The University of Sydney — School of Mathematics and Statistics
FMAT3888 Projects in Financial Mathematics
Semester 2, 2022

Interdisciplinary Project: Portfolio Optimisation with Market Data

In this project, you will work in groups to do portfolio optimisation based on real-world market data. You should complete the project in your group, but write up your report individually. Your group will also give an in-class presentation on your results (and everyone in your group must speak).

This is a deliberately open-ended project: there are a number of different questions here that you can investigate, and there may be multiple ways to approach any given question. You are welcome to go beyond the questions asked here (for example, also considering portfolios where short selling is not allowed), and the quantities in red can be changed to something else for your report, if you wish. More details about the report structure and grading are given at the end of this document.

- The report is due on Monday 31 October at 9am. Late submissions will not be accepted (except for special consideration).
- Your group's presentation slides are due on Monday 31 October at 9am.
- The presentations will happen during the week 13 lecture times (10am–12pm on Monday 31 October and 10–11am on Wednesday 2 November).

All submission is on Canvas.

Programming language: you are free to use any programming language you wish to complete this project. I recommend one where you are comfortable performing linear algebra and optimisation calculations (e.g. Python or MATLAB). Please note that you should give pseudocode for how you performed your calculations in an appendix (see report submission guidelines).

Data

The spreadsheet (available with this project on Canvas) contains monthly returns data for 8 different assets[1]:

1. Australian Equities (AEQ)
2. Developed Market Equities (DEQ)
3. Emerging Market Equities (EMEQ)
4. Australian Listed Property (ALP)
5. Hedge Funds (HF)
6. Australian Fixed Interest (AFI)
7. Global Fixed Interest (GOV)
8. Cash (CASH)

Let $S_i = (S_i)_{t\in\mathbb{N}}$ be the price of asset i (for $i = 1, \ldots, 8$) at the end of month t. The spreadsheet contains the monthly returns

$$R_t^i := \frac{S_t^i}{S_{t-1}^i} - 1,$$

for each month from $t = 0$ (January 2001) until $t = 258$ (July 2022).

[1] In reality each of these are an index based on multiple assets, but we will consider them to be a single asset here for simplicity.

Asset Dynamics We will assume that each asset follows a lognormal returns distribution. That is, given the log-returns (for asset i in month t)

$$X_t^i := \log\left(\frac{S_t^i}{S_{t-1}^i}\right),$$

we assume that the joint returns $\boldsymbol{X}_t := (X_t^1, \ldots, X_t^8)^T \in \mathbb{R}^8$ for each month t are i.i.d. multivariate normally distributed with mean $\boldsymbol{a} := (a_1, \ldots, a_8)^T \in \mathbb{R}^8$ and covariance $B = (b_{ij})_{i,j=1,\ldots,8} \in \mathbb{R}^{8\times 8}$.

The log-returns are related to the (simple) returns in the spreadsheet via

$$R_t^i = e^{X_t^i} - 1 \qquad \Longleftrightarrow \qquad X_t^i = \log(1 + R_t^i).$$

Note: since we assume X_t^i is normally distributed, this means that we expect $R_t^i > -1$ to always hold (why is this useful?). If we just took R_t^i to be normally distributed, then we could get $R_t^i \leq -1$ with positive probability.

Parameter Estimation

Question 1. Estimate the parameters $\boldsymbol{a}$ and B using market data for the two time intervals: (A) from January 2007 to December 2010 inclusive, and (B) from January 2016 to December 2019 inclusive.

Question 2. For $n \in \mathbb{N}$ the n-month return for asset i (from month t to $t+n$) is given by

$$R_{t,n}^i := \frac{S_{t+n-1}^i}{S_{t-1}^i} - 1.$$

Show that

$$R_{t,n}^i \sim e^{Y^i} - 1, \tag{1}$$

where $\boldsymbol{Y} := (Y^1, Y^2, \ldots, Y^8)^T \in \mathbb{R}^8$ has a multivariate normal distribution with mean $n\boldsymbol{a}$ and covariance matrix nB.

