Data instruction for "Portfolio Selection with Nonparametric Value-at-Risk: A Block Coordinate Descent Method"

June 28, 2017

All the data used in "Portfolio selection with nonparametric value-at-risk: A block coordinate descent method" can be downloaded from

http://www.se.cuhk.edu.hk/∼dli

under the item of 'Data set for "Portfolio optimization with nonparametric Value-at-Risk: A block coordinate descent method".'

This document indicates where readers can obtain the actual data values that were used in our paper, and explains how the data used in the computational experiments of our paper were created.

1. The data can be downloaded via a Thomson Reuters Eikon terminal, by searching "LS&Pcomp", which stands for S&P500 index constituents, setting the datatype as "price" and choosing the time period as follows for "daily" or "weekly" data. The daily historical price data we used in Sections 4 and 5 are from Nov. 19, 2004 to Nov. 8, 2012 and the weekly historical data we used in Section 6 are from Jan. 2, 2004 to Jan. 21, 2013. For a user guide of the Thomson Reuters Eikon database, please refer to the link of "http://libguides.lib.cuhk.edu.hk/c.php?g=500666&p=3530218".

Remark: The S&P 500 index generally carries out an annual reconstitution in its constituents. The constituent prices will be adjusted if a company takes a share split strategy. Thus, the downloaded price data of constituents in S&P 500 index could be slightly different from time to time.

2. In Section 4, we use 2039 daily returns of constituents of S&P500 index to randomly generate instances of models (P_m) and (P_c).

- Files "daily-price1.xls" - "daily-price8.xls" are the price files downloaded from Thomson Reuter database;
- Matlab code "data_input_day.m" is used to calculate the daily returns from the price files;

- All 2039 calculated return data are stored in file "Daily-return-in-Simulation.xlsx";
- Matlab code "random_instance.m" is used to randomly choose n constituents and N return samples for every instance;
- The random sample matrices in the instances in Tables 1-3 in Section 4 are recorded in instance files. Every instance file contains a return matrix with n rows and N columns which is the transpose of matrix R in (P_m) and (P_c), where n is the constituent number and N is the number of return samples.

3. In Section 5, to analyze the approximation effect of different VaR estimators, we consider market scenarios following two return distributions: (1) Multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ and (2) Mixed distribution defined as

$$(1 - I(\epsilon))\mathcal{N}(\boldsymbol{\mu}, \Sigma) + I(\epsilon)(Y\boldsymbol{e} + \boldsymbol{f}), \tag{1}$$

where $I(\epsilon)$ is a Bernoulli random variable with parameter ϵ, $\boldsymbol{e}$ is the n-dimensional all-one vector, $\boldsymbol{f}$ is a constant vector and Y is nonpositive exponential random variable with density

$$p(Y = y) = \begin{cases} \delta e^{\delta y}, & \text{if } y \leq 0, \\ 0, & \text{else,} \end{cases}$$

where we set $\epsilon = 0.05$, $\delta = 0.01$ and constant vector $\boldsymbol{f} = \boldsymbol{\mu} - \text{diag}(\Sigma)$.

- The parameters $\boldsymbol{\mu}$ and Σ in these distributions are estimated using the price data in file "Daily-return-in-Simulation.xlsx";
- The estimated parameters $\boldsymbol{\mu}$ and Σ are saved in files "mu.txt" and "sigma.txt".

4. In Section 6, we study the in-sample performance of portfolios generated from models (M_1) to (M_4) under the assumption of distribution (1).

- Files "week-price1.xls" - "week-price3.xls" are the weekly price files downloaded from Thomson Reuter database;
- The 473 weekly returns are calculated from price files "week-price1.xls" - "week-price3.xls" via MATLAB code "data_input_week.m";
- The parameters $\boldsymbol{\mu}$ and Σ in (1) are estimated using the 473 weekly return data of constituents in file "Week-return-Insample.xlsx",

- The estimated parameters $\boldsymbol{\mu}$ and Σ in this subsection are saved in "mu0.txt" and "sigma0.txt".

5. In out-of-sample analysis in Section 6, we use backtesting strategy to test the practical performance of portfolios from models (M_1) to (M_4) using real data samples. All 252 weekly returns required in this part are saved in file "Week-return-Outsample.xlsx". We use the first 200 weekly returns ("Week return 1" – "Week return 200" in the file) to generate portfolios, and calculate the portfolio values using the 201st weekly return ("Week return 201" in the file). Then we calculate new portfolios using updated 200 returns ("Week return 2" – "Week return 201" in the file), and evaluate the portfolio values using the 202nd return ("Week return 202" in the file). Repeat this procedure until the last week in out-of-sample period.