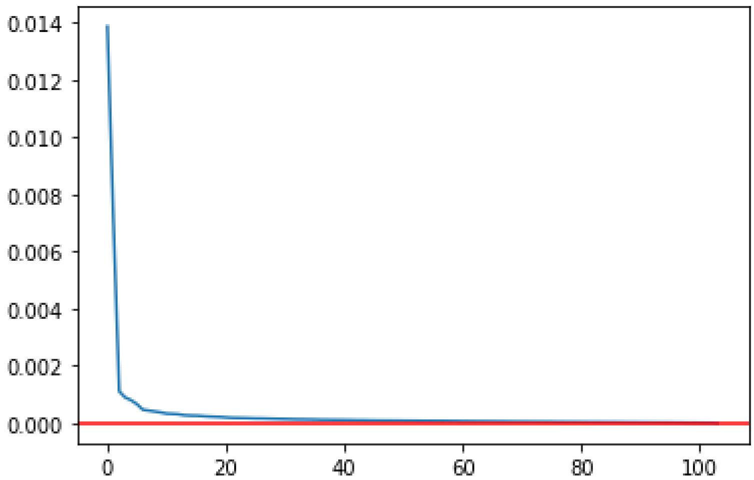
Problem 1: Covariance Matrix Decomposition: Download historical data from your favorite source for 5 years and at least 100 companies or ETFs. In this problem we will look at the covariance matrix for these assets and its properties.

1. Clean the data so that your input pricing matrix is as full as possible. Fill in any gaps using a reasonable method of your choice. Explain why you chose that particular method.

   By filling gaps of each column using the mean of each column, I clean the data to make the input pricing matrix as a full matrix. The reason why I chose that method is because filling gaps with the mean of the corresponding columns may reduce the effect of filling on the result of covariance matrix.

2. Generate a sequence of daily returns for each asset for each date.

3. Calculate the covariance matrix of daily returns and perform an eigenvalue decomposition on the matrix. How many positive eigenvalues are there? How many were negative? If any are negative, what happened?

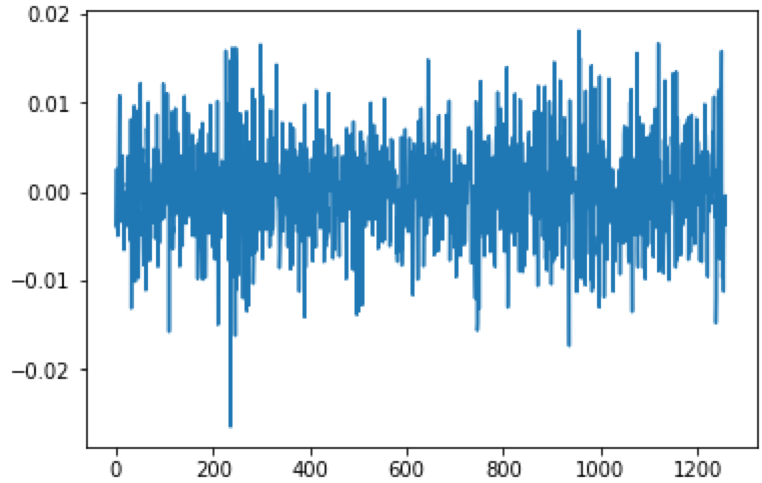
   

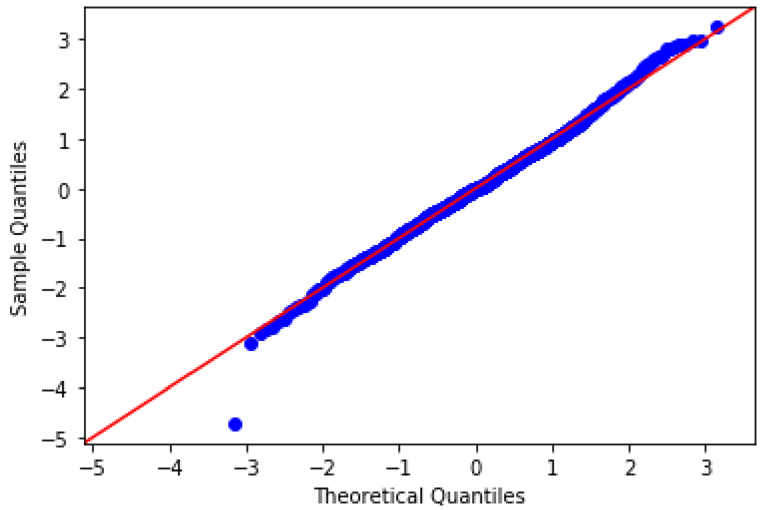
   As shown above, every eigenvalues is positive, no eigenvalue is negative, since the covariance matrix should be positive semi definite or even positive finite if it's full rank . If there is any negative eigenvalue, either the process of calculation or the data may go wrong.