Question 3. Let the random vector $\boldsymbol{R}^{(1)} := (R_1^{(1)}, \ldots, R_8^{(1)})^T$ (respectively, $\boldsymbol{R}^{(2)} := (R_1^{(2)}, \ldots, R_8^{(2)})^T$) model the joint annual (respectively, two-year) returns for the eight assets. For $k = 1, 2$ denote

$$\mu_i^{(k)} := \mathbb{E}\left[R_i^{(k)}\right], \quad c_{ij}^{(k)} := \mathrm{Cov}\left(R_i^{(k)}, R_j^{(k)}\right), \quad \rho_{ij}^{(k)} := \frac{c_{ij}^{(k)}}{\sqrt{c_{ii}^{(k)}}\sqrt{c_{jj}^{(k)}}}, \quad \forall i, j = 1, \ldots, 8.$$

Use the results in Q1 and Q2 to compute/estimate $\mu_i^{(k)}$, $c_{ij}^{(k)}$, $\rho_{ij}^{(k)}$ for all i, j and $k = 1, 2$ for the two time intervals (A) and (B) from Q1.

Simplifying Lognormal Distributions In (1), we are assuming that $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ follow a lognormal distribution. This may make the computations below more difficult; if so, you can approximate $\boldsymbol{R}^{(1)}$ and $\boldsymbol{R}^{(2)}$ as normally distributed using the below reasoning.

In general, if we have a normally distributed random variable X which tends to take small values (i.e. $X \approx 0$ with high probability), then the random variable $Y := e^X - 1$ satisfies $Y \approx X$ (since $e^x - 1 \approx x$ for $x \approx 0$).

So even though Y is actually lognormally distributed, for the purposes of computation it would be reasonable to simplify and assume Y is normally distributed instead. We could assume the mean/variance (or covariance for multivariate distributions) of Y is the same as X, since $Y \approx X$, but it would be better to directly compute $\mathbb{E}[Y] = \mathbb{E}[e^X - 1]$ and $\mathrm{Var}(Y) = \mathrm{Var}(e^X - 1)$ and use these instead.

Static Portfolio Optimisation

Question 4. Consider an investor who statically invests all their wealth in these eight assets for two years. Answer the following questions using both sets of parameters from Q1, namely for periods (A) and (B).

(a) Solve the utility maximisation problem:

$$\max \mathbb{E}\left[U(\boldsymbol{w}^T \boldsymbol{R}^{(2)})\right],$$
$$\text{s.t.} \sum_{i=1}^{8} w_i = 1,$$

where $\boldsymbol{w} = (w_1, \dots, w_8)^T$ is the vector of weights, and $U(x) = -e^{-\gamma x}$ with $\gamma = 1$.

(b) Comment on the differences of your results corresponding to the two datasets (A) and (B).

(c) Compare your result from (a) (with dataset (B)) with the realised return on her portfolio using the market data for the period from January 2020 to December 2021.

Question 5. Under the same setup of Q4, answer the following questions for both datasets (A) and (B) from Q1.

(a) Find the efficient frontier for the market consisting of these eight assets, using the estimated parameters $\mu_i := \mu_i^{(2)}$, $c_{ij} := c_{ij}^{(2)}$, $\rho_{ij} := \rho_{ij}^{(2)}$ for all $i, j = 1, \dots, 8$.

(b) Find the portfolio with minimum variance which yields at least 10% expected return. That is, solve:

$$\min \boldsymbol{w}^T C \boldsymbol{w},$$
$$\text{s.t.} \sum_{i=1}^{8} w_i \mu_i \geq 0.10,$$
$$\sum_{i=1}^{8} w_i = 1,$$

where $\boldsymbol{w} = (w_1, \dots, w_8)^T$ is the vector of weights and $C = [c_{ij}]$ is the covariance matrix for $\boldsymbol{R}^{(2)}$.

(c) Comment on the differences between your results for the two datasets.

(d) Compare your result from (b) (with dataset (B)) with the realised return on her portfolio using the market data for the period from January 2020 to December 2021.

(e) Compare your results to those from Q4.

