FSRM 588: Financial Data Mining Homework 4: Portfolio Management and Chapter 5

Fall 2015

Due 6:40pm on November 16

NOTES: Submit all your code with your assignment. Homework must be neatly written-up or typed for submission. Please read and follow the instructions carefully. I reserve the right to refuse homework that is deemed (by me) to be excessively messy.

Instructions:

- HW4 is intended to complement the materials in Chapter 5 and Lecture 8 and help you familiarize with all of the algorithms introduced in class.
- HW4 contains only one part: computer exercises. Your answers MUST be typed up and printed out. Submit the answers in class.
- Use set.seed(1) every time you start a new CV. We consider K-fold CV with $K = 5$.
- Submit the R code you use to the email address fsrm588@gmail.com as an attachment before class. The subject of the email should be "hw4 + your last name + your first name". For instance, "hw4 Yang Dan". Only .r files are allowed and the title of the attached document should be "hw4 + your last name + your first name.r". For instance, "hw4YangDan.r"
- The content of the email can be empty. There is no need to say thank me and I won't reply to the emails sent to fsrm588@gmail.com. fsrm588@gmail.com is only used for the purpose of submission of homework, project, and exam. Please reach me via dyang@stat.rutgers.edu if you have questions or need help.

PART I: Computer

1. This problem uses the `r6306_new.txt` dataset, which contains monthly returns of 2671 stocks from Jan 1963 to Dec 2006. The file `s&p6306.txt` gives the monthly return of the S&P500 index over this same time period. Each `PERMNO` in the `r6306_new.txt` corresponds to a unique stock from the CRSP database.

   (a) Repeat the following procedure for every year from $i = 1973$ to $i = 2005$.

   (1) At the end of April of year i, look at the collection of stocks (`PERMNO`'s) for which complete data from the previous 120 months *and* the following 12 months are available. Randomly select $p = 60$ stocks from this collection.

   (2) At the end of April of year i, use monthly data from the past 120 months to estimate the covariance matrix for the returns on the p stocks you selected in step (1). You should use each of the following 4 methods to estimate the covariance matrix (so you will have 4 estimated covariance matrices at the end of this step).

   (i) SAM: The sample covariance matrix (discussed in class).

   (ii) 1FA: One-factor model using the returns of the S&P500 index as the factor (discussed in class).

   (iii) 2FA: Statistical two-factor model with unobserved factors (discussed in class; use PCA to obtain this estimator).

(iv) IND: Independence model, where the diagonal entries of $\hat{\Sigma}$ are given by the sample variances of the returns and the off-diagonal entries are all set equal to 0.

(3) Using each of the covariance matrices computed in step (2), find the global minimum variance portfolio (GMVP) and the GMVP with no-short-selling (NSS). If your estimated covariance matrix $\hat{\Sigma}$ is non-invertible, replace $\hat{\Sigma}$ with $0.99 \cdot \hat{\Sigma} + 0.01 \cdot \text{diag}(\hat{\Sigma})$.

(4) Using the data from the 12 months following April of year i, compute the monthly returns from each portfolio found in step (3); then compute the variance of these returns. Also compute the variance of the returns from the equally weighted portfolio (EW) and the variance of the returns from the S&P500 (SP) over this same 12 month period.

Use a boxplot to display the portfolio variances for each of the portfolios from years $i = 1973$ to $i = 2005$. Which portfolio has the smallest variance, on average? Discuss your results.

(b) Repeat part (a) with $p = 120$, 250 and 500 stocks.

(c) Repeat parts (a)–(b) using the thresholded covariance matrix estimator $\hat{\Sigma}_T = (\hat{\sigma}_{ij} I\{|\hat{\sigma}_{ij}| \geq T\})$, where $\hat{\sigma}_{ij}$ is the ij-entry from the sample covariance matrix and T is the threshold level, in step 2 of the analysis. Choose the threshold level T by five fold cross-validation.

2. Read Section 5.2.2 and reproduce Figure 5.4 in the textbook. There is no need to shade the region, simply generate the standard error band.

3. Read Section 5.4.1 and reproduce Figures 5.7 and 5.8 in the textbook. Download the data from the website for the textbook.