

## MIE 1624 Introduction to Data Science and Analytics – Winter 2022

### Assignment 3

Due Date: 11:59pm, March 30, 2022

Submit via Quercus

#### Background:

Portfolio optimization is the process of determining the allocation of funds to a set of assets which yields the maximum performance. Performance can be thought of as maximizing return, minimizing risk, or maximizing the risk adjusted return (Sharpe ratio).

Markowitz (1959) mean-variance optimization (MVO) has been the standard for efficient portfolio construction for almost fifty years. MVO lowers the cumulative risk of a portfolio by blending anti-correlated data since equities which are positively correlated in the long term but anti-correlated in the short term cumulatively have equal returns but lower volatility. For a given set of equities, computing the MVO portfolio for the range of possible portfolio returns the efficient frontier. Thus, the efficient frontier is the set of optimal portfolios that offer the highest expected return for a defined level of risk or the lowest risk for a given level of expected return.

Computing the Mean-variance efficient frontier:

Inputs:

- $\mu$  - expected mean of returns for each asset
- $Q$  - Covariance matrix
- $\varepsilon$  - target return
- $R_{\min\text{Var}}$  - return of minimum variance portfolio
- $R_{\max\text{Ret}}$  - return of maximum return portfolio

$$\begin{aligned} \min_w \quad & w^T Q w \\ \text{s. t.} \quad & \mu^T w \geq \varepsilon \\ & \sum_i w_i = 1 \\ & w \geq 0 \\ & \varepsilon \in [\varepsilon_{R_{\min\text{Var}}}, \varepsilon_{R_{\max\text{Ret}}}] \end{aligned}$$

Output:

- $w$  - weights or investments for each stock

The drawback to conducting MVO on its own is that the algorithm is highly affected by sampling error in the input data. As well as being sensitive to uncertainty in risk-return estimates.

Resampling efficiency optimization and rebalancing technology (RE), developed by Michaud in 1998, introduces Monte Carlo resampling and bootstrapping methods into MV optimization. The portfolio built after the resampling method is applied is generally more stable, realistic, and investment efficient than MVO alone. Each resampled dataset will create a new set of Markowitz efficient portfolios to form a new efficient frontier. The resampled efficient frontier lies below the classical MV frontier. RE optimization expects less return and restricts risk to a narrower range. RE optimization can be thought of as a constraint which lowers in-sample expectations and improves out-of-sample performance.

### **Assignment:**

Given a data set of monthly closing prices for 25 assets from Jan 1st, 2018, to Dec 1st, 2020 (in-sample) and closing prices for the same assets in 2021 (out-of-sample), complete the following:

1. Load and clean data, in-sample data (2018-2020) and out-of-sample data (2021). From the in-sample data, compute the classical Markowitz and resampled Markowitz frontiers. **[45pts]**
  - a) Import necessary libraries and dataset, compute the matrix of monthly returns **[2.5pts]**
  - b) Compute the covariance matrix and means of the percent change matrix **[2.5pts]**
  - c) Solve for the minimum variance portfolio **[5pts]**
  - d) Solve for the maximum variance portfolio **[5pts]**
  - e) Compute the classical Markowitz efficient frontier **[10pts]**
  - f) Compute the resampled efficient frontier **[20pts]**
2. Using the frontiers, compute three portfolio weights (minimum variance, maximum return, and Sharpe ratio) along both the classical and resampled frontiers. Compare the performance of each of these portfolios on the out-of-sample data. Assume a total investment of \$1000 and a risk-free-return of 0.2% **monthly** (0.002). **[15pts]**
  - a) On the classical efficient frontier, find the minimum variance, maximum return, and Sharpe ratio portfolio weights **[5pts]**
  - b) On the resampled efficient frontier, find the minimum variance, maximum return, and Sharpe ratio portfolio weights **[5pts]**
  - c) For each of the 6 portfolios, calculate the expected return on the **2022** (out-of-sample) data for a \$1000 investment **[5pts]**
3. Plot the classical and resampled efficient frontiers **[10pts]**
4. Plot the portfolio compositions for classical and resampled frontiers. **[15pts]**
5. Discuss your findings from comparing the performance of all 6 portfolios and from observing the graphs from questions 3 & 4. **[15pts]**

### **Learning objectives:**

1. Learn how to implement optimization techniques to financial portfolio performance.
2. Understand how simulation techniques are implemented and their effect on computing optimal results.
3. Improve on abilities to develop, understand, and explain graphical data in the domain of financial portfolio construction

### **Submission:**

1. Produce an IPython Notebook (.ipynb file) detailing the analysis you performed to answer the questions for the given data set.
2. Produce a 2-page report containing the graphs for questions 3 & 4 and giving your response to question 5.

### Tools:

- **Software:**
  - Python Version 3.X is required for this assignment. Python Version 2.7 is not allowed.
  - Your code should run on the Google Colab cloud.
  - All libraries and built-ins are allowed but here is a list of the major libraries you might consider: Numpy, Pandas, Matplotlib, Cplex, Cvxpy.
  - No other tool or software besides Python and its component libraries can be used to touch the data files. For instance, using Microsoft Excel to clean the data is not allowed.
- **Required data files:**
  - **monthly\_closings\_2018\_to\_2020.csv**
  - **monthly\_closings\_2021.csv**

### What to submit:

1. Submit via Quercus a Jupyter (IPython) notebook containing your implementation and explanatory comments for all the steps of the analysis with the following naming convention:

**lastname\_studentnumber\_assignment3.ipynb**

Make sure that you **comment** your code appropriately and describe **each step** in sufficient detail. Respect the above convention when naming your file, making sure that all letters are lowercase and underscores are used as shown. **A program that cannot be evaluated because it varies from specifications will receive zero marks.**

2. Submit a report in PDF (up to 2 pages) including the findings from your analysis. Use the following naming conventions **lastname\_studentnumber\_assignment3.pdf**.

### Late submissions will receive a standard penalty:

- up to one hour late - no penalty
- one day late - 15% penalty
- two days late - 30% penalty
- three days late - 45% penalty
- more than three days late - 0 mark

### Other requirements and tips:

1. A large portion of marks are allocated to analysis and justification. Full marks will not be given for the code alone.
2. Output must be shown and readable in the notebook. The only files that can be read into the notebook are the files posted in the assignment **without** modification. All work must be done within the notebook.
3. You have a lot of freedom with how you want to approach each step and with whatever library or function you want to use. As open-ended as the problem seems, the emphasis of the assignment is for you to be able to ***explain the reasoning behind every step***.

### Tas:

Eric Floro, [eric.floro@mail.utoronto.ca](mailto:eric.floro@mail.utoronto.ca)

Please post your questions to Piazza, the TA will not answer clarification questions via email.