Dynamic Portfolio Optimisation

Question 6. Consider an investor who invests all their wealth in these eight assets for two years, during which they will adjust their portfolio weights at the beginning of the second year. For $k = 1, 2$, denote $\boldsymbol{V}^k := (V_1^k, \dots, V_8^k)$ the returns of the eight asset classes for the k-th year. Note that $\boldsymbol{V}^1$ and $\boldsymbol{V}^2$ are i.i.d. copies of $\boldsymbol{R}^{(1)}$. Let $\boldsymbol{w} = (w_1, \dots, w_8)^T$ (respectively $\boldsymbol{u} = (u_1, \dots, u_8)^T$) be the portfolio weights at the beginning of the first year (respectively second year). Then the return of the profolio over the two-year investment period is given by

$$G(\boldsymbol{w}, \boldsymbol{u}) = (1 + \boldsymbol{w}^T \boldsymbol{V}^1)(1 + \boldsymbol{u}^T \boldsymbol{V}^2) - 1. \qquad \text{(why?)}$$

Suppose the investor believes that parameters estimated using the dataset (B) are valid. Answer the following questions.

(a) Solve the utility maximisation problem:

$$\max \mathbb{E}[U(G(\boldsymbol{w}, \boldsymbol{u}))],$$
$$\text{s.t.} \sum_{i=1}^{8} w_i = \sum_{i=1}^{8} u_i = 1,$$

where $U(x) = -e^{-\gamma x}$ with $\gamma = 1$. Note $\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{V}^1)$ may depend on the realisation of $\boldsymbol{V}^1$.

(b) Compare your result with that for Q4(a).

Question 7. Under the setup of Q6, answer the following questions.

(a) Solve the portfolio optimisation problem:

$$\min \ \mathrm{Var}(G(\boldsymbol{w}, \boldsymbol{u})),$$
$$\text{s.t.} \ \mathbb{E}[G(\boldsymbol{w}, \boldsymbol{u})] \geq 0.10,$$
$$\sum_{i=1}^{8} w_i = \sum_{i=1}^{8} u_i = 1.$$

(b) Compare your result with that for Q5(b).

(c) Compare your results to those from Q6.

Possible Extensions

The above questions give you a chance to explore static and dynamic portfolio optimisation with real-world data. However there are many more real-world complications that you could consider, in addition to the analysis above. Some suggestions include:

- Vary the parameter γ in Q4 and Q6, and/or vary the minimum expected return in Q5 and Q7.
- No short selling allowed, $w_i \geq 0$ for all i.
- Limits on exposure to any given asset, $-L \leq w_i \leq L$ for some constant L.
- Alternative risk measures (other than portfolio variance) that only consider the risk of losses, such as 95% Value-at-Risk (i.e. the 5th percentile of the distribution). There are many other measures, such as Conditional Value-at-Risk and lower partial moments (e.g. semivariance).
- Change in market dynamics: suppose you are told by your CIO that the correlation between two assets is now to be different to the estimated value from Q3 because of some recent market disruption. Find a change which makes C no longer positive semidefinite.[2] Then calculate—it's up to you exactly how to do this—a positive semidefinite approximation to your new C, and see how this affects your portfolio optimisation in Q5.

This could be for either the static or dynamic case. I recommend that you consider at most one or two of these, but you should consider this a chance to explore questions that you find interesting.

[2] Of course, covariance matrices should always be positive semidefinite, but your CIO didn't do the calculations to confirm this. For example, the Basel Committee on Banking Supervision, the international banking regulatory organisation, proposed a mandatory risk calculation that has non-positive semidefinite covariance matrices [J. Zhan, Model Risk of FRTB Standardized Approach: I. Sensitivity Based Method (2021), available at `http://dx.doi.org/10.2139/ssrn.3934013`].

Submission Guidelines

Report

Your report must be a single pdf file, written in LaTeX with 12pt font and reasonable margins. It should be 8–12 pages (including references and appendices). You should work on the project questions as a group, but your report should be entirely written individually. Please include your group number and SIN on the first page of the report. Your report is due at 9am on Monday 31 October.

I recommend the following structure for your report:

1. Executive Summary
   - A short (< 1 page) overview of the key work and conclusions in your report. This must be *non-technical*, and should be understood by a finance professional with limited mathematical knowledge. This means you should avoid technical mathematical terms such as "vector" and "dynamic programming".
2. Introduction
   - Background of the task, e.g. popularity of portfolio optimisaton in investing; theory of portfolio optimisation (modern portfolio theory, utility theory, stochastic processes theory).
   - Summary of the work to be discussed in the report, including an explanation of the structure of the (rest of the) report.

