MIT Sloan School of Management

Advanced Analytics of Finance | 15.457
Hui Chen | Spring 2020

Final Project: Portfolio Optimization Hackathon

Instructions

- This hackathon starts at 5:00 pm EDT, Tuesday, April 28 and ends at 11:00 pm EDT, Thursday, May 7. Please submit your report and code on Canvas.
- You can work individually or in a team of up to 4 students on this project. Do not consult with anyone outside of your team about it.
- Prepare your report in a format that is concise and easy to read. Clearly state the methodology and highlight the key results. The report should be typed, double-spaced, with 12 point font and standard margin (at least 1.2 inches on all sides), and **no more than 3 pages long** (tables and graphs included). **No hand-written reports will be accepted.**
- **Honor code:** The final report should acknowledge clearly that "the team has not consulted with anyone outside the team or incorporated the work of others in this project" and be signed by each team member.

Project Description

It is your first day as the portfolio manager at A^{2}oF Capital. Your goal is to construct a quantitative portfolio strategy that delivers the best out-of-sample performance based on a mean-variance criterion. Specifically, the investment objective is:

$$U_t = \max_{w_t} E_t[r_{p,t+1}] - \gamma Var_t[r_{p,t+1}], \quad r_{p,t+1} = w_t' \mathbf{r}_{t+1} \tag{1}$$

where $\mathbf{r}_{t+1}$ is the vector of stock returns, w_t is the portfolio weight vector which satisfies $w_t'\mathbf{1} = 1$ (i.e., your portfolio should always be fully invested in stocks), and $\gamma \geq 0$ is the coefficient of risk aversion. For this project, assume $\gamma = 5$.

The analysts on your team have already finished stock selection. They have identified a list of 500 stocks for you to invest in. You need to build a portfolio with these stocks, which you are allowed to rebalance at the end of each month.

To solve for the optimal portfolio weights w_t^* in (1) (the solution is the tangent portfolio on the mean-variance frontier with slope γ), you need to first obtain estimates for the conditional mean and covariance of returns for the 500 stocks. Given your expertise, you came up with the idea of estimating a factor model of returns and then use it to estimate the expected returns and covariance matrix.

Specifically, suppose there is a multifactor asset pricing model for excess returns r_{it}:

$$r_{it} = \alpha_i + \beta_{i1} f_{1t} + \cdots + \beta_{ik} f_{kt} + \epsilon_{it}. \tag{2}$$

Assume ϵ_{it} in Eq. (2) are IID, independent of the factors as well as across stocks, i.e., $E_t(\epsilon_{t+1}) = 0$ and $var_t(\epsilon_{t+1}) = \Omega$ where Ω is diagonal. Then, the returns at time $t+1$ have conditional mean

$$E_t[r_{i,t+1}] = \alpha_i + \beta_{i1} E_t[f_{1,t+1}] + \cdots + \beta_{ik} E_t[f_{k,t+1}], \tag{3}$$

and conditional covariance

$$cov_t[\mathbf{r}_{t+1}] = \beta_i cov_t[\mathbf{f}_{t+1}] \beta_i' + \Omega. \tag{4}$$

Goal of the Hackathon: Your job is to find a way to produce the forecasts of $E_t[r_{i,t+1}]$ and $cov_t[\mathbf{r}_{t+1}]$ that will result in the best out-of-sample performance for your portfolio. To do so, you need to decide what factors to include in (2), as well as how to estimate the coefficients α, β, Ω. This step should be done using the training sample alone.

To implement (3)-(4), you will also need to estimate the conditional mean and covariance of the factors, $E_t[f_{1,t+1}]$ and $cov_t[\mathbf{f}_{t+1}]$. You should build the forecasting model for $E_t[f_{1,t+1}]$ and $cov_t[\mathbf{f}_{t+1}]$ using the training sample. You are allowed to update these forecasts each month in the testing period.

Data

- The training data can be downloaded at https://bit.ly/2SdLGpM. It contains two data files, "StockReturns.csv" and "Factors.csv". The description file "DataDescription.pdf" explains the variables included. No outside data are allowed.
- The testing data (not available now) will also contain two data files with exactly the same format as in the training set. The file "Factors.csv" will include data from 1989 to 2018 (the full sample). The sample period covered by the file "StockReturns.csv" will be post-2012; the exact period will be revealed later.

Report and Code Your final output should include two parts:

1. A report (no more than 3 pages) summarizing how your model is developed, the data used, and the analysis you have done for model selection.
2. The code (in R or Python). Your code should takes two data files, "StockReturns.csv" and "Factors.csv", as input, and produces two outputs:
   (a) The overall performance of your model in the test sample, determined by
   $$U_p = \overline{r}_p - \gamma \overline{\sigma}_p^2, \tag{5}$$
   where $\overline{r}_p$ and $\overline{\sigma}_p^2$ are the sample mean and variance of your portfolio's monthly returns in the testing period.
   (b) The monthly time series of your portfolio weights.