4. How many eigenvalues are required to account for 50% of the variance? How about 90%? Does this make sense to you?

   In my case, **2** eigenvalues are required to account for 50% of the variance, and **42** eigenvalues are required to account for 90% of the variance. It makes sense to me, as the underlying stock return are diverse and one or two eigenvalues are significant, seldom eigenvalues can account for 50% of the variance but the number increases obviously when it comes to 90% of the variance.

5. Using the number of eigenvalues in the 90% threshold above, create a return stream that represents the residual return after the principal components that correspond to these eigenvalues have been removed. Plot this return stream and comment on its properties.



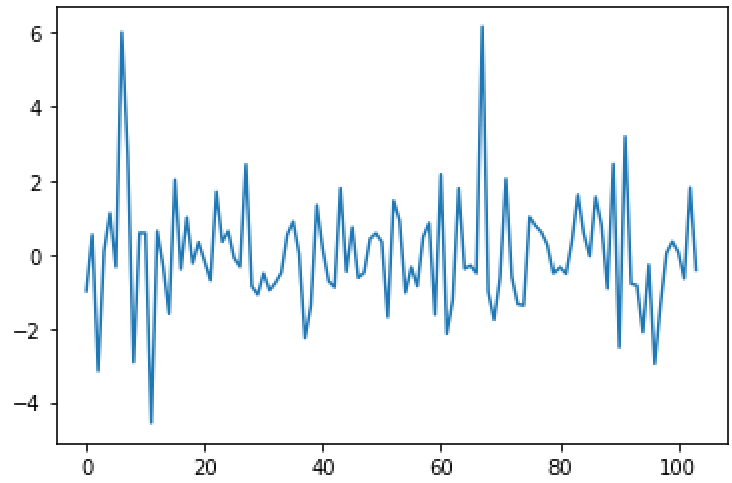


The return stream are shown above, since we choose the number of eigenvalues in the 90% threshold, we construct components which can account for most of the daily return. So the residual return almost exhibit a normal distribution with mean zero.

Problem 2: Portfolio Construction: In Lecture 7, we defined a Lagrangian for portfolio with constraints in matrix form by

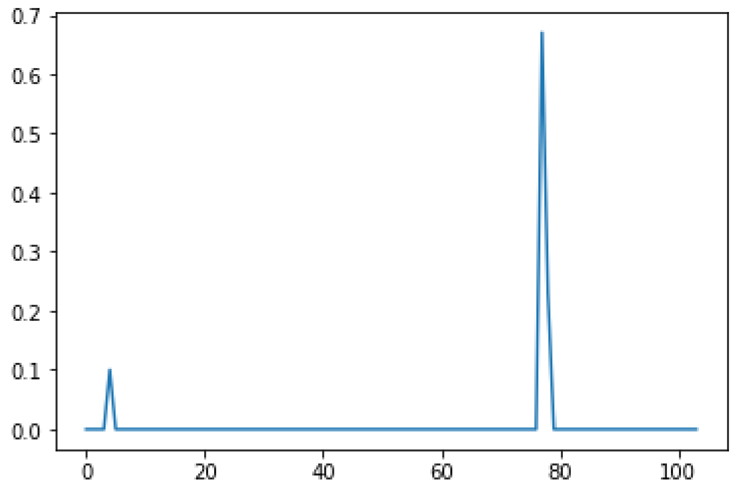
$$L(w, \lambda) = \langle R, w \rangle - a\langle w, Cw \rangle - \langle \lambda, Gw - c \rangle$$

1. Form the matrix G by imposing the budget constraint, which is , and another constraint that allocates 10% of the portfolio to the rest 17 securities (to simulate sector allocation). Using C from Problem 1, use your favorite method and the software package of your choice to invert in a nice, stable way. (Hint: consider my favorite method).



2. What does the resulting portfolio look like? Would it be acceptable to most mutual funds? If not, what would you to do to fix that?

   As shown above, the result weight is fluctuating between -4 to 6. It is not acceptable to most mutual funds since much of the weight is negative and usually mutual funds will not put 100% weight on equity. To fix the first issue, I set non-negative constraint on the weight, then I have the following result:



For the second issue, more assets such as bond shall be introduced to share the total weight.