3. Mathematical Setup
   - Give a technical mathematical formulation of the problems being discussed. The formulation should be consistent with your theoretical and computational analysis (e.g. don't show optimisation of returns if you later optimise portfolio value/wealth).

4. Theoretical Results
   - Details of your theoretical (i.e. pen-and-paper) calculations for the static & dynamic portfolio optimisation problems.

5. Computational Results
   - Explain your parameter estimation procedure.
   - Static & dynamic portfolio optimisation: interpret your results and compare static vs. dynamic.
   - Separate discussion of any extensions considered.
6. Conclusions
7. References
   - Any reasonable bibliography/citation formatting is acceptable.
8. Appendix
   - Provide brief pseudocode for your computational results (i.e. the main steps in your code), approx. 1–2 pages.

You may use a different structure if you wish (keeping in mind the grading criteria). For example, you could replace the 'mathematical setup' and 'theoretical analysis' sections with two sections on static & dynamic portfolio optimisation (where each section includes the relevant setup and theoretical analysis).

The report must be written as a self-contained document that tells a coherent story. You should not say things like "the answer to Q4 is...".

Presentation

As a group, you should prepare a presentation on your work which you will give during the week 13 lecture times.

- You will have a strict 9 minutes for your presentation, with 1 minute for questions afterwards (while the next group gets ready).[3]
- Every group member must speak during the presentation for roughly the same amount of time (i.e. you will each speak for approx. 1–2 minutes).
- Your must submit your presentation slides by 9am on Monday 31 October (everyone in your group must submit the same slides).
- Your slides must be in pdf form (e.g. by using LaTeX's beamer package or exporting a Microsoft Powerpoint presentation to pdf using "File → Export → Create PDF").
- Please be ready to give your presentation at the Monday lecture. I will use a random number generator to decide the speaking order (and I will not tell you in advance what the order will be). Most groups will speak on Monday, since we have 2 hours of class time.

There is not enough time to discuss everything you did: you should pick some of the most interesting ideas and results from your project to share. Remember that everyone in the audience has been working on this project, so there is no need to spend lots of time introducing the problem, data, etc.

Grading

Report

Your report is worth 40% of your FMAT3888 grade, and will be assessed individually. It will be graded based on:

- Completeness: are your methods fully explained? Did you interpret your results rather than just state them? Did you thoroughly answer each question?
- Mathematical correctness: mathematical reasoning is correct, efficient/elegant methods are used
- Writing: report is well-organised and easy to understand, graphs are properly formatted (axes labelled, font size not too small, etc.), spelling & grammar are correct, proper use of citations

The relative weight of each part of the report is as follows:

Report Part	Marks	Weighting
Executive (non-technical) summary	4	10%
Parameter Estimation	6	15%
Static Portfolio Optimisation	14	35%
Dynamic Portfolio Optimisation	10	25%
Quality of presentation & writing	6	15%
Total	40	100%

[3]There are 15 groups, and we have $3 \times 50 = 150$ minutes of class time, so we cannot have anyone's presentation go over time. For reference, when I am giving research talks, I usually prepare one slide per 1–2 minutes of speaking time. (but that is just me, you might be different)

Presentation

Your presentation is worth 10% of your FMAT3888 grade. You will be assessed as a group (i.e. everyone in your group receives the same grade). It will be graded based on similar criteria to the report:

- Completeness: there isn't time to show everything you did, but did you pick interesting/important results? Is your presentation detailed enough for everyone to understand what you did?
- Mathematical correctness: is your reasoning accurate, have you shown a good amount of detail (not every step in a calculation, but enough for us to understand), are your computational results relevant?
- Clarity of presentation: both slides (easy to understand, not too cluttered, figures/data well presented) and the live presentation (clear explanations, not rushed, keep to time)

The relative weight of each part of the presentation is as follows:

Presentation Component	Marks	Weighting
Slides	5	50%
In-class presentation	5	50%
Total	10	100%